A Cluster Analysis Applied to Determining Water Distribution Districts

by

Norio OKADA*

(Received May 31, 1980)

This paper deals with municipal water supply development and analyzes those factors which are considered most significant in zoning the service area into several water supply districts.

It has been shown that some cluster analysis methods may be effectively used in determining those significant factors qualitatively and quantitatively and in grouping the service area into some sub-areas as water supply districts. The methods may provide the water supply engineer with a powerful aid in this zoning practice, which otherwise involves much of intuitive assessment by the experienced engineer.

1. Introduction

Municipal water supply development involves a wide range of engineering practices. Various kinds of computer-aided techniques have been developed to systematize these relevant engineering practices. Much room has still been left, however, for the systematization of planning, designing and operating water supply facilities. A typical practical need for this type of development is seen in the planning and designing practice of zoning the sevice area into several subareas called water supply districts.

This zoning practice still involves much of intuitive assessment by the experienced water supply engineer. This intuitive work becomes necessary when there is much difference in topography on which hydraulic conditions and network pattern largely depend ; when there is much diversity in the areal distribution of population and industry which determines much of the water demand pattern ; and when the entire service area is excessively large in light of securing minimum level of service for each user ; or when the service area is expanding outwards step by step. such being the case, the zoning of the entire service area serves the purpose of offering each user a balanced level of service. It also works for securing a higher level of safety water supply against the breakdown of the total function.

* Department of Marine Civil Engineering, Tottori University.

The objective of this study is to develop some systematic tool to guide the engineer in carrying out this complicate practice, which otherwise had to undergo more subjective and unspecified assessment by the experienced engineer.

2. Study Area

To set up a concrete context to study this zoning practice we will take Tottori City as the study area.[1]

The water supply history of Tottori City dates back to 1915 when the first enterprise of water supply was completed and the then 50,000 population (some 70 percent of Tottori City) started enjoying its benefit. Since then there have been series of expansions of water supply facilities and service areas, as the population and the city area have gradually grown. As of 1975 the service population amounted to 111,5000 (93 percent of total population) and there is a maximal capacity of 63,330 m^3/day with a design per-capita supply of 630 *l*. **Fig. 1** shows the entire service area of the present municipal water supply system of Tottori City with seven water supply districts. **Fig. 2** shows the topography of the service area with the major road network. The areal distribution of water demand is sketched in **Fig, 1.**

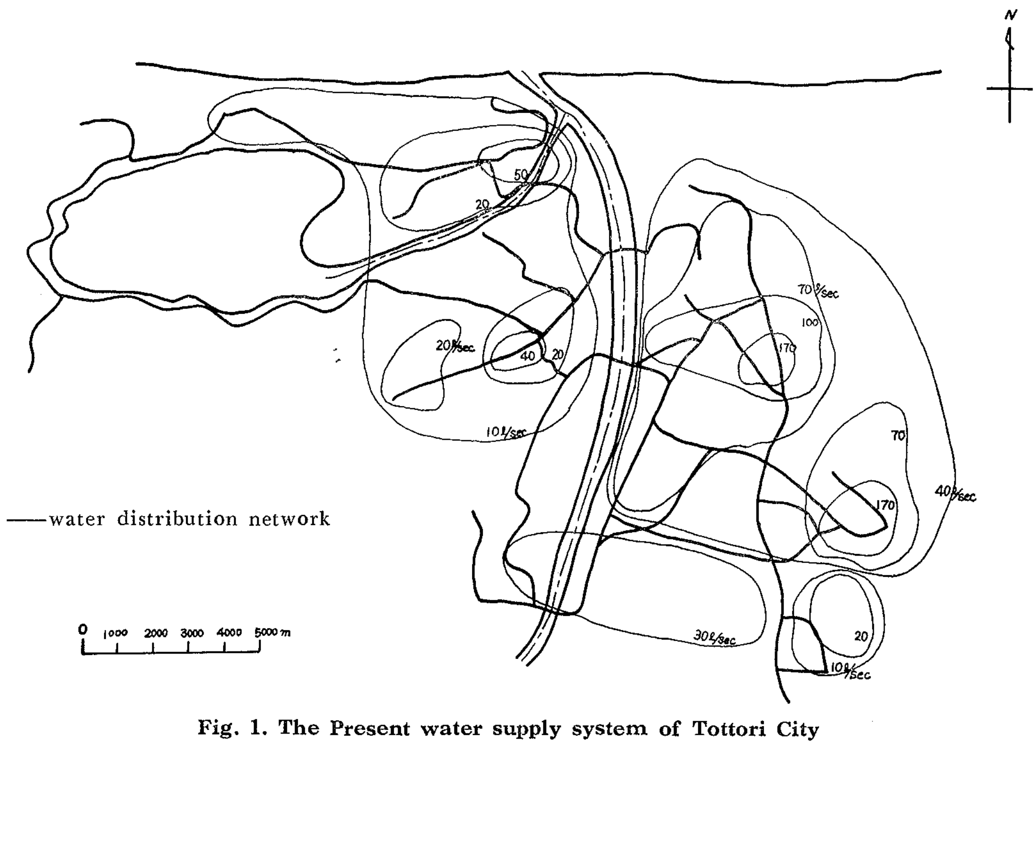

Fig. 1. The Present water supply system of Tottori City

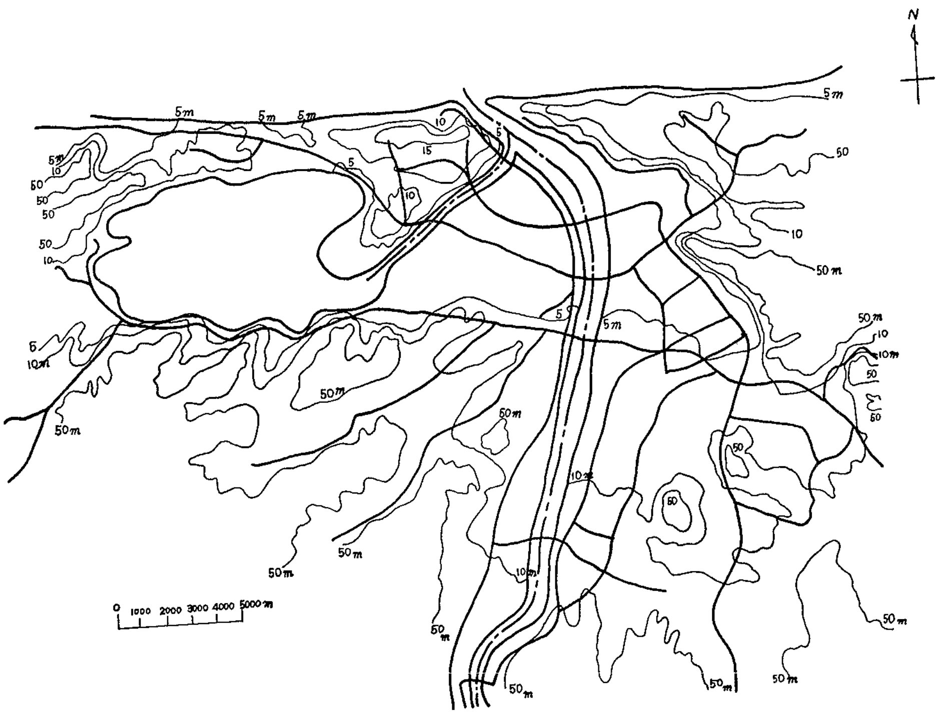

Fig. 2. The topography of the service area with the major road network

In expectation of the continuing growth of population and of the growing need for improved water supply service, the 7th expansion project is in progress with the target year of 1985. This project, when completed, will service another 24,000 population with an additional capacity of 66,000 m³/day. The focus of the development seems to be placed on the peripheral districts to meet with the outward expansion of the city area, especially the southmost district.

3. Cluster Analysis

We will see later that there is a good analogy between the practice of zoning a water supply service area and the clustering of statistical samples (items) by use of cluster analysis. Bearing this in mind let us first study cluster analysis briefly.

Given n items represented by p attributes (let each item be denoted by $\{X_{i\alpha}\}$ for $\alpha = 1, \cdots\cdots, n$; $i = 1, \cdots\cdots, p$), let us imagine that the n points are located on the p-dimensional space. The problem of our concern is to group the points into clusters by putting together those points which are considered "mutually close". By analogy let us define the degree of closeness by use of Eucledean distance. There are many variants of cluster analysis derived from different definitions of distance. The most basic type of distance is defined for a given pair of points as :

$$d_{\alpha\beta}=\sqrt{\sum_i (X_{\alpha i}-X_{\beta i})^2} \quad (i=1,\cdots,p), \cdots\cdots\cdots\cdots\cdots\cdots\cdots (1.1)$$

which is precisely identical to Euclidean distance.

With slight modification the weighted Euclidean distance is defined as :

$$d_{\alpha\beta}=\sqrt{\sum_i w_i (X_{\alpha i}-X_{\beta i})^{2u_i}} \quad (i=1,\cdots,p), \cdots\cdots\cdots\cdots\cdots (1.2)$$

where w_i and u_i (i = 1, ……, p) are weights to be assigned to ith attribute. Later we will apply this distance to our zoning (clustering) problem.

With the criterion of distance so defined we then proceed to the clustering process. In the first step all items are considered individual clusters, to get a cluster set C_0^1.

$$C_0^1=(A_1, A_2, \cdots, A_\alpha, \cdots, A_\beta, \cdots, A_n) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (1.3)$$

where $\{A_\alpha\} = \alpha$ (α = 1, …… 1, n), which means that there is only a single member in cluster A_α.

Let us assume that by calculating the distance between all possible pairs of points, we know that the distance between α and β is the smallest. Then we group α and β into a new cluster γ. Since the new cluster γ has multiple (two) members, we want to locate a representative member (reference point). Generally let us assume that cluster A_α with n_α members and cluster A_β with n_β have been aggregated into a new cluster A_γ. Although there are many possibilities to locate the reference point in a cluster, we let the centroid of the member points represent the cluster. That is, X_i (i=1,……,p), the reference point of the new cluster, is obtained by the formumula :

$$\begin{aligned} X\gamma i &= (n_\alpha X_{\alpha i}+n_\beta X_{\beta i})/(n_\alpha+n_\beta) \\ &= (n_\alpha X_{\alpha i}+n_\beta X_{\beta i})/n_\gamma \end{aligned} \quad (i=1,\cdots,p), \quad \cdots\cdots\cdots\cdots\cdots\cdots (1.4)$$

where n_γ is the number of the members of cluster A_γ.

Accordingly the distance detween cluster γ and any cluster ν with a single member or a reference point which has been synthetically located in the foregoing step is calculated as :

$$\begin{aligned} X_{\nu\gamma} &= \sqrt{\sum_i (X_{\nu i}-X_{\gamma i})^2} \\ &= \sqrt{\sum_i \{X_{\nu i}-(n_\alpha X_{\alpha i}+n_\beta X_{\beta i})/n_\gamma\}^2} \\ &= \sqrt{\sum_i \{(n_\alpha/n_\gamma)(X_{\nu i}-X_{\alpha i})^2+(n_\beta/n_\gamma)(X_{\nu i}-X_{\beta i})^2-} \\ &\quad \overline{(n_\alpha/n_\gamma)(n_\beta/n_\gamma)(X_{\alpha i}-X_{\beta i})^2\}} \quad (A_\nu \neq A_\gamma). \quad (1.5) \end{aligned}$$

The process repeats itself, thus producing a hierarchy of aggregated clusters as shown in **Fig. 3.** To teterminate this process, the minimum number of clusters needs to be a priori specified. We also need some exogeneous criteria to judge whether the produced zoning makes sense in engineerng terms. We will discuss this point in the next section.

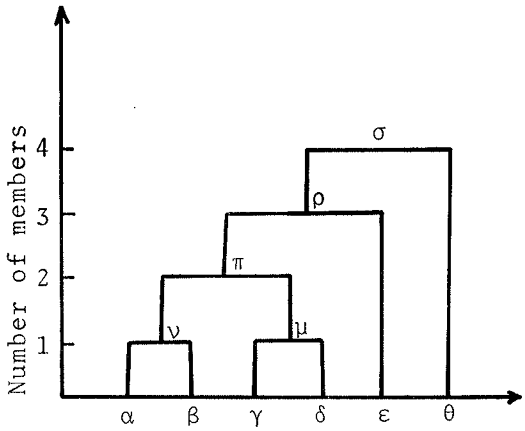

Fig. 3 Hierachical process of clustering

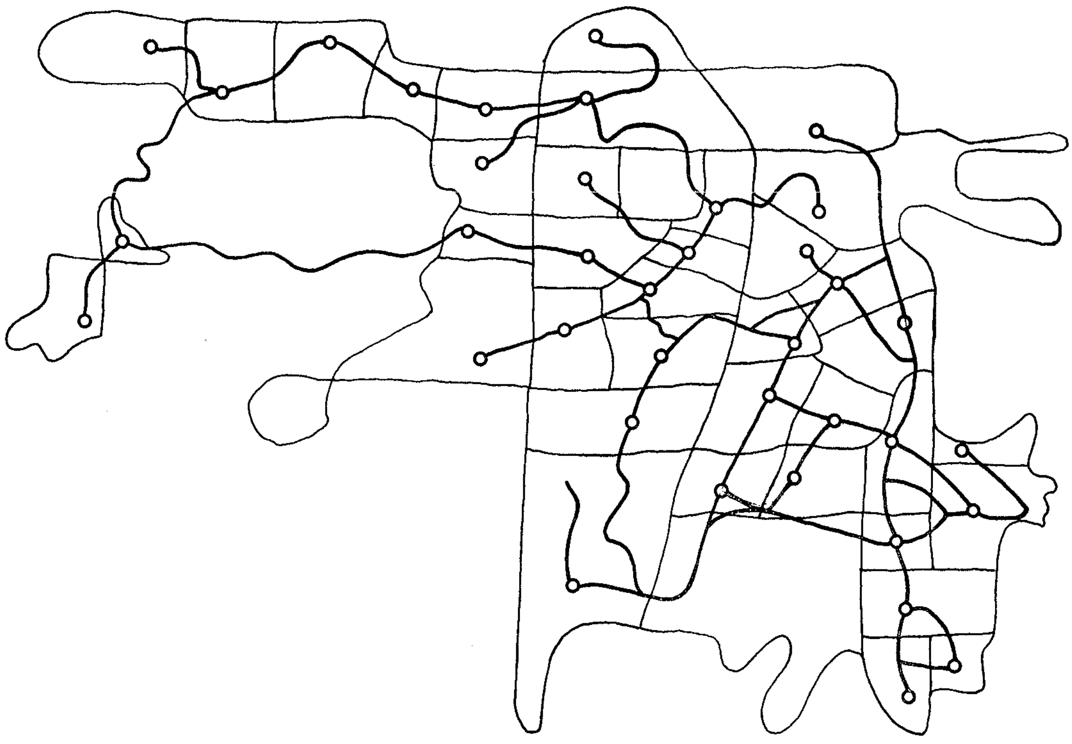

Fig. 4. Meshing of the service area

4. Modeling (Case Study of Tottori City)

By referring to the pattern of the water distribution network as well as that of the road network and topography, let us first throw meshes over the entire service area in the following manner.

1) As ahown in **Fig. 4** each mesh covers a node of the distribution network as a point representing the area unit within the mesh.

2) Each area unit is considered homogeneous in terms of all relevant factors to be considered in the practice of zoning the service area.

3) Accordingly each node within a given area unit can be regarded as a sample item in terms of cluster analysis.

4) Let us define water supply districts as a collection of those meshes (area units), the boundaries of which are consecutively connected each other ; thus forming those individual closed loops which are mutually exclusive and share boundaries.

5) Let attributes of each item (node) be defined as those factors characterizing topographic, engineering and socio-economic conditions of the associated area unit. The following $(R+2)$ factors (attributes) have been selected for node α ($\alpha = 1, \cdots \cdots, n$) : $X_{\alpha o}$ representing water demand as the aggregated data on population, industrial activity and water use patterns ; $X_{\alpha r}$ representing the distance between the node and distributing reservoir r ($r = 1, \cdots\cdots, R$; $R=8$) ; $X_{\alpha R+1}$ representing the difference in ground level between the node and the nearest distributing reservoir. The attributes $X_{\alpha r}$ ($r = 1, \cdots\cdots, R$) and $X_{\alpha R+1}$ represent hydraulic and topographic (hydro-topographic) conditions.

With the setting so postulated, let us now define three kinds of distances for a pair of noes (α, β).

1) water demand distance, $d^q_{\alpha\beta}$

$$d^q_{\alpha\beta} = \sqrt{w_0 (X_{\alpha 0} - X_{\beta 0})^{2u_0}} = \sqrt{w_0}\,|X_{\alpha 0} - X_{\beta 0}|^{\sqrt{u_0}} \quad (w_0,\ u_0 > 0) \quad \cdots\cdots \ (1.6)$$

2) two kinds of hydro-topographic distances, horizontal distance, $d^h_{\alpha\beta}$ and vertical one, $d^v_{\alpha\beta}$.

$$d^h_{\alpha\beta} = \sqrt{\sum_r u_r (X_{\alpha r} - X_{\beta r})^2}$$

$$(r = 1, \cdots, R) \quad \cdots\cdots\cdots\cdots\cdots\cdots \ (1.7)$$

$$d^v_{\alpha\beta} = \sqrt{u_{R+1} (X_{\alpha R+1} - X_{\beta R+1})^2}, \quad \cdots\cdots \ (1.8)$$

where u_r ($r=1, \cdots\cdots, R$) and u_{R+1} are unknown parameters.

The three kinds of distances are synthesized into an integral distance, D as illustrated in **Fig. 5 (1)**.

$$D = \sqrt{(d^q_{\alpha\beta})^2 + (d^h_{\alpha\beta})^2 + (d^v_{\alpha\beta})^2}. \quad \cdots \ (1.9)$$

By changing the values for the unknown parameters our concern is to identify (the range of) appropriate parametric values.

In addition to this integral distance D, let us develop another type of criterion to measure the "affinity of the nodes" or the "degree of reciprocal attraction between a pair of nodes". We observe that the demand distance $d^q_{\alpha\beta}$ orthogonal to the superplane spanned by the horizontal and vertical dimension, $d^h_{\alpha\beta}$ and $d^v_{\alpha\beta}$, respectively of hydro-topographic distance $d^t_{\alpha\beta}$(see **Fig. 5 (2)**). Then the demand distance $d^q_{\alpha\beta}$ may be reinterpreted as a load on one end of the beam spanned between a pair of nodes, α and β. In this context we now define a new "affinity", m (α, β) for a pair of nodes, α and β. That is,

$$m(\alpha, \beta) = d^q_{\alpha\beta} \cdot d^t_{\alpha\beta} = d^q_{\alpha\beta} \sqrt{(d^h_{\alpha\beta})^2 + (d^v_{\alpha\beta})^2}, \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (1.6)$$

where $d^t_{\alpha\beta} = \sqrt{(d^h_{\alpha\beta})^2 + (d^v_{\alpha\beta})^2}$

In the following section we will check this second criterion, momentum against the first one, distance.

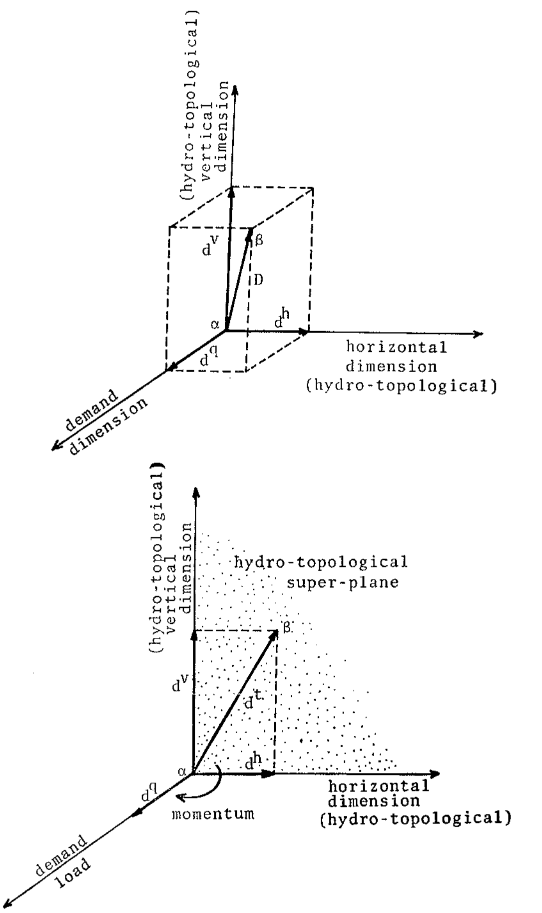

Fig. 5 (1). Geometric interpretation of integral distance
5 (2). Geometric interpretation of momentum

5. Model Operation

5.1) Input data

In the foregoing discussion we have already meshed the entire service area. In running the model we need some other input data as shown in **Tables I** to **III** and **Fig. 6**.

Table I. Capacity of distributing reservoir (as of 1978)

distributing reservoir	capacity (m^3)
Karo	800
Suegaki	1.134
Maruyama	1.000
Tokuno	4.000
Uemachi	10.000
Shimoajino	600
Omokage	3.763

5.2) Exogeneous criteria for checking validity

As referred to before, we need some exogeneous criteria to validate our model. Let us employ the following conditions as exogeneous criteria.

Table II. Data on distribution network (1)

nodeN.O.	ground level (m)	population	percapita demand(ℓ/sec)
1	8	900	7
2	5	900	7
3	11	1,000	11
4	4	1,200	12
5	10	2,000	21
6	8	4,900	51
7	11	1,000	10
8	23	900	10
9	3	400	5
10	5	700	73
11	4	1,500	16
12	4	1,100	29
13	7	2,700	28
14	4	1,300	14
15	2	800	14
16	2	500	5
17	15	200	16
18	8	200	2
19	10	2,500	26
20	7	2,200	49
21	8	1,800	19
22	12	3,200	33
23	6	9,000	137
24	4	4,000	41
25	20	3,900	41
26	4	7,100	137
27	5	1,700	173
28	6	6,500	74
29	6	7,500	85
30	5	4,800	81
31	11	3,000	172
32	8	1,200	14
33	8	7,200	55
34	8	7,400	77
35	11	7,200	172
36	10	2,600	27
37	19	1,000	11
38	7	300	3

Table III. Data on distribution network (2)

Line section N.O.	Line length	Line section N.O.	Line length	Line section N.O.	Line length
1	1960	19	1960	37	980
2	2450	20	2450	38	1960
3	2450	21	6200	39	1960
4	1960	22	2450	40	1960
5	1960	23	3430	41	1960
6	2940	24	1470	42	3430
7	2450	25	2450	43	2940
8	2450	26	2450	44	3920
9	4900	27	1960	45	1470
10	3900	28	4600	46	3430
11	1960	29	1960	47	1960
12	3920	30	4900	48	2450
13	980	31	2940	49	3430
14	2450	32	980	50	3430
15	2940	33	5880	51	2940
16	8820	34	6900	52	5880
17	2450	35	1470	53	2940
18	1470	36	2940		

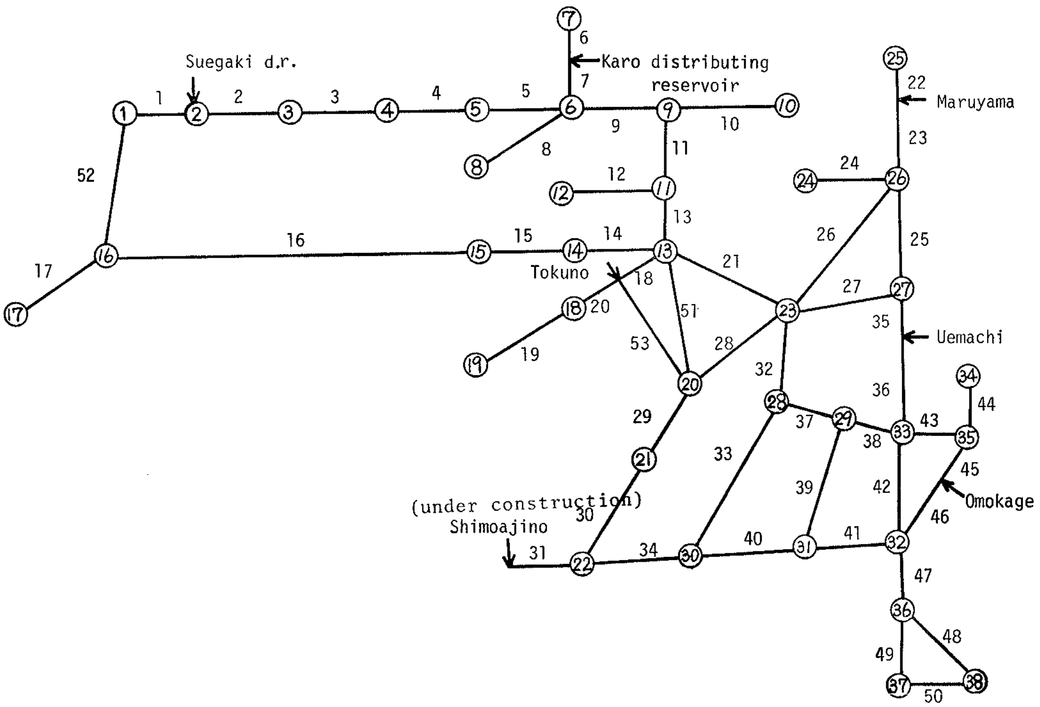

Fig. 6 Diagrammatic representation of the water distribution network

(i) Since water districts are mutually cxclusive in space and share boundaries, clusters (zones) to be made should be closed by a chain of meshes and should not intersect each other.

(ii) Each cluster (zone) should cover at least one distributing reservoir. This suggests that the appropriate number of zones may be between 6 and 8.

5.3) Zoning alternatives obtained

By changing the relevant parametric values on a trial-and error basis, the following four zoning alternatives have been found most appropriate. The parametric values for these four alternatives are listed in **Table IV**. **Figs. 7** to **10** illustrate the zoning patterns of alternatives **Z1, Z2, Z3,** and **Z4**, respectively. Alternative **Z4** turned out to be precisely identical to the actual zoning pattern as shown in **Fig. 4**.

Table IV. Major zoning alternatives obtained for different parametric values

parameter / zoning alternative	demand		hydro topographic horizontal	hydro topographic vertical	
	w_0	u_0	$w_1 \sim w_7$	w_8	p
z_1	0.5	1.0	1.0	1.0	0.01
z_2	0.0	1.0	1.0	1.0	0.01
z_3	1.0	1.0	1.0	1.0	0.01
z_4	0.8	1.0	1.0	2.0	0.01

Close examination of **Figs. 7** to **10** shows:

(i) **Z1** is the zoning alternative which takes account of both demand distribution and hydro-topographical conditions.

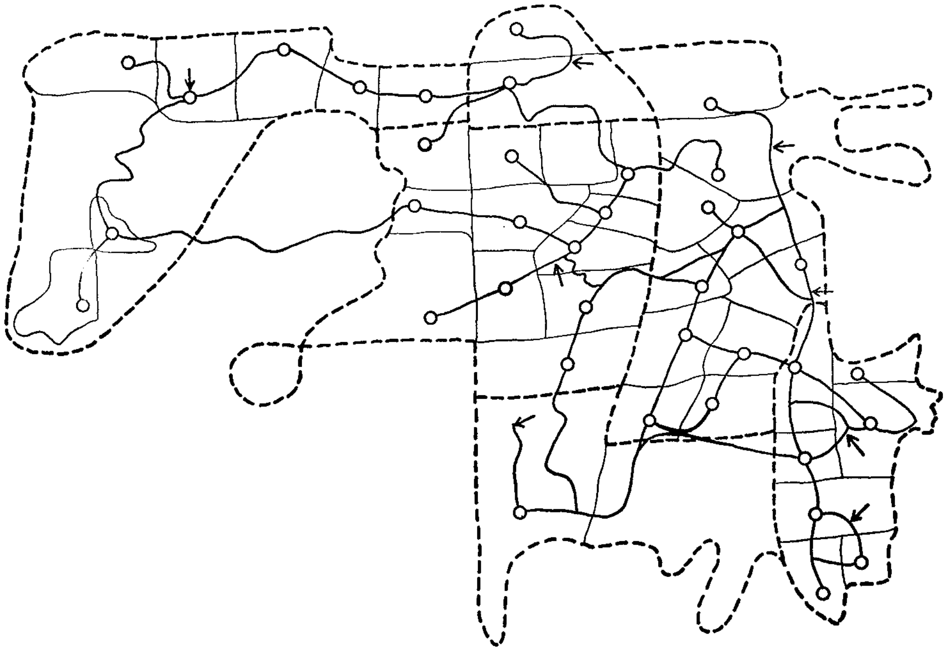

Fig. 7 Zoning alternative Z1

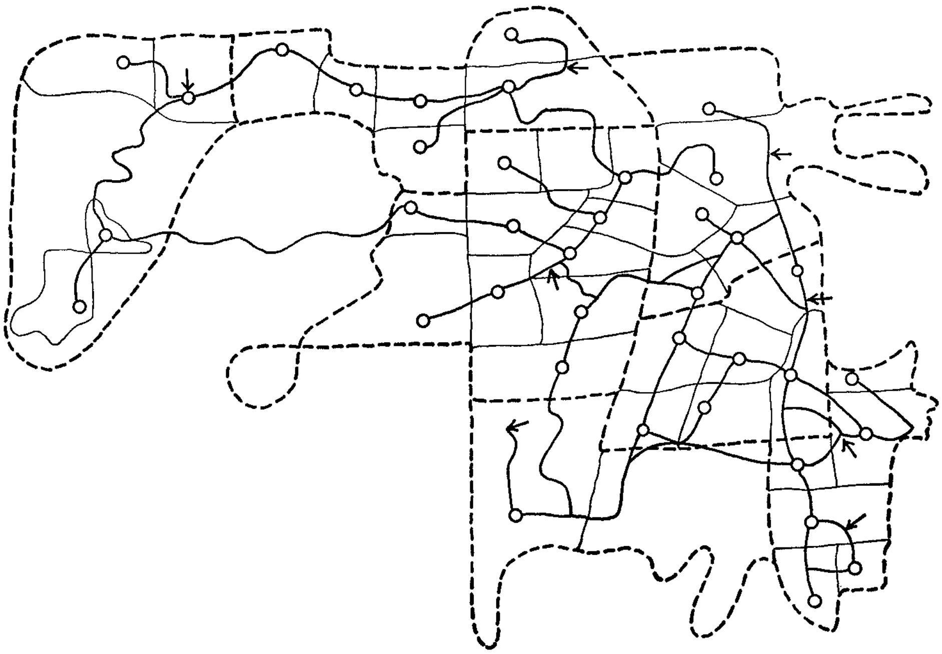

Fig. 8 Zoning alternative Z2

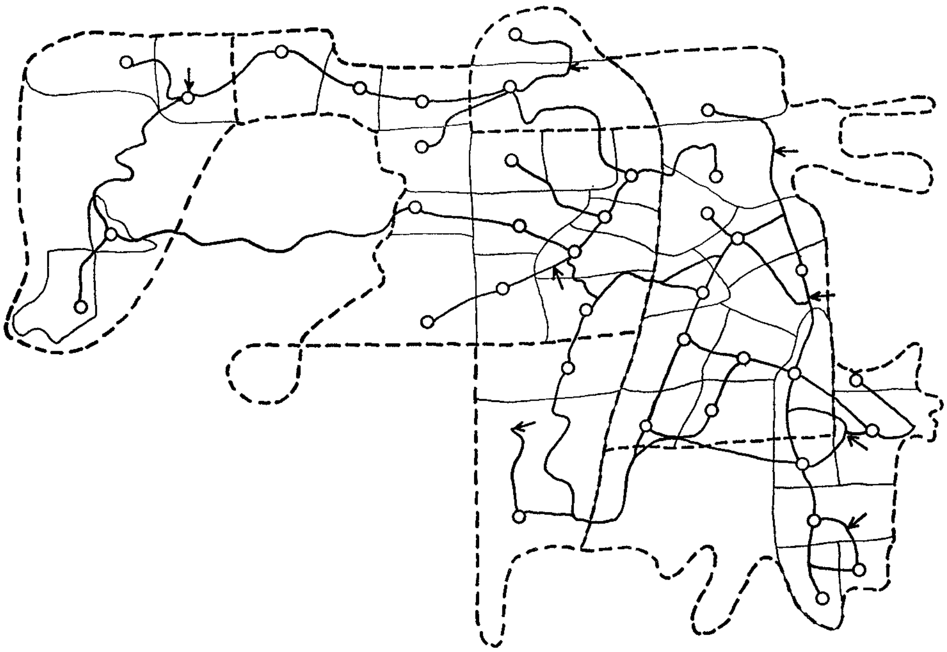

Fig. 9 Zoning alternative Z3

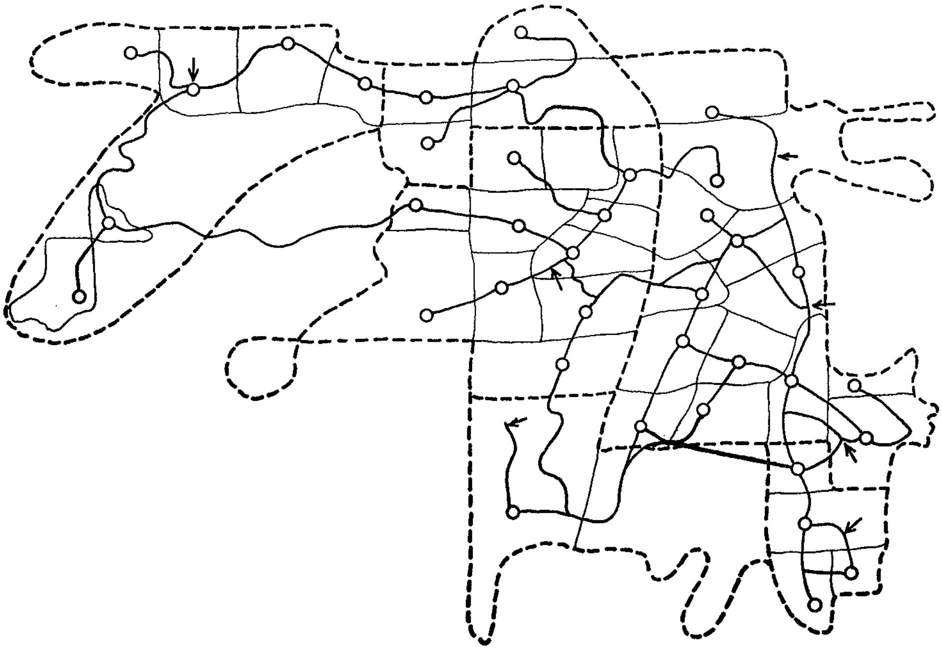

Fig. 10 Zoning alternative Z4

(ii) Zoning alterative **Z2** has been obtained by dropping the demand distribution conditions from **Z1**.
(iii) **Z3** takes account of both demand distribution and hydro-topographical conditions as Z1 does. Z3, however, attaches more weight to demand distribution conditions than **Z1** (w_0 being 1.0 for Z3 and 0.5 for Z1).
(iv) **Z4** which is identical to the actual zoning pattern of Tottori City has been generated with the criterion of momentum other than with that of distance as has been the case for alternatives **Z1**, **Z2** and **Z3**.
(v) Alternatives Z1 and Z2 are similar in zoning configuration. They are characterized by a relatively large size of the peripheral zones for the size of the capacity available of the distibuting reservoirs within the zones. Notably the peripheral areas are rapidly expanding whereas the capacity of the distributing reservoirs which are in use is still small for the expanded space of their cover area. This phenomenon does not seem to be less accounted for by alternatives **Z1** and **Z2** than by alternatives **Z3** and **Z4**. This may be explained by the fact that the existing water-demand distribution pattern has determined the capacity of the existing water supply facilities. Therefore it is natural that alternatives Z1 and Z2 which account for water demad less than **Z3** and Z4 have led to relatively large zones assigned to the north-western and southern periphery of Tottori City.

Then let us compare alternatives **Z3** and **Z4**. We observe that alternative **Z4** accounts for more of the water demand condition than Z3. In another word **Z4** handles the water demand distribution condition like the loads placed on the nodes of the watersupply network, whereas **Z3** deals with it as a kind of distance detween a pair of nodes. Bearing this point in mind, the following may be readily understood :
(i) In alternative **Z3** the Karo distributing reservoir has to cover a relatively large zone as its water supply area although its existing capacity falls short of the total demands in the cover area. This would not be feasible if the capacity remains unchanged and no capacity is added to the present one or if there is no complementary supply from outside of the zone.
(ii) In contrast to this problem of imbalance between supply and demand, there is another type of imbalance found in the water supply district to be covered by two distributing reservoirs, located on the southeastern part of Tottori City. The covered area seems to be too small for its total capacity of the two distributing reservoirs.
(iii) These problems are, however, not seen in alternative Z4, which is the actual zoning of the service area. This implies that the existing zoning proved to be relatively reasonable as compared to the other potential alternatives, provided that the present conditions of water deamand distribution and hydro-topography remain unchanged.
(iii) There is, however, some drawback to alternative Z4. The water supply district

of the Tokuno distributing reservoir is divided partially by the neighboring district of the Karo distributing reservoir. This may have caused an unorganized connection of the two districts, failing to keep up with the outward expansion of the northwestern part of the city area. Suppose we plan the expansion of the Karo distributing reservoir. Then alternative **Z3** may be rated before alternative **Z4.**

To sum up, we may state that alternatives **3** and **4** are identified as most appropriate from an engineering point of view. The choice of either of them may be conditional ; whether we intend to analyze the present pattern or to plan for the future determines much of the decision.

6. Conclusion

The problem of zoning the water supply service area has been discussed with Tottori City as the study area. Cluster analysis has been applied to the selected problem. It has been shown that :

1) There is a good analogy between the practice of zoning the water supply service area and clustering of sample items by use of cluster analysis.

2) The modeling of the problem by this technique may be conceived as the process of exposing some relevant factors which are intuitively and subjectively considered by the experienced engineer. Through the modeling practice it has been made explicit that : (i) In developing the existing patterns of Tottori City it has been shown that major factors to be considered are water demand distribution ; location, number and capacity of distributing reservoirs ; and the distance and ground level difference between nodes and distributing reservoirs ; (ii) the criteria to measure the degree of closeness or attraction between nodes and distributing networks may be defined as analogous to the notion of distance or momentum ; (iii) the parameters in the criteria employed represent the relative importance of the relevant factors considered in zoning.

3) The derived model may offer several benefits : (i) It may be used as the tool of learning the unknown mechanism of the zoning practice. In this sense we may use the information derived from the model as a guideline to locate bottelnecks of the existing zoning as well as to improve it ; (ii) the application of the model will relieve the engineer from much of the complicated routine practice by providing him with a systematic tool for producing basic zoning alternatives ; (iii) the benefits will become of more particular value if the service area is much larger than the study area of this paper or if we discuss the zoning practice in terms of planning.

Though there is still much room for the development of the study, we may as well state that the approach presented here has proved to be an ffective tool to handle the zoning of municipal water supply service area.

References

1) Matsui H ; A study on systematic methodology for zoning water service area, Graduation thesis submitted to the Faculty of Engineering, Tottori University, Mar. 1978 (in Japanese).
2) Okuno C. , et al ; Multi-variate quantification analysis, the Nikka Giren Pub. Comp., 1971 (in Japanese).