The CARE Algorithm

A Heuristic Method To Solve

Capacitated Set Covering Location Problems

On Large-Scale Networks

Wim de Rooij

Dissertation submitted in part fulfilment of requirements for the Degree of Master of Science in Geographical Information Systems

12 May 2000

To Astrid

ABSTRACT

The location of service centres determines for a larger part their effective and efficient functioning. Whether a certain location is optimal depends on the objective pursued. In this study the objective is a complete coverage of demand points along a road network with a miminal number of centres. The capacity of the service centres is assumed to be limited and the distance from a service centre to demand points allocated to it should not exceed a certain maximum value. This location-allocation problem is known as the capacitated set covering location problem (CSCLP).

In the last decades several heuristic methods have been developed to solve location-allocation problems. These methods are usually suited for small-scale networks (up to hundreds of nodes and arcs) to medium-scale networks (up to thousands of nodes and arcs), and capacitated centres were hardly considered. The CARE algorithm was developed to solve the CSCLP for large-scale networks (up to 100 000 nodes and arcs and more). The CARE algorithm is named after its main constituents: centre adding, centre repositioning, and centre elimination. The CARE algorithm has been implemented in a GIS environment, which facilitated rapid development, and made the visualisation of results straightforward.

The CARE algorithm has been applied to medium-scale model networks, and to large-scale (real-world) networks in a case study of Statistics Netherlands, where the problem is to determine the minimum number of interviewers required to cover all addresses in the Netherlands. A detail study indicated that by finding proper locations the number of service centres may be reduced without an increase of travel expenses, and with a decrease of total costs. The main results of the study was, that the CARE algorithm produces good results in a reasonable time, even for large-scale (real-world) networks.

Table of Contents

List of Table and Figures

DISCLAIMER

The results presented in this thesis are based on my own research for the Department of Geography at the University of Salford. All assistance received from other individuals and organisations has been acknowledged and full reference is made to all published and unpublished sources used.

This thesis has not been submitted previously for a degree at any institution.

Wim de Rooij

CHAPTER 1

Introduction and Outline

The problem of finding optimal locations for service centres has been studied for many centuries. Early in the 17th century Fermat posed one of the first location optimisation problems: given three points in the plane, find the optimal point of which has the sum of the distances to the other three points is the smallest. This problem was (partially) solved by Torricelli, who observed that the circles circumscribing the equilateral triangles constructed on the sides of the triangle, exterior to the triangle, intersect at the optimal point (cf. Love, Morris and Wesolowsky, 1988).

In the early 20th century Weber attempted to find the optimal location for a manufacturing plant with the objective to minimise the transportation costs of raw materials and finished goods. The geometrical procedures used where suitable only for simple single-facility problems, and little progress along these lines was made in the next decades (ReVelle, 1991).

With the advent of digital computers at the end of the 1950's location theory received a new impulse. Numerical methods for more complex problems and for multiple facility problems were proposed. With multiple service centres (or facilities) not only the location of the centres, but also the allocation of demand points to the centres has to be determined, which led to the term location-allocation modelling (cf. Gosh and Rushton, 1987).

In view of the economic impact of facility location, the basic versions of the location-allocation models may also be adapted in order to take other aspects into account. In the last few decade, such models may take into account micro-economic factors (e.g. the costs building and running a plant), competition for location, location-routing problems, and location of obnoxious facilities (see Daskin, 1995; Drezner, 1995; Domschke and Krispin, 1997). Since these

factor complicate the matter the size of the networks used is modest, or the network distances are approximated by continuous functions.

1.1 Location-allocation problems

Most private and public organisations have been faced with the problem of finding the best location for their facilities. Which locations are considered to be optimal depends on the pursued objectives, such as minimal transportation costs or maximal accessibility, which are reflected in the location-allocation models used. An important distinction in objectives is usually found between the private sector location models and the public sector location models. In the private sector the need to reduce costs and maximise efficiency is stressed, while in the public sector the provision of equitable service is of major importance, and maximising efficiency is less important (cf. Morrill and Symons, 1977).

Location-allocation models can be divided into planar location models and network models. In planar models demand and supply centres may be located anywhere on a plane, which is often an approximation to simplify the problem at hand, but also an appropriate approach in modelling noise levels (see e.g. Hurter and Martinich, 1989; Love, Morris, and Wesolowsky, 1988). In network models demand and supply centres are located on (or along) a network, which is, for example, a sound approach for handling freight-traffics on roads, railroads and waterways (see e.g. Handler and Mirchandani, 1979; Daskin, 1995).

1.2 Practical applications

There are many examples of the use of network location models in practice. One of the best known in the private sector is finding the optimal locations of warehouses in order to reduce transportation costs, which is an extension of Weber’s problem. A well-known problem in the public sector is the siting of ambulance posts. How many ambulance posts are needed, and

where should they be located, so that any place can be reached within a certain time period, say 15 minutes, by an ambulance from the nearest post? A similar problem faces a pizza service, which claims to deliver within 30 minutes after ordering. Another is the location of library branches within a municipality with the objective to minimise the maximum distance that an inhabitant has to travel to the nearest library branch.

Most of the network location-allocation problems are notoriously difficult to solve, because of the combinatorial nature of the problems. In practice exact solutions can only be obtained for small-scale networks. Several heuristic algorithms to find approximate solutions have been developed. These algorithms have been successfully applied to small-scale and medium-scale networks. Most heuristic methods are not suited for large-scale networks, especially when the capacity of the centres is limited. The problem that initiated the present study can be formulated as a large-scale network location-allocation problem with capacitated service centres.

1.3 The case of Statistics Netherlands

Statistics Netherlands conducts many surveys by personal interviewing of persons or households. The Department of Households Surveys, which is responsible for conducting the surveys, manages a group of interviewers of about 600 persons. For a number of years it is felt that the group of interviewers is undersized. Many interviewers complain that their workload is too high and many addresses in the sample which were allocated to the interviewers are returned unfinished. Some interviewers, however, complain that their workload is too low. In order to get a firm grip on the situation the following questions need to be answered:

- What is the minimum number of interviewers required to cover the entire Netherlands?
- What are the optimal locations for these interviewers?
- Which regions have a shortage or a surplus of interviewers?

The conditions are that both the maximum workload and the maximum travel distance (between interviewer and sample address) are fixed. The first two questions together define a so-called capacitated set covering location problem (Current and Storbeck, 1988), which needs to be solved in a road-network environment. The third question may be answered when the optimal locations of the interviewers are compared with the actual locations of the interviewers.

The purpose of the present study was to develop a heuristic method to solve the capacitated set covering location problem within a reasonable time for a large-scale road-network, which in principle should represent the entire road-network of The Netherlands.

1.4 Outline of the thesis

In Chapter 2 the capacitated set covering location problem and its relation to several other location-allocation problems will be discussed. A number of established heuristic methods to solve such problems will be reviewed. Further, the effects of data aggregation on the location of the service will be discussed. A new heuristic method to solve the capacitated set covering location problem in a large-scale network environment will be described in Chapter 3. The method is named CARE algorithm after the three basic elements: centre adding, centre repositioning, and centre elimination algorithms. An implementation of the CARE algorithm using the GIS environment provided by ArcInfo-software of ESRI is described in detail. In Chapter 4 the algorithm is applied to model networks in order to investigate the properties of the algorithm. One of the models represents the road-network of a small part of The Netherlands. The preparation of the road-network for the present purpose is described in detail. In Chapter 5 the CARE algorithm is applied to the case of Statistics Netherlands. The effects of maximum cover distance and maximum workload on an optimised interviewer configuration, and the effects of the boundaries of the study areas are discussed. The economic aspects of optimising locations are also considered. In Chapter 6 the properties of the CARE algorithm,

and possible extensions to related problems are discussed. The merits of the CARE algorithm are reviewed and conclusions are presented.

CHAPTER 2

Network Location-Allocation Problems

Location-allocation problems and models on networks may be classified in a number of ways (cf. Handler and Mirchandani, 1979, Section 1.3; Daskin, 1995, Section 1.5). Real-world road networks may be modelled as a connected general network, i.e. a network composed of nodes and links, that connect any two nodes (see Figure 3.1). In general there are several routes to connect two nodes. The connection between several groups of nodes may be weak, e.g. when large regions are separated by a river, which can only be crossed by bridges or ferries, which may be separated by distances of tens of kilometres. Frequently found in location-allocation literature are tree networks, which are special in the sense that in these networks there is a unique route connecting any pair of two nodes (see Figure 3.1).

This chapter proceeds with an overview of network location-allocation problems and their relationships. Next, the complexity of algorithms and a number of established methods to solve location-allocation problems are reviewed. The chapter ends with a discussion of the effects of demand aggregation on solving location-allocation problems.

2.1 Classification of problems and models

In the context of connected general networks several types of location-allocation models may be distinguished. The following distinctions will be useful:

- Private sector versus public sector models
- Number of locations: fixed number or model output
- Optimality criterion or objective function
- Coverage distance limited or unlimited
- Capacitated versus uncapacitated models

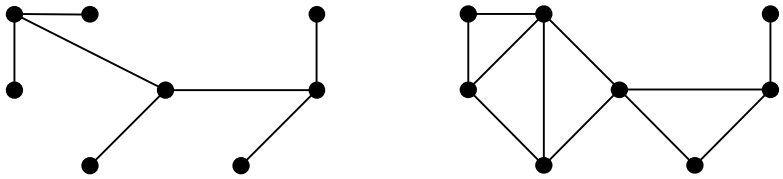

Figure 2.1 Examples of a tree network (left) and a connected network (right)

- Demand points only at nodes versus anywhere on the network
- Service centres only at nodes versus anywhere on the network

As mentioned before, in private sector models reduction of costs and increase of benefits is stressed, while in public sector models the provision of equitable service is most important. In public sector models complete coverage is usually required, which may require many service centres. In private sector models maximal coverage (or maximal profit) with a limited number of service centres is demanded. The number of service centres to be located may be specified in advance (optimisation for a fixed number of service centres) or has yet to be determined (the number of service centres required to obtain specific objective). The different types of objectives will be discussed later on. The coverage distance of a centre may be limited (e.g. a super market may deliver groceries at home if the distance is less than 5 kilometres) or it may be unlimited (furniture may be delivered through the entire country). The capacity of service centres may be limited (e.g. fixed capacity), unlimited or model output. In Section 2.3 the position of service centres and demand points (which require service) on nodes and links of the network will be discussed.

Other classifications of location problems and models are the following. The location models considered so far are static. The parameters of the models do not vary in time, but represent a typical or an average situation. In practice many location problems are dynamic. Police cars may be stationed such that any location may be reached within a certain amount of time. However, when one or more police cars are in action and leave their location a reloca-

tion of the remaining police cars is necessary in order to remain in a position to be able to reach any location within the specified time range. Centres may produce a single type of service or several types of services, e.g. centres may consist of a police station and/or a fire station. The problem may then be to determine type and location of the centre in order to satisfy the requirements of police stations and fire stations with a minimum of costs. Also location problems may pursue several objectives. A refuse dump should be sited such that both the transport costs are minimised and the inconvenience for the neighbours is minimised, which requires a trade-off between two objectives.

In practice many location problems occur in three different settings (cf. Hodgart, 1979):

1. There are no restrictions on the location of service centres and there are no existing service centres (general problem)
2. Locate new service centres in addition to existing centres (additional or incremental facility location problem)
3. Locate additional service centres and possibly close down some existing centres (reorganisation problem)

Of these problem settings the first has received most attention in the literature, perhaps because it is considered to be the hardest one to solve. This problem is, however, only applicable when an entirely new service is introduced, whereas the second and third problem are much more likely to occur in practical situations (Hodgart, 1979).

In the case of Statistics Netherlands sketched above the focus is on the short run on the first problem setting. One wishes to know the minimal number of interviewers required and the optimal locations of these interviewers. This optimal configuration of locations will be considered a primary goal to be reached in the next couple of years, a strategy close to the third problem setting. However, since dismissal of interviewers (often with considerable records) is

considered not appropriate, one has to rely on natural turnover and on selective (regional) recruitment of interviewers.

2.2 Relationships between problems and models

The set covering location problem (SCLP) can be formulated as follows. Determine the number of service centres and their locations such that all demand points are covered. A demand point is covered if the distance of the shortest path to a service centre is equal to or lower than a certain coverage distance (cf. Daskin, 1995, Chapter 4). The number of centres (and the costs involved) to require complete coverage may be very high. In that case the number of centres could be fixed and the limit on coverage distance removed. In that situation the objective could be to minimise the maximum distance of a demand point to its nearest service centre, which leads to the p-centre model (cf. Daskin, 1995, Chapter 5). Another objective may be to minimise the sum of the distances of the demand points to their nearest service centre, which leads to the so-called p-median model (cf. Daskin, 1995, Chapter 6). In the former case the stress is on equity of service, and in the latter on profitability (see Morrill and Symons, 1979). In some models a balance between the both objectives is sought, which may lead to network models such as the medi-centre models (see e.g. Handler and Mirchandani, 1979, Section 4.2).

If the number of centres is fixed and the coverage distance is kept limited two other models may be formulated. The primary objective in both is to maximise the demand covered. The secondary objective in one is to minimise sum of the travel distances, leading to the maximal covering location problem (Church and ReVelle, 1974). The secondary objective in the other would be to minimise the maximum travel distance. The latter has not been encountered in the literature.

Table 2.1 Network location-allocation models

Model	Objective(s)	Capacity of centres	Coverage Distance	Number of centres
Set covering location problem	Complete coverage of demand points	Unconstrained	Limited	Model output
P-median problem or minisum problem	Minimise sum of distances	Unconstrained	Unlimited	Fixed (P)
P-centre problem or minimax problem	Minimise maximum coverage distance	Unconstrained	Unlimited	Fixed (P)
Maximal covering location problem	Maximal coverage of demand points	Unconstrained	Limited	Fixed
Capacitated set covering location problem	Complete coverage of demand points	Constrained	Limited	Model output

In the set covering location model the capacity of each service centre is assumed to be unlimited. When the capacity is limited, the problem turns over into the capacitated set covering location problem (CSCLP), which may be formulated as follows. Determine the number of service centres and their locations such that all demand points are covered with the restrictions that the coverage distance and the capacity of the centres are limited (Current and Storbeck, 1988). Also in this case the number of centres required may be very high. Fixing the number of centres again leads to two additional models. The first one has as primary objective a maximal covering and as secondary objective minimisation of the sum of the distances, and is called the capacitated maximal covering location model (CMCLP) (see Current and Storbeck, 1988). The second one has as secondary objective the minimisation of the maximum travel distance. Also this problem has not been encountered in the literature.

Other types of location-allocation models may be found in the literature (cf. Hillsman, 1984). The most studied models are listed in Table 2.1 with a number of characteristics. In solving a particular problem, it may be worthwhile to keep an eye on these relationships. Models may be similar in some respect, but may also differ (considerably) in other aspects. A solution strategy suitable for one problem may also be suitable for solving a related problem. Problems may have properties in common, or may differ essentially in some.

A further distinction in network problems is the location of the demand points and the location of the service centres. Demand points may be located at the nodes and also anywhere along the links between the nodes. In the latter case the demand points can also be treated as nodes, which are sometimes called pseudo-nodes because they only connect two links. Usually the demand points are assumed to be at the nodes. Similarly the service centres may either be located at the nodes and along the links or only at the nodes. In the former case the term absolute location problem could be used and in the latter case the term vertex location problem. These are generalisations of the terms absolute centres and vertex centres used in the p-centre problem (Hakimi, 1964).

The p-median problem has at least one optimal solution where the service centres are located on the nodes of the network (Hakimi, 1965). In general the optimal locations of the service centres are not located at the nodes. However, if the set of nodes is supplemented by a finite set of so-called intersection points, then the uncapacitated set covering location problem has an optimum solution consisting of points in the augmented set (Church and Meadows, 1979). The intersection points are points chosen on the links such that the distance between an intersection point and at least one node is equal to the coverage distance. On a large scale network the number of intersection points on a link may be considerable, and the links may be rather densely populated with intersection points.

From now on it will be assumed that service centres may only be located at the nodes. Further, it will be assumed (as is common practice in the literature) that nodes may only be allocated to more than one service centre. In the case that the capacities of the centres are unlimited this not an important restriction, because nodes are allocated to the nearest centre, and when more than one centre is nearest, the node may be allocated to one of the nodes picked randomly. When the capacities of the centres are limited, allocation of a node to only one centre is an important restriction, because in that case a node may only be allocated to the nearest centre if the demand of the node does not exceed the remaining capacity of the centre. Otherwise, part of the demand could be assigned to this centre, and the remaining demand to other centres.

From now on it will be assumed that the demand points are located at nodes (or pseudo-nodes), or that the total demand along arcs is continuously distributed along the arcs. In the latter case, the model is similar to the so-called arc-covering models (cf. Church and Meadows, 1979; Hillsman, 1984). In Section 2.6 the effects of demand aggregation on the locations of the centres is briefly discussed.

2.3 Complexity of algorithms

An important property of problems is their complexity, or the complexity of the algorithm to solve the problem (see Larson and Odoni, 1972, Section 6.2.5; Evans and Minieka, 1978, Section 2.4; Daskin, 1995, Chapter 3). Algorithms belong to one of two classes: the polynomial-time algorithms or the NP-hard algorithms, where NP denotes non-deterministic polynomial. The p-median problem and the p-centre problem are in the class of polynomial problems (Daskin, 1995, pages 161 and 203). The three covering problems listed in Table 2.1 belong to the class of NP-hard problems (Daskin, 1995, pages 94 and 114), i.e. they belong to the same class as the notorious travelling salesman problem. The time to solve a polynomial problem is proportional to a polynomial of the input size. The size of network lo-

cation-allocation models may be characterised by the number of nodes in the network. When a problem is NP-hard no algorithm exist to solve the problem in a time proportional to the size of the input size. The time to solve NP-hard problems is proportional to an exponential function of the input size.

It will be clear that the time to solve NP-hard problems increases exponentially as the size of the problem increases. In the case of NP-hard location-allocation problems, such as the capacitated location set covering problem, the solution time increases exponentially as the number of nodes in the network increases. For large-scale networks the covering problems can not be solved in reasonable amount of computer time. Therefore, heuristic methods are necessary to obtain approximate solutions of the covering problem mentioned above. In general heuristic methods should have the following nice properties (see also Evans and Minieka, 1978):

- Solutions are generally close to optimal
- Solutions far from optimal are rare
- Reasonable computation time
- Reasonable storage requirements

In general, the accuracy of heuristic solutions is hard to give, although some heuristic algorithms provide worst-case error bounds. An example of the latter is the Lagrangean relaxation method (see Daskin, 1995, Section 4.5.2). The complexity of the heuristic algorithms can be determined in the same manner as for exact algorithms and should of course be polynomial. The computation time for Dijkstra's algorithm to compute the distance between two nodes on a network is proportional to n^2, where n is the number of nodes in the network (a derivation is given by Daskin, 1995, section 3.3).

2.4 Unified linear model

The network location-allocation models mentioned above (and several extensions of these) may be expressed as special cases of a unified linear model as follows (Hillsman, 1984). First, a matrix *X* is introduced to indicate at which nodes centres are located, and also to indicate to which centre a demand point is allocated. If a centre is located at node i

$$x_{ii} = 1 \qquad \text{(Eq. 2.1)}$$

If no centre is located at node i

$$x_{ii} = 0 \qquad \text{(Eq. 2.2)}$$

Demand points may only be allocated to nodes with centres, i.e.

$$x_{ij} = 1 \rightarrow x_{ii} = 1 \qquad \text{(Eq. 2.3)}$$

The location-allocation may now be formulated as follows. Minimise the objective function

$$\sum_{ij} a_{ij} x_{ij} \qquad \text{(Eq. 2.4)}$$

The matrix *A* is yet to be defined and specifies the problem at hand.

$$\sum_{j} a_{ij} = 1 \qquad \text{(Eq. 2.5)}$$

The matrix *D* is used to store the distance (or travel time, or travel costs) between two nodes, matrix element d_{ij} being the distance between node i and node j. Using this terminology the p-median problem may be derived when

$$a_{ij} = w_i d_{ij} \text{ ,} \qquad \text{(Eq. 2.6)}$$

where w_i is the demand at node i, and the number of centres is fixed:

$$\sum_{i} x_{ii} = p \qquad \text{(Eq. 2.7)}$$

The capacitated set covering location problem (CSLCP) is derived by substitution of the following matrix elements

$$a_{ij} = 1 \qquad \text{(Eq. 2.8)}$$

if i = j,

$$a_{ij} = 0 \qquad \text{(Eq. 2.9)}$$

if $i \neq j$ and $d_{ij} < S$,

$$a_{ij} = M \qquad \text{(Eq. 2.10)}$$

if $i \neq j$ and $d_j \geq S$. Here, S is the maximum coverage distance, and M an arbirary positive constant. The number of centres is now part of the solution of the model.

The ULM requires a precise definition of the objective function of location-allocation models. Further, the relations between the models become apparent. Several extensions of the basic location-allocation models may also be represented in this way. The way in which the ULM is presented can readily be cast in the canonical form of linear programming equations (cf. Daskin, 1995, Chapter 2)., which does not guarantee that linear programming techniques mat readily provide an answer. The ULM does not reveal whether the problem at hand is NP-hard or not.

2.5 Established solution methods

On small-scale networks location-allocation problems may be solved using enumeration, i.e. examining all possible solutions to determine the optimal solutions, or using a graph-theoretic approach. These methods are only suitable for very small networks, or when a single facility has to be located (Handler and Mirchandani, 1979). Graph-theoretic methods take advantage of the network structure and can be very efficient if the network is a tree, which implies that there is a unique route to travel from one node to another. Graph-theoretic methods may

also be more efficient than mathematical programming algorithms (Larson and Odoni, 1981, Chapter 6, Introduction).

Mathematical programming algorithms have been applied to many location-allocation models (see Daskin, 1995). In the ULM vertex location-allocation models are presented as integer programming problems. In solving these problems one usually relaxes one or more of the constraints to obtain a formulation with that be handled by one or more of the linear programming techniques. These approaches are in a strict sense heuristic, but often, when coupled with other heuristic methods, yield results that are provable optimal or very close to optimal (Daskin, 1995, Section 6.4). The linear programming approach is quite powerful in many applications (see Daskin, 1995), but is restricted to networks of a rather small scale.

In attacking medium-scale to large-scale network location-allocation problems several heuristic methods have been used with considerable success: (1) node partitioning algorithm (or alternating algorithm), (2) myopic (or greedy) algorithms, and (3) node-substitution (or bump-and-shift) algorithms. In discussing these methods the p-median problem will be used as an example.

Node partitioning was first used for the p-median problem (Maranzana, 1964), but can also be used for other problems such as the p-centre problem. In its application to the p-median problem this method runs as follows. Arbitrarily select p candidate centres and by assigning all demand nodes (points) to the nearest centre the node set is partitioned into p subsets. Now determine the median of each subset, thus obtaining a new set of p candidate centres. It can be shown that this procedure yields non-increasing values of the objective function. The procedure may, however, converge to a non-optimal solution (Maranzana, 1964). Little numerical work based on node partitioning has been reported (Handler and Mirchandani, 1979).

The greedy algorithms for the p-median problem have two variants: the add-algorithm and the drop-algorithm. Suppose m is the number of candidate centres. The add-algorithm starts with the selecting of the candidate centre that yields the lowest value of the objective function. Next a second centre is sought in the remaining set of (m-1) candidates such that the value of the objective function decreases most. This process is repeated until p centres have been found and (m-p) candidates remain. The drop-algorithm starts with all n candidate centres. Next, the centre is sought which upon removal increases the objective function the least. This procedure is repeated until p centres remain.

Several variants of interchange (or bump-and-shift) algorithms have been used. Again suppose that p centres have to be selected from m candidates. The interchange algorithm (Teitz and Bart, 1968) starts with a set of p arbitrarily chosen centres. In the next step the effect of an interchange between the first of the (m-p) non-centres and each of the p centres is evaluated. The interchange with the largest decrease of the objective function is executed. In the same way an interchange is effectuated for the remaining (m-p-1) non-centres, thus completing an entire iteration. In each of the (m-p) steps of an iteration the effect of p possible interchanges are evaluated, and the interchange with the highest decrease of the objective function is effectuated. The procedure is stopped when an iteration does not result in an interchange. In a variant of this procedure only one interchange is made during an iteration (Goodchild and Noronha, 1983). After evaluating the effect of all interchanges of between one of p centres and one of the remaining (m-p) candidates, only one interchange is made.

The global/regional interchange algorithm (GRIA) combines node partitioning and drop-and-add (Densham and Rushton, 1992). Start with an arbitrary selection of p centres. In the global phase first drop a centre, then add one of the remaining (m-p) candidates (greedy drop and add steps). In the regional phase: determine the median of each subset (subset median step). Repeat the two phases until no interchanges have to be made.

Enumeration and graph-theoretic approaches may be used for very small networks (up to tens of nodes), while the mathematical programming techniques seem appropriate for small-scale networks (up to hundreds of nodes). The interchange algorithms are suited for medium-scale networks (up to thousands of nodes), while myopic algorithms and node partitioning may be well suited for large-scale networks (up tens of thousands or hundreds of thousands nodes). The division of the methods in three groups is also based on the level on which the algorithms operate. In the small-scale group nodes are treated on an individual or microscopic level. The medium-scale group of algorithms operates on mesoscopic level. At the start the nodes are treated as collectively, and at the end (almost) individually. The large-scale group treats the nodes at a macroscopic level. In fact the nodes form partitions belonging to a certain centre, although individual nodes may switch during the repositioning phase.

In practice, fairly large regions are studied in location-allocation problems, which implies that the demand density only gradually changes on the relevant distance scale. As the size of the regions increases small-scale variations become less important, and the results of the methods operating on the mesoscopic and macroscopic levels will steadily improve.

2.6 Aggregation problem

In practice the data required to solve a location-allocation problem is often not available at the desired aggregation level. Even when the data is available, the size of the location-allocation model may be too large to handle, and working with aggregated data is inevitable. Aggregation of demand will have its effect on the solution of the problem, although there is some disagreement in the literature concerning the magnitude of the effect. In some cases data aggregation has little effect on the value of the objective function, but considerable effect on the location of the centres, especially in centre and covering models (Goodchild,

1979). In other cases the effect on the patterns of the locations of centres may be little (Casillas, 1987).

The aggregation problems to be dealt with in the present study are relatively minor ones. The first problem, when the CARE algorithm is implemented in ArcInfo (see Section 3.3.1), is the choice between aggregation of demand at the nodes or disaggregation along the arcs. In the first case, the total demand along an arc is evenly distributed between begin and end nodes. In the latter case, the total demand along an arc is continuously distributed along the arc. The resulting centre configurations of the two alternatives could be compared to get an impression of the effect of aggregation versus disaggregation of demand. In Section 4.6 the results of the two alternatives for model networks will be considered. The second problem is the effect of estimated demand at the nodes or along the arcs, that is based on aggregated demand data at a relative larger scale. Since demand data at a detailed level is as yet not available the magnitude of the effect can not be examined.

CHAPTER 3

The CARE Algorithm for the CSCLP

As discussed earlier, the capacitated set covering location problem (CSCLP) has many applications (see also Toregas and ReVelle, 1972; ReVelle, 1991). An objective function for the CSCLP may be formulated using the unified linear model (see Section 2.3). Several heuristic methods that make explicit use of the objective function, such as linear programming techniques and interchange heuristics, may in principle be used to attack the problems. Since the CSCLP is NP-hard it may be difficult to obtain a solution in reasonable time, therefore, these methods are in practice only employed to solve small to moderately sized problems. Many real-world problems, however, fall in the large-scale range, for which the development of efficient heuristics was (and apparently still is) an important area of research (Current and Storbeck, 1988; Daskin, 1995).

For large-scale networks an optimal solution of the CSCLP can (in general) not be found in an acceptable amount of time, because NP-hard problems require exponential algorithms. In this chapter a heuristic method is presented to obtain an approximate solution of the CSCLP for large-scale networks within a reasonable amount of time. Here, large-scale networks are networks with one hundred thousand nodes or more. For practical use it should be possible to obtain an acceptable solution for such networks within an hour of two using an average present-day personal computer.

The heuristic method presented was implemented using standard GIS software, instead of a modern software development environment, for two reasons. The first reason is, that present-day GIS software, such as the Windows NT version of ESRI's ArcInfo, support networks environments. This means that network data structures and a number of network techniques, e.g. an implementation of Dijkstra's algorithm to calculate network distances, are readily

available. In this way ArcInfo becomes a suitable software tool for rapidly developing and implementing the heuristic method, which is the main goal of the project. Reduced flexibility and slower software are part of the bargain. The second reason for using a GIS environment is the fact that the intermediary results and the end results of the calculations can readily be visualised, which has been of great help in getting insight in the behaviour of the algorithm. It appears that in the field of location modelling visualisation of results is not common practice. In the literature visualisations (maps) are only given in rare cases (e.g. Goodchild, 1979).

This chapter continues with a discussion of the node-partitioning algorithm. In the section thereafter, the CARE algorithm for the CSCLP is presented. Next, the main features of the NT ArcInfo implementation of the algorithm are explained. Technical details of the implementation have been incorporated in the comments in the source code (provided on the accompanying disk). The chapter ends with a discussion.

3.1 The node-partitioning algorithm

One of the simplest algorithms to solve the p-median problem is the node-partitioning algorithm (Maranzana, 1964). For networks with about 100 nodes this algorithm yields solutions which are within a few percent of the optimal solution. In fact the accuracy of the solutions increase when the number of nodes in the network increases (see Daskin, 1995, page 238).

The node-partitioning algorithm can also be used to solve the p-centre problem. Instead of determining the median of each subset one should determine the centre. In the same manner as for the p-median problem, it can be shown that the resulting series of coverage distances is non-increasing (see e.g. Hodgart, 1978, section VII). The set covering location problem could be solved by successively solving a series of p-centre problems: starting with one centre and increase the number of centres until the maximum coverage distance is smaller than the maximum coverage distance allowed (cf. Marianov and ReVelle, 1995).

When the node-partitioning algorithm is applied to p centres with a maximum coverage distance, it can readily be shown that the total number of demand points allocated is non-decreasing. The reasoning is as follows. In the first step of the algorithm each demand point within reach of the centres are allocated to the nearest centre. In the second step each centre is moved to the location where the maximum distance to the allocated demand points is minimal. After this repositioning of the centres all demand points already allocated will be allocated again, either to the same centre or to another centre which is nearer after the move. In both cases it will be allocated, since the distance to the nearest centre could not have increased. It may also be the case that demand points that were out of reach are now close enough to be allocated. After any two-step iteration the number of demand points allocated either remains the same or increases.

Intermezzo: allocation of nodes to centres

When the capacities of the centres are unlimited nodes are assigned to the nearest centre. Node-allocation can then be carried out in the following way. First, calculate the distances between the centres and the nodes, and arrange them in tabular form with three columns: node *i*, node *j*, and $D_{i,j}$ (the distance between the two nodes). Sort the table in descending order on the distance, and allocate the nodes as follows. Start with the centre-node pair with the shortest distance, i.e. the pair in the first row. Assign the node to the centre, then remove all rows in which this node occurs. Repeat the procedure until no more rows are left.

When the distance between node and centre is limited, remove all centre-node pairs with a distance exceeding the maximum value. Then use the allocation procedure as described above. In this case a number of nodes may remain unallocated.

When the capacity of the centres is limited, the total demand allocated to each centre has to be recorded during the allocation procedure. A node can only be allocated to a centre when - after the allocation - the capacity of the centre does not exceed the maximum value. Otherwise, the first row is removed without any further action, and the procedure is repeated as before. Also in this case a number of nodes may not be allocated to a centre.

When both the capacities of the centres and the node-centres distances are limited, the latter two procedures are combined. First, remove all rows with distances above the maximum value. Then apply the last procedure to the remaining rows in the table.

The nice properties of the node-partitioning algorithm are lost when the capacity of the centres is limited. When the coverage distance is unlimited and the capacity is limited the maximum coverage distance is no longer non-decreasing. This will be illustrated with the configuration depicted in Figure 3.1, where the numbers along the arcs are the lengths of the arcs, each node has a demand of one unit, and the centres have a capacity of three units. The positions of the centres are indicated by large dots. The centre and the demand point allocated have the same colour. Figure 3.1 (left) represents the situation after allocation of the demand point to the capacitated centres. It is easily seen that the coverage distance is equal to four. Figure 3.1 (right) represents the situation after one two-step iteration. The centres have been moved to their new positions and the demand points have been reallocated. The coverage distance has now increased to the value of five. This counter-example shows that, when the node-partitioning algorithm is applied to p capacitated centres the coverage distance is no longer non-increasing. Therefore, the node-partitioning algorithm can not be used to solve the capacitated p-centre problem. Consequently, the node-partitioning algorithm can not be used to solve the capacitated set covering location problem by successively solving a series of p-centre problems using the node-partitioning algorithm. It should be noted that also the global/regional interchange algorithm (GRIA) can not be used to solve the capacitated p-centre problem, because the regional interchange is a node-partitioning step. Examining the

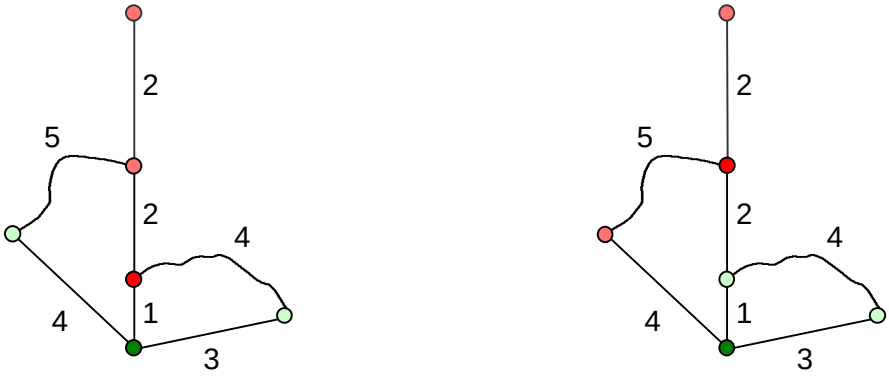

Figure 3.1 Repositioning of centres. The situation left is before respositioning, and the situation right after. The centres have bright colours. The allocated nodes have the same colour as the centre.

situations sketched in Figure 3.1, it is also easy to verify that the node-partitioning algorithm and, therefore, also the GRIA fail to solve the capacitated p-median problem.

The situation concerning the node-partitioning algorithm becomes even worse when both the capacities of the centres and the coverage distances are limited. Again consider the situations represented in Figure 3.1, when the capacities of the centres remain equal to three and additional maximum coverage distance of four is imposed. The initial situation will remain the same, but after the two-step iteration the leftmost demand point will remain unallocated. In this situation coverage distances are non-increasing and the total numbers of allocated demand points are non-decreasing.

The node-partitioning algorithm lost its nice properties because one of the centres reached its maximum capacity. Although these nice properties are lost on a global scale, on a local scale - in areas where the centres have not reached their maximum capacity - these properties will still show up. In fact, these properties will be used in the CARE algorithm to the capacitated set covering location problem, which is described in the next section.

3.2 The CARE algorithm

Although the nice properties of the node-partitioning algorithm are lost when the capacities of the centres are limited two properties may still be used. The number of demand points that have been allocated will in general increase. Also the centres are moved to positions where the coverage distance per partition is minimal, i.e. they are close to the middle of the partition. These two properties may be used in the following three-phase approach the CSCLP.

The CSCLP requires that the centres cover all demand points. Therefore, in the first phase of the algorithm enough starting centres should be generated in order to allocate (nearly) all demand points. In the second phase the fact that the node-partitioning algorithm tends to increase the number of demand points allocated may be used to reposition the centres and to allocate the remaining demand points. In the third phase centres with demand points that can be reallocated to neighbouring centres may be eliminated. Thereafter, the second and third phase should be repeated till no more centres can be removed. For this procedure to work two issues should be addressed. The first one is how to generate (in the first phase) as few starting centres as possible and still be able to allocate (nearly) all demand points. The second one is how to detect (in the third phase) the centres that may be eliminated

3.2.1 Generating initial centres

In principle, centres may be located anywhere on the network. In practice, the demand points are not randomly dispersed, but there is some form of spatial clustering. In some areas the demand density is low and in some areas it is high. In areas with a low density of demand points, the number of centres required will primarily depend on the maximum coverage distance. In areas with a high density of demand, the number of centres will primarily depend on

the capacity of the centres. However, concentrations of centres will be found in and around areas with a high density of demand. To determine a demand density a spatial average over a suitable neighbourhood has to be determined, where suitable refers to both size and shape. Since there is no preference of direction the shape of the neighbourhood should be circle symmetric. The size of the of the neighbourhood should be chosen in such a way that finer details are averaged out, but enough of the global structure is retained. If so desired, the characteristic size of the neighbourhood may be determined with techniques as kernel estimation (see Bailey and Gatrell, 1995, section 3.4).

After calculating the density of demand the peaks of the density surface may be determined. The number of peaks depends on the size of the averaging neighbourhood. As the size of the neighbourhood increases more detail will be averaged out and so the number of peaks will drop. In practice, usually more peaks than required will be found and only fraction of these should be used, e.g. a fixed percentage of the highest peaks. The selection of highest peaks will be used as initial centres to which demand points are allocated. The centres, allocated demand points and corresponding arcs are now removed from the network. The entire procedure of spatial averaging and selection of the highest peaks is now repeated for the remaining part of the network. The entire procedure is repeated as long as the initial centres do not (nearly) cover all demand points.

3.2.2 Elimination of centres

After the initial centres have been determined superfluous centres may be directly eliminated, or after the node-partitioning algorithm has been applied. Two procedures to eliminate centres will be presented. The first elimination procedure may always be used, even directly after determining the initial centres, whereas the second elimination procedure may only be used after the node-partitioning algorithm has been applied, i.e. when the centres are close to the middle of the partitions.

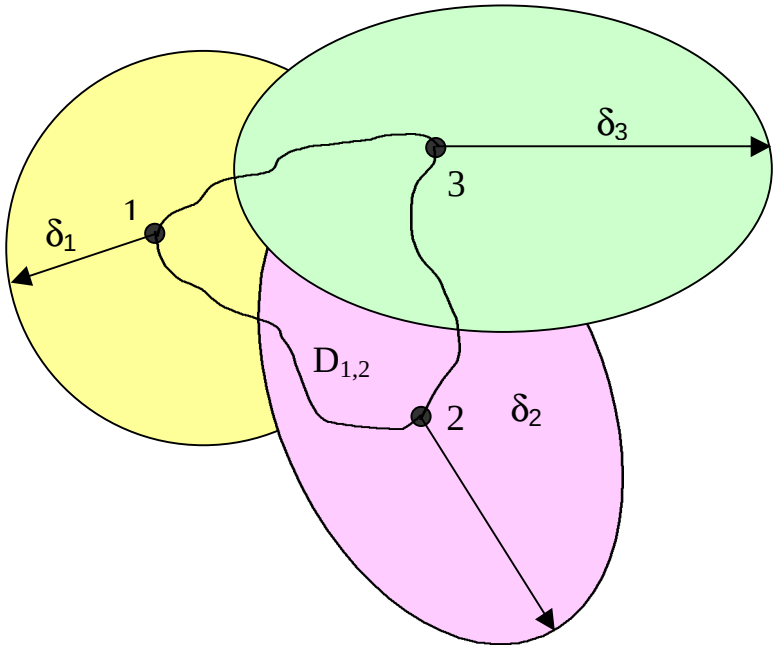

Figure 3.2 Three numbered centres with their partitions. Coverage distances d_i and the shortest path between the first and the second centre have been indicated.

3.2.2.1 General elimination procedure

A centre may be eliminated if all demand points allocated to it can be reallocated to neighbouring centres (adopting centres). All demand points can be reallocated only when two conditions are met. The first condition is that the surplus capacity of the adopting centres is sufficient to incorporate the total demand of the demand points to be reallocated. The second condition is that the distances between the neighbouring centres and the demand points do not exceed the maximum coverage distance. However, the second condition may be relaxed if repositioning of the centres by the node-partitioning algorithm can be taken into account.

To illustrate the first elimination procedure, consider the situation sketched in Figure 3.2, where centre 2 is the one to be eliminated. Assuming that the surplus capacity of the adopting centres (centres 1 and 3) is sufficient, the demand points are reallocated without taking account of the maximum coverage distance allowed. In view of the limited capacity of the

centres, it is not clear which demand point will be allocated to which centre. Therefore, it should be possible to reallocate any demand point to both centres. The maximum distance of these demand points to centre 2 is equal to its coverage distance, denoted by δ_2. One path leading from a reallocated demand point to centre 1 runs via centre 2, but there may be shorter ones. The length of the path via centre 2 does not exceed the sum of the distance between the two centres, denoted by $D_{1,2}$, and the coverage distance of the eliminated centre δ_2, i.e. the maximum distance between a reallocated demand point and centre 1 is equal to $D_{1,2} + \delta_2$. Now consider three separate cases:

- $\delta_1 > D_{1,2} + \delta_2$: all reallocated demand points are within coverage distance δ_1 of centre 1, which does not exceed the maximum coverage distance.

- $\delta_2 > D_{1,2} + \delta_1$: centre 1 may be moved to the position of former centre 2. In fact the roles centres 1 and 2 are interchanged. Now all (re)allocated demand points are within coverage distance δ_2 of centre 1, which again does not exceed the maximum coverage distance.

- In the remaining case the maximum distance between an original and a reallocated demand point (measures via centres 1 and 2) is most $\delta_1 + D_{1,2} + \delta_2$. Now centre 2 may be shifted along the shortest path between centres 1 and 2 to middle of the extremes. At that location the coverage distance is roughly equal to half the maximum distance between two extreme demand points, i.e. $\frac{1}{2}(\delta_1 + D_{1,2} + \delta_2)$. This distance should not exceed the maximum coverage distance.

In general, a centre *j* may be eliminated if the candidate adopting centres *i* in the set

$$S_j = \{\, i \mid d_i + D_{i,j} + d_j \leq 2\, d_{\max} \,\} \qquad \text{(Eq. 3.1)}$$

have enough surplus capacity, which is defined by constraint:

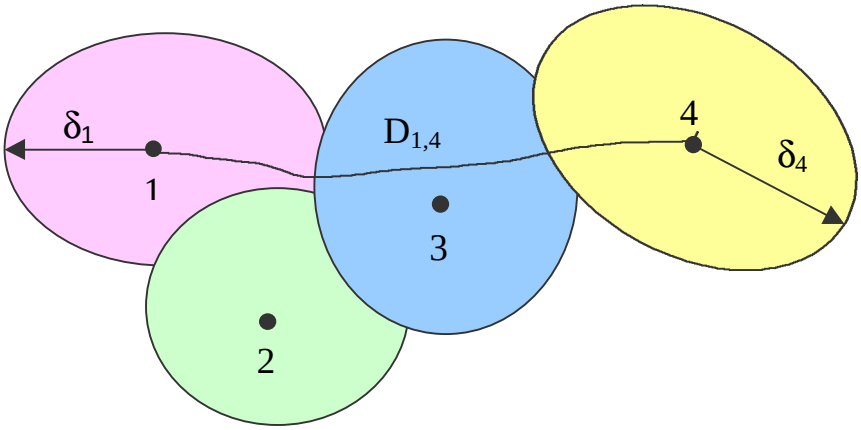

Figure 3.3a Group of four centres, where centre 3 is a candidate for elimination.

$$\sum_{i \in S_j} C_i \geq (N_j - 1)\ C_{\max} \qquad \text{(Eq. 3.2)}$$

The number N_j is the number of elements in the set S_j of eligible centres. Note that centre *j* is also a member of this set.

3.2.2.2 Special elimination procedure

The special elimination procedure may be applied only if the centres involved do not reach their maximum capacity or their maximum coverage distance. These restrictions have to be satisfied both before and after the elimination procedure has been applied to a certain centre. These are imposed to ensure that the node-partitioning algorithm may exhibit the following properties. All demand points will readily be allocated to the adopting centres, the adopting centres will be redistributed such that their coverage distances are roughly equal, and the adopting centres will end up close to the middle of their partitions.

Consider the situation sketched in Figure 3.3a, where one of the four centres may be eliminated, say centre 3. The largest distance between two demand points does not exceed the value of ($\delta_1 + D_{1,4} + \delta_4$). After elimination of centre 3 and repeated application of the node-

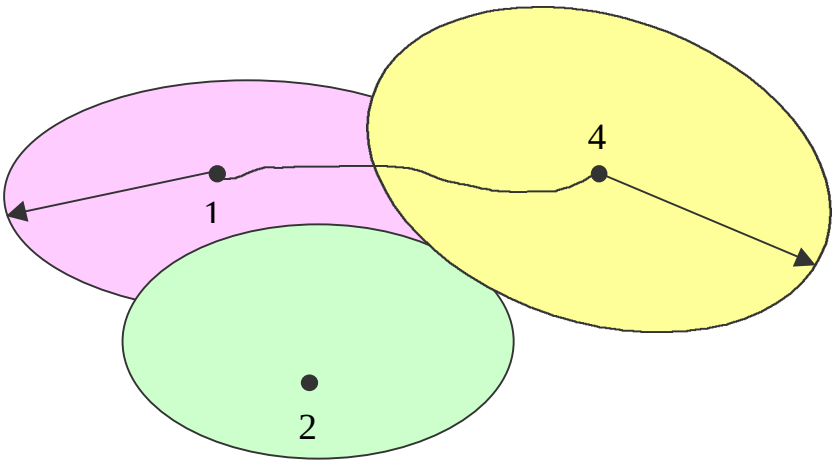

Figure 3.3b Same situation as in Figure 3.3a, but after elimination of centre 3.

partitioning algorithm, the remaining centres will be repositioned (see Fig. 3.3b). The largest distance between two demand points allocated by any of the centres is less than or equal to two times the average diameter of a partition or six times the average coverage distance. In fact the path between the two demand points with the largest distance ran through three partitions, and after the elimination runs through two partitions. The conclusion is that centre 2 may be eliminated when $(\delta_1 + D_{1,4} + \delta_4) \geq 4\ \delta_{max}$ and the surplus capacity of the adopting nodes is sufficient.

In general the number of partitions passed by the path connecting the two demand points with the largest distance should be estimated. As the coverage distance of each partition is roughly the same, the number of partitions passed is roughly equal to largest possible distance divided by the average coverage distance. When one centre is eliminated, each demand point should still be in reach of the repositioned centres.

In practice the neighbourhood of each centre could be scanned for candidate adopting centres. When the neighbourhood of centre *j* is scanned for centres within a distance range of D_{max}, the set of eligible sites S_j is given by

$$S_j = \{\ i \mid D_{i,j} \leq D_{\max}\ ;\ C_i \leq a\ C_{\max}\ ;\ d_i \leq b\ d_{\max}\ \} \qquad \text{(Eq. 3.3)}$$

The factors α and β (both positive numbers smaller than 1) have been included as safety margins. Other centres than the ones in the set S_j do not participate in the elimination process, and their role in the repositioning after the elimination will be virtually absent. In fact, the centres outside set S_j and their partitions act as blocks. Therefore, in calculating any distances between centres in the set S_j, the nodes allocated by centres not in that set should be treated as (road) blocks. As before N_j, is the number of elements in the set S_j of eligible centres. Note that centre *j* is also a member of this set. The surplus capacity of the centres in S_j is sufficient when

$$\sum_{i \in S_j} C_i \geq (N_j - 1)\ C_{\max} \qquad \text{(Eq. 3.4)}$$

The additional constraint derived from the allowed maximum distance between any pair of demand nodes allocated to the centres in S_j is given by

$$\underset{k,l \in S_j}{Max}\ (d_k + D_{k,l} + d_l) \leq\ 2\,(n_j - 1)\,d_{\max} \qquad \text{(Eq. 3.5)}$$

Here, n_j is the estimated number of partitions passed by the longest path:

$$n_j = \underset{k,l \in S_j}{Max}\ (d_k + D_{k,l} + d_l)\ /\ d_{ave} \qquad \text{(Eq. 3.6)}$$

Here, d_{ave} is the average coverage distance of the centres in S_j, given by

$$d_{ave} = \sum_{i \in S_j} d_i\ /\ N_j$$

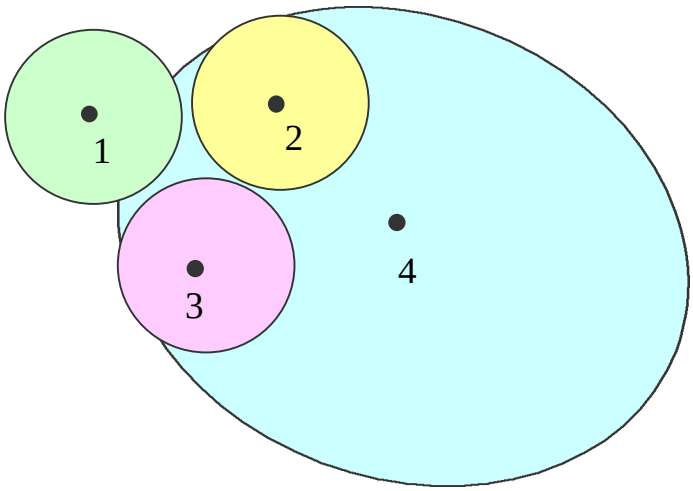

Figure 3.4 Three full-capacity centres. Centre 4 did not reach the maxumum capacity.

3.2.3 Full capacity centres

Centres that have reached their full capacities hardly move when the node-partitioning algorithm is applied. Groups of full capacity centres may be found in areas with high demand density. Often this will lead to a situation as sketched in Figure 3.4a, where slivers between the partitions of the full capacity centres are found. These slivers may be squeezed out when the centres are slightly moved in the direction of higher demand density. Since the centres have not reached by far their maximum coverage distance they will still be able to allocate the maximum demand. The coverage distance of centre 4 will now decrease considerable. After applying the node-partitioning algorithm centre 4 will shift to the lower right. Since its coverage distance has become smaller, centre 4 may take part in more elimination operations. In practice, the (slight) move of full capacity centres in the direction of higher demand density decreases the number of centres required. The direction from the centre of a partition towards the centre of gravity (or median) usually points to the direction of higher demand (see also Morrill and Symons, 1977).

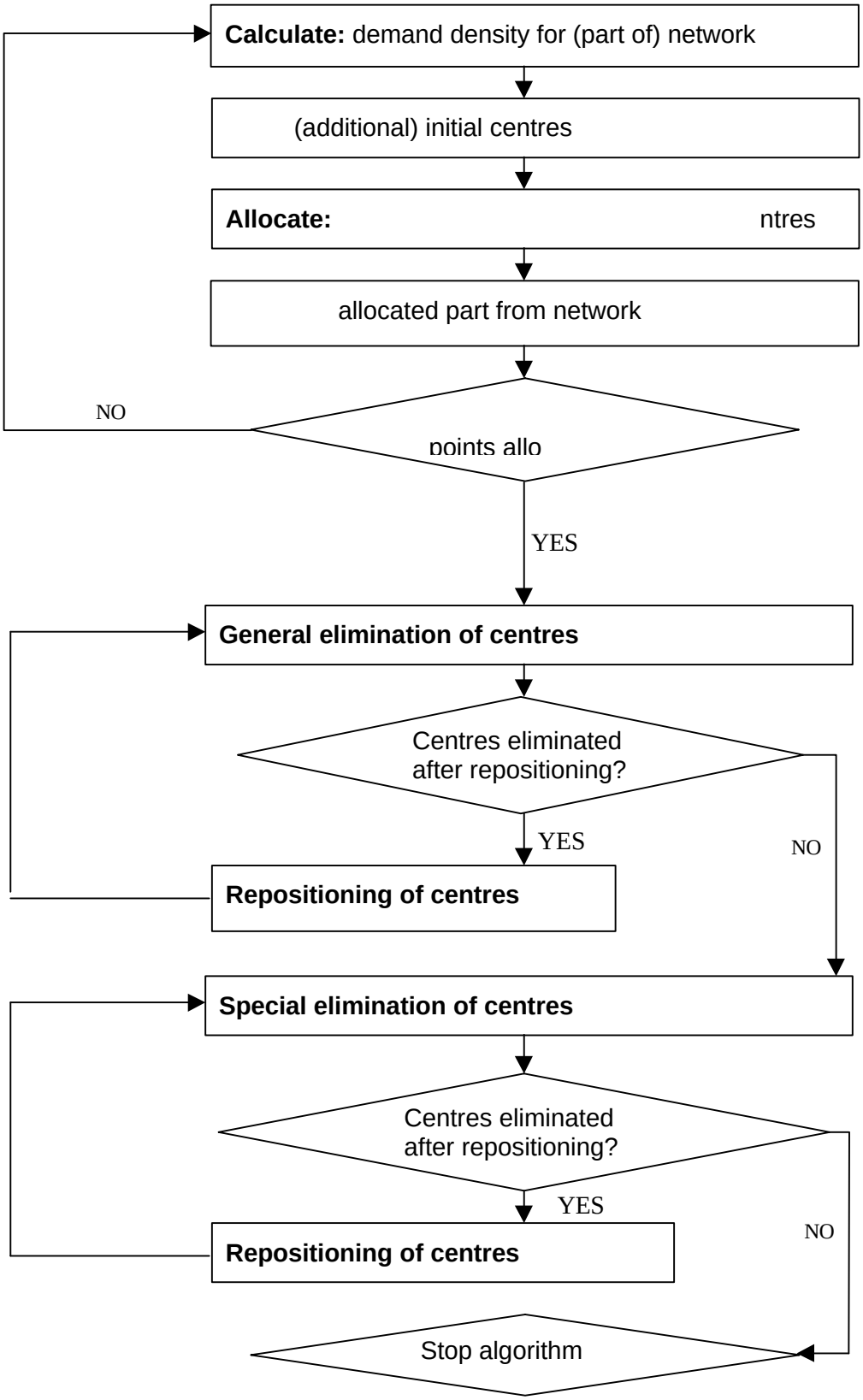

Figure 3.5 Flowchart for the processing in the CARE algorithm

In this section an implementation of the GIS approach using ESRI's ArcInfo Version 7.1 for Windows NT Workstations is described. The NETWORK module of ArcInfo has a set of ro - tines to solve a number of location-allocation models. However, in all models the number of

ALLOCATE for the allocation of demand points to capacitated centres with a maximum cov erage distance (see ESRI, 1992). The routine ALLOCATE will play a central role in the im plementation of the implementation of the GIS approach to the CSCLP.

3.3.1 Allocation of demand

assigned to arcs may be thought off as evenly distributed along the arcs from begin-node to end-node. ALLOCATE has the possibilities to assign part of an arc to one centre and part to

i.e. nodes can not be shared. Both nodes and arcs are assigned to the centres along least-impedance (most efficient) paths. Demand will be assigned as long as the capacity of the

s-

stance.

the demand along the arcs to both end nodes (that are referred to as from-node and to-node in ArcInfo terminology). In view of the application for Statistics Netherlands, and other appl - cations where demand is located at addresses along streets, it will be assumed that demand

located at crossing the demand at nodes will usually be negligible. Location of demand along arcs in stead of concentrating it at nodes has the following advantages:

- In practice road segments (represented as links, also called arcs) as long as five kilometres (with only a few addresses) are found in rural areas. Concentrating demand in nodes may then lead to a considerable reduction in the length of the shortest path to the demand. Since the maximum coverage distance is an important parameter, it should be taken into account as well as possible.

- When demand is located along arcs, part of the demand may be allocated to centres, which means that a centre may easily reach it maximum capacity. When the demand is only located at nodes, and only the entire demand can be assigned to a centre, the maximum capacity of a centre can usually only be approximated. In that case, centres that reached their maximum capacity or maximum distance (“finished centres”) can easily be identified, which facilitates the assessment of the results of the algorithm.

In view of these advantages demand is usually assigned to arcs, although demand located at nodes will be taken into account. The ALLOCATE routine generates a node-allocation table, that records the nodes assigned to each centre. However, no node-allocation table is generated when no demand is assigned to any node. In order to force the generation of that table a demand of one unit is assigned to each node where a centre is located. To compensate for this extra demand, the maximum capacity of the centres is raised with one unit.

3.3.2 Calculation of demand density

The locations of the initial centres are determined by finding the highest peaks in the demand density. Since the initial centres will be used as starting points in an iterative procedure the locations of the initial centres will not have to be determined with very high precision. The issue of the stability of the algorithm will be examined in the next chapter. The calculation of densities requires the spatial averaging of the demand using a “moving window”. Usually grid or raster GIS software provides many types of spatial averages, often used for processing remote sensing data. The spatial averaging routines ArcInfo offers are placed in the GRID

module. Raster or grid GIS software usually provides a number of techniques for spatial a - eraging.

procedures are available for this conversion, for instance the use of overlay techniques. First, the vector equivalent of the grid (called fishnet) and the network with demand along the arcs

Finally, the total demand per cell is converted to a grid. Since high accuracy in determining the initial centres is not required, high accuracy in calculating the demand density is not e - sential. Therefore, a simpler procedure may be used. In the first step, half of the total d -

node. The demand is now located ate the nodes only. Using the routine NODEPOINT the

in the GRID module may be used to aggregate the demand on a cell-by-cell basis and ca -

desired demand de sity grid.

Once the demand density grid is available, the local maxima can be calculated using the h -

cell values by –1 to turn local maxima into sinks (local minima). Since sinks may consist of

r

wise each individual sink cell would be identified as a local minimum. The function ZONA -

of the centroids corresponding to the highest demand density peaks may be used as initial

3.3.4 Scanning the network

Once the first initial centres have been determined, demand is allocated to these centres with the ALLOCATE routine in the ARCPLOT module. Demand is assigned to each centre along shortest paths, gradually increasing the distance from centre to demand, until the maximum capacity and/or maximum distance has been reached. Thereafter, only nodes and arcs that have not been (partially) assigned to initial centres are selected. The network formed by un-assigned nodes and arcs is the basis for the determination of additional initial centres. The procedure described earlier can be applied to the selected part of the network. The runs in which additional initial centres are determined should be repeated until (nearly) all demand has been assigned.

The number of initial centres that should be added in each run is difficult to estimate. When the fraction of local maxima designated as initial centres is increased, more demand may be assigned in each run, and fewer runs are necessary to cover the entire network. However, when designating more local maxima as initial centres may lead to undesired clustering of initial centres, which will have to be thinned out later using the elimination procedures. De-creasing the fraction will reduce the number of initial centres to be thinned out, but it will in-crease the number of runs required to cover the entire network. An appropriate number of local maxima to be designated should be determined by some experimentation, and may result in a suitable rule-of-thumb (see the source code on the accompanying disk).

In practice the search for additional initial centres should be stopped when nearly all demand has been assigned. The adjective “nearly” has been included for two reasons. As mentioned before not all demand has to be assigned, because the node-partioning algorithm tends to increase the total demand allocated. More important, however, is the situation where road networks are used which (should) represent the real-world situation in an accurate manner. In general, a real-world road network is not “a general connected networks”, but rather “a general almost connected network”. In principle, all nodes are connected to the main net-

Wadden Sea and the North Sea) in the Netherlands, but even these can be connected to the main network if regular ferry lines are included.

example, the National Roads File of Rijkswaterstaat, that represents the current complete road network of the Netherlands, as long as it concerns paved roads. An isolated (paved)

o-

e-

would be to track down these areas in advance, and include centres in these areas to cover their demand. However, this is not a practical solution, because the total demand of such an

coverage of the main network. The isolated parts of the network will have to be dealt with sep rately.

3.3.5 General and special elimination procedures

used, including the impedance of arcs and turns, and the specification of the demand at the nodes and total demand along the arcs. The impedance of arcs is represented by user-

be separately specified, and road-blocks in one or both directions may be indicated. It is also possible to specify the impedance of turns in the TRN (turn-table).

specified in a so-called centres table, which has attributes containing the node-number, the capacity of the centre, the maximum centre-demand (point) distance. The centres table also

o-

centres with the NODEDISTANCE routine, which produces a so-called origin-destination matrix (OD-matrix) in tabular form, here called OD-table.

After running the ALLOCATE and NODEDISTANCE routines, the centre tabled and the OD-table, contain enough information to start the elimination of superfluous centres. The general elimination procedure and the special elimination procedure may both be implemented using the TABLES utility of ArcInfo. When the origin-destination table is coupled (after a number of manipulations) on both sides with the centres table, a table $E_{i,j}$ with the following attributes can be constructed: i, j, $D_{i,j}$, C_i, C_j, d_i, and d_j. These attributes are sufficient in order to construct algorithms for the execution of both elimination procedures, which are formulated in a similar form, and can therefore be implemented along similar lines.

The general elimination procedure starts with removal of all rows (records) of table $E_{i,j}$ that do not satisfy the constraint on $(d_i + D_{i,j} + d_j)$, defining the set S_j. Next, the surplus capacity for all centres j are calculated with the STATISTICS routine of the TABLES utility. From the centres with sufficient surplus capacity the one with the largest surplus capacity (say centre k) is selected. Centre k is now removed from the centres table. All occurrences of the centres in the set S_k , or equivalently all centres which participated in the elimination of centre k, are removed from table $E_{i,j}$. The remaining centres occurring in table $E_{i,j}$ have not taken part in an elimination procedure, did not take over demand that was allocated to centre k, and may therefore participate in the elimination of another centre. Thus, the remaining part of table $E_{i,j}$ may now be subjected to the same procedure as before, starting with the calculation of the surplus capacity. The entire procedure may be repeated as long as there are sets S_j with sufficient surplus capacity. In Figure 3.6 a flowchart of the elimination procedure is presented.

The special elimination procedure is carried out in the same way, with one exception. In addition to the constraint on the surplus capacity, an additional constraint on the maximum dis-

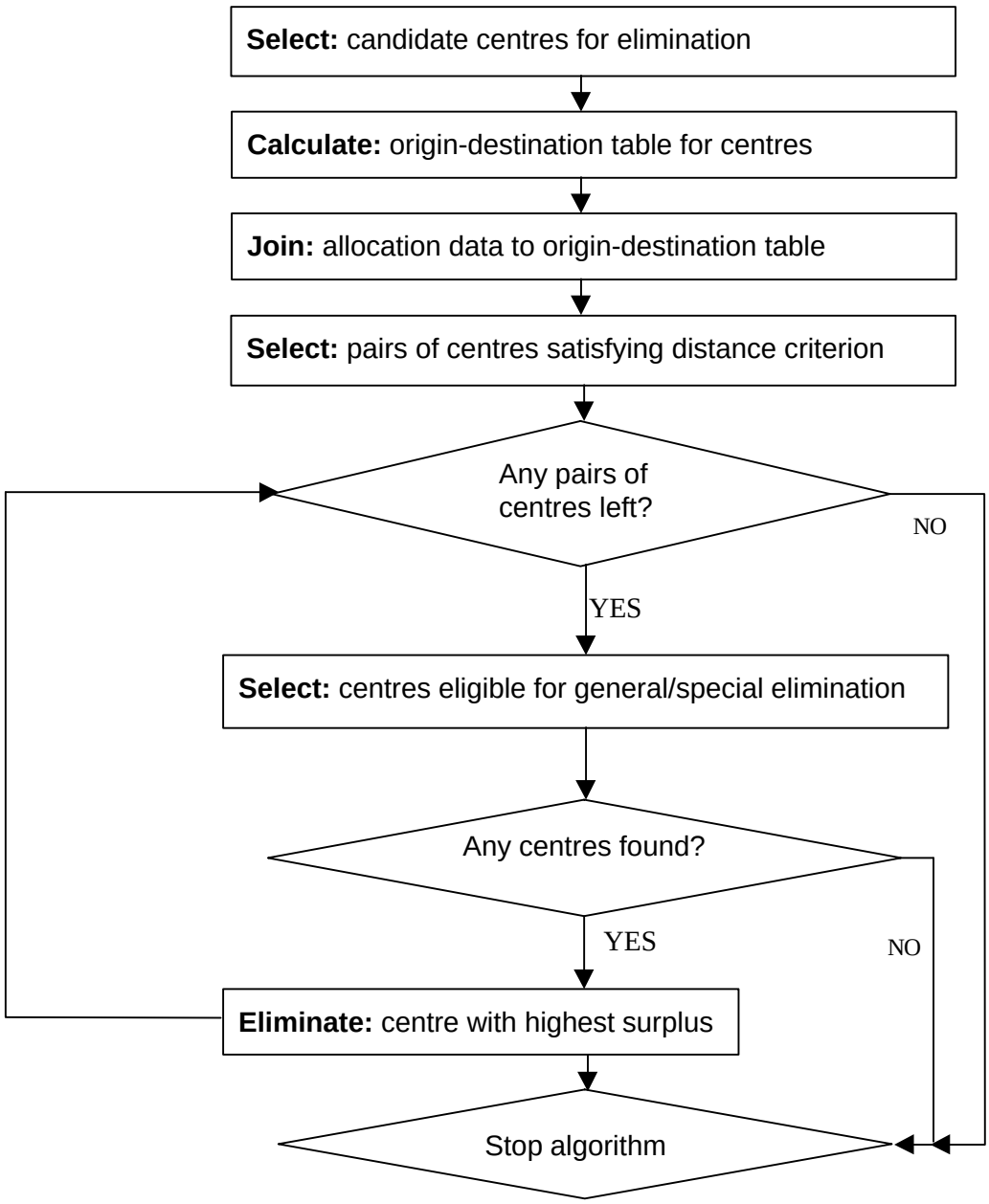

Figure 3.6 Flowchart for general and special elimination procedures

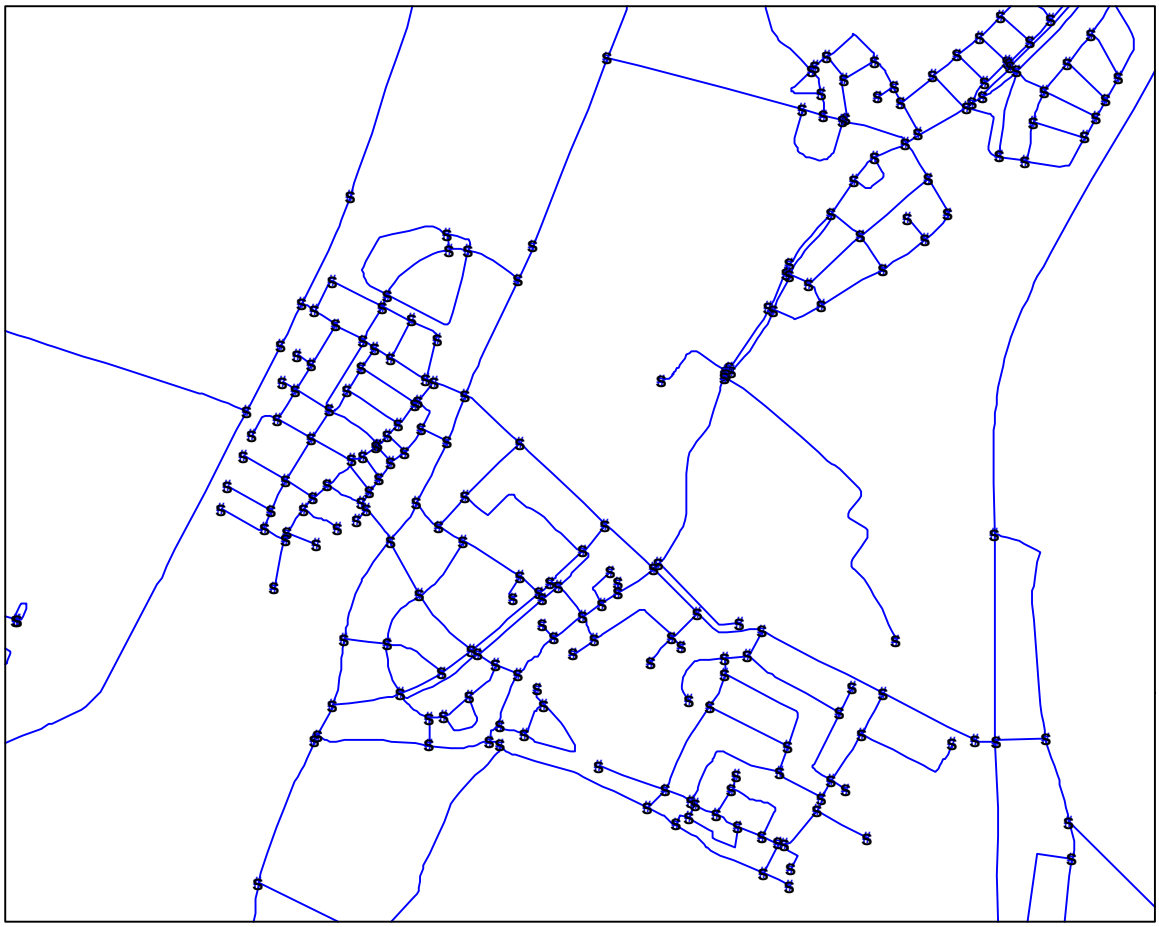

Figure 3. 7 Road segments and nodes in and around the municipality of Bennebroek

tance has to be satisfied. In practice the special elimination is run twice, with D_{max} equal to two and four times the maximum coverage distance. The parameters setting were = 0.8 and = 0.975. The flowchart in Figure 3.6 also applies to the special elimination proc edure.

3.3.6 Node-partitioning algorithm

The node-partitioning algorithm consists of two steps: allocation of nodes to centres and re-positioning of the centres to the middle of the partition. The allocation of nodes is carried outby the ALLOCATE routine. A ready-made routine for repositioning centres is, however,

not available. One way of finding the middle of a partition is by enumeration, which may be very time-consuming. It requires the calculation of the OD-matrix of all nodes in a partition, followed by the selection of the row where the maximum distance is the smallest. Since the OD-matrix is not kept in RAM-memory, but stored in an INFO-table, this becomes a tedious operation. A second way would be to estimate the position of the middle, then select candidate centres in the neighbourhood, and calculate the OD-matrix with the candidate centres as origins and all nodes in the partition as destination. The new centre is again found by searching the row where the maximum distance is the smallest. The computational effort may be reduced by an order of magnitude, but it will still be considerable.

In view of these considerations another more practical approach was used. As discussed in Section 2.4, the solutions set covering problems are not very sensitive to the exact locations of the centres. Therefore, without introducing considerable errors, the exact location of the centre of a partition may be replaced by a fair estimate. The exact centre of a partition will be replaced by the centre of gravity of the part of the region covered by the partition. To define the part of the region covered by a partition the following procedure is used. The study region is divided into Thiessen polygons constructed with the set of nodes of the network as basic points. Thiessen polygons are sometimes called proximal regions, because the points in the region are assigned to the nearest basic points (cf. Laurini and Thompson, 1992, Section 6.6). The situation is illustrated in Figure 3.7 for the municipality of Bennebroek and surroundings. (Bennebroek is the municipality with the smallest surface area.)The area allocated to a centre is an aggregate of the Thiessen polygons belonging to the nodes allocated by the centre. A reasonable estimate of the co-ordinates of the centre of gravity of the allocated area is given by the weighted averages of the co-ordinates of the nodes, where the weight of a node is equal to the area of its Thiessen polygon. Further, note that the areas of the Thiessen polygons are roughly inversely proportional to the node density. (When the number of nodes is doubled, the average area of the Thiessen polygons is halved.) There-

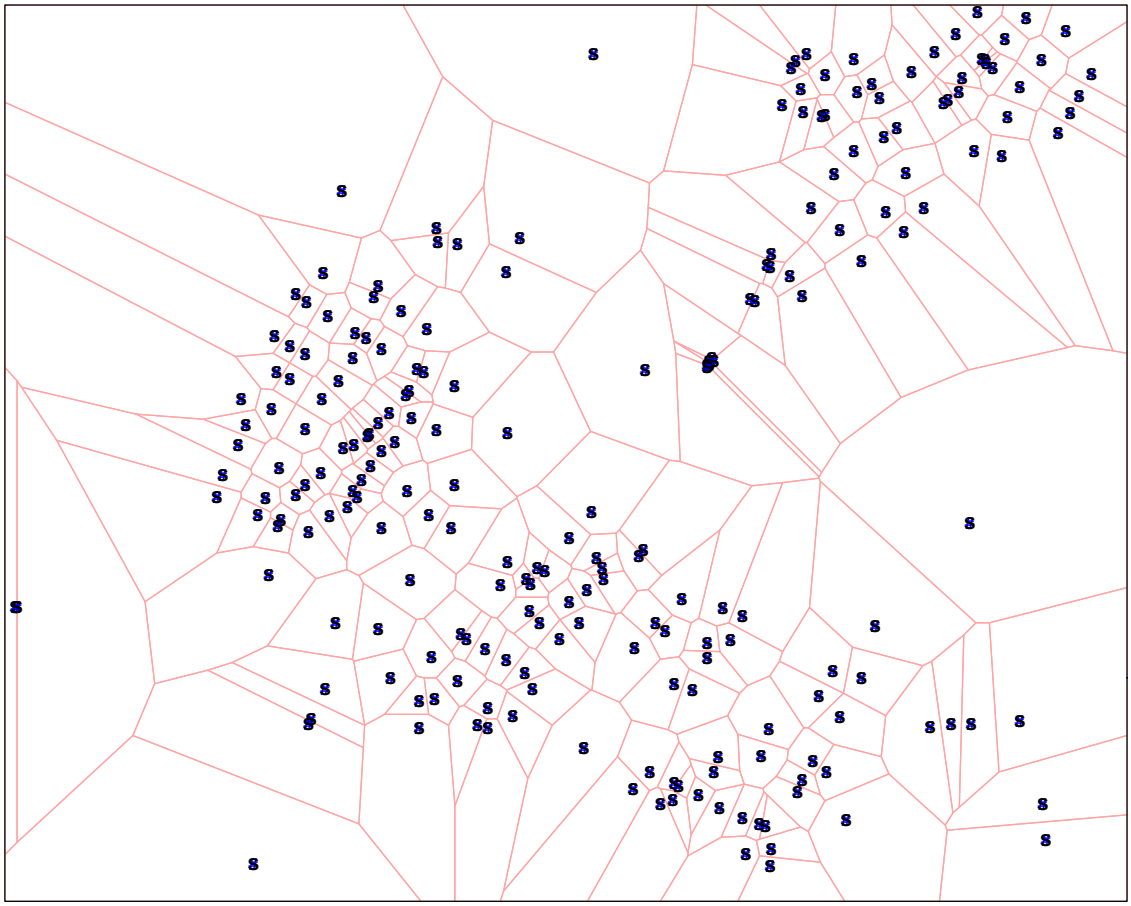

Figure 3.8 Nodes (blue dots) and Thiessen polygons (light red boundaries) in and around Bennebroek

fore, the weights of the nodes may be approximated (up to a constant factor) by the inverse of the node density.

For practical applications in The Netherlands, where characteristic variations in the node density occurs on a scale of about 1 kilometre, the node density is calculated for grid cells of 500 metres by 500 metres. The averaging neighbourhood used is a circle with a radius of 1 kilometre. First, the nodes are converted to a point coverage using the NODEPOINT routine. Next, the node density is calculated using the POINTSTATS function of the GRID module. The density grid is converted to a polygon coverage using the GRIDPOLY function, which is

overlain with the point coverage to assign node density values to each point. Finally the node density is transferred from the points to the corresponding nodes using the POINTNODE routine. Since the ALLOCATE routine is able to generate a node allocation table, the nodes belonging to a centre are easily retrieved and so the co-ordinates of the centres of gravity can be calculated in a fairly straightforward manner.

The co-ordinates of a full-capacity centre should be slightly shifted in the direction of the centre of gravity of the demand. Since the aggregated demand per node may vary by orders of magnitude, the biquadratic root (the square root of the square root) of the demand is used to calculate demand centre of gravity. The full-capacity centre is positioned at the weighted average of area centre and demand centre, with weight factors of 0.975 and 0.025, respectively.

3.3 Hardware, software, and data

The hardware used in this study was a personal computer with an Intel Celeron 400 MHz processor, 128 Mb RAM with a free disk space of 2 Gb. The operating system of the computer was MS-Windows NT Version 4.0. The GIS software used was NT ArcInfo Version 7.1.2 and ArcView 3.1 of ESRI Inc. The digital road database for The Netherlands that was used was the National Roads File (NWB) of Rijkswaterstaat, the Department of Public Works of the Ministry of Traffic and Transport (see AVV, 1999). The aggregated data from the Geografic Base File (GBR) was generously provided by Statistics Netherlands .

CHAPTER 4

Application of CARE Algorithm to Model Networks

In this chapter the CARE algorithm for the CSCLP is applied to two model networks. The first one is a fishnet with a mesh width of 500 metres and an arc demand that can be varied by a set of three parameters. Since the peculiarities of real-world networks can hardly be surpassed by simple models, the second model network is a part of the National Roads File of Rijkswaterstaat, the Department of Public Works of the Ministry of Traffic and Transportation. The arc demand for this model is equal to the number dwellings along the road segment represented by that arc. The latter model network is used to illustrate the intermediary and final results of the CARE algorithm. This model is also used to study the stability of the algorithm, and the effect of replacing arc demand by node demand. The chapter ends with conclusions.

4.1 The fishnet model network

The fishnet model network has a mesh width of 500 metres, which is a characteristic distance over which the population density varies in The Netherlands. The fishnet model network is a square with 100 rows and 100 columns, which means that the network has 10201 nodes and 10000 arcs. The middle of the network is placed at the origin of the co-ordinate system. The demand w_i at a node i is given by the following formula:

$$w_i = N \Big/ \left(1 + a\,[(x_i - b)^2 + (y_i - c)^2]\right) \qquad \text{(Eq. 4.1)}$$

Here, x_i and y_i are the co-ordinates of the node. The constants b and c are used to shift the peak of the demand distribution. The factor a determines the height of the peak in the demand distribution. Finally, N is a normalisation factor which is determines the total demand.

4.2 The Bommelerwaard network

The second model network represents the road network of the Bommelerwaard, a fairly large rather isolated region the province of Gelderland, between the major rivers Maas and Waal. Only a few bridges and ferries connect this area to other parts of the Netherlands. The road network is a small part of the December 1999 version of NWB. NWB is the part of the National Roads File of Rijkswaterstaat, which represents (in the very near future) all paved roads in The Netherlands. The roads of the Bommelerwaard are depicted in Figure 4.1.

4.3 Estimation of number of addresses

When the NWB is used for location-allocation modelling, the demand is distributed along the arcs. The total demand of an arc is equal to the number of dwellings along the corresponding road-segment. Road-segments in the NWB have among their attributes the numbers of the houses at the left side and at the right side at both begin and end of the road-segment. This information is not sufficient to calculate (or estimate) the number of dwellings, for instance, because the series of house numbers may not be complete and some house numbers may refer to office buildings. The required number of dwellings could be obtained by combining the NWB with the Geographical Base File (GBF), which contains all addresses in The Netherlands (where mail is delivered by PTT Post) with several attributes, such as a premises code indicating type and use of the building. Unfortunately, the house number ranges in the NWB are not yet up-to-date,

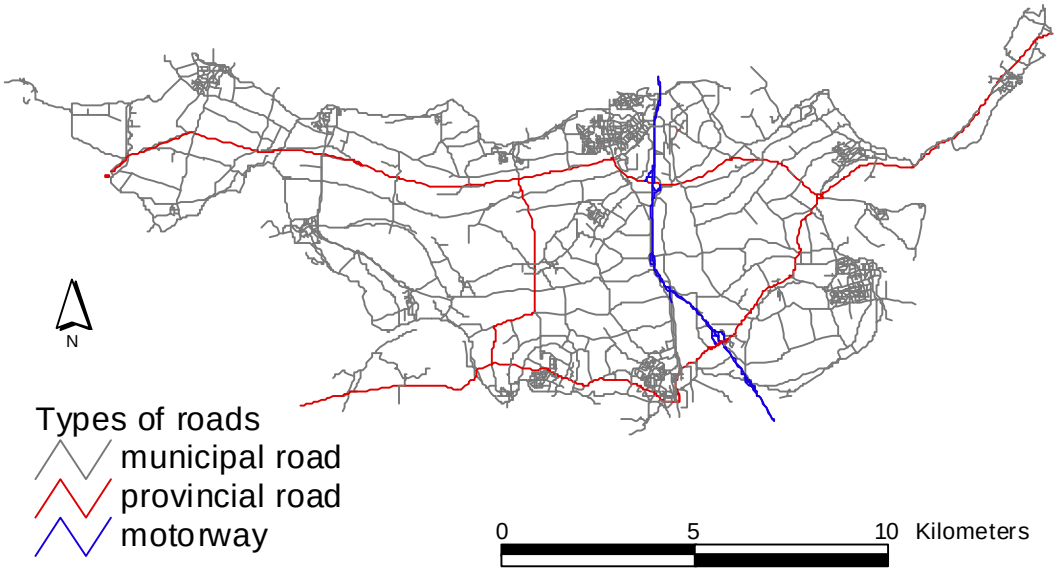

Figure 4.1 Road network of the Bommelerwaard

so that not all addresses in the GBF can be matched to a road segment. However, a fair estimate of the number of dwellings along each road-segment can be made in the following manner.

All addresses in the GBF are geo-referenced via their grid square code and their gemeente-wijk-buurt code (GWB code). The grid square code indicates to which grid square an address belongs, where The Netherlands are assumed to be divided in grid cells of 500 metres by 500 metres according to the rectangular co-ordinate system of the Rijksdriehoeksmeting (see e.g. Ormeling and Kraak, 1993, page 105). Further, the GWB code indicates the municipality, census district (in Dutch: wijk) and census tract (in Dutch: buurt) to which an address belongs. (Municipalities are divided into census districts, which in their turm are divided into census tracts.) When grid square code and GWB code are combined, the position of an address can be estimated, if

Figure 4.2 Roads, grid quares, municipal boundaries and census tract boundaries in and around Bennebroek

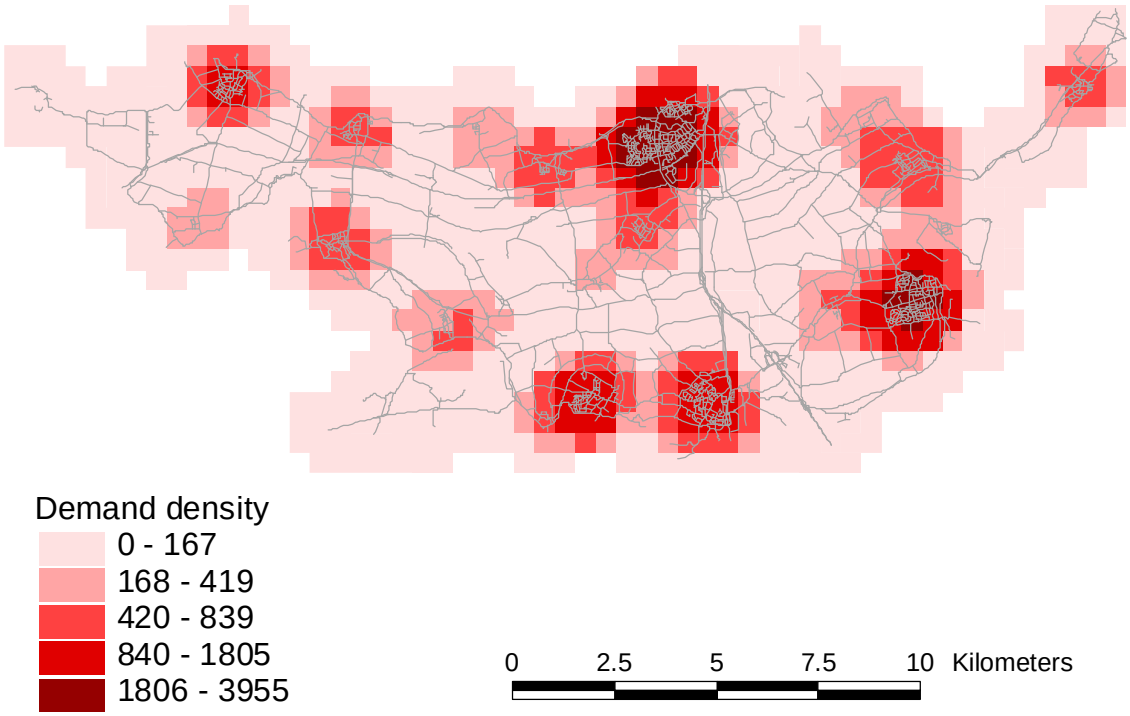

Figure 4.3 Demand density in the Bommelerwaard

the boundaries of the grid squares and the census tracts are known. The boundaries of the grid squares can easily be constructed. The To pografische Dienst (TDN) - the Dutch national mapping agency - maintains digital boundary files of the municipalities. Statistics Netherlands maintains digital boundary files of the census districts and the census tracts of all municipalities. Starting with the edition of 1997, the digital boundaries of municipalities, census districts and census tracts are combined in the CBS/TOPgrenzen files, which are co-produced by Statistics Netherlands and the TDN. The result of an overlay of the grid square boundary file and the CBS/TOPgrenzen file produces a small areas file (SAF). The small areas are referenced by the combination of grid square code and GWB code.

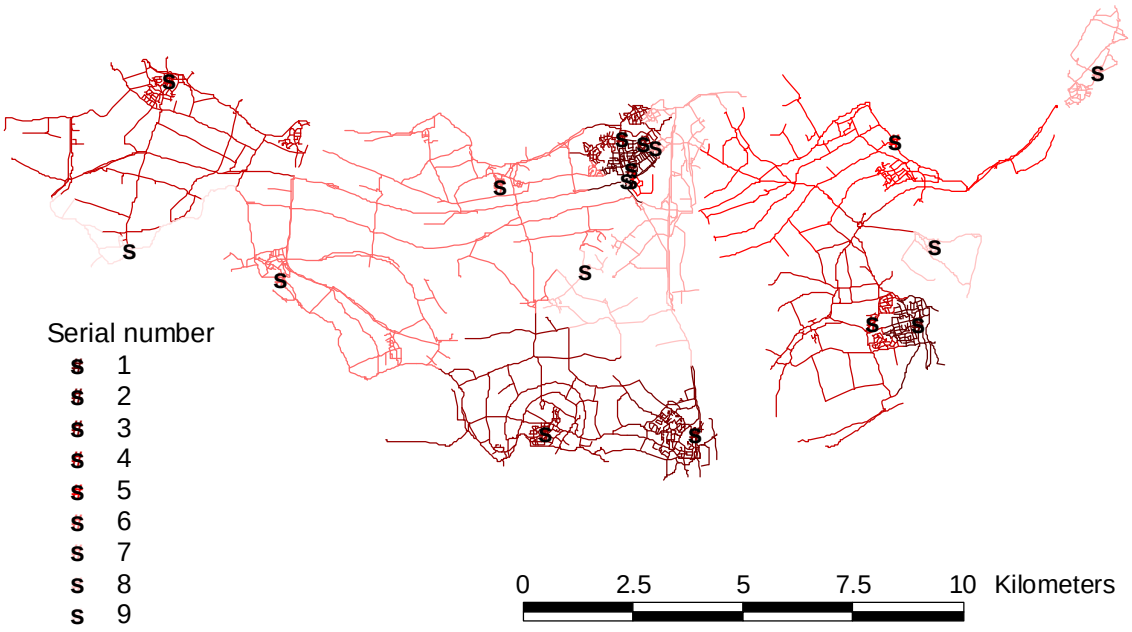

Figure 4.4 Initial centres and partitions in the Bommelerwaard. The serial number refers to loop in which the initial centres were selected

An estimate of the number of dwellings can now be made as follows. The NWB contains roads maintained by several authorities, e.g. municipalities and water boards. With very few excep tions, dwellings are only found along municipal roads. Now assume that the dwelling density along the municipal roads within each small area is constant. First, calculate the number of dwellings for each combination of grid square code and GWB code using the GBF. Next, over-laythe NWB and the SAF (see Figure 4.2 for the result for the surroundings of Bennebroek), and calculate the total length of the municipal roads per small area. With the dwelling density along municipal roads known, the number of dwellings along each municipal road-segment in a small area is readily calculated. Finally, the estimated numbers of dwellings should be aggregated by road-segment, because road-segments may be in one or more small areas.

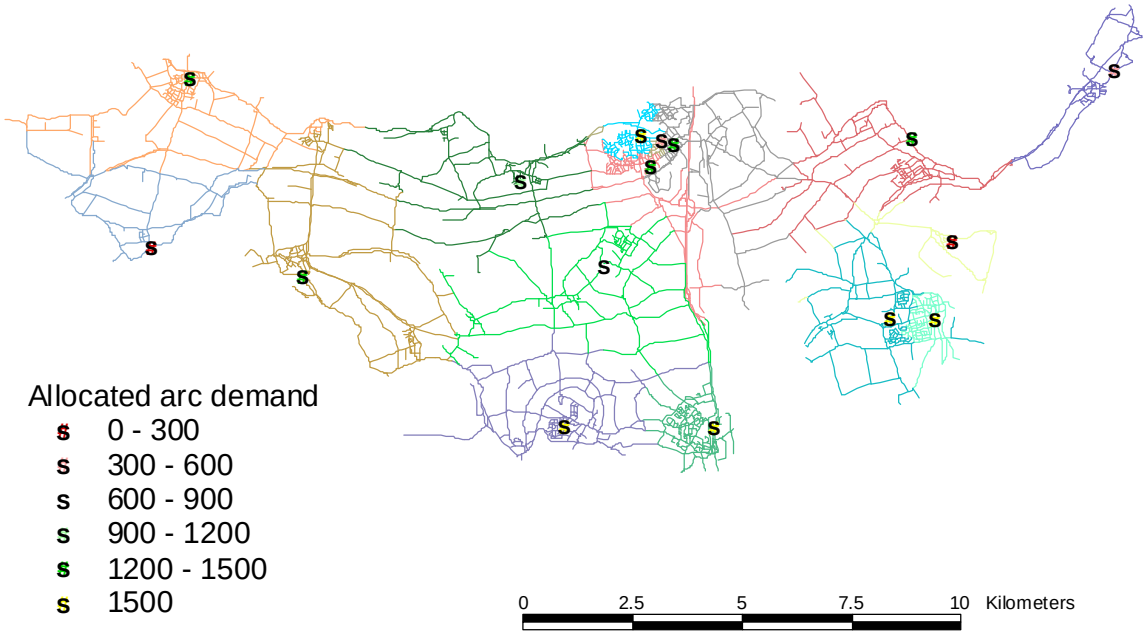

Figure 4.5 Centres and partitions after two general eliminations

The NWB version of December 1999 was used for the model calculations. Unfortunately, the quality of the assigned GWB code is as yet below the standard, and only the grid square codes could be used in estimating the number of dwellings along each road segment.

4.4 Results in the Bommelerwaard

The successive steps in the algorithm will now be illustrated for the Bommelerwaard network, which is large enough to be representative and small enough to show relevant details. The first step is the straightforward calculation of the demand density for the grid with cells of 500 metres by 500 metres. The result of this calculation is shown in Figure 4.3, where the peak in the upper centre reflects the position of Zaltbommel, after which the Bommelerwaard is named. The initial

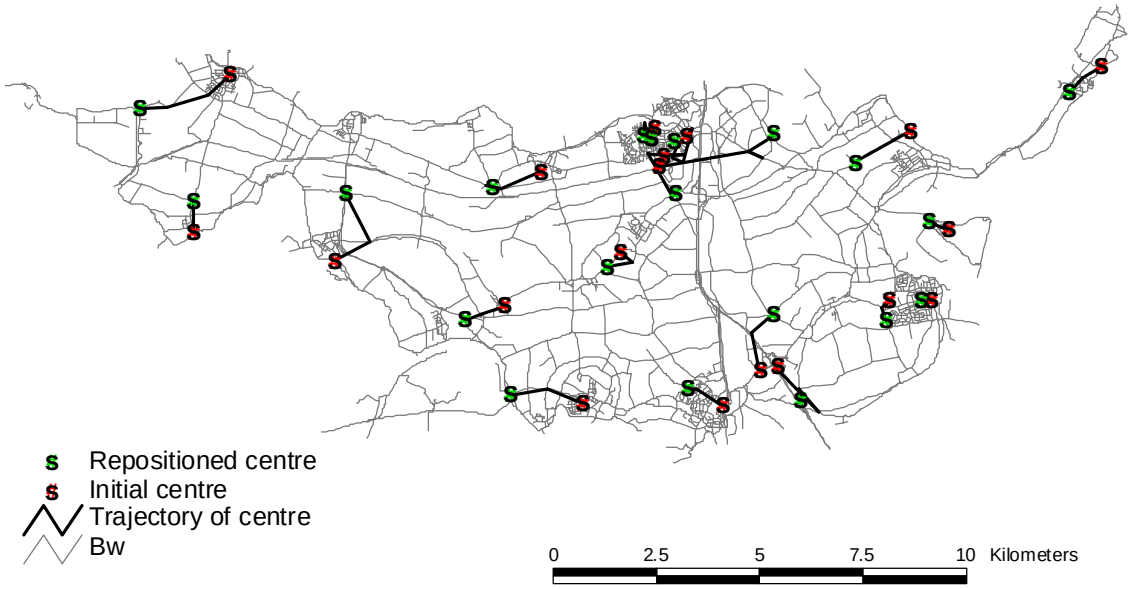

Figure 4.6 Trajectories of centres during the repositioning

centres and their partitions are shown in Figure 4.4, where the areas covered in the successive runs have been indicated in separate colours. The maximum coverage distance was set to 5 km and the demand capacity of the centres to 1500 addresses. Its is obvious that the first centres appear near demand concentrations, and that gradually the entire Bommelerwaard is covered.

When all demand has been allocated, the general elimination procedure is invoked to eliminate superfluous centres. The reduced configuration of centres is shown in Figure 4.5, which shows that the centres are not always in the middle of their partition. The centres may be repositioned by invoking the node-partitioning algorithm, which is repeated until the shift of the centres is smaller than 1oo metres. In Figure 4.6 the trajectories of the centres, showing the successive positions of the centres, and the final positions of the centres and their partitions are shown. In

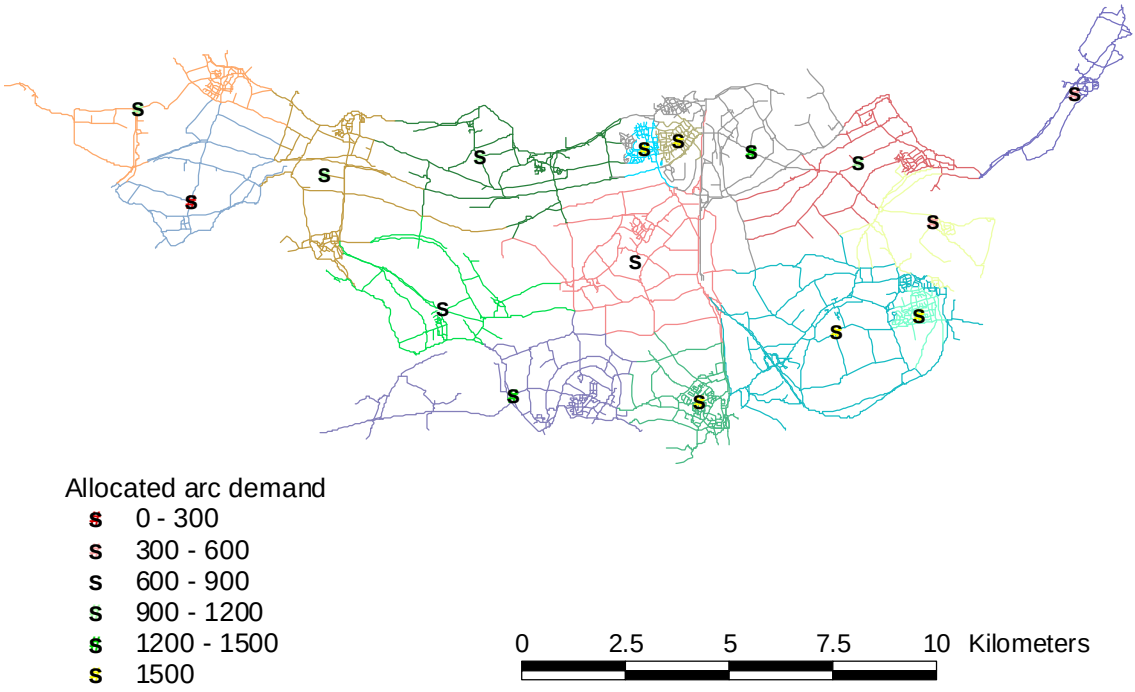

Figure 4.7 End configuration of CARE algorithm in the Bommelerwaard

the rural province of Friesland the centres may even travel distances of 10 kilometres or more. In Figure 4.7 the end result of the algorithm (repeated application of the elimination procedures and the node-partitioning algorithm) is presented.

4.5 Fishnet model network results

The fishnet network has several symmetries (rotation and reflection symmetries), which may be used to investigate the properties of the CARE algorithm. If the demand distribution has similar symmetries, some of the symmetries should be reflected in the distribution of the centres and in the corresponding partitions. The simplest case is the one where demand is only at the nodes and each node has the same demand. In Figure 4.8 the centres and partitions are presented for

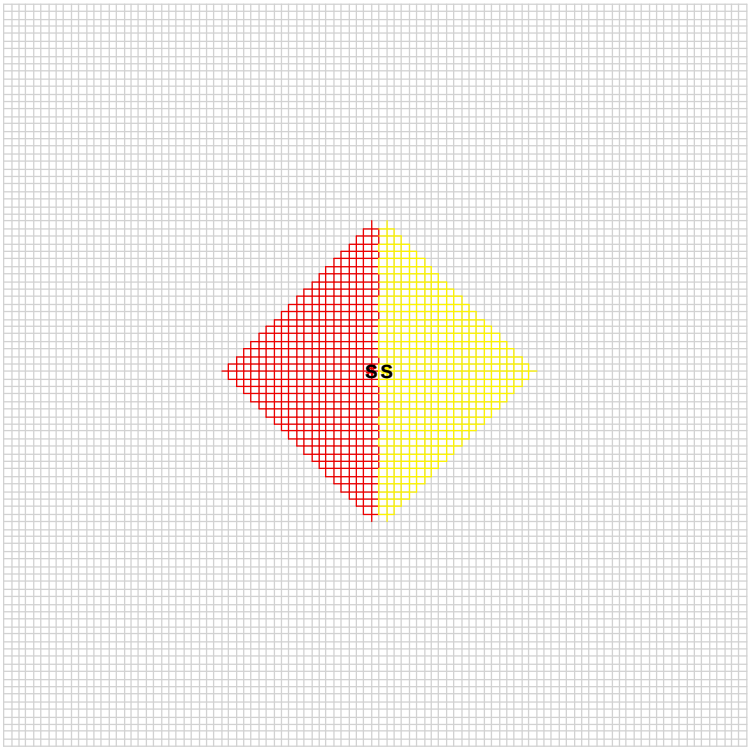

Figure 4.8 Allocation of nodes by two centres for the fishnet model network

two close centres. The capacity of the centres is unlimited, and the maximum coverage distance is 10 kilometres. The resulting partitions should in principle be symmetric. In fact this is not the case, and also the allocated demand slightly differs. To explain these deviations a precise description of the allocation algorithm used in the ALLOCATE routine is required. Unfortunately, only a global description is available (see ESRI, 1992). The limits of the ALLOCATE routine will influence the results of the CARE algorithm, but probably these will only show up in the details.

The same fishnet model can also be used to demonstrate the behaviour of the centres under the node-partitioning algorithm. Figure 4.9 shows the trajectories of two centres, and also shows the partitions at the end of the repeated application of the node-partitioning algorithm. As long as the

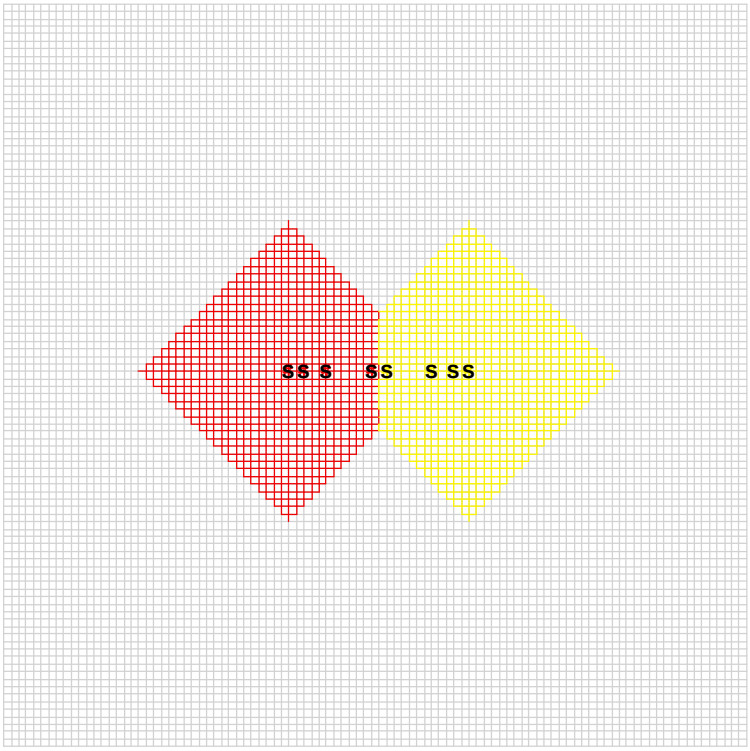

Figure 4.9 Trajectories of the centres (see Figure 4.8) during the repositioning and their final partitions (same color as centre)

centres are not in the middle of their partitions, they will steadily drift apart by the repositioning. In fact, the drifting apart of two centres will always occur when the centres compete for demand points. When two or more centres are close together the drifting apart may be viewed upon as mutual repulsion. With the repulsion view in mind the trajectories of the centres in Figure 4.6 are more easily interpreted.

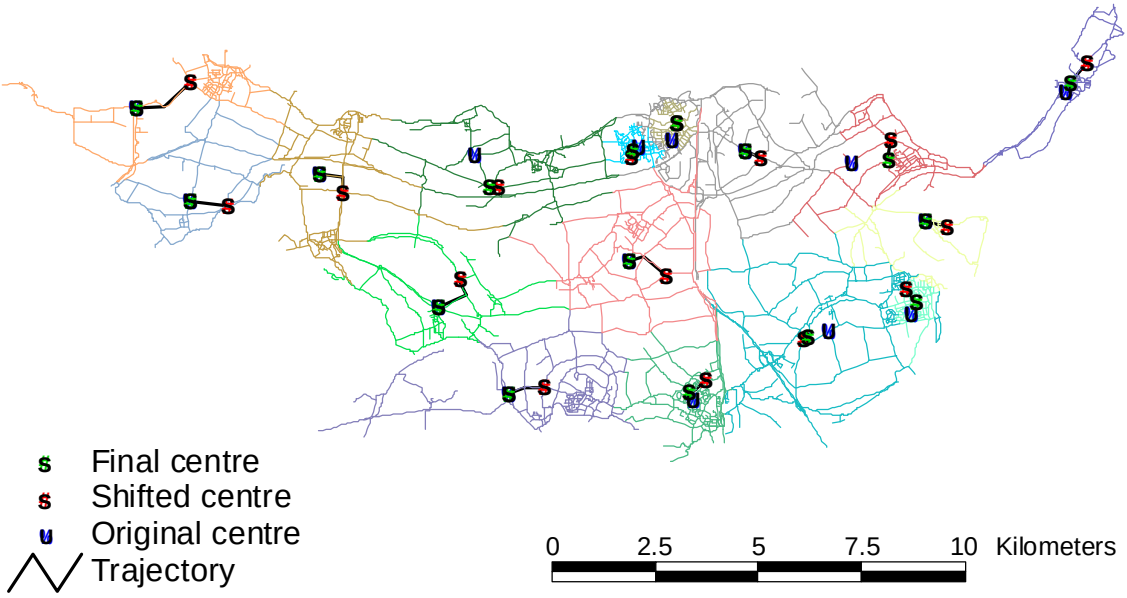

Figure 4.10 Trajectories of shifted centres moving to their final positions during repositioning

4.6 Stability of the algorithm

In addition to providing good solution in an acceptable (computation) time with reasonable storage requirements, heuristics algorithms should also produce stable solutions. In general, the SCLP has more than one optimal solution, i.e. several configurations of centres may yield a complete coverage. Therefore, the solutions produced by the CARE algorithm need not be stable, i.e. a slightly modified configuration of centres may also be a solution. Therefore, application of repositioning (repeated node-partitioning) to a slightly modified configuration need not lead to a return to the original configuration. Figure 4.10 shows the trajectories of slightly moved centres to their final positions in the Bommelerwaard. In areas without full-capacity centres, the centres tend to return to their original position. The full-capacity centres are less mobile, as was discussed in Section 3.2.3, and do not return to their original position, and the resulting distortion of the original configuration propagates in the surroundings.

Although the initial configuration of centres is not restored, the centres move to position close to the initial ones. In the stability calculations the repositioning was stopped when the maximum displacement of the centres, measured as Euclidean distance, was less than 100 metres. The distances between original and final positions may be considerably higher than 100 metres, and are therefore not caused by a premature stop of the repositioning procedure. It should be noted, however, that in the agricultural areas (outside towns and villages) the distances between the adjacent nodes may be considerably higher than 100 metres, which means that centres may have to jump a considerable distance to change position. This threshold could prevent a centre to return to its original position, but this will only be the case in exceptional circumstances.

4.7 Arc demand versus node demand

To examine the effect of demand aggregation results of the CARE algorithm for demand aggregated at the nodes, and demand distributed along the arcs will be compared. In Figure 4.11 the resulting centres and partitions in the Bommelerwaard are shown, when the demand is aggregated at the nodes. This configuration is to be compared with the corresponding results for demand distributed along the arcs shown in Figure 4.7. The global distribution of centres in both cases is similar, and again local differences are remarkable. These local differences may have several causes. As discussed in Section 3.3.1 the allocation algorithm differs considerably. This may cause differences in the locations, and even number, of initial centres. It also causes differences in the nodes that are allocated to the centres. Since there is generally no unique optimal configuration, these small differences may result in a different configuration of centres. The results for the Bommelerwaard, and the results for the province of Friesland shown in Figures 5.6 and 5.7, indicate that the differences in the arc demand and node demand configurations of centres are within the margins allowed by the non-uniqueness of the optimal configuration.

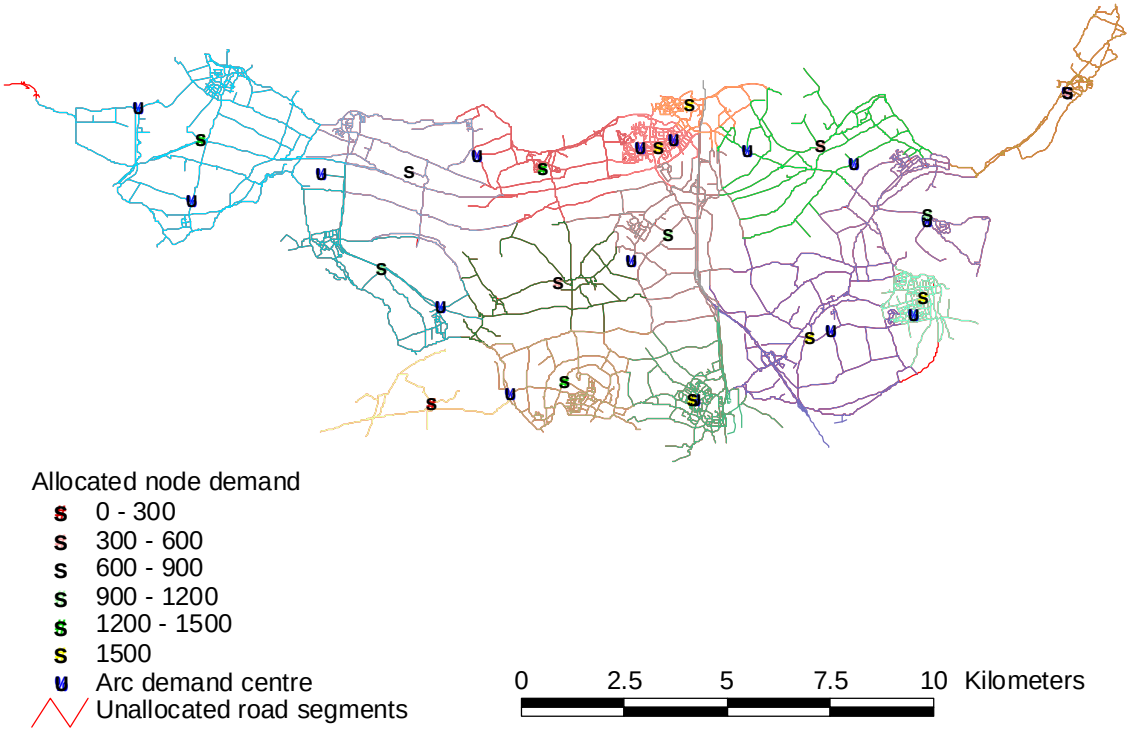

Figure 4.11 End configuration of the CARE algorithm (based on node demand) in the Bommelerwaard

4.8 Discussion

The results thus far obtained demonstrate that the CARE algorithm produces, from a practical point of view, good centre configurations. Whether these configurations are (close to) optimal can not be decided, because exact algorithms require exponential-time algorithms. Even a comparison with GRIA results was not feasible, because the computation times were excessive, even for the fishnet model. The CARE algorithm, on the other hand, produced the fishnet model results in a few minutes, and the Bommelerwaard results in about 20 minutes. Where it should

be noted that about the half of this time was spent in writing (intermediary) results for visualisation to disk.

The configurations of centres for node demand and arc demand calculations are different, but the differences appear to be well within the margins allowed by the non-uniqueness of optimal solutions. Arc demand calculations more easily produce complete coverage, while node demand calculations sometimes miss a few nodes in the distant corners. Since arc demand is continuously distributed is probably easier to detect distant demand. In general, however, the quality of the results for arc demand and node demand calculations is comparable.

CHAPTER 5

Application of CARE Algorithm to Interviewer Selection

Statistics Netherlands conducts many surveys by personal interviewing of persons or households. The Department of Household Surveys, responsible for conducting these surveys, manages a group of interviewers of about 600 persons. For a number of years it is felt that the group of interviewers is undersized. Many interviewers complain that their workload is too high and many addresses in the sample that were assigned to interviewers were returned unvisited. Some interviewers, however, complain that their workload is too low.(Vosmer, 1998). This paradox can only be solved with a tool to estimate the required number of interviewers and their optimal locations. In fact, the absence of such a tool within Statistics Netherlands was the reason that the present research project was initiated.

In the next section the problem is sketched and the translation of the problem into a specific CSCLP given. Next, representative regions for the application of the CARE algorithm are also selected. The algorithm is applied to the study regions to examine boundary effects, and the influence of variations in maximum cover distance and capacity of the centres. Further, the costs and benefits for the present configuration of interviewers and the configuration produced by the CARE algorithm are compared. A discussion ends the chapter.

5.1 Problem description

To determine the number of interviewers required and their optimal locations two parameters have to be known. One is the maximum coverage distance, which is fixed at 20 kilometres. The other parameter is the monthly workload of an interviewer, which is determined by the number of addresses to be visited and by the local response rate. All addresses assigned to an interviewer have to be visited. A non-response visit lasts only a few minutes, whereas a

successful visit may last up to 45 minutes. The response rate is, therefore, an important factor in determining the monthly number of addresses to be assigned to an interviewer. The average contribution of an address to the workload can be estimated if the three relevant factors are known. The factors are response rate, the average duration of an interview, and the sampling fraction are known. It should be realised that all three factors are may depend on the specific region. In addition to the response rate, also the types of surveys may be different in different regions, and so the average interview duration, but also the sampling fraction may vary, e.g. when some regions are over-sampled.

The response rate may depend on a number of regional and socio-economic factors. The response rate in rural areas, for instance, is usually higher than the response rate in urban areas. The response rate may, however, also depend on the workload of interviewers. If the workload is too high not all addresses can be visited, or revisited. If the workload is too low, the number of interviewers is higher than required. Since interviewers are scarce, determination of the optimum workload of interviewers is of very important. At the moment the determination of the optimum workload for interviewers and the analysis of models to estimate the response rates (on the basis of socio-economic and regional parameters), are important research topics at Statistics Netherlands.

The problem of determining number and locations of interviewers can be cast into a CSCLP in the following way. The demand at an address is equal to the average workload multiplied by the (regional) sampling fraction, i.e. the fraction of all addresses to be visited, is equal to the demand. The optimum workload for interviewers is equivalent to the maximum capacity of (service) centres. Once the optimal workload and the response model are available, the CARE algorithm can be used to solve the problem. Since a model to estimate the response rate is not yet available, regional differences in the response rate are ignored. In practice, variations in the response rate of about a factor of 1.2 around its average value are found. Since the optimum workload is not yet available, the workload of an interviewer is assumed

to be equivalent to a work area of 15 000 addresses. This workload is based on the average number of addresses presently assigned to an interviewer and on the average sampling fraction, which is about 3½ percent.

5.2 Representative study regions

Before application of the CARE algorithm to the entire network of The Netherlands, if possible, an application to two representative regions seemed appropriate. One of these regions should be a typical urban region and the other one a typical rural region. In consultation with the Department of Household Surveys the regions chosen are the municipality of Rotterdam (urban region) and the province of Friesland. Because the shape of the municipality of Rotterdam is rather elongated, and also because many interviewers live at a considerable distance of Rotterdam, it was decided to include the municipalities in the neighbourhood of Rotterdam to get a better representation of the situation. Finally, the relevant region nearly took up the entire province of Zuid-Holland. Then, it was decided that the representative urban region should be the entire province of Zuid-Holland, which also contains the urban region formed by The Hague and its neighbouring municipalities. The rural region was in the course of the study expanded to include the largely rural provinces Groningen, Friesland, Drenthe, and Overijssel. The networks of the study regions then have 95 000 nodes and 134 000 arcs, and 120 000 nodes and 165 000 arcs, respectively.

In both study regions many topographical obstacles are found, which is characteristic for The Netherlands. Major lakes are found in the province of Friesland, and in the north-west part of the province of Overijssel. Many major rivers are found in the province of Zuid-Holland. Obstacles like canals and railroads are found in all Dutch provinces. These obstacles are one of the main reasons that network location models are necessary, and planar location models do not suffice. A planar location model has been used, however, to conduct a feasibility study for the eastern part of the province of Overijssel, the region of Twenthe.

Figure 5.1 The Netherlands: provinces (black text), and prominent areas (blue text)

In Figure 5.1 a simple map of The Netherlands is presented, which shows the Dutch provinces, and the positions of the Bommelerwaard and the municipality of Bennebroek, which were used as examples in the previous chapter. Also the positions of the major cities of Rotterdam and Den Haag (The Hague) in the study regions have been indicated.

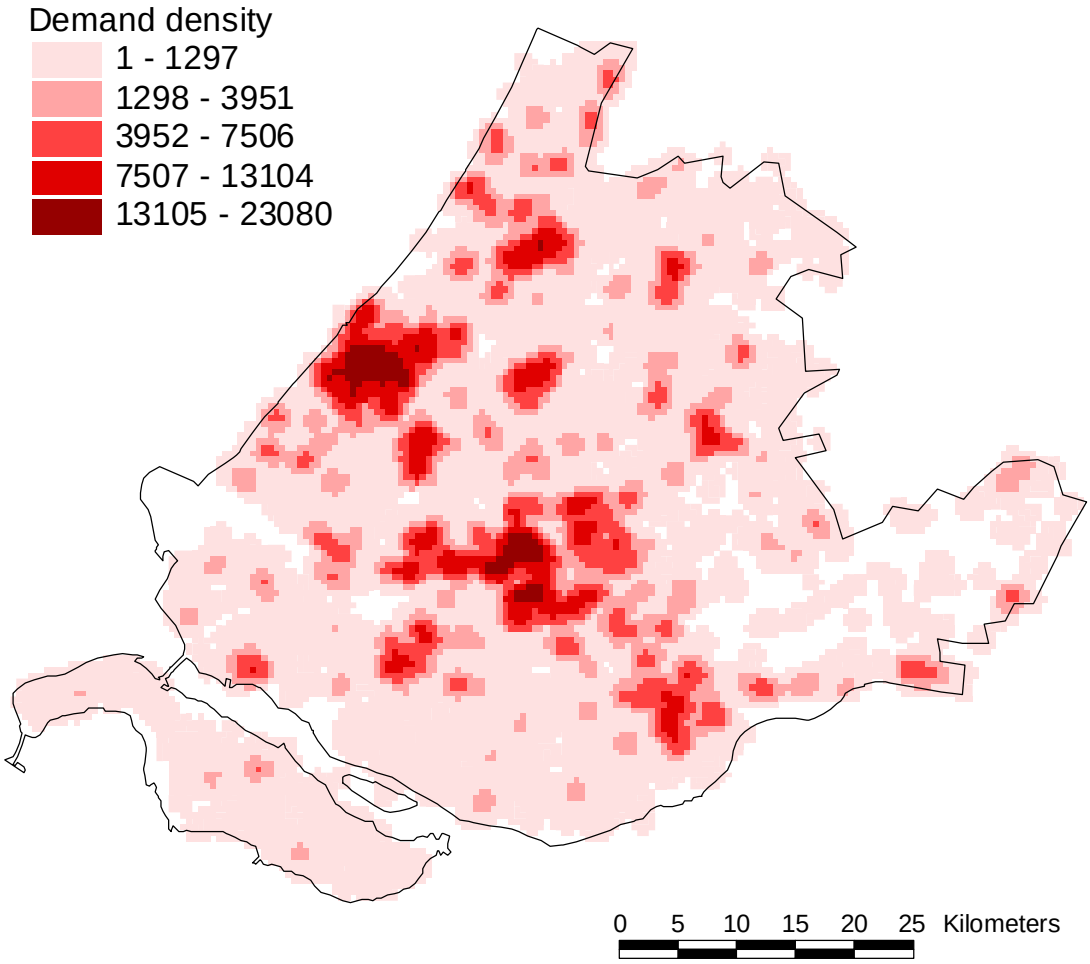

Figure 5.2 Demand density in the province of Zuid-Holland

5.3 Application to study regions

In Figures 5.2 and 5.3 the demand density, as calculated in the CARE algorithm, is shown for the study regions. It is self-evident that the demand density, based on numbers of dwellings, and the population density are strongly correlated. In Figure 5.2 the urban agglomerations of Rotterdam and The Hague are clearly visible. Rural regions are found south of Rotterdam, on the Zuidhollandse Eilanden, i.e. the former isles of the province of Zuid-Holland, and in the central and eastern part of the province of Zuid-Holland, part of the Groene Hart (i.e. Green Hart) of The Netherlands. In Figure 5.3 the major cities in the rural study region are

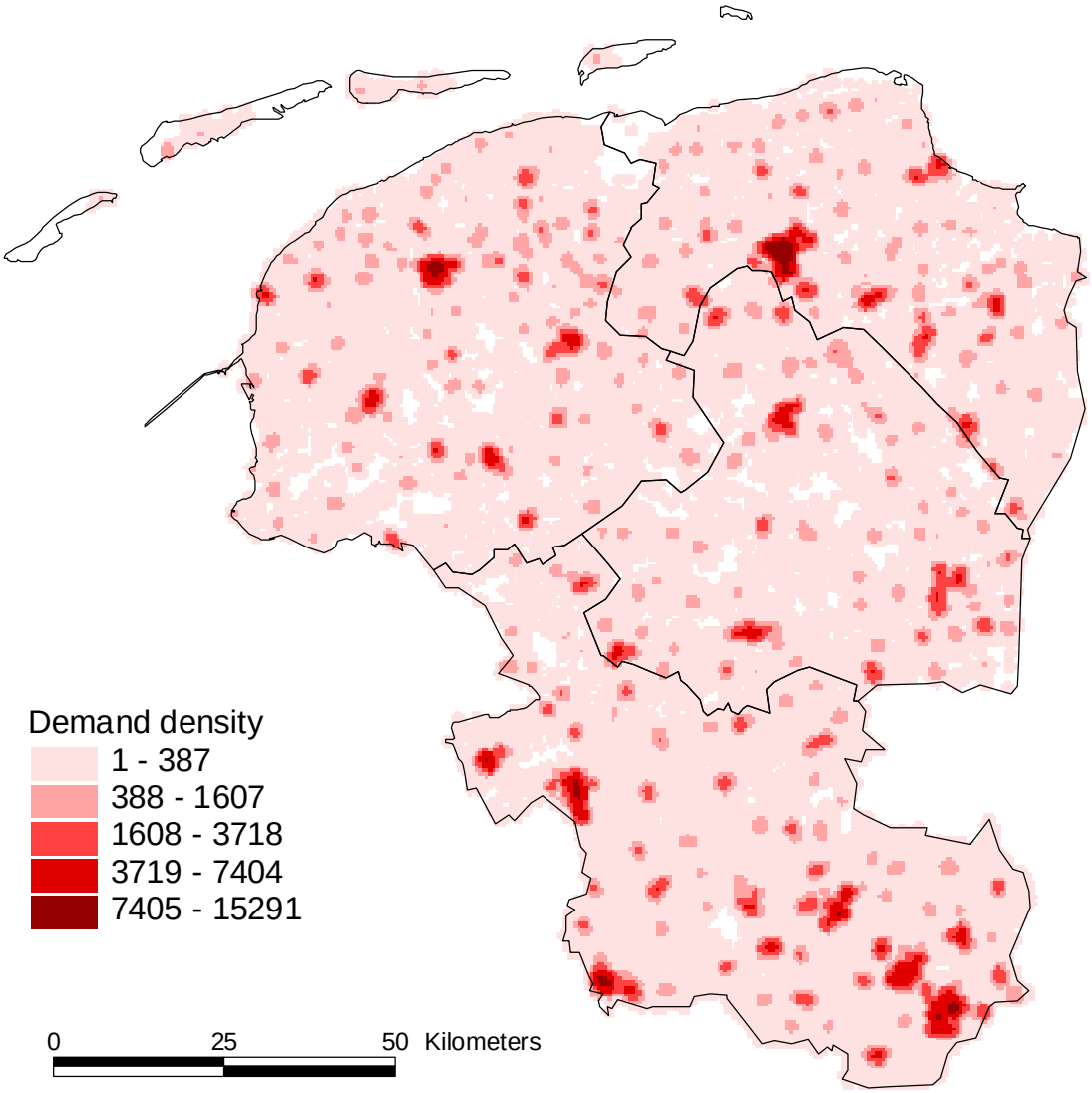

Figure 5.3 Demand density in the four north-east provinces

easily recognised, as is the region of Twenthe in the south-east, an industrial area with Hengelo and Enschede as most prominent cities.

The CARE algorithm has been applied to both study regions, using a maximum coverage distance of 20 kilometres and a maximum workload equivalent to 15 000 addresses. The repositioning was stopped when the maximum displacement of the centres was less than 1 kilometre. In Figures 5.4 and 5.5 the resulting configuration of interviewers and their work

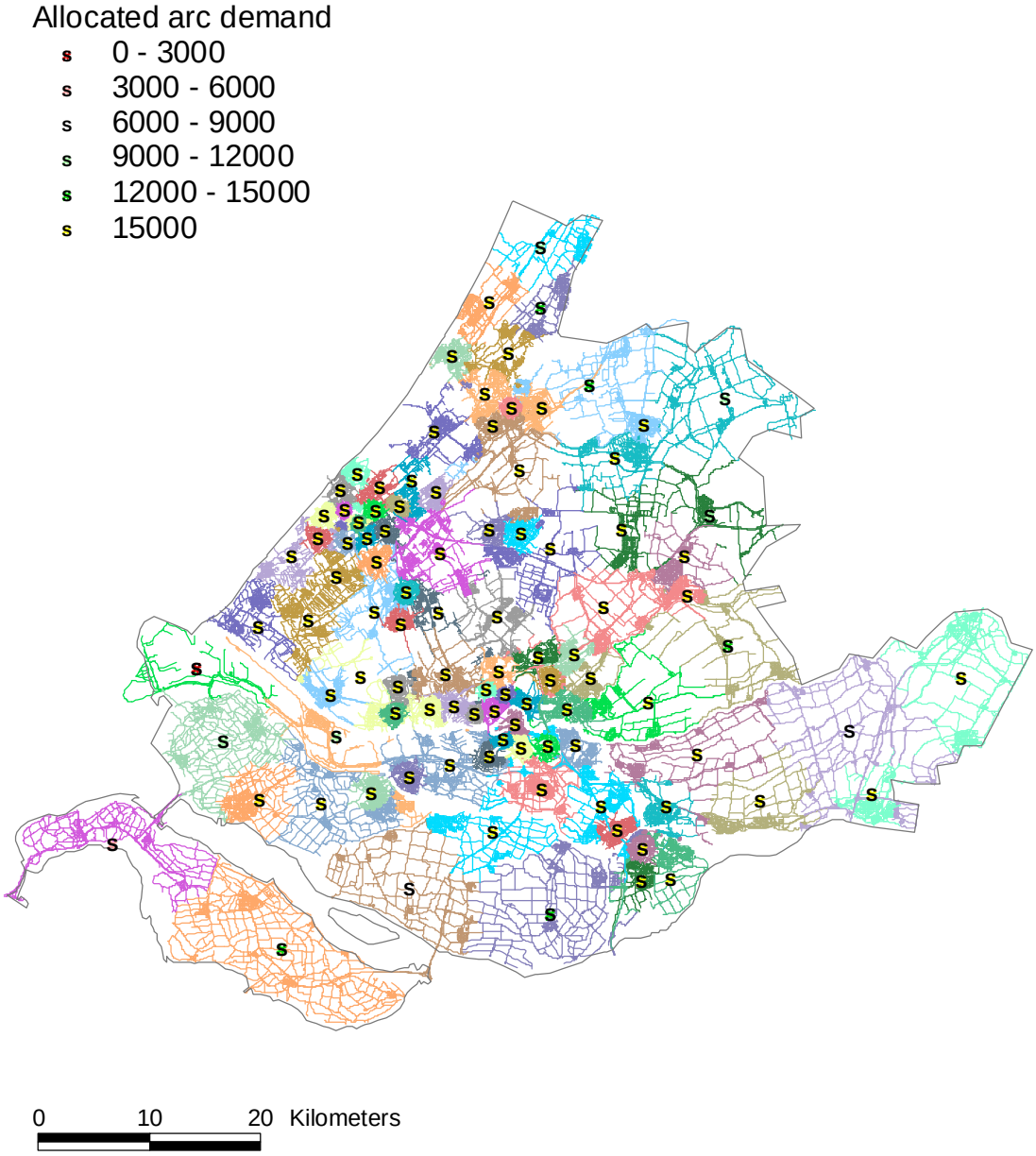

Figure 5.4 End configuration of the CARE algorithm in the province of Zuid-Holland

areas is shown. In Zuid-Holland most interviewers have a workload equal to the maximum value. Only interviewers in large rural areas do not reach the maximum value. Some peculiarities may be explained if the topographical situation (see Figure 5.1) is taken into account. For instance, the Alblasserwaard in the south-east part is connected to the rest of the province by a few bridges and ferries. Connections by ferries have not been taken into account because interviewers are not supposed to use ferries to cross over rivers and canals (ferri

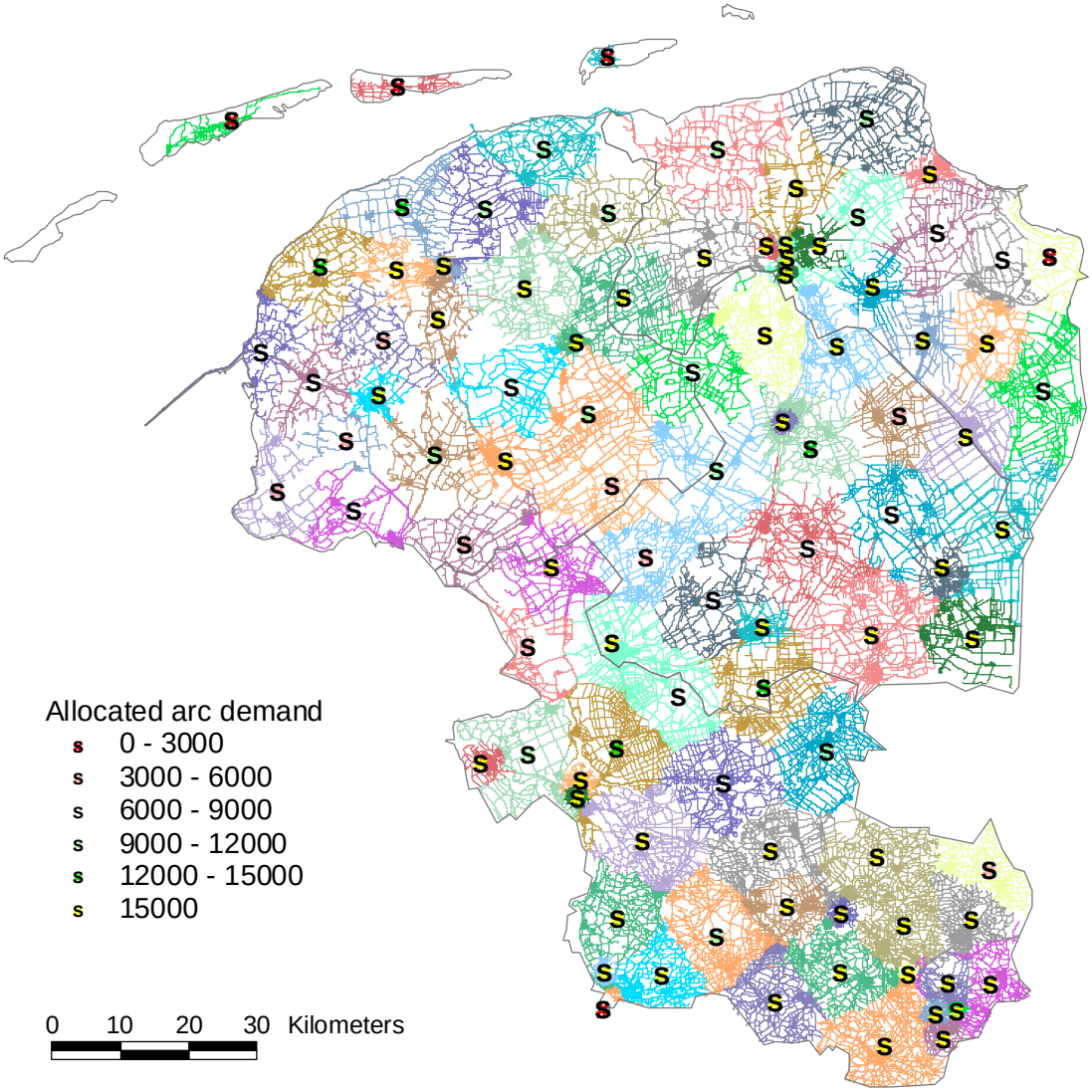

Figure 5.5 End configuration of centres and partitions in the north-east of The Netherlands

ages are not reimbursed). The Zuidhollandse Eilanden also have few connections to the main to the rest of Zuid-Holland, in this casea few dams, bridges, tunnels, and ferries.

Prominent in Figure 5.5 are the Waddeneilanden, the isles in the north on the border of the Wadden Sea and the North Sea. The population of the Waddeneilanden is rather small, but surveys samples should represent the entire population of The Netherlands. Therefore, each Waddeneiland is included in the sample. For reasons of efficiency, however, the sample ad

dresses of these isles are visited during one month, whereas other regions may be visited all the year round. In view of the special treatment of the Waddeneiland we will focus our attention to the mainland of the study region. Since the major part of the study region is rural area, both the maximum coverage distances and the fraction of interviewers that do not reach their maximum number of addresses will be considerable. The most prominent topographical obstacles (see Figure 5.1) are the lakes in the south part of Friesland and in the north-west part of Overijssel. The lakes are the reason that in these lake districts substantial distances haveto be travelled to reach the farthest corners. In these areas the elimination of centres, which might seem likely at first sight, is often not possible, because these farthest corners become out of reach.

When the numbers of interviewers produced by the CARE algorithm are compared with their absolute minimum, i.e. when the maximum coverage distance is ignored, the following results are found. In predominantly urban study region of Zuid-Holland 99 interviewers are required, while the absolute minimum is 95. In the rural study region formed by the four north-east provinces 95 interviewers are required, while the absolute minimum is 73. In urban regions, where the maximum coverage distance is not reached, the number of interview required is about equal to the absolute minimum. In rural areas the number of interviewers required is considerably higher than the absolute minimum. The reason is that when the coverage distance increases the influence topographical obstacles increases. Such obstacles are also found in urban areas, e.g. the river Nieuwe Maas flowing through the centre of the city of Rotterdam. Because the demand density is high, an obstacle can usually be avoided by allocating demand points elsewhere, and moving the centre away from the obstacle in the repositioning phases of the CARE algorithm. The number of interviewers required is considerably increased above the absolute minimum by two separate factors: a low demand density, so that the maximum number of addresses is out of reach, and/or topographical obstacles, so that more interviewers are necessary in order to reach addresses in the farthest corners. In

addition to these external factors, also the influence of the internal factors, maximum work-load and maximum travel distance, is of importance.

5.4 Influence of maximum workload and maximum travel distance

The configuration of interviewers and work areas produced by the CARE algorithm depend on the two internal factors, maximum travel distance and maximum workload, which are not entirely independent. As the maximum coverage distance increases, in rural areas more time may have to be spent on travelling, which reduces the time available for interviewing and so reduces the maximum workload that can be imposed. Variations of the internal factors will change the configuration of the centres. Recalculation with the CARE algorithm the only sure way to reveal the effects, although the global effects can be indicated beforehand.

Variations in the maximum coverage distance have little effect on the centre configuration in the urban areas, which are usually covered by full-capacity centres. In the rural areas the effects on the centre configurations are usually small. Considerable changes appear when certain thresholds are passed. As the maximum coverage distance decreases centres will have to be added when complete coverage with the same number of centres is no longer possible. As the maximum coverage distance increases centres will be eliminated when the demand points of a certain centre can be completely taken over by neighbouring centres.

Variations in the maximum capacity will have little effect on the configuration of the centres in rural areas, which in practice do not reach their full capacity by far. The effect of variations in the maximum capacity will mainly be found in urban areas, where many full-capacity centres are found. Therefore, in urban areas the number of centres required will decrease when the maximum capacity increases, and will increase when the maximum capacity decreases.

5.5 Boundary effects

For very large networks running the CARE algorithm may become troublesome. Although the software will run for networks of any size (see ESRI, 1992, page 9), above a certain size, which depends on the computer configuration, not all data will fit into the internal memory of the computer and swapping of data between internal memory and disk will be necessary. The swapping process dramatically increases the run-time of the software (by factors of 15 and more), and results will no longer be produced in reasonable time. A solution may be to divide the region of interest into smaller regions and, in one way or another, fit the results for the smaller regions together. In that case we will have to deal with boundary effects. Since addresses across a boundary are not within reach, the centres tend to move away from the boundaries. To reduce the effects of boundaries the region could be subdivided along topographical obstacles such as rivers and canals, which limit the traffic between the areas they separate.

To study the magnitude and the range of the boundary effects, the rural study region will be subdivided into four separate provinces. The CARE algorithm is applied to each province separately, and the resulting interviewer configuration and work areas are compared with those of the entire study region. In Figure 5.6 the separate results for the four provinces are combined. The resulting centres for the entire study region are also depicted. Comparing Figures 5.5 and 5.6 it is clear that boundary effect are considerable, and almost at the other side of the provinces. Only the urban centres and the centres along outer boundaries remain at about the same position. Remarkable is the fact, that the total number of centres and the number of centres in each province remain almost the same. The effect of the subdivision appears to be a distribution of the centres near the provincial boundaries among the provinces concerned. The distribution of centres is followed by a repositioning procedure, which may influence the position of centres at a considerable distance.

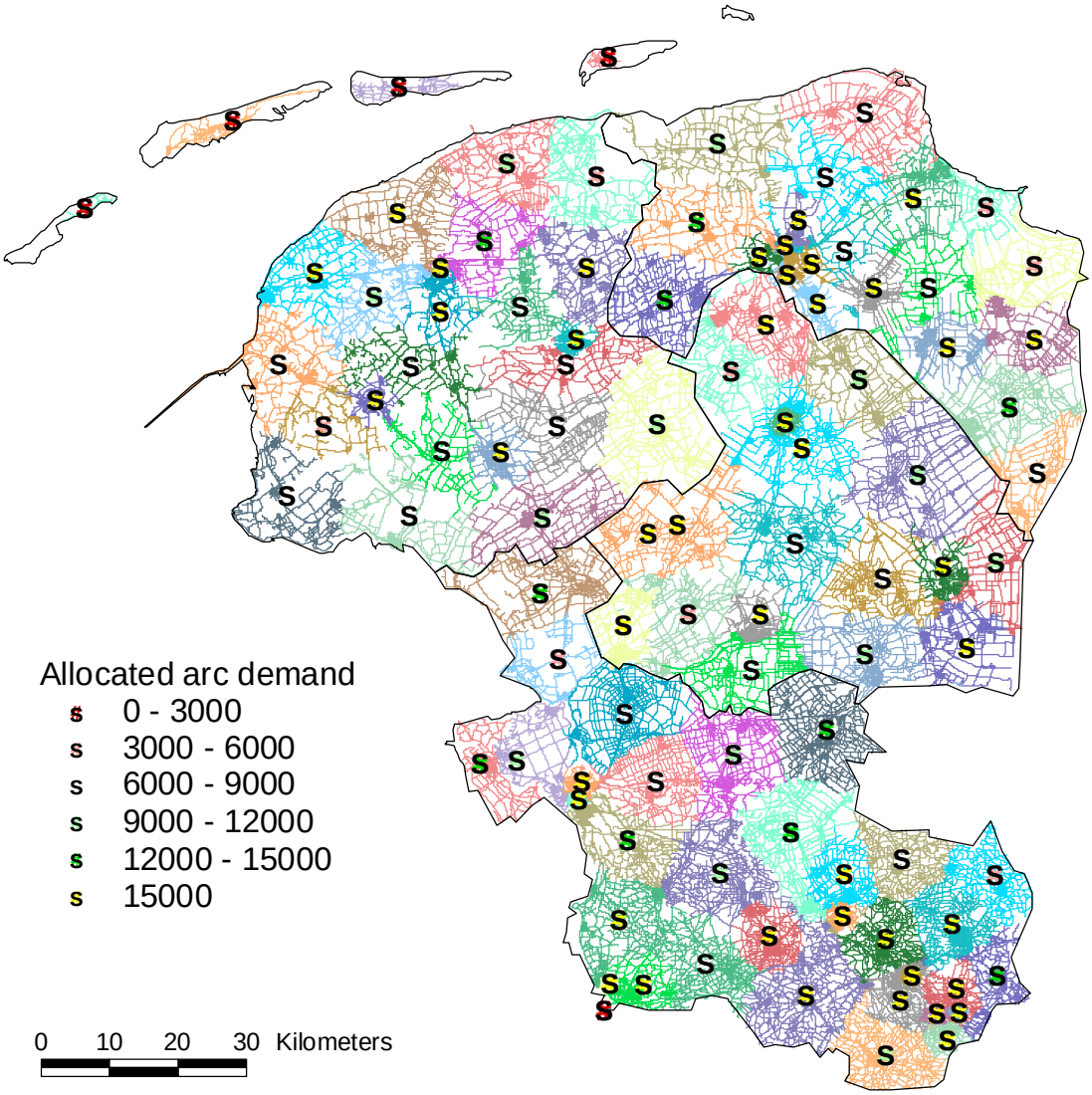

Figure 5.6 End configuration of centres and partitions for four separate provinces in the north-west of The Netherlands

In appears that in order to estimate the number of interviewers in a large region, the region may be subdivided into a number of smaller regions. The positions of the centres, however, will differ considerably, especially near the boundaries. Therefore, if the positions of the centres are very important, the larger region should be subdivided along natural boundaries. The configuration of centres in Zuid-Holland, for instance, shows that this may very well be done (see Figure 5.4) without considerable disturbances.

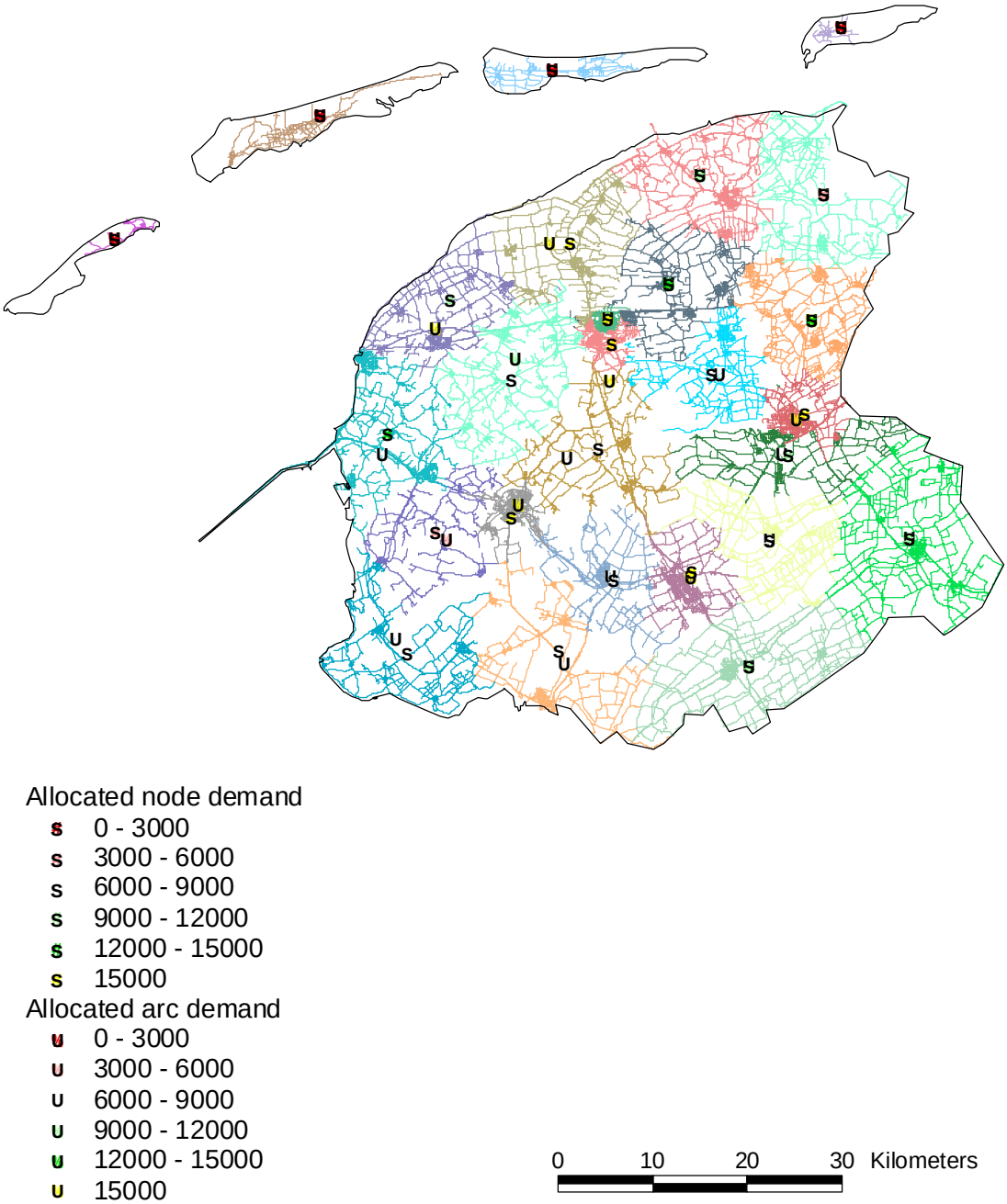

Figure 5.7 Comparison of centre configurations for the province of Frieland. One configuration is based on node demand, the other one on arc demand.

5.6 Cost reduction by optimised configuration

Economic aspects of location-allocation modelling are important, and often the primary motive. Since interviewers are hard to recruit, in the case of Statistics Netherlands the primary

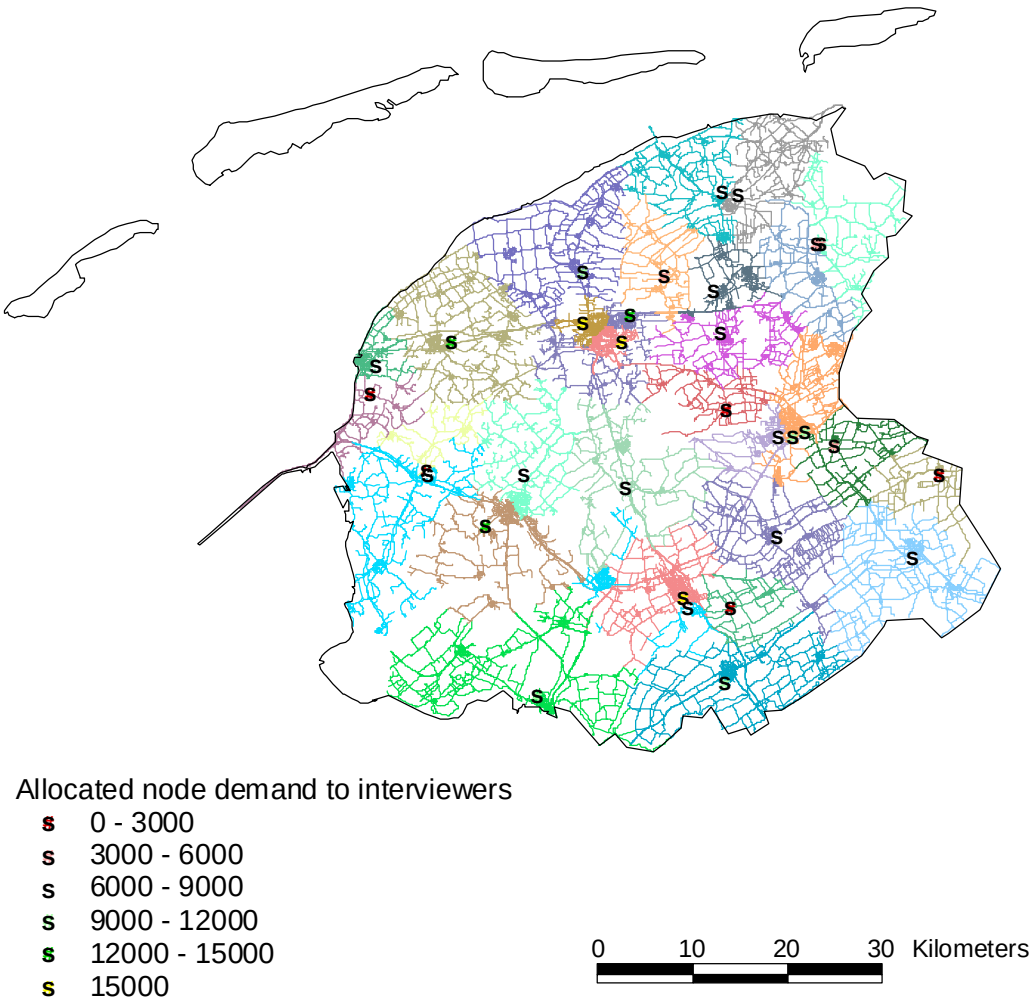

Figure 5.8 Allocation of node demand to interviewers in the province of Friesland. The maximum travel distance used was 20 kilometres.

motive is to reduce the number of interviewers required by recruiting them in the most appropriate regions. The desired configuration in Friesland is shown in Figure 5.7, while the present situation – in fact ,roughly the situation about a year ago – is depicted in Figure 5.8. In the real situation 32 interviewers are required, and in the desired configuration only 27, including 4 interviewers for each of the Waddeneilanden. So, in fact, the real 32 interviewers should be compared with the 23 interviewers on the mainland. When properly located the number of interviewers could be considerable reduced. Less interviewers may mean that the distances travelled by the interviewers may increase. Calculations have shown that the total travel distance remains roughly the same, depending on the model an increase of 10% (sum

of distances to demand nodes),or a decrease of 7% (sum of distances to demand nodes weighted with demand) is found.

These results should be carefully interpreted, because the aim of the CARE algorithm is to determine a minimum configuration. This minimum configuration may require interviewers on location where recruitment is difficult, e.g. the south-west part of Friesland. Further, a surplus of interviewers is desirable, because they are not always available. Nevertheless, the present configuration of interviewers in Friesland can be considerably improved recruitment in the proper regions, without an expected increase in the travel expenses to be paid.

5.7 Discussion

The results for the urban and rural study regions also demonstrate that the CARE algorithm produces good centre configurations, which are may be of direct practical use. The computation times are about two hours (including the time required to write results to disk), which is considerable, but quite acceptable considering the size of the networks (see Section 5.2). Probably, not one of the established methods mentioned in section 2.5 is be able to provide an acceptable solution in a reasonable amount of time.

A large study region may be divided in smaller regions in order to estimate the number of interviewers required. When the configuration of centres is also required, division along topographical obstacles, such as rivers, is a prerequisite. On the other hand, there is usually more than one configuration of centres able to cover the entire region, a fact that should be taken into account when assessing a particular solution produced by the CARE algorithm.

With the CARE algorithm a minimum interviewer configuration for Friesland was calculated. It was found that, without increasing the travel expenses, the number of interviewers could be

reduced, if interviewers can be recruited in specific regions. A comparison of actual and desired configurations will show which regions have too few or too many interviewers.

CHAPTER 6

Discussion and Conclusions

In this final chapter various issues related to the CARE algorithm for solving the capacitated set covering location problem (CSCLP) will be discussed. Some of these issues directly concern the present algorithm, some concern extensions to related problems found in practice. For clarity each of these issues will be considered in separate sections. In the last section of this chapter conclusions with respect to the application of the CARE algorithm will be drawn.

6.1 Discussion

The present implementation of the CARE algorithm yields good solutions for the capacitated set covering location problem (CSCLP). As discussed in Section 2.3, the CSCLP is NP-hard, and exact solutions can only be found using exponential-time algorithms, which are not useful for large-scale problems. The CARE algorithm is a heuristic method, which may be applied to large-scale problems, but will yield only approximate solutions. It has been shown in the previous two chapters, that the solutions provided by the CARE algorithm are “good”, i.e. (probably) not far from optimal and useful in practice. In general, solutions of a CSCLP are not unique, i.e. many optimal configurations of centres with the same optimal number of centres are possible. The same applies to the solutions of the CARE algorithm. Therefore, in interpreting these solutions the available “elbow room” of the centres has to be kept in mind.

Location-allocation modelling focuses on the optimisation of certain pre-defined objectives. One should realise, however, that “(t)he value of many of these models is rarely the optimal solution they provide, but rather their ability to explore and compare solutions” (Bailey and Gatrell, 1995, page 374).

6.1.1 Computation times

The computation time needed by the CARE algorithm is acceptable when present-day personal computers are used, and when the size of the model is taken into account. Results for model networks with about 100 000 nodes and arcs can be obtained within 1½ hours, using the combination of hardware and software described in Section 4.7.

The methods to solve location-allocation problems reviewed in Section 2.5 require the calculation of the complete origin-destination matrix. The computation time needed for the calculation of the complete OD-matrix is proportional to n^4, where n is the size of the network (the number of nodes or arcs). Since the complete OD-matrix has n^2 elements, and the computation of distances using Dijkstra's efficient heuristic algorithm is proportional to n^2 (see Daskin, 1995, Section 3.3), the total computation time is proportional to n^4.

In the CARE algorithm only the distances in the gradually expanding neighbourhood of the centres have to be calculated. Since each node is allocated to one centre, only n distances have to be calculated. Actually the number of distances calculated is slightly higher, because at the moment neighbourhoods touch for nodes near the boundary the distance to more than one centre will be calculated. The total computation time needed to calculate the distances for the CARE algorithm is proportional to n^3.

In the CARE algorithm the distances are calculated anew for each allocation run, whereas in the other methods the complete OD-matrix is calculated only once. However, for large networks the repeated calculation of parts of the OD-matrix is much more efficient than calculating and storing the complete OD-matrix in advance. Therefore, for large networks the CARE algorithm will be much faster than the methods reviewed in Section 2.5, with an exception for the node-partitioning algorithm that could use the same approach. However, as

stated in Section 3.1, the node-partitioning algorithm and the GRIA may not perform properly for capacitated centres.

6.1.2 Improving the repositioning implementation

In the present implementation of the CARE algorithm the position of the centre of a partition is estimated on a geometrical basis. An improvement to the algorithm would be to develop a procedure to determine the actual centre. An implementation of this procedure based solely on ArcInfo routines and AML statements will be very slow. More appropriate for an efficient implementation of this procedure would be the use of 3GL-code, which was outside the scope of this study. Here, 3GL is a term used to denote third generation programming languages, such as PL/1, FORTRAN, Pascal, C, and C++. The goal of this study was the development of an effective algorithm to solve the CSCLP with a reasonably efficient implementation. In implementing the CARE algorithm improvements in the algorithm had a considerably higher priority than improvements in the efficiency of the source code. The present source code may be streamlined, however, and a reduction of 30% in the execution time seems feasible.

To reduce the execution time drastically, the CARE algorithm should be implemented entirely in 3GL-code, so that essential data may be kept in memory during execution and time consuming swapping of data between memory and disk may be avoided. An implementation in 3GL-code also has the advantage of increased flexibility, which may lead to further tuning of the algorithm. For instance, in the case of capacitated centres the allocation of nodes on the basis of shortest distances may be adjusted to the special needs this case needs.

6.1.3 Improving the node-allocation algorithm

The allocation algorithm of ArcInfo roughly follows the allocation procedure described in the Intermezzo of Section 3.1. As has been discussed in Section 4.4 centres will be repositioned by the node-partitioning as long as they compete for demand point. Centres in an area of high demand density will reach their full capacity sooner than centres further on in areas with a low demand density. In that case, no competition for demand points will occur when the present allocation algorithm is used. The priority for allocation is now purely based on the shortest distance between demand points and centres as long as there is capacity available at the centres. If priority is based on the remaining capacity of the centres, the competition for demand points will increase, and so will the mobility of the centres. In this way more centres may participate in the repositioning, and also more centres may take part in the special elimination procedures. Consequently, the total number of centres eliminated may increase which will yield a further improvement of .the CARE algorithm.

6.2 Related location-allocation problem

In this section a number of problems related to the CSCLP will be considered, and possible extensions of the CARE algorithm for these problems will be discussed.

6.2.1 Dealing with existing centres

The CSCLP assumes that there are no existing service centres, i.e. it represents the ideal situation. In practice, however, many centres may already exist and the problem is to determine the number of centres that have to be added, and to determine the optimal locations (additional facility location problem). The CARE algorithm may be adapted to this situation when following constraints are taken into account: (1) existing centres may not be reposi-

tioned, and (2) existing centres may not be eliminated. It should be noted that the repositioning of centres is also part of the elimination procedures. After elimination of a centre its demand points have to be reallocated to neighbouring centres, which may shift during the repositioning steps in the CARE algorithm. In that case the fact that existing centres have fixed positions will have to be taken into account in selecting centres which may be eliminated. In the general elimination procedure existing centres have to be excluded altogether, because their immobility may prevent these centres from allocating enough demand points, and part of the area could remain uncovered.
In principle, the existing centres should also be excluded form the special elimination procedure for the same reason. However, since for the special elimination procedure repositioning is less critical, an increase of the distance margins may suffice in practical situations.

Another situation that may occur in practice is the location of addition of new service centres and the possible closing down of existing centres (reorganisation problem). To examine the possibilities of closing down an existing centre, the general and special elimination procedures could be applied to existing centres and neighbouring additional centres, which may still be repositioned. Neighbouring existing centres may also be included in the elimination as long as their inability to move is taken into account.

6.2.2 Dealing with off duty centres

In determining the minimum number of service centres, it is assumed that these centres are permanently available. In practice one or more centres may be temporarily out of order. In the case of Statistics Netherlands interviewers may be on holiday or may be down with flue and other interviewers have to take over their work. Many ways to deal with such situations are available. The simplest one is a (temporary) increase of both the maximum workload (which is inevitable) and the maximum travel distance of the neighbouring interviewers.

Another option is a (temporary) increase of the maximum workload, while the maximum travel distance remains fixed. In that case the number of interviewers required will increase, because any sampling address has to be within reach of at least two interviewers. In that situation it should be possible to eliminate an arbitrary centre without loosing complete coverage. If this is not the case a centre has to be added in the neighbourhood of this indispensable centre. After which the centres may be repositioned to improve their distribution over the area. Indispensable centres may be detected with the elimination procedures, which are now applied in the opposite way.

In the previous case it was assumed that the centres are fixed. In several cases, however, mobile centres appear, e.g. police cars stationed in such a way that the maximum time for the nearest police car to arrive at any location in the district does not exceed a certain maximum time. When one car is called away, the neighbouring police cars will have to be repositioned to fill in the gap. To minimise the number of police cars to be moved, only the nearest police cars around the gap should be repositioned. In this particular case the capacity of the centres is of minor interest, the maximum travel distance being the most important factor. To determine the new positions of the remaining police cars, the repositioning procedure of the CARE algorithm could be used, which will do the job in a simple and efficient way.

6.2.3 Alternative location of centres

The principle objective of set covering models is to determine the minimum number of centres necessary to cover all demand points. In general there is more than one optimal configuration of centres. This non-uniqueness leaves room to add a secondary objective, e.g. the minimisation of the maximum coverage distance (e.g. to minimise the maximum travel time between centre and demand point), or the minimisation of the weighted sum of distances (e.g. to minimise transportation costs). In the CARE algorithm the service centres are placed in the centres of their partitions, which is a local minimisation of the coverage distance.

When minimising the number of centres is by far the most important objective, one of the consequences may be that centres are located at undesirable places. In the case of Statistics Netherlands, for instance, Figures 5.2 and 5.4 show that the centres located on the Zuidhollandse Eilanden, former islands south of Rotterdam, are found in areas with a very low population density. Recruitment of interviewers in the direct neighbourhood of the desired locations will be difficult.

It is possible to shift locations towards more densely populated areas, for instance, when in the CARE algorithm the centres are repositioned to the (demand) centres of gravity. In fact, this was the case in the first implementation of the CARE algorithm, which lead to fast repositioning of centres and a very stable algorithm. In that approach, repositioning is not only guided by the mutual repulsion of the centres, but also by the attraction of the centres by the (concentrations of) demand points. The drawback is that the centres are not in the middle of their partitions, so their coverage becomes smaller, and as a consequence more centres are required to cover all demand points. The main advantage is that fewer centres are located in thinly populated areas.

Shifting the centres towards the medians, taking care that the entire remains covered, may be a second objective in the postprocessing of the centre configuration found by the CARE algorithm. The postprocessing may provide a reduction in the sum of distances (cost reduction), and will also results in stabilty, since the ”elbow room” of the centres is reduced.

6.2.4 Clustering of demand points

The CARE algorithm may also be used for less obvious applications. After the centres and work areas have been determined, the algorithm can also be used to cluster demand points. Suppose that all demand points in a work area have to be serviced once a month, and that

about 17 a month available for work in the field. Using the CARE algorithm is possible, to divide the work area in 17 clusters of demand points which can each be serviced in a single work day. By tuning the maximum daily workload (capacity of a centre) and the maximum coverage distance of a cluster an acceptable clustering of demand points will be the result.

6.3 Conclusions

The goal of the present study was the development of an algorithm to solve the capacitated set covering location problem for a large-scale network in an acceptable amount of time. For large-scale networks exact algorithms require an excessive amount of time. The established heuristic location-allocation algorithms, which could be applied to large-scale networks, also require vast amounts of computing time. The CARE algorithm, however, which was developed during the present study, can handle the CSCLP for large-scale networks in acceptable amount of time using present-day personal computers.

The CARE algorithm has provided good solutions to the CSCLP for networks with more than 100 000 nodes and arcs. The computation time (excluding the time required to write visualisation data to disk) for those networks was usually less than 1½ hours. For medium-scale networks with about 10 000 nodes and arcs the computation time is less than half an hour.

Very large networks, such as the roads network of The Netherlands (800 000 nodes and arcs), may have to be subdivided into several large networks. In that case the minimum number of service centres required hardly changes. The locations of the centres in the distant neighbourhood of the boundaries, however, may change considerably. A division along topographical obstacles, which reduce cross boundary traffic, can remedy this.

Application to the case of Statistics Netherlands has shown that by recruiting interviewers in the specific regions, the number of interviewers required may be reduced considerable without causing an increases in the travel expenses to be paid. Since it is (at the moment) very difficult to recruit interviewers, in order to be most effective, the recruiting efforts should be focussed on the “optimal regions”. By keeping the group of interviewers as small as possible also the costs for training, equipment, and support by supervisors will be reduced.

The applicability of the CARE algorithm can be also extended, for instance, to include existing centres (additional facility location problem), or to take into account off duty centres.

References

AVV, 1999, Handleiding Nationaal Wegen Bestand (NWB), Adviesdienst Verkeer en Vervoer, Ministerie van Verkeer en Waterstaat, Rotterdam, The Netherlands.

Bayley T.C., Gatrell A.C., 1995, Interactive Spatial Data Analysis, Longman Scientific and Technical, Harlow, Essex, England

Casillas, P.A., 1987, Data aggregation and the p-median problem in continous space, in: Ghosh A., Rushton G. (eds.) , 1987, Spatial Analysis and Location-Allocation Models, Van Nostrand Rheinhold Company Inc, New York, U.S.A.

Church, R.L., Meadows, M., 1979, Location Modeling Utilizing Maximum Service Distance Criteria, Geographical Analysis, Vol. 11, pages 358-379.

Church, R.L., and ReVelle, C., 1974, The Maximal Covering Location Problem, Papers of the Regional Science Association, Vol. 32, pages 101-118.

Current, J.R., Schilling, D.A., 1987, Elimination of Source A and B Errors in p-Median Location Problems, Geographical Analysis, Vol. 19, No. 2, pages 95-110

Current, J.R., Schilling, D.A., 1990, Analysis of Errors Due to Demand Aggregation in the Set Covering and Maximal Covering Location Problems, Geographical Analysis, Vol. 22, No. 2, pages 116-126

Current, J.R., Storbeck, J.E., 1988, Capacitated covering models, Environment and Planning B: Planning and Design, volume 15, pages 153-163

Daskin M.S., 1995, Network and Discrete Location: Models, Algorithms, and Applications, John Wiley & Sons, Inc., New York, U.S.A.

Densham, P.J., Rushton, G., 1992, A More Efficient Heuristic for Solving Large P-Median Problems, Papers in Regional Science, Vol. 71, No. 3, pages 307-329

Dijkstra, E.W., 1958, A Note on Two Problems in Connexion with Graphs, Numerische Mathematik, Vol. 1, pages 269-271.

Domschke, W., Krispin, G., 1997, Location and planning layout. A survey, Operations Research-Spektrum, volume 19, pages 181-194.

Drezner, Z. (ed.), 1995, Facility Location: a Survey of Applications and Methods, Springer Verlag, New York, U.S.A.

ESRI, 1992, Network Analysis – Modeling network systems, Arc/Info User's Guide, Release 6.1, Environmental Systems Research Institute, Inc., Redlands, California, U.S.A.

Evans, J.R., Minieka, E., 1992, Optimization Algorithms for Networks and Graphs (second edition, revised and expanded), Marcel Dekker, Inc., New York, U.S.A.

Ghosh A., Rushton G. (eds.) , 1987, Spatial Analysis and Location-Allocation Models, Van Nostrand Rheinhold Company Inc, New York, U.S.A.

Goodchild, M.F., 1979, The Aggregation Problem in Location-Allocation, Geographical Analysis, Vol. 11, pages 240-255

Goodchild, M.F., Noronha, V., 1983, Location-allocation for small computers, Monograph 8, Department of Geography, University of Iowa, Iowa City, Iowa, U.S.A.

Hakimi, S.L., 1964, Optimum Location of Switching Centers and the Absolute Centers and Medians of a Graph, Operations Research, Vol. 12, pages 450-459

Hakimi, S.L., 1965, Optimum Distribution of Switching centers in a Communication Network and Some Related Graph Theoretic Problems, Operations research, volume 13, pages 462-475.

Handler G.Y., Mirchandani, P.M., 1979, Location on Networks, MIT Press, MA, U.S.A.

Hillsman, E.L , 1984, The p-Median Structure as a Unified Linear Model for Location-Allocation Analysis, Environment and Planning A, Vol. 16, pages 305-318

Hodgart, R.L., 1978, Optimizing access to public services: a review of problems, models and methods of locating central facilities, Progress in Human Geography, Vol. 2, pages 17-48.

Hurter A.P., Martinich J.S., 1989, Facility Location and the Theory of Production, Kluwer, Boston, Massachusetts, U.S.A.

Ormeling, F.J., Kraak, M.J., 1993, Kartografie. Visualisatie van ruimtelijke gegevens (Studenteneditie), Delftse Universitaire Pers, Delft The Netherlands.

Larson, R.C., Odoni, A.R., 1981, Urban Operations Research, Prentice-Hall, New Jersey., U.S.A.

Laurini, R., Thompson, D., 1992, Fundamentals of Spatial Information Systems, Academic Press Inc., London, U.K.

Love R.F., Morris J.G., Wesolowsky G.O., 1988, Facility Location: Models and Methods, North-Holland, New York, U.S.A.

Maranzana, F.E., 1964, On the location of supply points to minimize transport costs, Operational Research Quarterly, Vol. 15, pages 261-270.

Marianov, V., ReVelle, C, 1995, Siting Emergency Services, in: Drezner, Z. (ed.), 1995, Facility Location: a Survey of Applications and Methods, Springer Verlag, New York, U.S.A.

Mirchandani, P.M., Francis R.L. (eds.), 1990, Discrete Location Theory, John Wiley & Sons, Inc., New York, U.S.A.

Morrill, R.L., and Symons, J ., 1977, Geographical Analysis, Efficiency and Equity Aspects of Optimum Location, Vol. 9, pages 215-225

ReVelle, C., 1991, Siting Ambulances and Fire Companies. New Tools for Planners, Journal of the American Planning Association, volume 57, no. 4, pages 471-484

Rushton G., Goodchild, M.F., Ostresh, L.M., Computer Programs for Location-Allocation Problems, Monograph Number 6, Department of Geography, University of Iowa, Iowa City, Iowa, U.S.A.

Teitz, M.B., Bart, P., 1968, Heuristic Methods for Estimated Generalized Vertex Median of a Weighted Graph, Operations Research, Vol. 16, pages 955-961

Toregas, C., ReVelle, C., 1972, Optimal Location under Time or Distance Constraints, Papers of the Regional Science Association, volume 28, pages 133-143

Vosmer, M., 1998, Personal Communication, Department of Household Surveys, Statistics Netherlands, Heerlen, The Netherlands,

Appendix

Using The Software Provided on Disk

The accompanying disk, attached at the back cover, contains a number of AML scripts. AML denotes ArcInfo Macro Language, the scripting language of ArcInfo. Except for two scripts, these scripts are the ArcInfo implementation of the CARE algorithm for the SCLP. The scripts called “splitadr.aml” and the script “sc0_node.aml” may be helpful in preparing the network coverage for application of the algorithm.

The AML scripts as provided on the disk are suitable for the case of demand along the arcs only. With minor alterations the script become suitable for the case of demand in the nodes only. For that case, the statement “DEMAND ADRESSEN NODETRIC” should be replaced by the statement. “DEMAND # ADRESSEN”, and in the ALLOCATE statements “BOTH” should be replaced by “NODE”.

It is assumed that the demand is assigned to the arcs in an item called “ADRESSEN” in the AAT (arc attribute table). The script called “splitadr.aml” may be used to split the demand along the arcs in halves, and assign one half to the from-node and one half to the to-node. The total demand assigned to the nodes will be stored in an item called “ADRESSEN” in the NAT (node attribute table).

In calculating the centres of gravity the node density at the nodes are required. The node density is calculated in the script called “sc0_node.aml”, that stores the node density in an item called “PDENS” in the NAT.

The NAT should also contain items called “NODETRIC” and “ADRES_W”. The former is an integer number item used in the allocation to force the creation of a node-allocation table.

The latter is used to calculate the position of full-capacity centres, and is a real number item. The values stored in “ADRES_W” are the root of the root of the node demand stored in the item called “ADRESSEN”, and is calculated in the TABLES module with the statement “CALCULATE ADRES_W = ADRESSES ** 0.25”.

When the AAT and the NAT are prepared in this way, the CARE algorithm may be applied. The AML scripts should be run in the following order:

1. SC1_VARS (to initialise the global variables)
2. SC2_INIT (to calculate the initial centres)
3. SC3_MRT (general elimination of centres)
4. SC4_MRT (special elimination of centres; $\gamma = 4$)
5. SC5_MRT (special elimination of centres; $\gamma = 6$)

Intermediary and final results are stored as shape files, which may be examined and further processed with ESRI’s ArcView GIS software.