Pergamon

PII: S0038-0121(96)00016-X

Socio-Econ. Plann. Sci. Vol. 31, No. 1, pp. 11–28, 1997

0038-0121/97 $17.00 + 0.00

Political District Determination Using Large-scale Network Optimization

JOHN A. GEORGE, BRUCE W. LAMAR† and CHRIS A. WALLACE

Department of Management, University of Canterbury, Christchurch, New Zealand

Abstract—Political redistricting involves the amalgamation of a large number of geographic or statistical units into a smaller set of electoral districts such that the districts satisfy various criteria prescribed by law, judicial mandate, or historical precedent. This paper describes a network-based optimization model that is currently being used by governmental decision-makers in New Zealand to assist in the preliminary stages of determining Parliamentary district boundaries. The model is capable of solving large-scale problems. In the New Zealand case, the procedure has been applied to problems involving over 35,000 geographic units (called "meshblocks") and 95 Parliamentary districts. The model generates electoral districts that have roughly equal population and are generally compact in shape. One of the key features of the model is its flexibility. On the one hand, it can take into account currently existing features such as topographic terrain or previously existing electoral boundaries. On the other hand, it can also be used to generate an entirely new set of political boundaries as was mandated by the voters in a recent referendum in New Zealand.

INTRODUCTION

This introduction is divided into four parts. The first part summarizes the standards used in determining electoral district boundaries, the second part describes the redistricting process in New Zealand (which is used as a case study in this paper), the third part reviews the literature on operations research models for political redistricting, and the fourth part points out the contributions that this paper makes to the field.

Criteria for determining electoral districts

The process of political redistricting involves the grouping together of a relatively large number of small statistical or geographic units (such as census tracts or block numbering areas) into a much smaller set of electoral districts (such as Congressional or Parliamentary districts). Balinski and Young [3], Busteed [6], Butler and Cain [7], Dixon [11], Jewell [22], Lijphart [25], Morrill [28, 29], Saxbe [32], and others have identified a number of criteria commonly used in determining the size and shape of electoral districts. Specifically, the boundaries should be drawn such that the electoral districts:

(a) have equal population to within a specified tolerance;
(b) are contiguous and generally compact in shape;
(c) avoid crossing natural barriers (such as mountain ranges or major waterways);
(d) follow established routes of transportation and communication;
(e) avoid unnecessary splitting of "communities of common interest";
(f) adhere to the boundaries of other political constituencies (such as cities, counties, or local wards); and
(g) match, as closely as possible, the boundaries of the previously existing electoral districts.

Although this list is not all-inclusive (and, in fact, purposely omits the covert or partisan activities sometimes present during reapportionment), it does express many of the major objective issues considered during the redistricting process. Clearly, some of these criteria (such as equalizing population) are easier to define and measure than other criteria (such as maintaining communities

†Author for correspondence. *E-mail*: B.Lamar@canterbury.ac.nz; *Tel*: 64 + 3-364-2941; *Fax*: 64-3-364-2020.

of common interest). Although, as pointed out by Balinski and Young [2] and Ernst [13] even the specification of equal population can be challenged. Depending upon the country or jurisdiction involved, these criteria may be enforced by legislative directive, judicial mandate, historical precedent, or some combination of the three.

New Zealand situation

The determination of Parliamentary electoral districts in New Zealand is under the auspices of a quasi-judicial body known as the Representation Commission. The Commission is convened by the Surveyor General (the head of the New Zealand Department of Survey and Land Information (DOSLI)) following each quinquennial (i.e. 5 y) Census of Population and Dwellings. The Surveyor General is required to prepare a set of "provisional" district boundaries. Using these provisional boundaries as a starting point, the Representation Commission prepares a set of "proposed" Parliamentary district boundaries. After deliberating on the objections and counterobjections to the proposed boundaries, the Commission then produces a set of "final" boundaries that are released to the public.

To produce the set of Parliamentary district boundaries, population data from the most recent census is used. In New Zealand, the geographical units used are called "meshblocks". A meshblock is analogous to a census tract used in the United States, but much smaller in size. (A meshblock usually contains no more than 200 people, whereas a census tract typically contains between 2,500 and 8,000 people.) There are currently in the order of 35,000 separate meshblocks defined in New Zealand. In contrast, in the most recent election (in 1993), there were a total of 95 Parliamentary districts on the general electoral roles. Thus, the Representation Commission has the formidable task of combining this enormous number of meshblocks into a much smaller set of Parliamentary districts.

The rules governing the Representation Commission's determination of Parliamentary district boundaries parallel the criteria given in the previous (sub) section. Specifically, the Commission is mandated by the Electoral Act of New Zealand to allocate meshblocks to electoral districts so as to achieve equal population among electorates (as in criterion (a)). However, deviations of $\pm 5\%$ from the equal population target are permitted in each electoral district in order to take into account criteria (c) through (g). (Criteria (e) and (f) are combined under the single category of "community of interest".) Also, although not formally stated in the Act, it is implicit from the intent of the Act that Parliamentary districts are to be contiguous and compact in shape as in criterion (b) (see, for example, Ref. [31]).

To facilitate the redistricting process, a geographic information system (GIS) is used by the Commission. The GIS allows the commission to see instantly the effects of alternative electoral boundary options. It permits the movement of meshblocks from one district to another either by "pointing and clicking" or by database query; further, it automatically recalculates the population totals for each Parliamentary district [17].

In addition, the New Zealand parliamentary electoral system has recently undergone several far-reaching changes. Parliamentary elections are typically conducted every three years. The most recent one was held in November, 1993. In that election (and in previous ones), the candidate who received a simple plurality of votes in a district was chosen to represent the district in the unicameral (i.e. single-house) Parliament. (Members to the United States House of Representatives and Senate are also elected by using the simple plurality rule.) In New Zealand, this system is referred to as the FPP or "first past the post" system.

In August of 1993, however, the New Zealand Parliament voted to adopt a new electoral system referred to as the "mixed member proportional" (MMP) system. This new system was ratified by the voters in a national referendum in November of 1993. It will take effect in the next general election. Under the MMP electoral system, there will be a total of 120 members in Parliament. Approximately half of these members will be elected from single-member Parliamentary districts (as in the FPP system) whereas the remaining half will be selected at large from lists of candidates put forward by registered political parties and chosen by voters nationwide (see Ref. [27] for further discussion of the MMP system).

A significant consequence of the MMP system is that, compared with the previous FPP system, the number of electoral districts will be reduced by roughly one-third. Thus, rather than simply adjusting the boundaries of the electoral districts that existed under the FPP system, an entirely new set of electoral districts will need to be delineated throughout the entire country. These changes have brought about a new set of opportunities—as well as challenges—for the application of operations research models to the political redistricting process. In order to place the modelling procedure discussed in this paper in perspective, we next briefly survey districting models in the operations research literature.

Operations research redistricting models

The application of operations research models to redistricting problems dates back to the early 1960s. These models can be grouped into three different approaches. The first uses an analogy to the location of facilities (representing electoral districts) and the allocation of customers (representing the small geographic units) to these facilities. This approach—first proposed by Hess *et al.* [20]—enforces criterion (a) using a transportation model and represents criterion (b) using a facility location model that minimizes an objective function that is convex with respect to distance. Other criteria are incorporated using ad hoc rules. Related approaches are described in Refs [4, 9, 15, 26]. See Ref. [21] for a recent survey.

The second approach—introduced by Garfinkel and Nemhauser [16]—uses a set partitioning framework to express the allocation of geographic units to electoral districts. The problem is formulated as a 0–1 integer programming problem. Criterion (a) is expressed as the objective function which minimizes the maximum population deviation from the mean; while criterion (b) is enforced using a proximity measure between geographic units which produces districts that are compact in shape. The other criteria are included as side constraints. This approach has also been applied to school redistricting and municipal land allocation problems [18, 35]. Note, however, that the practicality of solving large-scale 0–1 integer programs limits the size of problems that can be solved using this approach.

Finally, the third approach uses heuristic rules to construct and/or alter electoral districts using the criteria established above. Many of these approaches are most appropriate when only slight modifications of previously established electoral districts are needed. See Refs [6, 14, 19, 24] for a discussion of heuristic redistricting approaches and their effects.

Contribution and organization of this paper

The New Zealand Department of Survey and Land Information (DOSLI) has decided to adopt the optimization procedure described in this paper to assist in its preparation of preliminary Parliamentary electoral district boundaries used by the Representation Commission. The results of the optimization model are being incorporated into the GIS used by the Commission. Therefore, the emphasis throughout this work has been on the development of a practical tool that can assist the decision-makers (in this case, DOSLI and the Representation Commission) in developing electoral boundaries for Parliamentary districts. The model had to solve large-scale problems (namely, the allocation of some 35,000 meshblocks) yet, at the same time, be able to be solved very efficiently in order to be effective in a decision support role. For these reasons, we have adopted the location/allocation approach pioneered by Hess *et al.* [20].

However, we have also extended and modified this model in three significant ways. First, we have taken advantage of the large number of geographical units (i.e. meshblocks) and used a piecewise-linear objective function to resolve the undesirable effect of having "split" geographical units (i.e. meshblocks allocated to more than one electoral district). The principal output of a districting model is a set of binary variables indicating whether or not a given meshblock is allocated to a given electoral district. The model by Hess *et al.* [20], however, does not constrain these variables to be integer. Hence, it is possible for a meshblock unit to be split between two or more electoral districts. Fleischmann and Paraschis [15] suggested a postprocessing procedure for correcting geographic units that have been split. However, in this paper we show an alternative method for resolving the issue of split geographic units.

The second contribution of his paper involves the use of "barriers" to represent topographical features such as major mountain ranges or waterways. This is especially important in New Zealand, which consists of two major, and a number of smaller, islands. The coastline is very indented in places and the landmass is very mountainous, particularly the South Island with a massive mountain chain (the Southern Alps) running almost the entire length of the island. Our optimization model takes these topographical features into account by including circuity factors in the distance between meshblocks and electoral districts. In this way, the network structure (and therefore the efficiency) of the model is maintained thus allowing the procedure to provide useful results in a real-time support role.

Finally, the third, and perhaps most important, characteristic of the modelling procedure presented in this paper is its flexibility. On the one hand, previously existing electoral district boundaries can be taken into account (as in criterion (g), above) when electoral districts are being updated following the release of new population census data. But, on the other hand, the model can also be used to generate an entirely new set of electoral districts independent of the previously existing districts, as is required for the electoral district boundaries to be formed under the new MMP electoral system (described earlier). Thus, the modelling procedure can assist the electoral boundary decision-makers in a wide range of scenarios.

The organization of the remainder of the paper is as follows. The next section briefly reviews the Hess *et al.* model. We then illustrate the flexibility of the model by showing how alternative versions can be used to incorporate various boundary selection criteria Finally, we summarize the paper and comment on future work in this area.

LOCATION/ALLOCATION DISTRICTING MODEL

This section summarizes the location/allocation political districting model developed by Hess *et al.* [20] (see Fig. 1).

Initialization

In this step, the input data are specified. Let $I = \{i: i = 1, \ldots, m\}$ be the set of meshblocks, let $J = \{j: j = 1, \ldots, n\}$ be the set of electoral districts, let $(\hat{x}_i, \hat{y}_i)$ denote the x-y coordinates of the ith meshblock, and let b_i denote the population of the ith meshblock. In addition, let u_{ij}^0 denote the initial number of people in the ith meshblock that are allocated to the jth electoral district. (By construction, $\Sigma_j u_{ij}^0 = b_i$.) Also, the iteration counter, denoted by k, is set to 1.

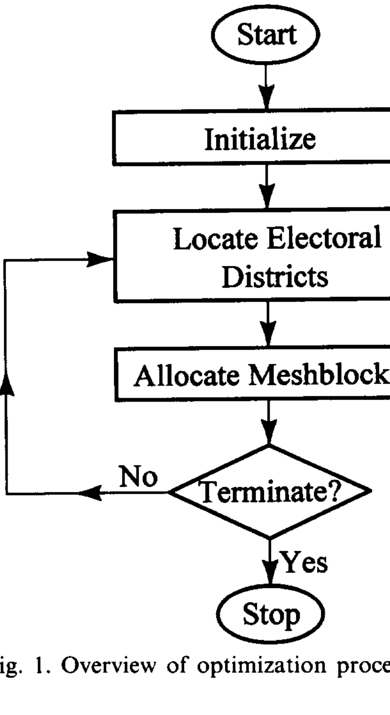

Fig. 1. Overview of optimization procedure.

Locate electoral districts

Given the coordinates of each meshblock and given the current allocation of population of the meshblocks to electoral districts, this step determines the location of district centroids using a "center of gravity" location model [34]. The reason for choosing this form of location model is 2-fold. First, this method locates electoral district centroids so as to favor compact electoral districts. Second, the x-y coordinates for each electoral district can be determined very efficiently. Let (x_j, y_j) denote the x-y coordinates of the centroid of the jth electoral district. These coordinates are the decision variables in this step. The Euclidean distance between the ith meshblock and the jth electoral district is given by

$$d_{ij} = \sqrt{(\hat{x}_i - x_j)^2 + (\hat{y}_i - y_j)^2}. \tag{1}$$

Given the current population allocation u_{ij}^{k-1}, the electoral district coordinates are determined by minimizing the population weighted squared distances between electoral districts and meshblocks. That is, the centroid for each electoral district j is found by solving the following unconstrained optimization problem:

$$\underset{(x_j, y_j)}{\text{Minimize}} \sum_i u_{ij}^{k-1} d_{ij}^2. \tag{2}$$

For each electoral district j let (x_j^k, y_j^k) denote the value of (x_j, y_j) that minimizes eqn (2). Eqn (2) strictly convex and the first order conditions can be solved explicitly for x_j^k and y_j^k. This yields

$$x_j^k = \frac{\sum_i u_{ij}^{k-1}\hat{x}_i}{\sum_i u_{ij}^{k-1}} \; \forall j, \quad y_j^k = \frac{\sum_i u_{ij}^{k-1}\hat{y}_i}{\sum_i u_{ij}^{k-1}} \; \forall j. \tag{3}$$

Once the electoral district centroids have been determined by eqn (3), the distance d_{ij}^k is computed by substituting x_j^k for x_j and y_j^k for y_j in eqn (1).

Allocate Meshblocks

Given the electoral district locations, this step determines which meshblocks should be allocated to each of the electoral districts using a minimum cost network flow model (see Fig. 2). Each meshblock is represented as a node indexed by $i \in I$ with the supply at the ith meshblock node set at b_i (the meshblock population). Each electoral district is represented as a node indexed by $j \in J$. All of the electoral district nodes are connected to a "super sink" node, denoted as node "0". The demand at this node is denoted as B, where $B = \Sigma_i b_i$.

There are two sets of decision variables measuring flows in the network. Specifically, u_{ij} represents the population flow from meshblock i to electoral district j, while v_j represents the population flow from district j to the super sink node "0". The cost functions for these two sets of variables are denoted by $f_{ij}(u_{ij})$ and $g_j(v_j)$, respectively. The specific form of these functions (used to represent the selection criteria given earlier) is described later in the paper.

Using the notation given above, the following minimum cost network flow model is used to determine how meshblocks are allocated to electoral districts.

Minimize

$$z^k = \sum_{i,j} f_{ij}(u_{ij}) + \sum_j g_j(v_j). \tag{4}$$

Subject to:

$$\sum_j u_{ij} = b_i \; \forall i, \quad \sum_i u_{ij} = v_j \; \forall j, \quad \sum_j v_j = B \tag{5}$$

$$u_{ij} \geqslant 0 \; \forall i, j, \; v_j \geqslant 0 \; \forall j. \tag{6}$$

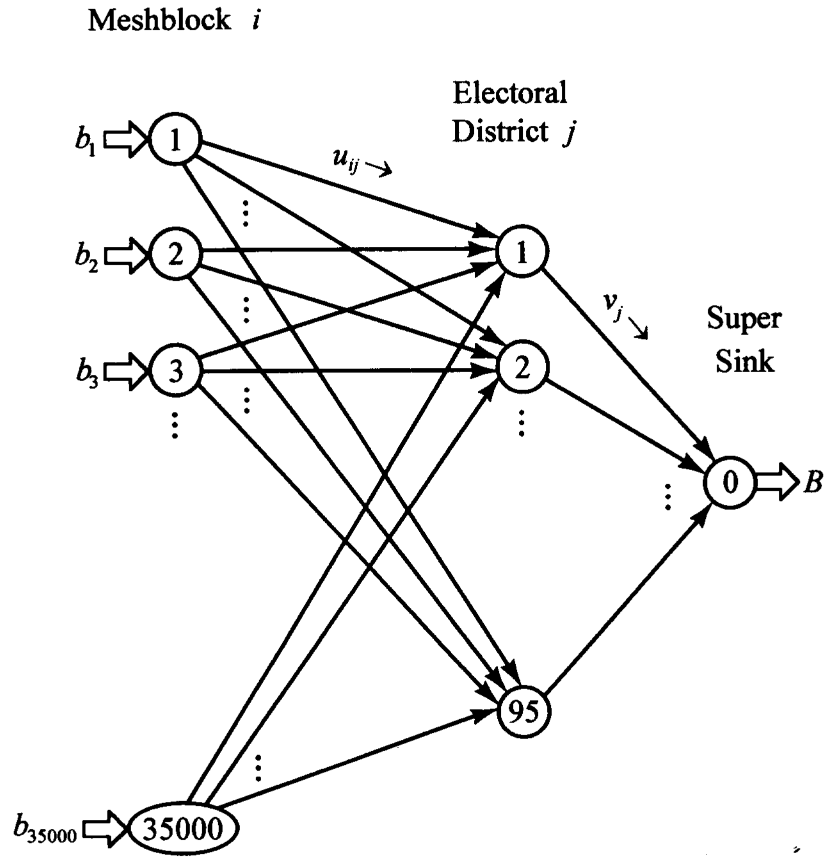

Fig. 2. Network representation for meshblock allocation step.

The objective function (4) minimizes total cost in the network; constraints (5) are flow balance equations; and constraints (6) ensure that all arc flows are nonnegative. Assuming that the cost functions $f_{ij}(u_{ij})$ and $g_j(v_j)$ satisfy certain mild convexity conditions, this network flow problem can be solved very efficiently (see, for example, Refs [1, 5]). The optimal objective function value for this problem is denoted by z^k, and the optimal arc flows are denoted by u^k_{ij} and v^k_j.

Termination

In this step, the objective function values z^k and z^{k-1} from the current and previous iterations, respectively, are compared. If the difference between these two values is sufficiently small, then the procedure stops. Otherwise, the iteration counter k is incremented by 1 and the algorithm returns to the electoral district location step for another iteration of the procedure.

If the value of u^k_{ij} is nonzero, then meshblock i is allocated to electoral district j. Note that, with this model, it is possible that several meshblocks will be "split"; that is, allocated to more than one electoral district. However, because of the large number of meshblocks compared to the number of electoral districts, such splits are very rare. To see this, note that each of the m meshblock nodes must have at least one arc emanating from it that is basic in the optimal solution to the network flow problem. Yet, the total number of basic arcs in the optimal solution is $m + n$. Thus, there are at most n meshblock nodes with multiple basic arcs emanating from that node. This, in turn, means that there are at most n meshblock nodes with multiple nonzero arcs emanating from that node. Therefore, the proportion of split meshblocks is at most n/m. For the New Zealand situation, with m approximately 35,000 and n less than 100, this ratio is less than one-half of one percent.

For the few meshblocks that are split among more than one electoral district, such meshblocks can be reassigned solely to the electoral district having the highest proportion of population for that meshblock.

The final allocation of meshblocks to electoral districts, after resolving splits, is summarized as a binary variable, denoted a_{ij}, where

$$a_{ij} = \begin{cases} 1 & \text{if meshblock } i \text{ is assigned to district } j \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

These variables are the output of the optimization procedure shown in Fig. 1. They are used as input to a GIS in order to facilitate a visual display of the electoral boundaries determined by the model. This allows alternative versions of the models to be easily assimilated and compared by decision-makers. The alternative versions of the model are discussed next.

APPLICATION OF THE MODEL

The purpose of this section is to compare the electoral districts produced by alternative versions of the optimization procedure presented in the previous section with the districts that were developed manually by the New Zealand Representation Commission. Because of the timing of this research, our evaluation of the optimization model was done in parallel with the work of the Representation Commission for the determination of the 1993 Parliamentary electoral districts. This meant that we were not privy to the boundaries being determined by the Commission; and the Commission was not influenced by the output of our optimization procedure. Thus, although the results described in this section were not part of an organized experimental plan, they still provide a very useful comparison between the manual and automated approaches.

To illustrate the similarities and differences between the two approaches, the "Tamaki" and surrounding Parliamentary electoral districts in the Auckland metropolitan area are discussed in this section. These districts were chosen for study for two reasons. First, because of significant population shifts in recent years, the 1993 electoral districts could not be simply a "carbon copy" of the previous electoral boundaries. Second, because there are pronounced topographical features in this region (namely, waterways), the determination of appropriate electoral boundaries is challenging for the manual as well as the computerized methods.

The presentation in this section is as follows. First, the previously existing electoral district boundaries are summarized. In this case, the previously existing boundaries are those that were based on the 1987 census. We next describe the electoral districts determined manually by the Representation Commission for the 1993 general election. The subsequent four (sub)sections present the results of four alternative versions of the optimization model described above. The different versions are used to illustrate how the selected boundary criteria, listed in the Introduction, can be incorporated into the modelling procedure. The input data used in these versions is comparable to that used by the Representation Commission in determining the 1993 electoral district boundaries.

Previously existing electoral districts

The data for the 1987 Parliamentary electoral boundaries were used as the previously existing districts. These data consisted of approximately 35,000 meshblocks combined into 93 electoral districts: 68 in the North Island and 25 in the South Island. A view of the 1987 "Tamaki" and surrounding electoral districts is shown in Fig. 3. The "Tamaki" district is situated about 5 km southeast of central Auckland (Fergusson Wharf and other aspects of Auckland Harbour can be seen in the upper left-hand corner of Fig. 3). This map, as with all of the maps contained in this paper, were produced using the GIS package TRANSCAD [8].

As noted earlier, subsequent to determination of the 1987 electoral district boundaries, there have been some significant population movements in New Zealand. To document these changes, information from the population census that was used to determine the 1993 electoral district boundaries was mapped onto the 1987 electoral districts. These individual district populations were then compared with a district-wide equal population target value. As mandated in the New Zealand Electoral Act, the target population value (denoted as t in this paper) was determined by dividing the total population in the South Island by 25 (the number of electoral districts in the South Island).

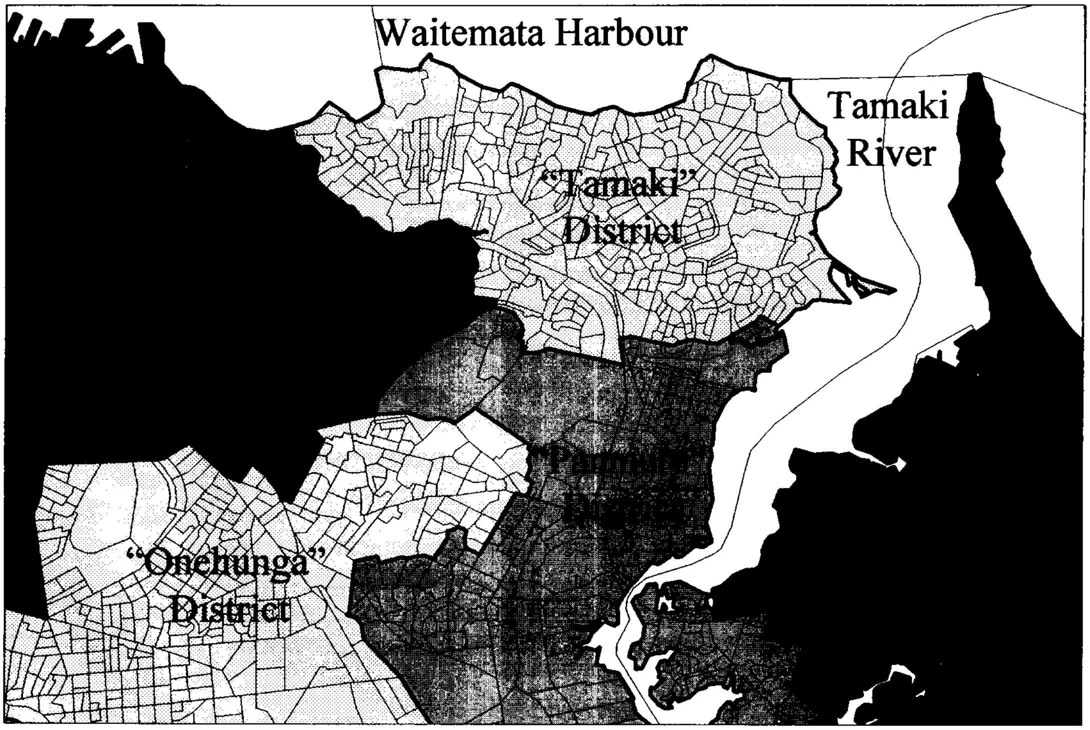

Fig. 3. Manually generated 1987 electoral boundaries for the "Tamaki" and surrounding districts.

This yielded an equal population target value of 33,336 people per district. To see the extent of the population changes between 1987 and 1993, the difference between the current population of each 1987 electoral district and the target population, expressed as a percent, was computed. Figure 4 gives the distribution of the percent deviation from the equal target value for each of the previously existing electoral districts. Figure 4 also shows the allowable deviation ($\pm 5\%$) from the

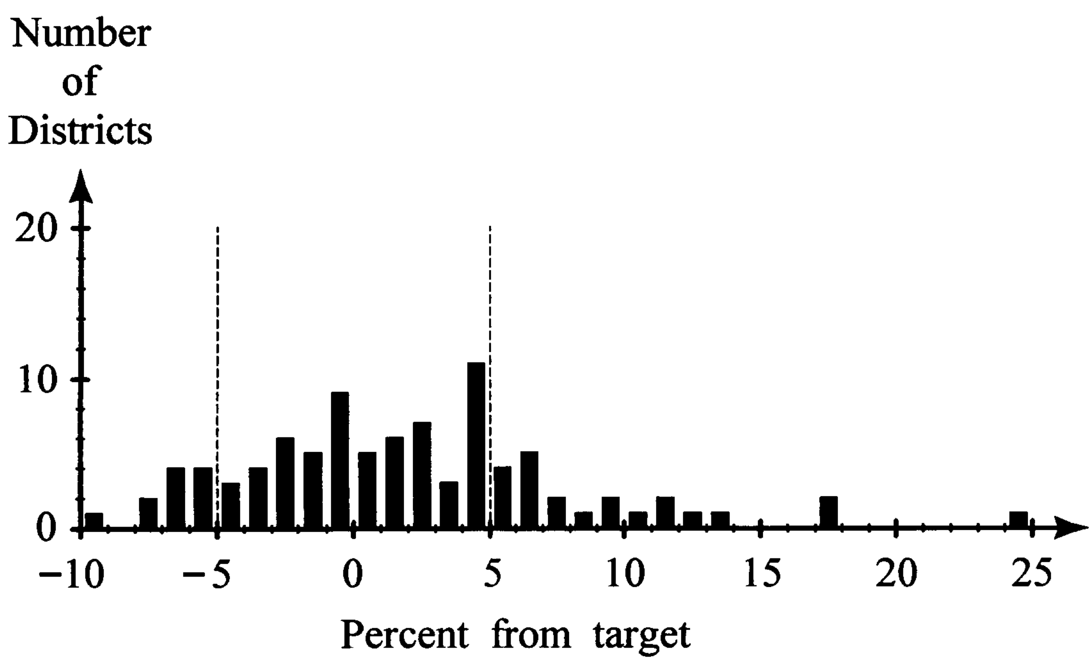

Fig. 4. Population distribution of 1987 electoral districts.

target value set by the Electoral Act. This barchart illustrated that, if the 1987 electoral district boundaries were used for the 1993 districts, then fully one-third of these districts would have a population that was outside the allowable $\pm 5\%$ tolerance. In fact, the population in several of the 1987 electoral districts had increased so substantially that by 1993 these districts had a population that was more than 20% above the equal population target value. In short, this meant that significant boundary changes were needed in a number of the 1993 electoral districts in order to produce districts that were within $\pm 5\%$ of the target value, as required by law.

Manually generated 1993 electoral districts

Because of these population changes, not only were alterations to the previously existing electoral district boundaries required, but also several new districts had to be created in the North Island. To determine the exact number of 1993 electoral districts required in the North Island, the total population in the North Island was divided by the equal population target value, t (computed previously). This ratio, rounded to the nearest integer, indicated that there should be 70 electoral districts in the North Island in 1993. In other words, compared to the previously existing electoral districts, two additional districts were required. These two additional districts were located in the Auckland metropolitan area near the areas where the greatest increases in population had occurred.

The 1993 "Tamaki" and surrounding electoral districts determined manually by the Representation Commission are shown in Fig. 5. Compared with the 1987 district boundaries (shown in Fig. 3), the southern portion of the "Tamaki" district (and, correspondingly, the northern boundary of the "Panmure" district) was altered. A principal reason for this change is that if the 1987 boundaries had remained intact, then the "Tamaki" district would have had a population nearing the $+5\%$ population limit. Hence, for the 1993 electoral district, a portion of the southeastern corner of the 1987 "Tamaki" district was transferred to the "Panmure" district. In turn, a portion of the northwestern corner of the 1987 "Panmure" district was transferred to the "Tamaki" district. This latter transfer, however, consisted of sparsely populated land (including the Remuera Golf Course). Thus, the net result was that the 1993 "Tamaki" district had a population of 32,164 individuals (well within $\pm 5\%$ of the target value of 33,336).

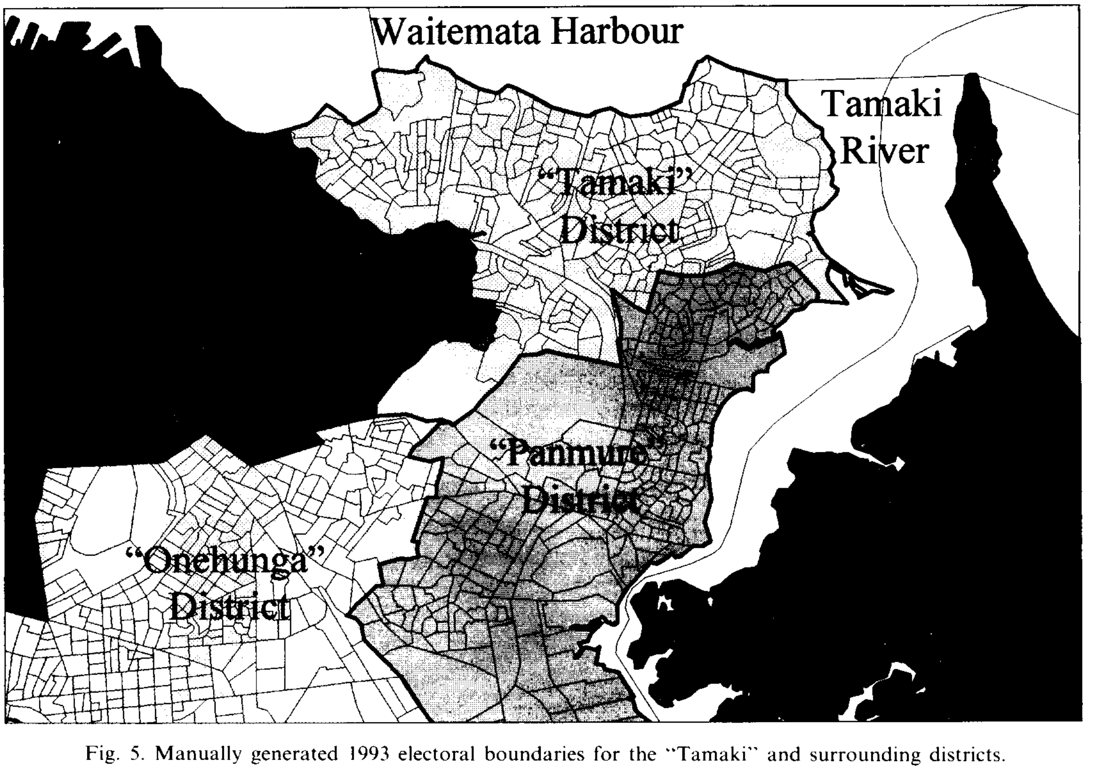

Fig. 5. Manually generated 1993 electoral boundaries for the "Tamaki" and surrounding districts.

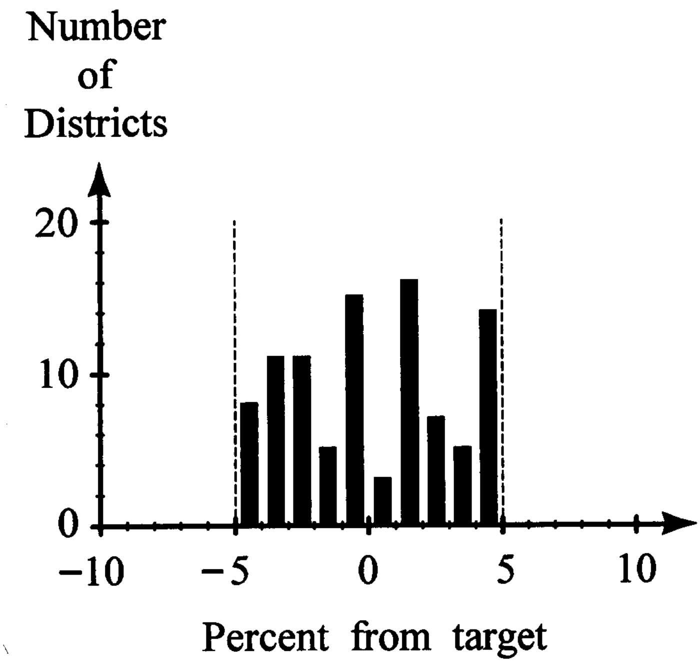

Fig. 6. Population distribution of manually generated 1993 electoral districts.

The distribution of population for all of the 1993 electoral districts produced by the Representation Commission is shown in Fig. 6. This figure shows that, for 1993, each of the 95 electoral districts is within $\pm 5\%$ of the target population value, as required by the Electoral Act. However, a number of districts have a population that is close to the allowable limit, with eight districts 5% below the target population value and fourteen districts 5% above the target population value.

Model generated 1993 electoral districts: Version 1

As a starting point for the computerized procedure, the original political redistricting model suggested by Hess *et al.* [20] was investigated (see "Location/Allocation Districting Model",

Table 1. Arc cost functions for alternative versions of the model

Version	Form of $f_{ij}(u_{ij})$	Form of $g_j(v_j)$
1	$f_{ij}(u_{ij}) = D_{ij}u_{ij}$ where $D_{ij} = \begin{cases} d_{ij} & \text{if } d_{ij} \leqslant R_j \\ \infty & \text{otherwise} \end{cases}$	$g_j(v_j) = 0$
2	Same as Version 1	$g_j(v_j) = \begin{cases} r_3 - p_3(v_j - q_1) & \text{if } q_1 \leqslant v_j \leqslant q_2 \\ r_2 - p_2(v_j - q_2) & \text{if } q_2 \leqslant v_j \leqslant q_3 \\ r_1 - p_1(v_j - q_3) & \text{if } q_3 \leqslant v_j \leqslant q_4 \\ 0 + p_1(v_j - q_4) & \text{if } q_4 \leqslant v_j \leqslant q_5 \\ r_1 + p_2(v_j - q_5) & \text{if } q_5 \leqslant v_j \leqslant q_6 \\ r_2 + p_3(v_j - q_6) & \text{if } q_6 \leqslant v_j \leqslant q_7 \end{cases}$
3	$f_{ij}(u_{ij}) = (\alpha C_{ij} + (1 - \alpha)D_{ij})u_{ij}$ where $C_{ij} = \begin{cases} \sqrt{D_{ij}} & \text{if meshblock } i \text{ was in existing district } j \\ D_{ij} & \text{otherwise} \end{cases}$	Same as Version 2
4	$f_{ij}(u_{ij}) = (\alpha C_{ij} + (1 - \alpha)D_{ij} + E_{ij})u_{ij}$ where $E_{ij} = \begin{cases} M & \text{if allocating meshblock } i \text{ to district } j \text{ crossed barrier} \\ 0 & \text{otherwise} \end{cases}$	Same as Version 2

above). For this version, the arc cost functions $f_{ij}(u_{ij})$ and $g_j(v_j)$ were specified as shown in Table 1. Specifically, $f_{ij}(u_{ij})$ was set equal to the Euclidean distance d_{ij} (see eqn (1)) between meshblock i and electoral district j times the population flow from meshblock i to district j for meshblocks within a prespecified radius R_j of electoral district j. By limiting meshblocks to be within a prespecified radius R_j, the model did not allocate a meshblock to an electoral district if the distance between the meshblock and the district was too great. This has the effect of reducing the number of arcs (i.e. decision variables) in the problem. For our problems, R_j ranged from around 25 km for metropolitan areas to as much as 600 km for rural and mountainous regions. In addition, for Version 1, $g_j(v_j)$ was set equal to zero and lower and upper bounds were set on each electoral district population flow variable v_j. Specifically,

$$0.95t \leqslant v_j \leqslant 1.05t \ \forall j, \tag{8}$$

where the equal population target value, t, was computed using the procedure described in the section "Previously Existing Electoral Districts". These bounds ensured that the population in each electoral district was within the $\pm 5\%$ allowable tolerance.

For this version of the model (as well as the other versions), the algorithm flowcharted in Fig. 1 was coded in FORTRAN [10]. The meshblock allocation procedure was solved using RELAXT, a capacitated, minimum cost, linear network flow code, as a subroutine [5].

Version 1 of the model was used to generate 95 electoral districts (70 in the North Island and 25 in the South Island). In the initialization step of the procedure, the location of the electoral district centroids was taken as the geographic center of the 1987 electoral districts. The initial location of the centroids of the two new electorates in the North Island was taken as the midpoint between the 1987 electoral districts with the greatest population increase. Because the North and South Islands in New Zealand do not have any electoral districts in common, the model was run separately for each island. However, this still resulted in a fairly large model, with 25,176 meshblocks being allocated to 70 electoral districts in the North Island and 9,960 meshblocks being allocated to 25 districts in the South Island. The solution times for the North and South Islands were approximately 15 and 5 min, respectively, on a VAX 7610. The output of the models were the binary variables a_{ij} indicating which meshblocks had been allocated to which electoral districts (see eqn (7)).

The population distribution of electoral districts generated by Version 1 (computed as $\Sigma_i a_{ij} b_i$) is summarized in Fig. 7. This figure shows the distribution of the percent deviation of the district population from the equal population target. Note that, because of the linear form of the objective function, this version of the model produced electoral districts whose population tended to "cluster" around either the $+5\%$ tolerance level or the -5% tolerance level. Moreover, once the "split" meshblocks were resolved (as discussed previously), a number of electoral districts actually exceeded the allowable population tolerance. This was true in spite of the fact that only a small percent of meshblocks were actually split (approximately one-quarter of one percent).

The difficulty of having the population actually exceed the allowable tolerance could be corrected by making the lower and upper flow bounds in eqn (8) somewhat tighter than $\pm 5\%$. Then, after resolving the "split" meshblocks, the population of each electorate would remain within the $\pm 5\%$ tolerance. However, even if the tolerances in eqn (8) were made tighter, an inherently undesirable characteristic of this version of the model is the fact that the population of the electoral districts generated by the procedure will tend to "cluster" at either the lower or upper population limit (again, refer to Fig. 7). Thus, although this version provided a useful starting point, more sophisticated versions of the modelling procedure were considered necessary.

Model generated 1993 electoral districts: Version 2

In order to overcome the "clustering" effect associated with the model in Version 1, O'Donoghue and Wallace [30] proposed a graduated penalty function for electoral districts whose population deviated above or below the target population. We refer to this extension as Version 2 of the model. For this version, the form of $f_{ij}(u_{ij})$ remained the same as in Version 1. However, to incorporate the graduated penalties, the arc cost function $g_j(v_j)$ shown in Table 1 was used. A representative form of the function $g_j(v_j)$ is shown in Fig. 8. Here, q_4 is the same as the equal population target value t; q_1 through q_3 are population flow limits below the target value; and q_5 through q_7 are

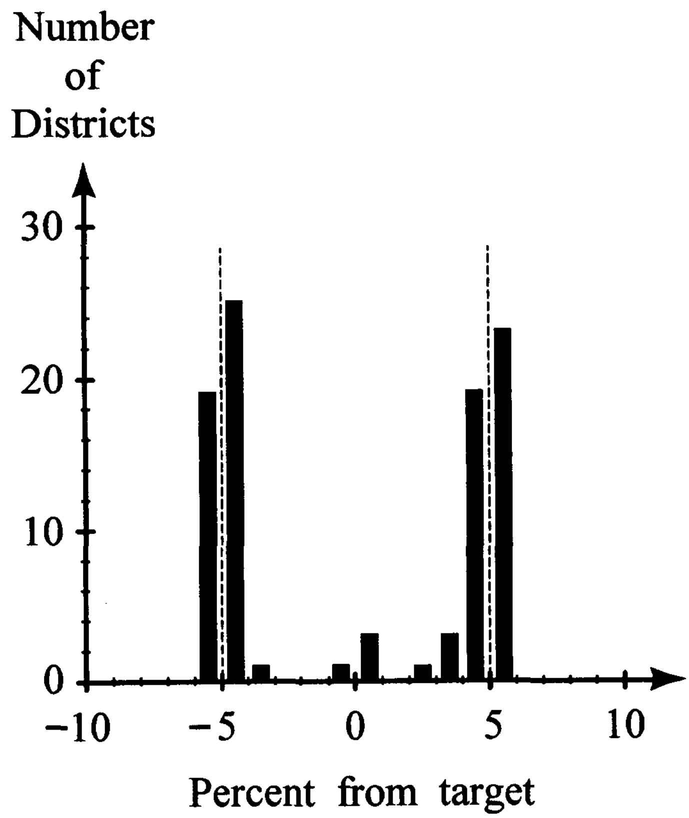

Fig. 7. Population distribution of model generated 1993 electoral districts: Version 1.

population flow limits above the target value. The slopes p_1 through p_3 are successively increasing penalty costs associated with having the population of an electoral district deviate from the target level.

Although $g_j(v_j)$ is not linear, it is well known (see, for example, Ref. [1]) that the network in Fig. 2 can be converted to a linear form by replacing each arc from an electoral district node j to the super sink node "0" with a set of parallel arcs as shown in Fig. 9. The costs of the arcs in Fig. 9 are commensurate with the marginal costs in function $g_j(v_j)$; and the capacities of the arcs in Fig. 9 are commensurate with the flow intervals in $g_j(v_j)$.

Version 2 of the model was solved to determine the 95 electoral districts using the same procedure as described in the previous section. The solution times for the North and South Islands were approximately the same as for Version 1 of the model.

The distribution of the percent deviation of the population of the electoral districts generated by this version of the model is shown in Fig. 10. Note that, with this version of the model, the population of each of the 95 electoral districts is no more than $\pm 1\%$ from the target population value, well within the allowable $\pm 5\%$ tolerance.

In short, the problem of "split" geographical units inherent in the Hess *et al.* [20] model has been eliminated by (i) solving a problem in which the number of geographic units (i.e. meshblocks) is large compared to the number of districts, and (ii) by using a piecewise linear penalty cost function to assign population flows to districts.

To see the specific results of this model, the 1993 "Tamaki" and surrounding electoral districts produced by this version of the model are shown in Fig. 11. It is interesting to compare Fig. 11 with the 1987 electoral districts in Fig. 3. There are some distinct differences in the electoral boundaries in the two figures. This is not surprising since the model in Version 2 did not have any information as to which meshblocks were assigned to which of the 1987 electoral districts. In fact, it is surprising that there is not a greater difference between the districts generated by this model

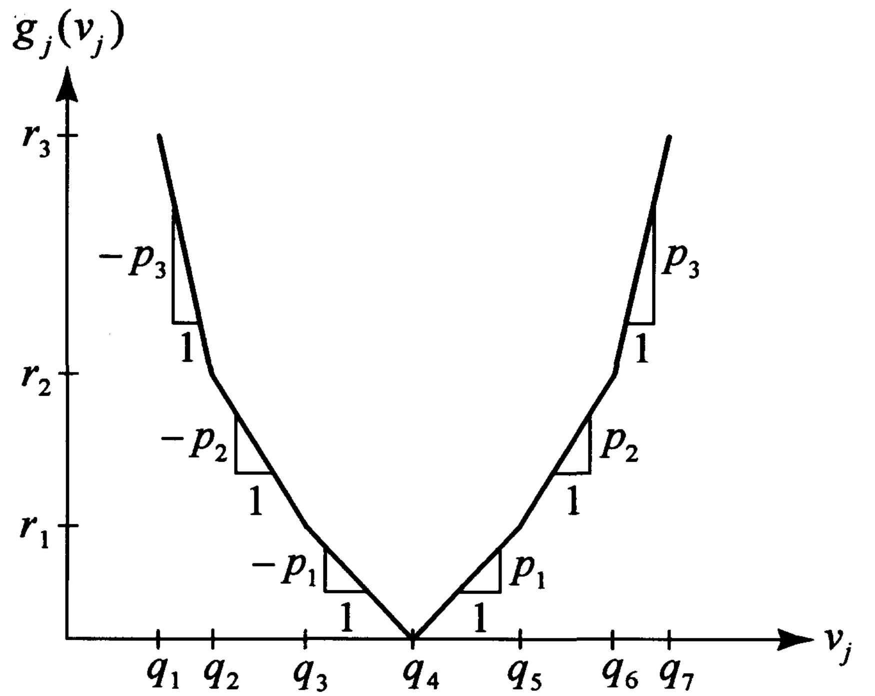

Fig. 8. Form of penalty cost function.

and the 1987 electoral boundaries, given that the model did not use any information from the 1987 electoral districts (other than the initial location of the electoral district centroids).

Figure 11 highlights two important features that are not taken into account in Version 2 of the model. First, the electoral districts generated by Version 2 have what could be described as "rough edges". One method of producing less jagged boundaries is to use information from the previous electoral districts which were produced manually. This issue is addressed in Version 3 of the model. The second aspect of Version 2 of the model brought out in Fig. 11 is the fact that certain electoral districts have crossed over a topographical feature, in this case the Tamaki River. This phenomenon occurs because Version 2 of the model allocates meshblocks to electoral districts based on the Euclidean distance between the meshblock and the district centroid irrespective of any topographical features in place. In some cases, this may be acceptable. For instance, it is reasonable for the "Panmure" district to cross over to the eastern bank of the Tamaki River since the river is less than 200 meters wide at that location, and there are two major roads crossing the river (highways 5 and 10A). In fact, the manually produced 1987 electoral districts also included land on both sides of the Tamaki River in the "Panmure" district (see Fig. 3). It is less reasonable, however, to have an electoral district cover both banks of the Tamaki River near the mouth of this river (which is more than a kilometer wide there). Thus, it may be undesirable to have the "Pakuranga" district include both the eastern and western sides of the Tamaki River as is the case in Fig. 11. Methods for controlling these "cross-over" effects are considered in Version 4 of the model.

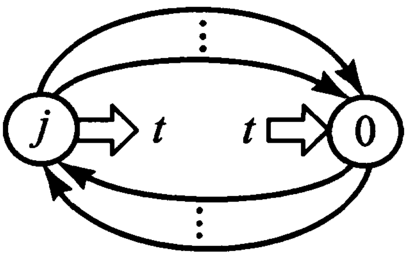

Fig. 9. Parallel arc representation of piecewise linear arc cost function.

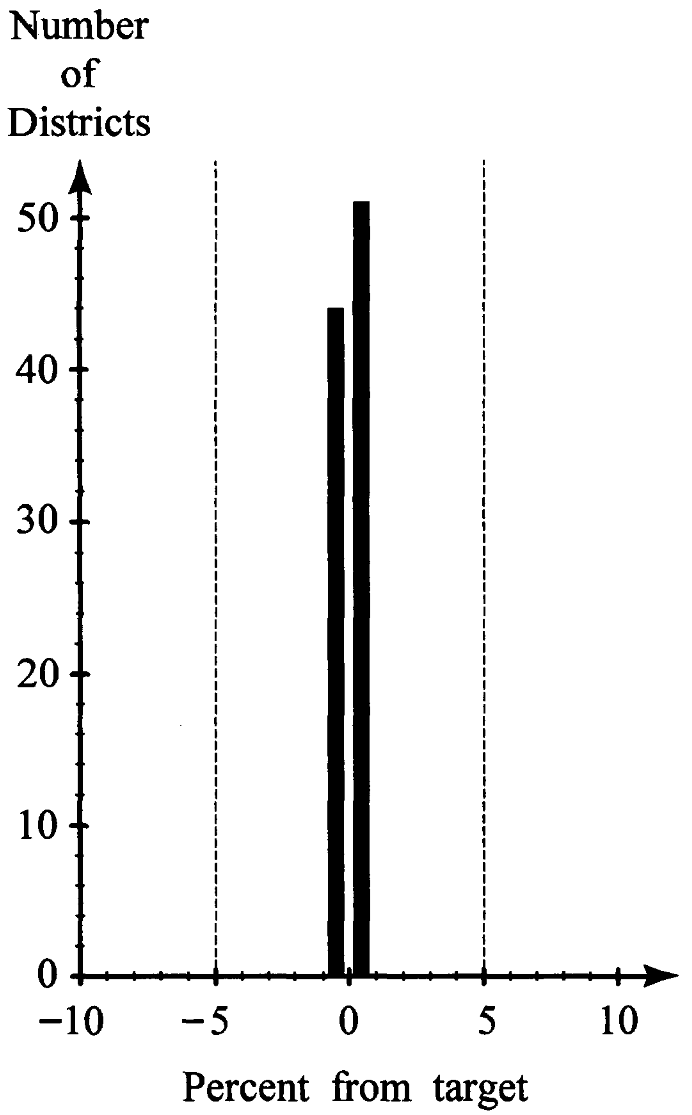

Fig. 10. Population distribution of model generated 1993 electoral districts: Version 2.

Model generated 1993 electoral districts: Version 3

Version 3 of the model extended the previous version by explicitly incorporating information about the previously existing (i.e. in this case, the 1987) electoral boundaries (see criterion (g), in the Innovation). Here, each meshblock in the model received an incentive if the population from that meshblock was assigned to its previously existing electoral district. This incentive structure was incorporated within the model by using the forms of $f_{ij}(u_{ij})$ and $g_j(v_j)$ shown in Table 1 for Version 3. Here, the parameter α specified in $f_{ij}(u_{ij})$ was a weighting factor between 0 and 1. Small values of α provided little incentive for a meshblock to be assigned to its existing electorate, while large values of α provided a strong incentive for the model to assign a meshblock to its existing electorate. Note though, that for any value of α, $f_{ij}(u_{ij})$ remains a linear function of u_{ij}. Thus, the network model for this version of the procedure can also be solved very efficiently.

A series of computational tests was performed to determine the most appropriate value for the weighting factor, α. These tests indicated that α should be set very close to 1 in order to ensure that the electorates generated reflected the shape of the previously existing electoral districts. The results given in this paper are for a value of $\alpha = 1$.

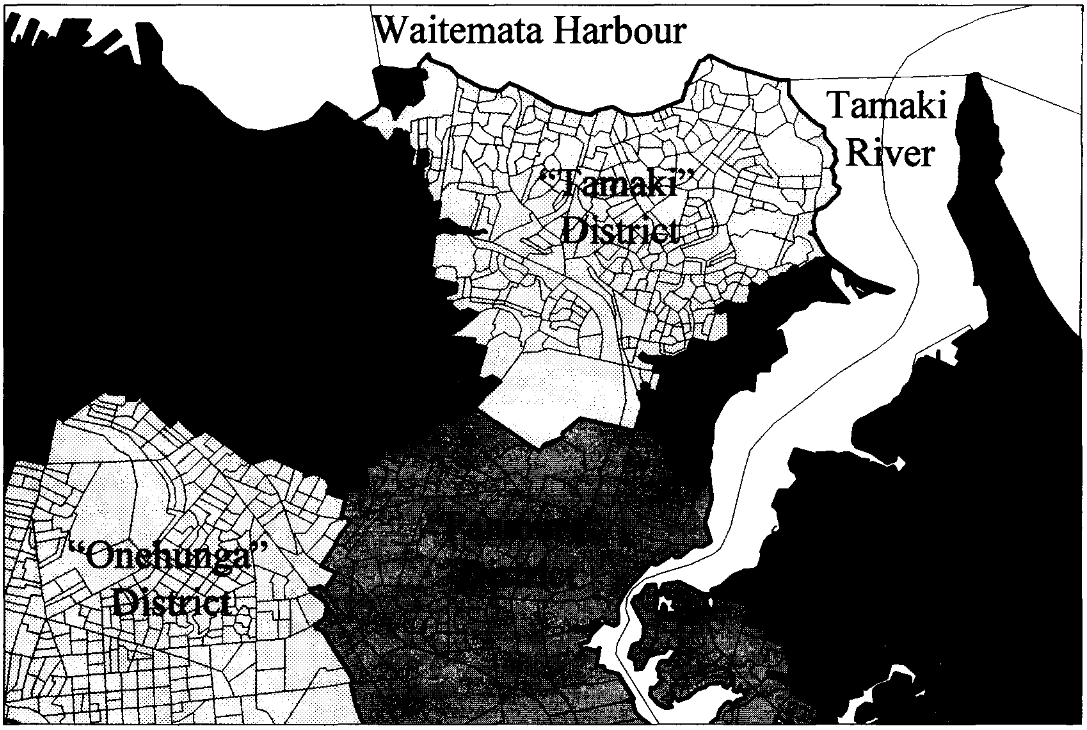

Fig. 11. Model generated 1993 electoral boundaries for the "Tamaki" and surrounding districts: Version 2.

The solution time for this version was approximately the same as for the previous versions of the model. Moreover, as with Version 2, the population of each electoral district was well within the $\pm 5\%$ tolerance, as typified by Fig. 10.

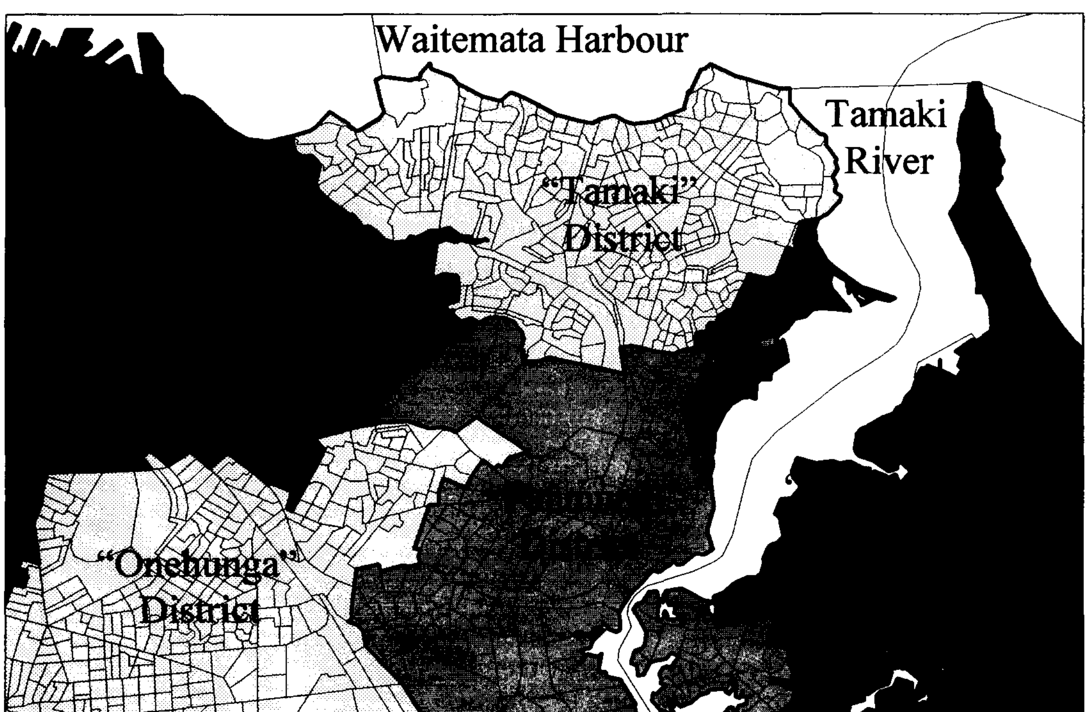

Fig. 12. Model generated 1993 electoral boundaries for the "Tamaki" and surrounding districts: Version 3.

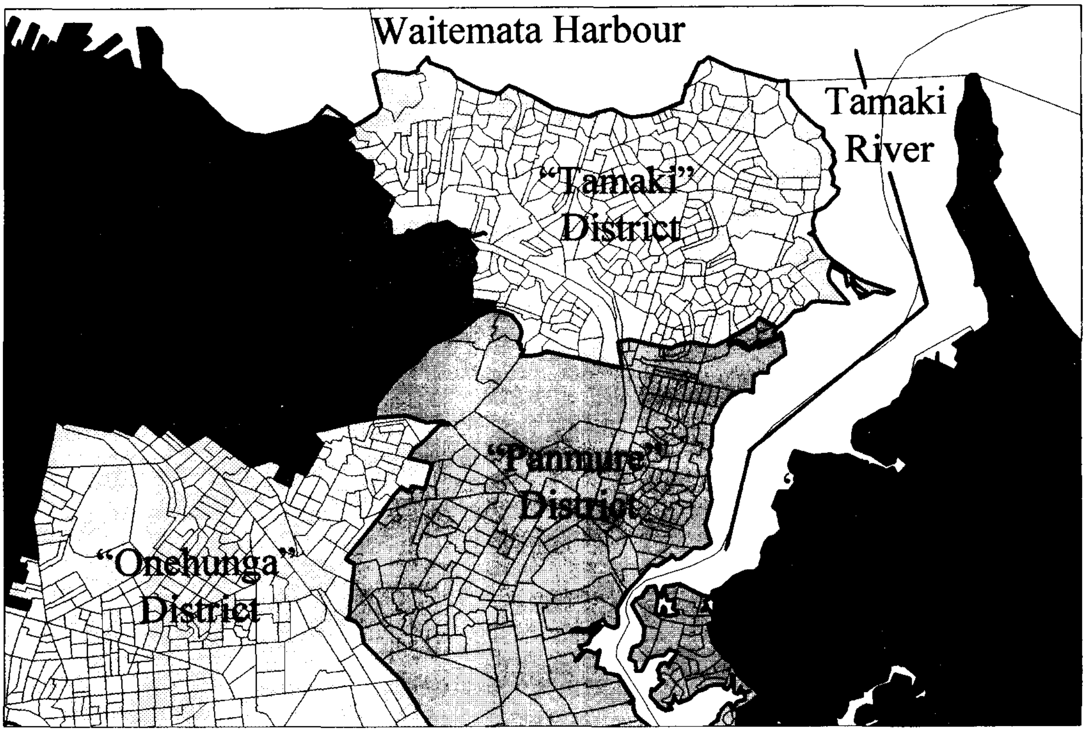

Fig. 13. Model generated 1993 electoral boundaries for the "Tamaki" and surrounding districts: Version 4.

Figure 12 shows the electoral district boundaries generated by Version 3 for the 1993 "Tamaki" and surrounding electoral districts. Comparing the results of Version 3 (Fig. 12) with the previous existing electoral districts (Fig. 3), it can be seen that the electoral districts generated by the model have boundaries that adhere fairly closely to the 1987 boundaries while, at the same time, have populations that are well within the $\pm 5\%$ of the target population value.

The one pronounced difference between Figs 12 and 3 is the crossing of the "Pakuranga" district onto the western bank of the Tamaki River. This issue is considered next.

Model generated 1993 electoral districts: Version 4

The purpose of this final version of the modelling procedure is to include information about topographical features such as major waterways and mountain ranges. This information was included in the modelling procedure by representing these features as "barriers". Additional examples of the use of barriers are contained in Refs [17, 30, 33].

In Version 4, if the line segment between a meshblock centroid and an electoral district centroid crossed a barrier, then the distance between that meshblock-district pair was increased in order to discourage the model from assigning that meshblock to that electoral district. Specifically, $f_{ij}(u_{ij})$ was changed to the form shown for Version 4 in Table 1. In this equation, M is a large positive constant.

Version 4 of the model was also used to generate the 95 electoral districts. The running time and the percent deviation from the target population were similar to those for Versions 2 and 3 of the model. Thus, once again, the electoral district populations generated by the model were well within allowable tolerances.

The electoral district boundaries generated by Version 4 for the 1993 "Tamaki" and surrounding districts is shown in Fig. 13. The barriers used to represent the Tamaki River are also shown in the figure. The use of the barriers has prevented the unwanted "cross over" effects that were present in the model used in Versions 2 and 3. Yet, at the same time, the existing 1987 boundaries are still fairly closely adhered to (compare Fig. 13 with Fig. 3).

It is also quite interesting to compare the 1993 districts determined by the Representation Commission (Fig. 5) with those generated by Version 4 of the model (Fig. 13). Even though the two sets of electoral boundaries were prepared independently—one manually and the other by an optimization model—there are a remarkable number of similarities. The principal differences are in the southern border of the "Tamaki" district and the extent to which the "Panmure" district crosses over the Tamaki River.

In general, alternative versions of the model can be used to represent the various boundary criteria given at the beginning of the paper. We acknowledge, however, that some of these criteria are hard to quantify or express explicitly. Therefore, the results produced by the models described in this paper should be considered only as a starting-point. Further, it is noteworthy that the use of this modelling procedure is in a real-time decision support role to assist with the development of appropriate electoral district boundaries.

CONCLUSIONS AND FUTURE WORK

This paper has described a method for automating parts of the initial stages of the political redistricting process. The procedure involved a set of network-based optimization programs that were used to generate electoral districts that were roughly compact in shape and had approximately equal populations. Alternative versions of the model were considered that took into account the boundaries of previously existing electoral districts as well as topographical features such as major waterways and mountain ranges. The procedure was used to generate a set of electoral districts in New Zealand based on data for the 1993 Parliamentary elections. Although the problem was quite large (involving over 35,000 meshblocks and nearly 100 electoral districts), the optimization procedure was able to generate the electoral districts quite efficiently (approximately 20 min on a VAX 7610 computer). The districts generated by the model were displayed using a geographic information system (GIS).

The districts produced by the optimization procedure were compared with the actual electoral districts for the 1993 Parliamentary elections, which were determined manually with the assistance of a GIS system. Although the districts generated by the manual and the automated approaches were not identical, there was enough similarity between the two sets that the New Zealand Department of Survey and Land Information (DOSLI) has decided to use the optimization procedure in its preparation of preliminary electoral district boundaries produced following each five-year census. Moreover, under a new "mixed member proportional" (MMP) electoral system, the number of electoral districts in New Zealand will be reduced by roughly one-third. DOSLI is thus using the procedures described in this paper to assist in the determination of an entirely new set of electoral districts.

Our future work in this area will focus on adapting the model to reflect other electoral district boundary criteria of the type listed earlier in the paper. For instance, the current version of the model does not take into account the boundaries of cities, counties, or local wards. To reflect these local boundaries in the model, a concave objective function can be used in place of the linear function described here. The concave objective function favors the assignment of a meshblock within a local ward to a single Parliamentary district but does not specify any given district *a priori*. Initial testing of this method has been favorable, and we are currently investigating solution methods for this class of nonconvex network flow problems [23].

Acknowledgements—The authors wish to thank the New Zealand Department of Survey and Land Information, and the New Zealand Department of Statistics for their assistance. We also wish to thank the two anonymous referees. Their comments have resulted in an improved paper.

REFERENCES

1. R. K. Ahuja, T. L. Magnanti and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, New Jersey (1993).
2. M. L. Balinski and H. P. Young. Criteria for proportional representation. *Operations Research* **27**, 80–95 (1979).
3. M. L. Balinski and H. P. Young. *Fair Representation: Meeting the Ideal of One Man, One Vote*. Yale University Press, New Haven, CT (1982).
4. S. Benabdallah and J. R. Wright. Multiple subregion allocation models. *Journal of Urban Planning and Development* **118**, 24–40 (1992).

5. D. P. Bertsekas and P. Tseng. The relax codes for linear minimum cost network flow problems. *Annals of Operations Research* **13,** 125–190 (1988).
6. M. A. Busteed. *Geography and Voting Behaviour*. Oxford University Press, London (1975).
7. D. Butler and B. E. Cain. *Congressional Redistricting: Comparative and Theoretical Perspectives*. Macmillan Publishing, New York (1992).
8. Caliper Corporation. *TransCAD Transportation Workstation Software, Reference Manual, Version 2.1.* Caliper Corporation, Newton, MA (1990).
9. J. T. Diamond and J. R. Wright. Efficient land allocation. *Journal of Urban Planning and Development* **115,** 81–96 (1989).
10. Digital Equipment Corporation. *VAX Fortran Language Reference Manual, Version 5.7.* Digital Equipment Corporation, Maynard, MA (1991).
11. R. G. Dixon, Jr. Fair criteria and procedures for establishing legislative districts. In *Representation and Redistricting Issues* (Edited by B. Grofman, A. Lijphart, R. B. McKay, and H. A. Scarrow), pp. 7–20. Lexington Books, Lexington, MA (1982).
12. S. Eilon, C. D. T. Watson-Gandy and N. Cristofides. *Distribution Management: Mathematical Modelling and Practical Analysis*. Hafner Publishing, New York (1971).
13. L. R. Ernst. Apportionment methods for the house of representatives and the court challenges. *Management Science*. **40,** 1207–1227 (1994).
14. J. A. Ferland and G. Guénette. Decision support system for the social districting problem. *Operations Research* **31,** 15–21 (1990).
15. B. Fleischmann and J. N. Paraschis. Solving a large scale districting problem: a case report. *Computers and Operations Research* **15,** 521–533 (1988).
16. R. S. Garfinkel and G. L. Nemhauser. Optimal political districting by implicit enumeration techniques. *Management Sci.* **16,** B495–B508 (1970).
17. J. A. George, B. W. Lamar and C. A. Wallace. Determining New Zealand electoral districts using a network-based model. *Proceedings of the Operational Research Society of New Zealand*. Auckland, New Zealand, August 1993, vol. 29, pp. 276–283.
18. K. C. Gilbert, D. D. Holmes and R. E. Rosenthal. A multiobjective discrete optimization model for land allocation. *Management Science* **31,** 1509–1522 (1985).
19. M. Hale. Computer methods of districting. In *Reapportioning Legislatures: A Consideration of Criteria and Computers* (Edited by H. D. Hamilton), pp. 96–124. Merrill Books, Columbus, OH (1966).
20. S. W. Hess, J. B. Weaver, H. J. Siegfeldt, J. N. Whelan and P. A. Zitlau. Nonpartisan political redistricting by computer. *Operations Research* **13,** 998–1006 (1965).
21. R. S. Howick and M. Pidd. Sales force deployment models. *European J. Operational Res.* **48,** 295–310 (1990).
22. M. E. Jewell. Criteria reflected in recent apportionments. In *Reapportioning Legislatures: A Consideration of Criteria and Computers* (Edited by H. D. Hamilton), pp. 14–23. Merrill Books, Columbus, OH (1966).
23. B. W. Lamar. An improved branch and bound algorithm for minimum concave cost network flow problems. *Journal Global Optimization* **3,** 261–287 (1993).
24. S. P. Leach and A. Kandel. GROUPER: an expert system for redistricting, *Proceedings of the 1990 ACM SIGSMALL/PC Symposium on Small Systems*. Arlington, Virginia, March (1990) pp. 6–12.
25. A. Lijphart. Comparative perspectives on far representation: the plurality–majority rule, geographical districting, and alternative electoral arrangements. In *Representation and Redistricting Issues* (Edited by B. Grofman, A. Lijphart, R. B. McKay, and H. A. Scarrow), pp. 143–159. Lexington Books, Lexington, MA (1982).
26. P. G. Marlin. Application of the transportation model to a large-scale "districting" problem. *Computers and Operations Research* **8,** 83–96 (1981).
27. A. McRobie. Electoral system options: three proposals for reforming the electoral process. In *New Zealand Politics in Perspective* (Edited by H. Gold), pp. 452–472, third edn. Longman Paul, Auckland, New Zealand (1992).
28. R. L. Morrill. Ideal and reality in reapportionment. *Annals of the Association of American Geographers* **63,** 463–477 (1973).
29. R. L. Morill. *Political Redistricting and Geographic Theory*. Association of American Geographers, Washington, DC (1981).
30. I. P. O'Donoghue and C. A. Wallace. The Automatic Generation of Electorates. Project report, Department of Management, University of Canterbury, Christchurch, New Zealand, November (1992).
31. J. B. Ringer. *An Introduction to New Zealand Government*. Hazard Press, Christchurch, New Zealand (1991).
32. W. B. Saxbe. Criteria established by court decisions. In *Reapportioning Legislatures: A Consideration of Criteria and Computers* (Edited by H. D. Hamilton), pp. 24–37. Merrill Books, Columbus, OH (1966).
33. C. A. Wallace and B. W. Lamar. Automatic Electoral Boundary Generation: Project Report. Report prepared under contract no. 6500-11-2 to the Department of Survey and Land Information, Wellington, New Zealand, September (1993).
34. A. Weber. *Theory of the Location of Industries*. Tübingen, Germany (1909).
35. J. R. Wright, C. ReVelle and J. Cohon. A multiobjective integer programming model for the land acquisition problem. *Regional Science and Urban Economics* **13,** 31–53 (1983).