

# 회귀모형에 의한 상수도 1일 급수량 예측에 관한 연구

## A Study on the Prediction of Daily Urban Water Demand with Multiple Regression Model

박 성 천\* · 문 병 석\*\* · 오 창 주\*\*\* · 이 병 조\*\*\*\*  
Park, Sung Chun · Moon, Byoung Seok · Oh, Chang Ju · Lee, Byoung Jo

### Summary

The purpose of this paper is to establish a method estimating the daily urban water demand using statistical analysis that is used for developing the efficient management and operation of the water supply facilities, and accuracy of the model is verified by error rate and F-value. The data used in this study were the daily urban water use, the weather conditions such as temperature, precipitation, relative humidity, etc, and the day of the week. The case study was taken placed for the city of Namwon in Korea. The raw data used in this study were rearranged either by month or by season for analysis purpose, and the statistical analysis was applied to the data to obtain the the regression model. As a result of this study, the linear regression model was developed to estimate the daily urban water use with weather condition. The regression constant and coefficients of the model were determined for each month of a year. The accuracy of the model was within 3% of average error and within 11% of maximum error. The resulting model was found to be useful to the practical operation and management of the water supply facilities.

### I. 서 론

급격한 공업화와 경제성장에 따른 산업구조의 변화와, 국민의 생활수준이 크게 향상되고, 인구가 도시에 편중되어 도시의 비대화를 초

래하고 있다. 이러한 현대사회는 상수도계통 뿐만 아니라, 도시하수나 공장폐수, 분뇨 등에 의하여 자연수역이 오염되고, 그 결과 보건위생상 안전을 보장할 수 없는 상태를 초래하기에 이르렀다. 또한 물 수요량도 가정, 공업,

\* 동신대학교 공과대학

\*\* 서남대학교 공과대학

\*\*\* 전남대학교 대학원(박사과정)

\*\*\*\* 전남대학교 대학원(석사과정)

키워드: 1일 급수량, 예측, 요일 및 기후인자, 수정결정계수, 다중회귀분석, 회귀모형식, F검정

상업용 등 여러 방면에서 수요량이 증대하여 종래의 용수로 시설로는 수량면에서 수요를 충족시키지 못하고 있는 실정이다.

따라서 시민이나 국민이 안심하고 먹을 수 있는 음용수를 안정적으로 공급하기 위하여 상수도 시설의 확충과 개량도 중요하지만 이에 앞서 기존의 시설을 수요량에 맞춰서 질 높은 상수를 생산공급할 수 있는 합리적인 운용방법을 개발하고, 이에 따라서 상수도시설을 운용하는 것이 시급한 일이 아닐 수 없다. 이를 위해서는 먼저 수요량을 예측하고 이에 따라서 상수도 시설을 유효·적절하게 운용하는 것이 중요하다.

상수도 1일 급수량을 예측하는 방식은 과거의 자료로부터 통계학적 방법에 의한 다중회귀분석법과 시계열분석법으로 급수량을 예측할 수 있다.<sup>3,9)</sup> 또한 예측기간에 따라 단기예측과 장기예측으로 구분할 수 있다. 단기예측은 예측기간을 시간단위나 하루단위로 구분하여 급수량을 추정하는 것을 말하고, 이는 급·배수량을 실시간 제어하는데 유효하며, 장기예측은 예측구간을 연별, 계절별 또는 월별로 구분하여 평균급수량을 예측하고, 예측된 자료는 급·배수시설의 효율적인 운영계획에 필수적인 자료이다. 급수량 산정에 관한 연구는 미국이나 일본 등 선진국에서 선행되어 왔다. 급수량과 기후인자와의 상관관계는 1967년 How와 Linawever<sup>1)</sup>가 기온과 강우량 등의 기후인자를 도입하여 급수량 연구를 처음 시도할 때 확인되었다. 그 이후 Young<sup>2)</sup>(1973)은 연별로 변화되는 급수량의 자료를 해당도시의 인구, 세대의 평균수입, 수도요금과 더불어 강우량, 기온, 증발량을 연관시키려고 시도하였다. R. J. Frankel과 P. Shouvanavirakul<sup>3)</sup>(1973)은 급수수요모형을 1차선형의 다중회귀분석에 여름철과 겨울철의 두 요소속에 포함하여 급수수요량이 결정된다고 하였다. D.E. Agthe & R.B. Bilings<sup>4)</sup>(1980)는 Young이 급수량의

변화에 영향을 주는 인자로 선택한 변수를 사용하여 월별로 변화하는 급수량을 계산하는 방식을 제안하였다. D. R. Madidment & Parzen<sup>5)</sup>은 월별 평균급수량을 기후변동에 따라 산정하였으며 기후변동인자로 계절적 변동, 강우량, 증발량, 기온을 이용하였다.

국내의 연구동향을 이경훈 외 2인<sup>6,7,8)</sup>은 상수도 1일 급수량의 시간별, 요일별, 계절적 분포에 관한 연구, 기후인자와 경년변화 등을 고려한 회귀식 및 시계열 분석법인 ARIMA 모형을 이용한 상수도 단기예측을 시도하여 좋은 결과를 얻고 있다. 그러나 급수량예측과 이의 활용에 대한 국내에서 연구는 아직 미흡하여 앞으로 이 분야의 연구가 더욱 활발히 추진되어야 할 것으로 판단된다. 따라서 본 연구에서는 남원시의 1일 상수 수요량을 추정하기 위하여 급수량에 영향을 주는 변동요인으로 요일 및 기후인자를 사용하여 1일 급수량 예측모형식을 제시하여 상수도 시설을 합리적이고 과학적으로 운용하는 데 필요한 자료확보와 운영방법의 개선을 목적으로 연구하였다.

## II. 급수량 예측 이론

### 1. 다중회귀분석

회귀분석법(regression analysis)은 종속변수와 설명변수간의 함수적인 관련성을 규명하기 위해서 임의의 수학적 모형으로 가정하고, 이 모형을 측정된 변수들의 자료로부터 실측된 자료에 적합한 회귀방정식에 접근하여 추정하는 통계분석의 한 방법이다.

종속변수가 하나이며 여러개의 설명변수가 있을 때 설명변수들이 종속변수에 어떻게 영향을 미치고 있는가를 분석하는 것이 다중회귀분석(multiple regression analysis)이다. 1개의 설명변수 X와 연관지을 수 있는 종속변수 Y가 1개 존재하는 모집단을 생각해 보면

다음과 같이 나타낼 수 있다.

$$\begin{aligned} Y_1 &= C_0 + C_1 X_{11} + \varepsilon_1 \\ Y_2 &= C_0 + C_1 X_{21} + \varepsilon_2 \\ &\vdots \\ Y_n &= C_0 + C_1 X_{n1} + \varepsilon_n \dots\dots\dots (1) \end{aligned}$$

로 쓸 수 있으며  $k$ 개의 설명변수로 확장하면  $X$ 와 연관지을 수 있는 종속변수  $Y$ 가  $n$ 개 존재하는 모집단을 생각해 보면 다음과 같이 나타낼 수 있다.

$$\begin{aligned} Y_1 &= C_0 + C_1 X_{11} + C_2 X_{12} + \dots + C_k X_{1k} + \varepsilon_1 \\ Y_2 &= C_0 + C_1 X_{21} + C_2 X_{22} + \dots + C_k X_{2k} + \varepsilon_2 \\ &\vdots \\ Y_n &= C_0 + C_1 X_{n1} + C_2 X_{n2} + \dots + C_k X_{nk} + \varepsilon_n \\ &\dots\dots\dots (2) \end{aligned}$$

여기서  $C_0$ 는 회귀상수,  $C_1, C_2, \dots, C_k$ 는 회귀계수이며  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 는 오차항이다.

식 (3.1)을 행렬을 사용하여 표기하면

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} C_0 \\ C_1 \\ \vdots \\ C_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \dots (3)$$

이다. 따라서

$$\underline{Y} = \underline{XC} + \underline{\varepsilon} \quad (4)$$

$(n \times 1) \quad \{n \times (k+1)\} \{ (k+1) \times 1 \} \quad (n \times 1)$

오차항 조건은  $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$ , 그리고  $Cov(\varepsilon_i, \varepsilon_j) = 0$ 이므로 독립적인 정규확률 변수이다.  $E(\underline{Y}) = \underline{XC}$ 이며 표본자료에서 구할 수 있는  $\underline{b}$ 를  $\underline{C}$ 의 추정벡터라 하면

$$\hat{\underline{Y}} = \underline{Y} = \underline{Xb}$$

$$\underline{Y} = \underline{Xb} + \underline{e} \dots\dots\dots (5)$$

이며  $\underline{e}$ 는 실측치와 추정치의 차이인 잔차벡터이다.

모집단에서의 회귀계수 벡터의  $\underline{\beta}$ 와 오차분산  $\sigma^2$ 의 값은 표본자료를 이용하여 잔차자승의 합을 최소로 하는  $\underline{b}$ 로부터 추정할 수 있다.

$$\begin{aligned} \sum e_i^2 = e' e &= (\underline{Y} - \underline{Xb})' (\underline{Y} - \underline{Xb}) \\ &= \underline{Y}' \underline{Y} - 2\underline{b}' \underline{X}' \underline{Y} + \underline{b}' \underline{X}' \underline{Xb} \\ &= \underline{Y}' \underline{Y} - \underline{Y}' \underline{Xb} \dots\dots\dots (6) \end{aligned}$$

이다. 이것을  $\underline{b}$ 에 관하여 미분한 후에 0으로 놓으면,

$$\frac{\partial e' e}{\partial \underline{b}} = -2\underline{X}' \underline{Y} + 2\underline{X}' \underline{Xb} = 0$$

이다. 따라서 다중회귀모형의 계수벡터  $\underline{b}$ 에 관하여 정리하면 다음과 같다.

$$\underline{b} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{Y} \dots\dots\dots (7)$$

## 2. 회귀모형의 적합성

통계학적인 적합성검정 방식은 계산된 회귀방정식이 통계적으로 유의한가를 검정하는 것이다. SST(Total Sum of Squares)은 SSE(Sum of Squares due to Errors)와 SSR(Sum of Squares due to Regression)로 분류되며, 적합성검정을 위한 분산분석(analysis of variance)은 SSE 변동평균(MSR)과 SSR 변동평균(MSE)의 비율을 계산한  $F$ 값을 임계치와 비교하여 결정한다.

자료의 SST를 계산하면,

$$\begin{aligned} SST &= \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n(\bar{Y})^2 \\ &= \underline{Y}' \underline{Y} - n(\underline{Y})^2 \dots\dots\dots (8) \end{aligned}$$

이다. 다음으로 SSR을 구하면,

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 = \sum \hat{Y}_i^2 - n(\bar{Y})^2$$

$$= \underline{\hat{Y}}' \underline{\hat{Y}} - n(\bar{Y})^2$$

이다. 여기에 (5)식과 (7)식을 대입하여 정리하면,

$$SSR = \{ (X'X)^{-1}XY \}' X' X \{ (X'X)^{-1}XY \}$$

$$- n(\bar{Y})^2 = \underline{b}' X' \underline{Y} - n(\bar{Y})^2 \dots\dots\dots (9)$$

이며

$$SSE = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{Y}_i)(Y_i - \hat{Y}_i)$$

$$= \underline{Y}' \underline{Y} - \underline{b}' X' \underline{Y} \dots\dots\dots (10)$$

이다. 자유도를 계산하면, SST는 일반적으로  $n-1$ 개의 자유도를 갖는다. SSE는  $n-(k+1)$ 의 자유도를 가지며, SSR은 독립변수  $k$ 개의 갯수만큼 자유도를 갖는다.

따라서  $MSR = \frac{SSR}{k}$ ,  $MSE = \frac{SSE}{n-k-1}$ ,  $F = \frac{MSR}{MSE}$ 로 나타낼 수 있다.

다음은 종속변수  $Y$ 와 한 세트의 설명변수  $X_1, X_2, \dots, X_k$  사이의 관계가 있음을 검정하는 과정을 모색하기로 한다. 회귀모형을 검정하기 위해서는 먼저 회귀모형에 관한 가설을 세우고 표본에 대한  $F$  값을 계산한 후 자유도를 고려한 임계치와 비교하여 가설의 채택 여부를 결정한다.

유의수준  $\alpha$ 에서,  
만일  $F \leq F_{(\alpha:k, n-k-1)}$ 이면 귀무가설(null hypothesis)이 채택되고

만일  $F > F_{(\alpha:k, n-k-1)}$ 이면 귀무가설(null hypothesis)이 기각된다.

### 3. 회귀모형의 정도 및 상관도

회귀모형이 어느정도로 실측된 자료를 잘

설명하고 있는가 하는 문제는 회귀모형의 다중결정계수( $R^2$ : coefficient of multiple determination)를 살펴 봄으로서 알 수 있다. 그런데  $R^0$ 값에 대한 해석에 주의할 필요가 있다.  $R^2$ 이 크다고 해서 반드시 적합한 모형이라고 할 수 없다. 왜냐하면 어떤 예측은,  $R^2$ 이 크다 하더라도, MSE가 너무 커서 추론을 할 수 없을 수도 있기 때문이다. 그러므로 이러한 폐단을 막기 위하여 결정계수를 수정 변경하여 측정한, 수정결정계수( $R_a^2$ : adjusted coefficient of determination)를 써야 한다.

$$R_a^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

$$= 1 - \left( \frac{n-k}{n-k-1} \right) \left( \frac{SSE}{SST} \right) \dots\dots\dots (11)$$

$R_a^2$ 는 설명변수의 수가 늘어난다 하더라도, 이를 보완하여 결정계수값에 영향을 주지 않는 장점을 갖고 있어 모형의 적합성을 판단하는 중요한 인자로 인식되고 있다.

다음은 종속변수와 설명변수, 설명변수간의 상관성을 조사하는 방식에 대하여 기술하기로 한다. 주관심이 되는 두 변수의 선형관계의 방향과 정도는 상관계수(correlation coefficient)를 이용하는데 상관계수  $\rho_{XY}$ 는 식 (12)과 같다

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, -1 \leq \rho \leq 1 \dots\dots\dots (12)$$

여기서,  $\rho_{XY}$ 는  $X$ 와  $Y$  두 변수의 공분산이며  $\sigma_X$ 와  $\sigma_Y$ 는 각각  $X$ 와  $Y$ 의 표준편차이다. 지금  $\sum (X_i - \bar{X}) = \sum X_i^*$ ,  $\sum (Y_i - \bar{Y}) = \sum Y_i^*$ 라 하면, 모집단상관계수는 식 (13)과 같은 표본상관계수로 구할 수 있다.

$$r = \frac{\sum X_i^* Y_i^*}{\sqrt{\sum X_i^{*2}} \sqrt{\sum Y_i^{*2}}} \dots\dots\dots (13)$$

### III. 대상지역 및 연구자료

본 연구의 대상지역으로는 남원시를 선정하여, 1일 급수량자료와 기후자료를 이용하여 본 연구를 하였다. 남원시의 1995년 12월말 현재 급수구역내 인구는 56,229명, 급수인구는 50,450명으로 급수보급율은 89.7%에 달하고 있다.

본 연구에서 1일 급수량을 예측하기 위하여 사용한 자료는 남원시에서 측정한 1995년부터 1996년 까지의 2년간 자료를 사용하였다. Fig. 1은 1996년 남원시 1일 급수량 자료를 표시하고 있다.

1일 급수량의 추정 설명변수인 기후자료는 최고기온, 평균기온, 최저기온, 상대습도, 강우량, 운량, 일조시간 등을 사용하였다. 기후자료는 남원지방 기상청에서 관측한 자료를 사용하였다.

기후이외의 변수로는 요일계수를 사용하였고, 요일계수는 월요일부터 금요일(휴일은 제외)까지는 1을, 토요일은 2, 일요일 및 휴일은 3의 요일계수를 사용하였다. 1일 급수량의 추정에 중요한 변수인 급수인구는 남원시의 인구증가가 등차적으로 증가한다고 가정하여 계산하면 하루 약 2명 정도의 인구가 증가하고 있어 인구수가 그렇게 큰 의미를 가지고

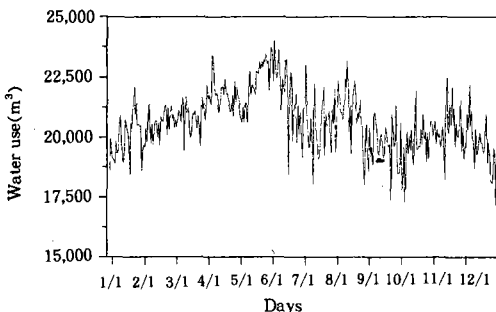


Fig. 1. The variation of daily water use in the city of Namwon for the year of 1996

있지 않고, 상관분석을 하여도 상관계수가 0.2를 넘지 못하므로 본 연구에서는 제외하였다. 또한 생활수준과 문화수준의 향상에 따라 1인 1일 급수량은 매년 증가를 반영할 수 있는 경년변화는 관측된 자료의 연수가 2년밖에 되지 않으므로 고려한다는 것은 무리라고 판단되어 본 연구에서는 제외하였다.

### IV. 분석결과 및 고찰

#### 1. 상관분석

상수도 1일 급수량의 회귀모형을 개발하기 위해서 회귀모형의 구성에 필요한 최고기온, 평균기온, 상대습도, 강우량, 평균운량, 일조시간, 등의 기후인자를 설명변수로 해서 상호간에 상관관계와 독립성을 확인하고자 상관분석을 하였다. 상관분석에 이용한 자료는 시계열 구간을 월별, 계절별 구간으로 분류하여 상관계수를 Fig. 2와 Fig. 3에서와 같이 구하였다.

Fig. 2에서, 먼저 기온에 대해서 살펴보면 1일 급수량과 기후인자 중에서 기온과의 상관성이 가장 높게 나타났으며, 계절별, 월별구간에서 최고기온, 평균기온, 최저기온의 상관계수는 -0.21~0.87 정도로 나타났고, 최고기온, 평균기온, 최저기온의 순으로 나타났다. 또한 설명변수간의 상관계수는 최고기온, 평균기온 및 최저기온 사이에 높은 상관성(0.63~0.90)이 존재하였다. 따라서 급수량 예측모형에서는 최고기온을 설명변수로 채택하였다.

Fig. 3에서, 강우량과 상대습도, 운량은 하나의 계열로 간주하고 이들 인자가 급수량에 미치는 영향과 인자상호간의 상관성을 조사하였다. 강우량, 습도와 운량이 급수량에 미치는 영향을 서로 비교하면 강우량의 상관계수는 -0.1~-0.44로 상대습도와 운량의 영향보다 적게 나타났으며, 상대습도는 -0.1~-0.40, 운량은 -0.1~-0.60으로 이들 인자와 급수량과의 상관도는 역상관관계를 이루고 있으며, 계

Table 1. Example of input data

Water use	Average temp. (°C)	Maximum temp. (°C)	Minimum temp. (°C)	Relative humidity (%)	Cloudiness	Precipitation (mm)	Sunshine duration (hr)	A Day of the week coefficients
19,155	2.0	12.3	-4.0	63	10.0	12.2	0.0	3
17,177	-3.3	-0.8	-4.6	58	8.6	8.6	0.5	3
19,247	-3.8	1.9	-8.6	56	2.1	2.1	7.8	1
19,506	-2.9	4.1	-8.7	58	2.9	2.9	6.3	1
18,282	-5.5	0.7	-7.6	69	9.8	9.8	0.0	1

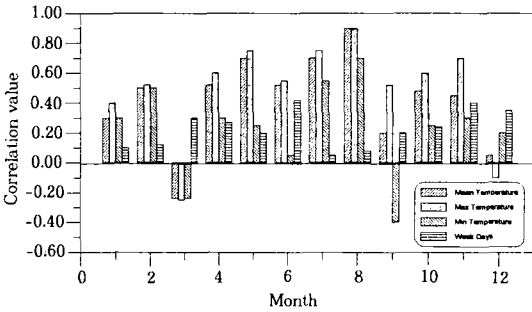


Fig. 2. Correlation value between water use and air temperature, a week days

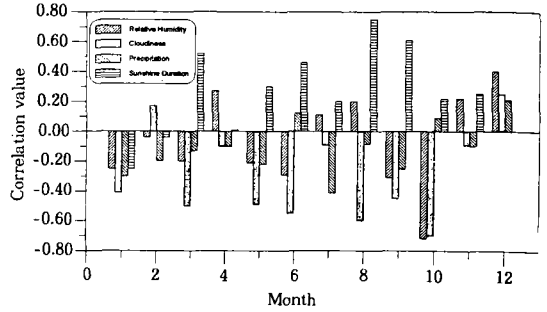


Fig. 3. Correlation value between water use and used parameters

절적으로 다르게 영향을 미치고 있는 것으로 나타나고 있다. 봄, 가을, 겨울철에는 이들 인자와 급수량과의 상관도는 상관계수가 -0.2보다 작게 나타나고 있으나, 여름철에는 이들 인자와 급수량과의 상관도가 높아지는 것으로 나타나고 있다.

급수량에 대한 습도의 상관도를 보면 계절에 상관없이 상관도가 거의 -0.3 정도로 작게 나타나고 있다. 또한 강우량과 습도 및 운량의 상관계수는 0.6과 0.8사이로 높은 상관성을 보였다.

일조시간의 상관계수를 보면 1~2월까지의 미미한 상관성(0.2정도)을 보이다가, 3월부터 상승하기 시작하여 8월달에 최고 0.72정도로 높아지다가 9월부터 다시 감소하는 경향을 보이고 있다.

요일에 대한 상관계수를 보면 1월부터 5월까지의 0.1정도의 상관을 보이다가 6월에 0.30

으로 상승하였다가 다시 감소하였다가 11월 12월에는 약간 높은 0.41과 0.349를 보이고 있다. 이와 같이 11월과 12월에 급수량과 요일간의 상관계수가 다른 달보다 높은 것은 남원시는 소도시이기 때문에 겨울 김장철과 관계가 있는 것으로 판단된다.

## 2. 회귀모형식

다중회귀 모형식의 주요인자에 대한 특성과 인자상호간의 상관분석을 통하여 모형식을 제안하고 모형식을 이루는 인자와 모형식의 계수를 결정하였다. 모형식에 사용된 주요인자는 8개의 변수로서 급수량에 영향을 주지 않는 변수나, 설명변수 상호간의 상관도가 높은 변수를 제외하는 조합으로 4개의 회귀모형을 설정하고 이 지역에 가장 적합한 회귀모형을 도출하고자 하였다. 4개의 회귀모형식은 Table 2와 같이 구성하였다.

**Table 2. Regression model for predicting the municipal water use with weather condition**

	Regression model
Model No. 1	$C_0 + C_1 X_1 + C_2 X_2 + C_3 X_3 + C_4 X_4 + C_5 X_5$
Model No. 2	$C_0 + C_1 X_1 + C_2 X_2 + C_4 X_4 + C_5 X_5 + C_6 X_6$
Model No. 3	$C_0 + C_1 X_1 + C_3 X_3 + C_4 X_4 + C_5 X_5 + C_6 X_6$
Model No. 4	$C_0 + C_1 X_1 + C_2 X_2 + C_5 X_5$

**Table 3. Regression coefficients of model 1 obtained for each month and season**

Months & Season	C <sub>0</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>
March	17,780	-51.3	21.1	21.2	7.7	140	110.0
August	14,844	205	11.8	29.1	-6.76	22.1	-109.0
October	201,805	122.0	-33.7	8.0	22.7	5.0	120.0
December	13,216	26.9	71.3	222	19.1	220	80.0
Spring	17,211	81.5	22.6	-93.3	-2.8	1.1	241.0
Summer	13,517	249	11.4	-11.9	7.81	-12.7	27.0
Fall	19,369	91.3	-11.9	29.0	3.1	36.0	120.0
Winter	17,961	45.1	12.9	89.1	-46.8	90.4	1.0

Table 3은 시계열자료를 월별(3, 8, 10, 12월) 계절별로 구분한 모형 1에 대한 회귀식이다.

### 3. 모형의 적합성 검증

급수량 예측을 위한 회귀모형식의 적합성 검증은 실측치와 예측치를 비교하고 통계적으로 유의한지 여부를 검증하기로 한다. 자료의 시계열구간은 월별, 계절별 단위로 하였으므로 시계열구간 단위가 모형에 미치는 영향과 회귀모형식에 대한 적합성을 검토하여야 한다. 시계열구간에 따른 모형의 정도를 비교하기 위하여 수정결정계수(adjusted coefficient determination)와 모형별 오차율을 살펴보았다.

시계열구간을 월별구간으로 회귀모형을 구성하였을 때 수정결정계수의 최저값은 최저 88.9%(6월달)에서 최대 94.8%(7월달)로 나타났다.

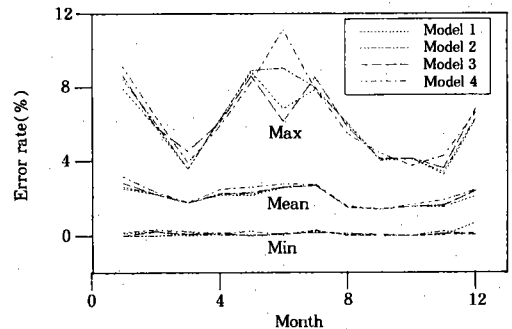
시계열구간을 계절별로 회귀모형을 구성할

시에선 수정결정계수의 최저값은 83.3%이며 겨울철의 모형 4에서 나타난다. 또한 최대값은 92.8%로 가을철에 대하여 모형 2에서 나타나고 있다.

월별, 계절별에 따라 구성한 회귀모형에 대한 수정결정계수를 비교하면 회귀모형의 정확도는 월별, 계절별 단위로 구성된 모형식의 순서임을 알 수 있다. 따라서 급수량 예측에 관한 회귀모형은 월별단위로 구성하는 것이 바람직하다고 판단된다.

다음은 회귀모형의 오차율에 대하여 검토하기로 한다. 여기에서는 회귀모형을 월별단위, 계절별로 구성하였을 때의 오차율을 비교하기로 한다. Fig. 4는 회귀모형을 월별단위로 구성할 때 발생하는 회귀모형의 오차율이며 Fig. 5는 회귀모형을 계절별로 구성할 때 발생하는 회귀모형의 오차율이다. Fig. 4와 5에 나타난 오차율을 비교하면 시계열구간을 월별단위로 구분하여 회귀모형을 구하는 것이 오차율을 줄일 수 있다는 것을 확인할 수 있다. 따라서 수정결정계수에 이어서 오차율의 비교에 의하여 회귀모형은 월별단위로 구성하는 것이 정도가 좋다는 것이 확인되었다.

Fig. 4의 모형별 오차율을 비교하여 보면 4개의 모형별 오차율은 평균 및 최소오차율은 모든달에 대해 대동소이 하지만 최대오차율은



**Fig. 4. Error rate of the regression model obtained using the monthly data for the year of 1996**

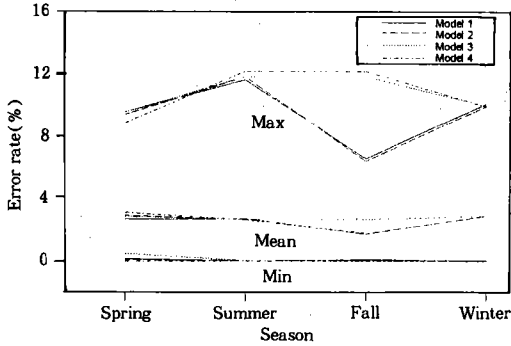


Fig. 5. Error rate of the regression model obtained using the seasonal data for the year of 1996

6월달에 상당한 오차(약 5%)를 보이고 있다.

Fig. 5의 각 월별 모형에 따른 최대오차율을 보면, 1월, 2월, 3월, 7월, 9월, 11월, 12월은 모형1이 우수하지만, 4월, 5월, 6월, 8월, 10월은 모형4가 모형1보다 최대오차율면에서 더 정확한 것을 알 수 있다. 그러나 모형 4가 4, 5, 8, 10월에서는 모형 1보다도 더 정확한 예측값을 가지고 있기는 하지만 모형 1과 비교하여 최대 오차율의 차가 1%을 넘지 않았고 다른 달은 모형 1이 모형 4보다 최대 오차율에서 최대4% 정도의 오차율을 보이고 있으므로 모형 1이 모형 4보다 우수한 모형이라고 말할 수 있다.

평균오차율 및 최소 오차율을 보면, 모형 1 및 모형 4 모두 거의 비슷한 오차율을 보이고 있고. 따라서 모형을 구성하는데는 최고기온, 상대습도, 운량, 강우량, 일조시간, 요일을 고려한 모형 1로 회귀식을 구성하는 것이 가장 바람직 할 것으로 판단된다.

통계학적인 검정방식은 계산된 회귀모형이 통계적으로 유의한가(statistical significant)를 검정하는 것이다. 적합성 검정은 분산분석(analysis of variance)의 F 값을 계산한 후 계산된 F값에 대한 분자와 분모의 자유도에 대응하는 임계치가 유의수준에 적합한지를 판

Table 4. Analysis of variance of January

Model No.	Source	DF	F-ratio	Critical value	Significance level
1	Regression	6	20.53	3.11	0.01
	Residual	54			
	Total	60			
2	Regression	5	26.71	3.34	0.01
	Residual	54			
	Total	59			
3	Regression	5	26.79	3.34	0.01
	Residual	54			
	Total	59			
4	Regression	3	30.48	4.13	0.01
	Residual	54			
	Total	58			

Table 5. Analysis of variance of Spring

Model No.	Source	DF	F-ratio	Critical value	Significance level
1	Regression	6	34.41	2.80	0.01
	Residual	178			
	Total	184			
2	Regression	5	36.13	3.22	0.01
	Residual	177			
	Total	182			
3	Regression	5	38.28	3.22	0.01
	Residual	177			
	Total	182			
4	Regression	3	40.58	3.78	0.01
	Residual	177			
	Total	175			

단하는 방식을 말한다.

월별모형 모두 분산검정을 수행하고 그 중에서 1월달에 해당되는 모형 1, 2, 3, 4의 4개의 모형에 대하여 F검정 절차로 검정한 것을 Table 4에 나타냈다.

실제로 F-검정의 기각치(critical value)는 모형 1인 경우  $F_{(6,54;0.01)}=3.11$ 인데  $F=20.53$ , 모형 2는  $F_{(5,54;0.01)}=3.34$ 인데  $F=26.71$ , 모형 3은  $F_{(5,54;0.01)}=3.34$ 인데  $F=26.79$ , 모형 4는  $F_{(3,54;0.01)}=4.13$ 인데  $F=30.48$  이므로 도출한 회귀식은 급수량을 유의하게 설명하는 것으로 나타났다. 모형들도 따라서 월별 모형식 모형 1, 2, 3, 4 모두 적합한 모형이라



고 판단이 되었다.

Table 5은 계절별(봄)의 F검정의 값을 나타내 것이며, Table 5를 보면 알 수 있듯이 계절별 모형에 대해 모두 적합한 것으로 판명되었다.

Fig. 6은 2월의 실측된 급수량과 모형에 의한 예측치를 보여주고 있다. 모형간의 오차율을 비교하면 모형 4개의 오차율, 즉, 최대오차율은 모형 1에서 5.90%, 모형 4에서 6.47%, 평균오차율에서도 모형 1이 모형 4보다 적은 오차를 보이고 있지만 그 차는 미미하다. 따라서 모두 모형 1로 모형식을 구성하는 것이 바람직하다고 판단된다.

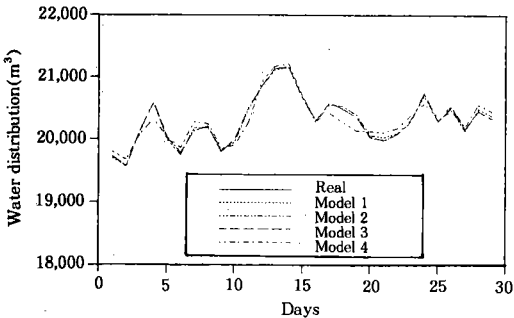


Fig. 6. The comparison between the measured and the predicted municipal water use for the month of February of 1996

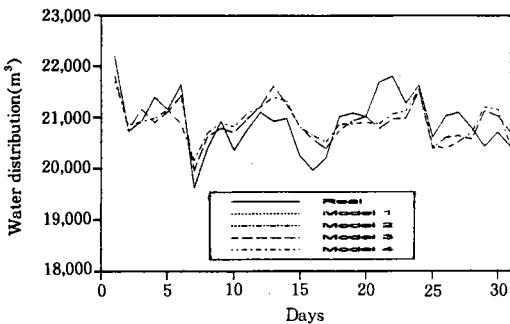


Fig. 7. The comparison between the measured and the predicted municipal water use for the month of October of 1996

Fig. 7은 10월의 급수량과 예측치를 보여주고 있다. 모형간의 오차율은 비슷하나 최대오차율은 모형 3에서 발생하였으며 최대오차율은 4.19%로서 타 모형에 비하여 0.5% 정도의 오차가 많은 것으로 나타나고 있다. 평균오차율은 2% 이내로 나타나고 있으며 F검정 결과 회귀모형식이 유의수준 1% 범위 내에서 적합한 것으로 판정되었다.

앞에서 검토한 2, 4, 10월을 제외한 나머지 1, 3, 4, 9, 11~12월에 대하여 조사한 결과 나머지 달에 대해서도 평균 오차율은 2% 이내이며 최대 오차율은 4%~11% 이내에 들어가는 모형으로 판정이 되었다.

단지 본 모형에서 미흡한 점이라면 급수량이 급증하거나 급감할 때 오차율이 커진다는 점을 들 수 있으나 오차율이 최대로 발생한 7월의 경우에도 모형의 평균오차율이 2.5%내외이고 최대오차율이 10%내외이었다. 겨울철 모형은 계절별 모형 중에서 오차가 가장 크게 발생하고 있는데 여름철 모형에 비하여 모형 1은 최대 0.594%, 모형 2는 2.83%, 모형 3과 모형 4에서는 각각 여름철의 오차가 0.5%, 1.2%정도 크며, 평균오차율은 0.3~1.2%정도 여름철의 오차가 크다. 1일 급수량 예측에 있어서 최대 10% 오차율은 2.4시간의 급수량에 해당하는 오차로서 시설을 운영하는 측면에서 배수지용량에 견주어 비교하여 보면 수용할 만한 오차율이라고 할 수 있다.

## V. 결 론

본 연구에서는 남원시를 대상으로 상수도 1일 급수량자료, 및 기후자료, 요일계수를 이용하여 회귀분석을 실시하였다. 또한 1일별로 변화되는 급수량을 예측하기 위한 회귀모형식을 제안하였으며, 다음과 같은 결론을 얻었다.

1. 요일 및 기후인자를 주요인자로 이용하여 상수도 1일 급수량의 예측을 위한 선형 다중

회귀모형을 구성할 수 있었으며, 본 연구에서 제시한 남원시에 대한 모형은 평균오차율이 평균 3% 이내이며 최대오차율이 10% 이내로서 비교적 좋은 정확도를 갖는 모형으로 판단하였다.

2. 상수도 1일 급수량은 회귀모형의 시계열 구간을 연별, 월별, 계절별 구간으로 분류하여 예측한 결과, 연별, 계절별구간 보다 월별구간의 장도가 높게 나타나 월별구간의 회귀모형으로 예측하는 것이 적절하다고 판단하였다.

3. 상수도 1일 급수량을 예측하는 회귀모형의 설명변수는 본 연구에서 사용한 7개의 기후인자(1일 평균기온, 1일 최고기온, 1일 최저기온, 상대습도, 강우량, 운량, 일조시간)중에서 1일 최고기온, 상대습도, 운량, 강우량, 일조시간과 기후이외의 인자로 요일로 분석하였다.

4. 본 연구에서 제시한 모형은 상수도 시설의 운영시 1일 계획 급수량을 추정하는 데 유용하게 사용할 수 있으나 생산량과 급·배수지역의 수요량을 실시간으로 비교한 단위 시간별 생산량을 추정하는데 미흡하므로 이에 대한 추가적인 연구가 필요하다고 판단되며, 보다 더 정확한 모형의 개발을 위하여 신경망이론 및 퍼지이론 등의 최신이론을 이용한 연구가 더 진행되어야 할 것으로 판단된다.

## 참 고 문 헌

1. C. W. Howe and F. P. Linaweaver, "The Impact of Price on Residential Water Demand and It's Relation to System Design and Price Structure", Water Resources Research, Vol.14, No. 4, pp. 13~32, 1967.
2. R.A. Young, "Price Elasticity of Demand for Municipal Water : a Ease Study of Tuscon, Arizona", Water Resources Research, Vol.9, No.4, pp. 1068~1072, 1973.
3. R. J. Frankeland P. Shouvanavirakul, "Water Consumption in small Communities of Northeast Thailand", Water Resources Research, vol.9., No.5, pp. 1196~1207, 1973.
4. D. E., Agthe and R. B. Buildings, "Dynamic Models of Residential Water Demand", Water Resources Research, vol. 16, No. 3, pp. 476~480, 1980.
5. D. R. Maidment and E. Parzen, "Cascade Model of Monthly Municipal Water Use", Water Resources Research, vol. 20, No. 1, pp. 15~23, 1984.
6. 이 경훈 외 2인, "상수도 1일 급수량의 시간적 변화의 특성에 관한 연구", 한국수문학회지, 제27권 제2호 pp. 135~143, 1994.
7. 이 경훈 외 2인, "상수도의 급수량 및 배수지 용량변화에 관한 연구", 대한상하수도학회지, 제9권 제2호 pp. 118~126, 1995.
8. 이 경훈 외 2인, "기후조건에 의한 상수도 1일 급수량의 변화에 관한 연구", 한국수자원학회지. 제 28권 제 6 호, pp. 147~158, 1995.
9. 강변서, "행렬과 SPSS/PC+ 이용 다변량 통계분석", 학현사, pp. 257~303, 1993.
10. 안상형 외 1인, "현대통계학", 학현사, pp. 335~462, 1993.
11. 남원시, "남원시 통계연보", 1996.
12. 기상청, "기상월보", "기상연보", 남원지방기상청, 1982~1994.