

## 주성분분석과 군집분석을 이용한 장기 물수요예측과 활용

구자용 · 유명진 · 김신결<sup>†</sup> · 심미희 · 小泉明\*

서울시립대학교 환경공학부 · \*東京都立大學 土木工學科

(2005년 6월 16일 접수, 2005년 8월 17일 채택)

### Estimation of Long-term Water Demand by Principal Component and Cluster Analysis and Practical Application

Koo, Ja-Yong · Yu, Myung-Jin · Kim, Shin-Geol<sup>†</sup> · Shim, Mi-Hee · Koizumi Akira\*

Department of Environmental Engineering, The University of Seoul

\*Department of Environmental Civil Engineering, Tokyo Metropolitan University

**ABSTRACT** : The multiple regression models which have two factors(population and commercial area) have been used to forecast the water demand in the future. But, the coefficient of population had a negative value because proper regional classification wasn't performed, and it is not reasonable because the population must be a positive factor. So, the regional classification was performed by principal component and cluster analysis to solve the problem. 6 regional characters were transformed into 4 principal components, and the areas were divided into two groups according to cluster analysis which had 4 principal components. The new regression models were made by each group, and the problem was solved. And, the future water demands were estimated by three scenarios(Active, moderate, and passive one). The increase of water demand are 89,034 m<sup>3</sup>/day in active plan, 49,077 m<sup>3</sup>/day in moderate plan, and 19,996 m<sup>3</sup>/day in passive plan. The water supply ability as scenarios is enough in water treatment plant, however, 2 reservoirs among 4 reservoirs don't have enough retention time in all scenarios.

**Key Words** : Principal Component Analysis, Cluster Analysis, Forecasting Water Demand, Assessment of Water Supply Capacity

**요약** : 서울 중심부인 종로구와 중구를 대상으로 하여 거주인구수와 영업용 연상면적을 독립변수로 하는 물수요 예측식을 만들었다. 그런데 명확한 물수요 증가요인인 거주인구수가 감소요인으로 나오는 적절하지 못한 결과를 보여주었으며, 이것은 지역특성별 분류가 이루어지지 않았기 때문이었다. 이러한 모순점을 해결하기 위해 6가지의 지역특성을 이용한 주성분분석과 군집분석으로 지역을 분류하였다. 6가지 인자들을 대상으로 한 주성분분석결과 4번째 주성분까지의 고유값 누적이 92.6%로서 원래 인자들이 가지고 있던 정보량을 대부분 표현할 수 있었으며, 군집분석은 워드방법(Ward's method)으로 대상지역을 주거와 상업지역으로 분류하였다. 이에 각각의 지역에 대해 중회귀 모델을 구성하였으며, 모델결과 이전에 발생하였던 모순점이 해결되었다. 또한 이 모델을 대상으로 세 가지의 장래 시나리오인 적극적인 개발, 중간적인 개발, 소극적인 개발로 나누어서 장래 물수요량을 예측하여 보았다. 이에 적극적인 개발이 89,033 m<sup>3</sup>/일, 중간적인 개발이 49,077 m<sup>3</sup>/일, 소극적인 개발이 19,996 m<sup>3</sup>/일의 증가량을 보여주었다. 이에 대해 관할 정수장과 배수지를 대상으로 시설용량을 평가하였으며, 관할 D정수장의 경우 운전율을 85%로 높여주면 시나리오에 따른 물수요 증가량을 충분히 공급하여 줄 수 있었다. 배수지에 있어서도 D, A, N, B 등 네 개의 관할 배수지에 대해서 체류시간을 계산한 결과, D와 A배수지는 모두 기준시간 12시간을 충족시켜주었으나, N, B 배수지는 만족시켜 주지 못하였다. 하지만 현재 수도정비기본계획에 의해 용량이 충족되고 있으므로 큰 문제는 없을 것으로 판단된다.

**주제어** : 주성분분석, 군집분석, 물수요예측, 시설용량평가

## 1. 서론

장기 물수요예측은 막대한 예산이 소요되는 상수도 시설의 용량과 위치를 결정하기 때문에 그 정확성과 타당성이 높게 요구되는 분야로서 장기에 걸친 예측의 안정성이 요구된다.<sup>1-3)</sup> 이러한 장기 물수요예측을 위하여 과거에는 인구수만을 고려한 원단위를 사용하였으나 현재는 그 정확도

를 더욱 높이기 위해 거주인구수와 영업용 연상면적을 독립변수로 하는 중회귀 모델을 이용하고 있다. 서울의 경우에는 각 구단위의 행정구역별로 동별 거주인구와 영업용 연상면적을 독립변수로 하는 중회귀식을 장래의 물수요 예측식으로 사용하고 있다.<sup>4-6)</sup> 그런데 실제 행정구역별로 중회귀 모델을 작성한 결과 명확한 물수요량의 증가요인인 거주인구가 물수요량을 감소시킨다는 결과가 나왔다. 이것은 거주인구수는 적지만 전체 물수요량이 많은 상업지역과 이와 반대로 거주인구는 많지만 전체 물수요량이 적은 주거지역의 자료가 적절히 분류되지 않고 함께 중회귀 모델

<sup>†</sup> Corresponding author  
E-mail: sgkim75@uos.ac.kr  
Tel: 02-2210-2946

Fax: 02-2244-2245

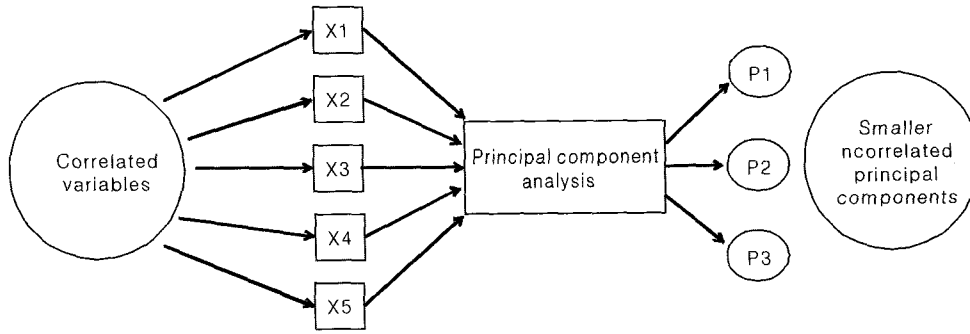


Fig. 1. Summary of Principal component analysis

화되었기 때문이다. 따라서, 본 연구에서는 이러한 모순점을 해결하기 위해 우선 합리적인 지역구분이 이루어져야 한다고 판단하였다.<sup>7-10)</sup> 이에 합리적인 지역구분을 이루기 위해서는 지역특성에 따른 군집분석(Cluster analysis)이 적합한 방법이라고 결정하였다. 그런데 군집분석은 특성상 공간상의 거리에 기초하여 입력자료들 군집하는 방법으로써 자료들간의 독립성이 우선 보장되어야 하기 때문에 본 연구에서는 군집분석의 입력자료들간의 독립성을 확보하기 위해 먼저 주성분분석(Principal component analysis)을 거치도록 하였다.<sup>11-13)</sup> 주성분분석과 군집분석을 통하여 분류된 군집들에 대해 다시 물수요예측식을 구성하였으며 앞에서 언급한 모순점이 해결되었는가를 고찰하였다. 또한 작성된 물수요예측식의 활용방안으로써 해당지역의 장래 시나리오별 물수요 증가량을 추정하고 물수요증가량을 해당 상수도시설물인 정수장과 배수지에서 원활히 공급할 수 있는가에 대한 타당성을 살펴보았다.

## 2. 연구 방법 및 순서

### 2.1. 연구방법

본 연구에서는 해당지역에 대해 적절한 특성인자들에 기초한 지역구분을 하고자 주성분분석과 군집분석을 이용하였으며, 이들 주성분분석과 군집분석을 통해 분류된 지역들에 대해 다시 중회귀 분석을 이용하여 물수요예측식을 구성하였다.

#### 2.1.1. 주성분분석

주성분분석은 Fig. 1과 같이 다수의 서로 상관된 반응 변수들을 소수의 개념상 의미있고, 서로 독립인 주성분으로 변화시키는 분석이다. 즉, 상호 상관있는 p개의 특성치가 갖는 정보를 상호 상관없는 m개의 종합특성치(이를 제1, 제2, ..., 제 m 주성분이라 한다)로 요약하는 것이다.

이러한 주성분은 기존의 좌표축이 아닌 새로운 좌표로 정렬하여, 서로 독립인 새로운 값들을 갖도록 하는데 여기에 이용되는 방법은 라그랑제(Lagrange)의 미정승수법에 의한 것이다. 이 방법은 미지의 값을 구하기 위해 고유방정식(Characteristic equation)을 도입하여 p개의 고유값 해들 고유값  $\lambda_i$  ( $i = 1, 2, \dots, p$ )를 구한다. 여기에서 각 고유값들을 통하여 각 주성분과 그 주성분이 가지는 정보량의 크기가 결정된다.<sup>14)</sup>

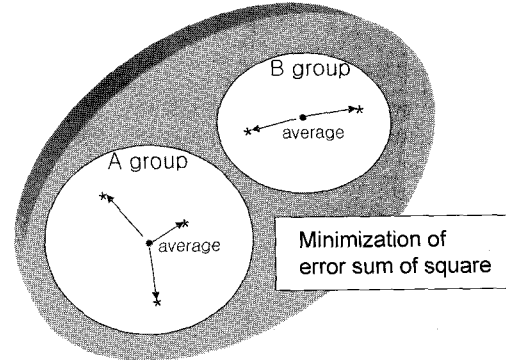


Fig. 2. Ward's method

#### 2.1.2. 군집분석

군집분석은 각 객체의 유사성을 측정하여 유사성이 높은 대상 집단을 분류하고, 같은 군집에 속한 객체들의 유사성과 서로 다른 군집에 속한 객체간의 상이성을 규명하는 통계분석방법이다. 이 때에 유사성은 측정된 변수들을 이용하여 모든 개체들 간의 거리(distance) 또는 비유사성(dissimilarity)을 통하여 계산된다. 군집분석에서 이러한 유사성은 유클리드(Euclid) 또는 마할로비스(Mahalanobis) 거리를 이용하여 측정되며 거리가 가까울수록 유사성이 높다.

군집화하는 방법에는 공간상의 거리에 기초한 여러 가지 방법이 있는데 여기에서는 그 중에 워드법(Ward's method)을 사용하였다. 이 방법은 Fig. 2와 같이 오차의 제곱합을 최소화시켜 정보의 손실량이 최소화가 되도록 하는 방법이다.<sup>15)</sup>

#### 2.1.3. 중회귀 분석

중회귀 모델은 하나의 종속변수 Y에 대해 여러개의 독립변수  $X_i$  ( $i=1, 2, \dots, k$ )들이 선형적인 영향을 미칠 때에 정형화 되는 방법이다.

$$Y = A_0 + A_1X_1 + A_2X_2 + \dots + A_kX_k + \epsilon \quad (2)$$

여기에서,  $A_0, A_1, A_2, \dots, A_k$ 는 모델에서 추정되는 모수로서 최소자승법(Method of Least square)에 의한 정규방정식(Normal equation)에 의해 결정되며,  $\epsilon$ 은 오차항이다. 중회귀 모델은 인자의 선택, 다중공선성(Multicollinearity)과 오차에 대한 검정 등을 통하여 모델화가 이루어진다.<sup>16)</sup>

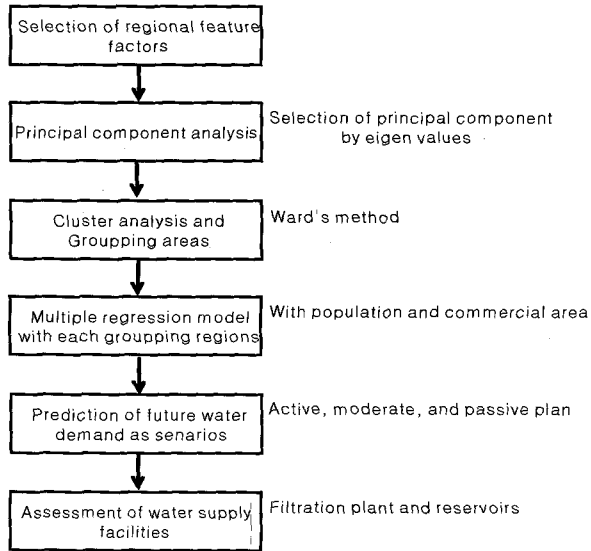


Fig. 3. Procedure of the study.

2.2. 연구의 진행순서

본 연구의 순서는 Fig 3과 같다. 첫 단계는 지역특성을 나타내는 인자들을 선정하는 것으로 6가지의 인자가 선택되었다. 이후 이들 인자들은 주성분분석을 통하여 통계적으로 유의한 주성분들로 변환되며, 선택된 주성분들은 군집분석의 입력자료가 된다.

군집분석을 통하여 성립된 군집들을 대상으로 다시 거주 인구수와 영업용 연상면적을 독립변수로 하는 중회귀 모델을 작성한다. 그리고, 작성된 물수요 예측식을 통하여 장래 개발시나리오에 따른 물수요 증가량을 추정한다. 추정된 물수요 증가량과 정수장과 배수지의 용량을 비교하여 장래 개발시나리오에 따른 상수도 시설용량을 고찰한다.

3. 연구대상지역 물수요예측

3.1. 연구대상지역

본 연구대상 지역은 Fig. 4~5와 같이 서울시의 중심부인

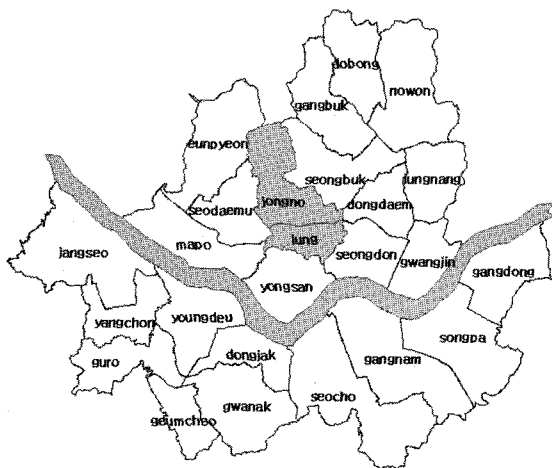


Fig. 4. Seoul metropolitan city.

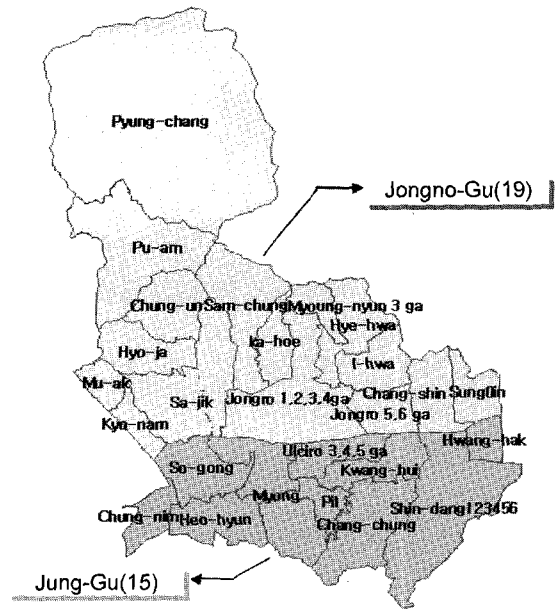


Fig. 5. Area of study(Jung-Gu and Jongro-Gu).

Table 1. Dongs of Jongro-Gu and Jung-Gu

Autonomy	Dong
Jongro-gu (19)	Chung-un, Hyo-ja, Sa-jik, Sam-chung, Pu-am, Pyung-chang, Mu-ak, Kyo-nam, Ka-hoe, Jongro 1, 2, 3, 4 ga, Jongro 5, 6 ga, I-hwa, Hye-hwa, Myoung-nyun 3 ga, Chang-shin 1, Chang-shin 2, Chang-shin 3, Sung-in 1, Sung-in 2
Jung-gu (15)	So-gong, Hoe-hyun, Myong, Pil, Changchung, Kwang-hui, Ulciro 3, 4, 5 ga, Shin-dang 1, Shin-dang 2, Shin-dang 3, Shin-dang 4, Shin-dang 5, Shin-dang 6, Hwang-hak, Chung-nim

Table 2. Water supply facility on study area

Water supply facility		Capacity
Filtration plant	D Filtration plant	750,000 m <sup>3</sup> /day
Reservoir	D Reservoir	200,000 m <sup>3</sup>
	A Reservoir	30,000 m <sup>3</sup>
	N Reservoir	40,000 m <sup>3</sup>
	B Reservoir	40,000 m <sup>3</sup>

종로구와 중구를 대상으로 하고 있다. 이들 지역은 서울시의 전통적인 도심부로서 물공급의 안정성과 정확성이 더욱 요구되는 지역이다.

이들 행정구역들은 Table 1과 같이 각각 종로구가 19개, 중구가 15개의 행정동으로 구성되어 있다. 이 종로구와 중구를 관할하는 정수장과 배수지는 Table 2와 같으며 해당 자료는 2001년의 자료를 이용하였다.

3.2. 기존 물수요예측식과 문제점

기존의 물수요예측식은 행정구역별로 인구나 영업용 연상면적을 이용하여 물수요예측식을 구성하였으며 결과는 Table 3과 같다.

**Table 3.** Formalization of the multiple regression model

Autonomy	Multiple regression model	R <sup>2</sup>
Jongro-gu	$Y = 0.2794X_1 + 0.0055X_2 - 131.5$	0.9659
Jung-gu	$Y = -0.0979X_1 + 0.0030X_2 + 4767.7$	0.8720

X<sub>1</sub>: Population, X<sub>2</sub>: Commercial area

이 예측식들은 높은 결정계수값을 보여주었지만 중구지역의 경우에는 거주인구수의 계수값이 음의 값을 가지고 있다. 이것은 거주인구수가 증가할수록 물수요량이 감소한다는 것을 의미하며 그 이유는 Fig. 6과 같다. 상업지역에서는 전체 물사용량은 많지만 거주인구가 적고, 거주지역에서는 전체 물사용량은 상업지역에 비해 적지만 거주인구가 많기 때문에 두 지역을 적절히 분류하지 않고 함께 중회귀 모델을 구성하게 되면 Fig. 6과 같이 인구가 증가할수록 물수요량이 감소한다는 엉뚱한 결과를 얻게 된다.

따라서, 본 연구에서는 적절한 지역구분을 통하여 이러한 모순점을 해결하여 통계적으로 더욱 의미있는 물수요 예측 모델을 구성하고자 한다. 이를 위해 다음 4.에서 합리적인 지역구분을 통하여 통계적으로 의미있는 모델로 다시 구성한다.

### 4. 연구결과

본 연구는 3.2절에서 보여준 것과 같은 문제점을 해결하고 유의한 모델을 구성하여 장래의 시나리오별 물수요량의 증가와 이때 관한 상수도 시설물의 공급능력을 평가하는 것을 목적으로 한다. 이를 위해 우선 적절한 지역인자를 선택하고 주성분분석과 군집분석을 통하여 지역을 합리적으로 구분하고자 한다.

#### 4.1. 주성분분석과 군집분석을 이용한 지역의 구분

##### 4.1.1. 지역의 구분

본 연구에서는 문제점을 해결하기 위한 최초의 작업으로서 지역특성을 나타내는 적합한 인자를 선택하고자 한다. 이를 위해 행정구역별 자료 및 상수도 통계자료 등을 참조

하였지만 필요한 자료는 동단위로 집계된 자료들이지만 지역특성을 나타내는 많은 자료들이 구지역 단위로만 집계 이루어진 경우가 많아서 자료를 많이 취득할 수 없었다. 또한 본 연구에서의 목표는 바로 상업지역과 주거지역을 분류하는 것이기 때문에 이에 초점을 맞추어 지역의 특성을 나타내는 자료들을 최대한 활용하였으며, Table 4와 같이 6가지를 선정하였다.

##### 4.1.2. 주성분분석

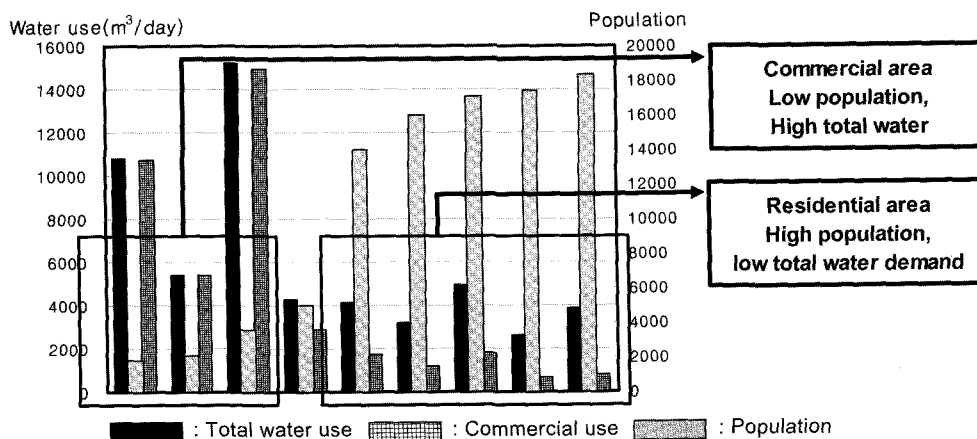
지역구분을 위한 군집분석은 공간에서의 거리에 기초하여 군집을 나누기 때문에 먼저 서로 독립적인 인자들로 변환이 이루어져야 한다. 이를 위해 지역특성을 나타내는 6개의 인자들에 대해 주성분분석을 실시하여, Table 5와 같은 결과를 얻었다.

**Table 4.** Regional characteristic factor

Regional characteristic factors	Contents
X <sub>1</sub>	Domestic usage/Population (m <sup>3</sup> /capita)
X <sub>2</sub>	Commercial usage/Area of commercial district(m <sup>3</sup> /m <sup>2</sup> )
X <sub>3</sub>	Employees/Population(capita/capita)
X <sub>4</sub>	Population/Total area of dong(capita/m <sup>2</sup> )
X <sub>5</sub>	Number of business/Total area of dong(number/m <sup>2</sup> )
X <sub>6</sub>	Area of commercial district/total area of dong(m <sup>2</sup> /m <sup>2</sup> )

**Table 5.** The results of principal component analysis

Item	Eigen value	Difference	Proportion	Cumulative
Prin. 1	2.9076	1.7866	0.4846	0.4846
Prin. 2	1.1210	0.1765	0.1868	0.6714
Prin. 3	0.9445	0.3638	0.1574	0.8289
Prin. 4	0.5807	0.2824	0.0968	0.9256
Prin. 5	0.2983	0.1504	0.0497	0.9754
Prin. 6	0.1479	-	0.0246	1.0000



**Fig. 6.** Problem of existing model.

**Table 6.** Principal component scores for six factors

Item	Prin. 1	Prin. 2	Prin. 3	Prin. 4
X <sub>1</sub>	-0.3405	0.1515	0.7457	-0.2730
X <sub>2</sub>	-0.3194	-0.6371	0.0909	0.6144
X <sub>3</sub>	0.4857	-0.2434	-0.2585	-0.1000
X <sub>4</sub>	-0.3134	0.6294	-0.4102	0.2954
X <sub>5</sub>	0.3968	0.3355	0.3771	0.6702
X <sub>6</sub>	0.5390	0.0565	0.2417	-0.0400

주성분분석결과 6개의 주성분인자로서 재편이 되었으며, Table 5에서 각 주성분에 해당하는 고유값을 보여주고 있다. 여기에서 고유값의 누적값을 보면 4번째까지의 주성분으로 전체 정보량의 92.6%를 표현할 수 있다고 나타나 있다. 전체의 92.6%의 정보량이면 이 정보량을 바탕으로 결론을 내더라도 왜곡이 발생하지 않으리라 판단되기 때문에 이 4개의 주성분을 선택하였다. 이 네 번째까지의 주성분들에 대해 해당 고유값들을 각각 넣어 원래의 변수들의 계수를 구하였으며, 결과는 Table 6과 같다.

**4.1.3. 군집분석**

주성분분석결과 지역을 나타내는 전체 6개의 인자들은 4개의 주성분으로 다시 재편되었다. 재편된 4개의 주성분들을 입력변수로 하여 군집분석을 실시하였다. 군집분석의 방법은 정보의 손실량이 최소가 되도록 하는 워드 방법을 이용하였으며 결과는 Table 7과 Fig. 7과 같다.

Fig. 7에서와 같이 해당군집들은 명확하게 두 개의 군집으로 나누어 졌으며, 이들 군집에 대해 실제의 위치를 살펴본 결과 대부분이 각각 상업지역과 주거지역으로 명확하게 나누어진 것을 확인할 수 있었다.

**Table 7.** Regression models for Jongro-Gu and Jung-Gu using the results of principal component analysis

Group (Group's feature)	Dong(code)
Group 1 (Residential district)	Chung-un(1), Hyo-ja(2), Sa-jik(3), Sam-chung(4), Pu-am(5), Pyung-chang(6), Mu-ak(7), Kyo-nam(8), Ka-hoe(9), I-hwa(12), Hye-hwa(13), Myoung-nyun 3 ga(14), Chang-shin 2(16), Chang-shin 3(17), Sung-in 1(18), So-gong(20), Pil(23), Chang-chung(24), Shin-dang 2(28), Shin-dang 3(29), Shin-dang 4(30), Shin-dang 5(31), Shin-dang 6(32), Chung-nim(34)
Group 2 (Commercial district)	Jongro 1, 2, 3, 4 ga(10), Jongro 5, 6 ga(11), Chang-shin 1(15), Sung-in 2(19), Hoe-hyun(21), Myong(22), Kwang-hui(25), Ulciro 3, 4, 5 ga(26), Shin-dang 1(27), Hwang-hak(33)

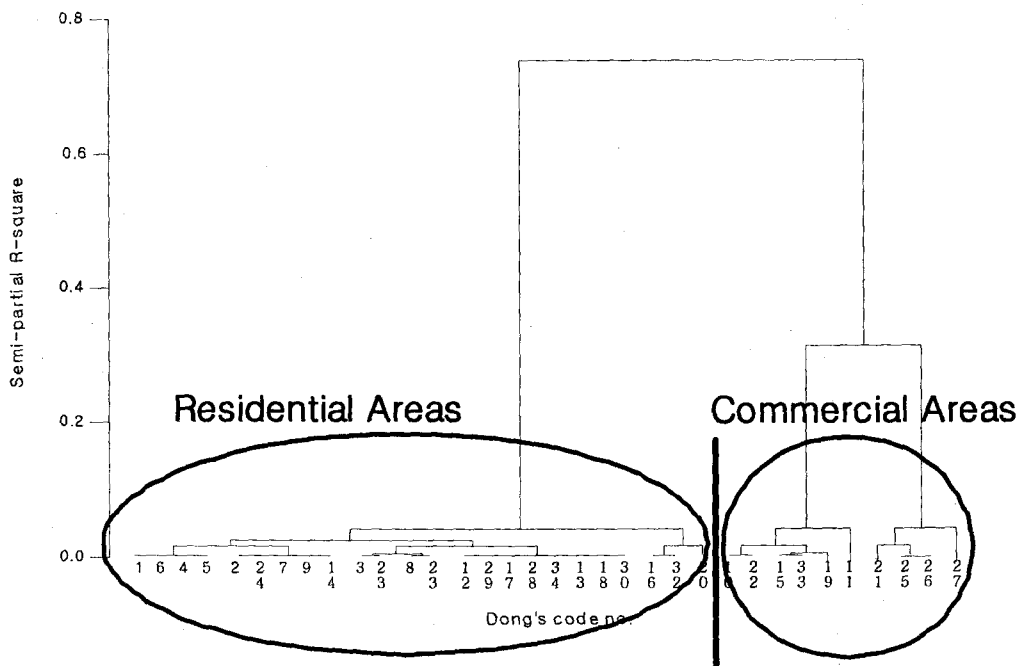
**4.2. 장래 물수요량의 추정 및 응용**

본 절에서는 군집분석에 의해 분류된 각 지역들에 대해 다시 물수요 예측식을 구성한다. 이후 장래 시나리오에 따른 물수요량의 증가분을 추정하고, 이 추정된 증가분을 기존의 상수도 시설로써 공급할 수 있는가를 살펴본다.

**4.2.1. 물수요 예측식의 구성**

본 절에서는 Table 8과 같이 그룹화 된 지역들을 대상으로 각각 거주인구수와 영업용 연상면적을 독립변수로 하는 중회귀식을 작성하였다.

Table 8에 나타난 것과 같이 행정구역별 모델에서 거주 인구 앞의 계수에서 음의 값을 가지는 문제는 해결되었으며, 결정계수의 값도 0.7903과 0.8660으로 양호한 값을 나타내었다. 이 값은 기존의 모델에 비해서는 낮은 값이다. 하지만, 행정구역별 모델의 결정계수 값이 높게 나온 이유



**Fig. 7.** Dendrogram constructed using cluster analysis.

**Table 8.** Regression models for Jongro-Gu and Jung-Gu using the results of principal component analysis

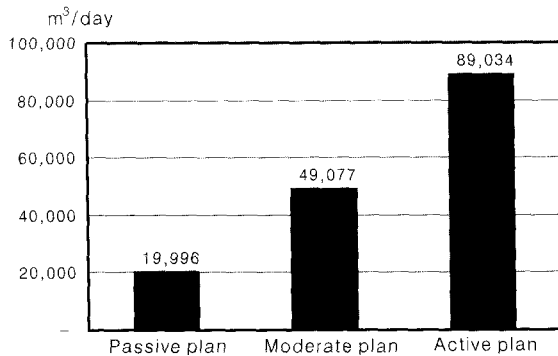
Group	R-square	Estimated formula
Group 1 (Residential district)	0.7903	$Y = 0.1414X_1 + 0.0054X_2 + 1145.20$
Group 2 (Commercial district)	0.8660	$Y = 0.3785X_1 + 0.0041X_2 + 491.96$

$X_1$ : Population (capita),  $X_2$ : Area of commercial district ( $m^2$ )

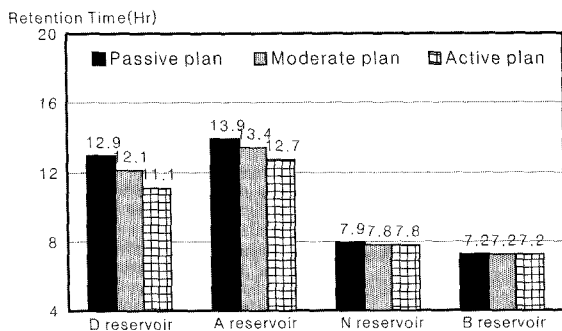
는 결정계수의 값이 변동범위에 기초하여 결정되기 때문이다. 즉 이전 모델에서는 변동범위가 상업지역과 주거지역을 모두 포함하고 있었기 때문에 전체 변동에 대한 오차의 비율이 작게 나왔지만, 이번 모델에서는 전체 물사용량이 다른 상업지역, 주거지역이 따로 분리되어 더 적은 변동범위를 가지게 된 상태에서 결정계수값이 결정되었기 때문이다. 따라서, 결정계수가 서로 다른 변동범위를 가지고 결정되는 특성상 현재 모델의 정확도가 더 떨어진다고 할 수 없으며, 객관적으로 이 정도의 결정계수값은 높은 수준으로서 현재의 모델은 유의한 모델이라 할 수 있다.

**4.2.2. 시나리오에 따른 장래 물수요량 예측**

본 연구에서는 향후 발생하게 될 지역의 개발 시나리오를 작성하였다. 이 작성된 시나리오는 재개발 지구를 적극적으로 개발을 하는 경우와 현재의 계획대로 하는 경우, 재개발지역의 가급적 줄이는 경우 등으로 각각 적극적, 중간적, 소극적인 개발로 나누어서 본 물수요 예측식에 대입하였다. 대입한 결과는 Fig. 8과 같다.



**Fig. 8.** The increase of water demand as scenarios.



**Fig. 9.** Retention time of reservoirs as scenario.

결과는 해당 지역에 대해 소극적인 개발을 하는 경우에 19,996 m³/일, 중간적인 개발을 하는 경우에 49,077 m³/일, 적극적인 개발을 하는 경우에 89,034 m³/일의 물수요량이 증가하는 것으로 나타났다.

**4.2.3. 물수요예측식의 활용**

본 절에서는 장래에 시나리오에 따라 증가하는 물수요량에 관한 정수장과 배수지의 용량이 이를 뒷받침할 수 있는가에 대해 논의한다. 이를 위해 우선 정수장에 대해서 살펴보면 관할 정수장은 D 정수장으로서 시설용량은 75만m³/일이며 2004년 평균 생산량은 54만m³/일이며 평균 운전율이 72%이다. 이 운전율을 85%까지 높인다면 추가로 97,500m³/일을 생산할 수 있어서 적극적인 개발을 하는 경우에 증가분인 89,034m³/일을 감당할 수 있다.

또한 이 지역을 관할하는 네 개의 배수지에 대해서 관할 행정구역을 확인한 다음에 해당 배수지에 물수요 증가량을 대입하여 체류시간을 구하였으며, 결과는 Fig. 9와 같다. A 배수지는 모두 적극적인 개발을 하더라도 배수지 체류시간이 12시간을 확보하고 있으며, D배수지는 적극적인 개발인 경우에 12시간을 만족시키지 못하는 것으로 나타났다. N 배수지와 B 배수지는 모두 12시간에 이르지 못하고 있다. 하지만, 본 연구에 사용된 자료들은 배수지의 체류시간을 8시간을 확보하도록 하였을 때의 자료들로서 과거의 8시간 기준이었을 때에 버퍼용량인 6시간은 모두 넘는 값이고, 현재에는 체류시간 12시간을 확보하기 위해 해당 배수구역에 배수지의 용량이 증설되고 있기 때문에 큰 문제는 없을 것으로 판단된다.

**5. 결론**

본 연구는 서울 도심부인 종로구와 중구를 대상으로 물수요 예측식 산정시 발생하는 문제점을 해결하기 위하여 시작하였으며, 다음과 같은 결과들을 얻었다.

첫째, 물수요 예측식에서 발생한 문제점을 해결하기 위해 6개의 지역특성을 나타내는 인자들을 선택하였으며, 이들에 대해 주성분분석을 하여, 전체 정보량의 92.6%를 표현할 수 있는 4개의 주성분을 선택하였다.

둘째, 이들 4개의 주성분을 바탕으로 군집분석을 실시한 결과 대상지역 내에서는 두 개의 군집으로 명확히 나누어졌으며, 실제 위치를 확인해 본 결과 주거지역과 상업지역으로 나누어진 것을 확인할 수 있었다.

셋째, 각 나누어진 군집에 대해서 다시 중회귀 모델을 작성한 결과 명확한 물수요량 증가요인인 거주인구수의 파라미터가 음의 값에서 양의 값으로 바뀌었으며 결정계수에 있어서도 0.8이상의 양호한 값을 가졌다.

넷째, 작성된 물수요 예측식을 바탕으로 장래의 개발시나리오에 따른 물수요 증가량을 산정한 결과 적극적인 개발에서 89,034 m³/일, 중간적인 개발에서 49,077 m³/일, 소극적인 개발에서 19,996 m³/일의 물수요량이 증가하는 것으로 나타났다.

다섯째, 이러한 장래 개발 시나리오별 물수요증가량에 대해서 해당 정수장과 배수지를 대상으로 용량을 검증해 본 결과 정수장은 운전율을 85%까지 높이면 수요량을 감당할 수 있는 것으로 나타났으며, 배수지의 경우 12시간 기준으로는 만족스럽지 못한 결과이지만, 배수지 증설 계획에 따라 해결될 것으로 판단된다.

## 사 사

본 연구는 21세기 프론티어연구개발사업인 수자원의 지속적 확보기술개발사업단의 연구비지원(4-2-2)에 의해 수행되었습니다.

## 참 고 문 헌

1. Billings, R. B. and Jones, C. B., Forecasting Urban Water Demand. Americal Water Works Association, Denver, Colorado(1996).
2. 환경부, 환경백서(2003).
3. 한국수도협회, 상수도 시설기준(1997).
4. 具滋茸, “都市の短期水需要豫測に關する研究,” 東京都立大學大學院 工學研究科 博士學位論文(1997).
5. 최재호, 주충남, 구자용, 유명진, “동적계획법을 이용한 장기 상수도 시설계획의 최적화연구,” 서울시립대학교 산업기술연구소, 7, 145~152(1999).
6. 서울특별시 상수도 사업본부, 서울특별시 수도정비기본 계획(2000).
7. 서울특별시, 서울통계연보(1996, 2001).
8. 변신숙, 시스템 다이내믹스 모델을 이용한 장기물수요 예측, 서울시립대학교 석사학위논문(2003).
9. 서울시립대학교 도시과학연구원, “서울 도심부 하부기반시설의 수용용량분석”(1998).
10. 서울시립대학교 도시과학연구원, 도심부 및 청계천 주변지역 하부기반시설 수용용량분석(2003).
11. Grove, D. M. and Roberts, C. A. Principal component and cluster analysis of 185 large towns in England and Wales, Urban Studies, 17, 77~82(1980).
12. 김영주, “주성분분석에 의한 도시호수의 수질평가에 관한 연구,” 환경관리학회지, 9(2), 197~203(2003).
13. 남영우, 성은영, “인자분석과 군집분석에 의한 세계도시의 유형화,” 한국도시지리학회지, 4(1), 1~12(2001).
14. Jolliffe, I. T., Principal Component Analysis, 2nd ed, Springer Verlag, Heidelberg(2002).
15. Brian, S. E., Sabine, L., and Morven, S., Cluster analysis, 4th ed, Edward Arnold, London(2001).
16. Goldberg, M. A. and Cho, H. A., Introduction to Regression Analysis, Springer Verlag, Heidelberg. Heidelberg (2003).