

시계열 회귀모형에 기초한 욕실 내 용수 사용량 추정

명성민*, 김동건**, 조진남**

Estimating Bathroom Water-uses based on Time Series Regression

Sungmin Myoung*, Donggeon Kim**, Jinnam Jo**

요약

신뢰성 있는 물 수요예측을 실시하기 위해서는 실측자료를 이용하여 다양한 수요구조의 변화를 합리적으로 반영할 수 있는 수요예측모형을 개발·활용하는 것이 필요하다. 본 연구에서는 가정에서 사용하고 있는 욕실 내 용수사용량 특성을 파악하기 위하여 전국 140여개 가구를 대상으로 전자식 유량계와 무선송신시스템이 결합된 원격측정시스템을 이용하여 실측자료를 취득하고, 이를 이용하여 각 사용량의 기준이 되는 원단위를 도출하였다. 향후 사용량 예측을 위하여 욕실 내 용수를 욕조용수와 세면용수로 구분하여 시계열 모형을 적용함으로써 물 수요관리 및 정책 수립을 위한 정보로서 활용할 수 있도록 하였다.

▶ Keywords : 용수 수요, 시계열, 자기회귀오차모형, 회귀모형, 원격측정시스템

Abstract

Analysis of influential factors on water consumption in households will help predicting the water demand of end-use in household and give an explanation to cause on the change of trend. In this research, the data are gathered by radio telemetry system which is combined electronic flow-meter and wireless communication system in 140 household in Korea. Using this data, we estimate for each residential type to determine liter per capita day. we used real data to predict bathtub and washbowl water-uses and compared the ordinary least square regression model and autoregressive regression error model. The results of this study can be applied in the planning stages of water and waste water facilities.

▶ Keywords : Water conservation, time series, autoregressive error model, regression model, radio telemetry system

•제1저자: 명성민 •교신저자: 조진남

•투고일 : 2014. 7. 4. 심사일 : 2014. 7. 19. 게재확정일: 2014. 8. 20.

* 중원대학교 의료정보행정학과(Dept. of Medical Information and Administration, Jungwon University)

** 동덕여자대학교 정보통계학과(Dept. of Statistics & Information, Dongduk Women's University)

I. 서론

생활용수는 사용용도에 따라 크게 가정용수와 비가정용수로 구분되는데, 비 가정용수는 다시 영업용, 업무용, 욕탕용, 전용공업용, 기타 등으로 구분한다. 생활용수의 이용량은 환경부에서 매년 '상수도 통계'를 공표함으로써 제시하는데, 2013년에 발표한 상수도 통계에 따르면 2011년 유수수량 기준 (분수량 제외)으로 가정용이 3,160백만 m^3 (62.9%)으로 가장 많고, 영업용 1,259백만 m^3 (25.1%), 업무용 307백만 m^3 (6.1%), 공업용 146백만 m^3 (2.9%), 욕탕용 88백만 m^3 (1.8%), 기타 30백만 m^3 (0.6%) 순이며, 1인당 물 사용량은 279 ℓ 로 나타났다[1]. 이와 같이 생활용수 중 가정용수가 차지하는 비중이 가장 크며, 생활수준의 향상, 산업구조의 변화, 수도와 관련된 가전제품이나 설비 등의 변화, 기후변화에 대한 온난화, 가뭄 등의 자연적인 요인을 이유로 가정에서의 물 소비량은 계속 증가하는 추세에 있다.

가정용수 사용량에 영향을 미치는 요인을 파악하는 문제는 가정용수의 변동 특성을 이해하고 향후 사용경향을 예측할 수 있다는 점에서 중요한 의미를 가진다[2]. 기존에 제시된 연구에서 가정용수 사용량에 영향을 미치는 요인을 내부적 요인과 외부적 요인으로 구분하였는데, 내부적 요인은 주택구조, 주거형태, 거주자 생활양식 등이며, 수도요금, 온도, 날씨 등은 외부적 요인으로 기술하였다[3]. Cochran 등은 물 소비에 영향을 미치는 환경적, 경제적, 사회적 요인을 고려함으로써, 정책입안자들에게 도시의 장기 생활용수 수요를 예측할 수 있는 방법을 제안하였고, 용수수요함수를 도출하여 예측모형으로 이용하였다[4].

신뢰성 있는 물 수요예측을 실시하기 위해서는 실측자료를 이용하여 다양한 수요구조의 변화를 반영할 수 있는 합리적인 수요예측방법을 개발, 활용하는 것이 필요하다[5]. 미국 등 선진국에서는 일찍부터 용수의 수요관리, 수도시설의 적정규모 산정을 위하여 생활용수 중에서도 가정용수의 용도별 사용에 대한 실측조사 및 장/단기적 변화특성과 영향인자 분석 등이 수행하여 용수관리의 효율성 및 과학적 수요예측을 제시하고 있다[6,7,8].

실측자료를 이용한 물 수요량 경향성을 분석한 선행연구 사례를 살펴보면, 가정용수의 용도별 사용량 모니터링을 통한 물 수요 특성분석, 가정용수 용도별 사용량의 통계적 특성분석, 비가정용수의 업종별 사용량 원단위 및 통계적 특성분석, 가정용수의 수요량 예측을 위한 통계적 모형비교 등으로 연구되었다[2,3,9,10]. 해외 사례로는 Zhou 등이 호주 멜버

른에서 물 수요량에 영향을 미치는 요인으로 경향성, 계절적 요인, 기후 상관성, 자기상관성의 4가지를 고려하여 Box-Jenkins 시계열 분석에 기초한 예측모형을 제시하였고, 기초된 자료는 1989년 1월부터 1996년 6월까지의 전체 용수사용량 원단위를 대상으로 일일데이터에 대하여 모형을 구축하였다[8]. 또 다른 연구로서, Ghiassi 등은 San Jose Water Company에서 취득한 용수자료에 시계열 예측기법으로 동적 신경망 모형(DAN: Dynamic artificial neural network)을 적용하였고, 이를 자기회귀이동평균법과 비교하여 모형의 타당성을 확인하였다[11]. 그리고 Wong 등은 홍콩 지역의 597 가구를 대상으로 물 소비현황과 유량에 영향을 미치는 요인을 결정하기 위한 시계열 회귀모형을 고려하였다. 주요 요인은 추세요인, 계절요인, 날씨, 캘린더 효과(calendar effect), 자기회귀의 5가지 요인으로 가정하였으며, 1990년부터 2007년까지의 자료를 대상으로 적용한 결과, 물 소비량에 가장 큰 영향을 미치는 요인으로 계절요인이라 주장하였다[12].

본 연구는 욕실 내 용수인 욕조용수와 세면용수를 전자식 유량계와 무선송신시스템에 결합된 원격측정시스템을 이용하여 실측하고, 이에 대한 1인 1일 용수 사용량을 종속변수로, 월별 가변수 및 선형추세를 독립변수로 하여 이를 시계열 모형에 적용시켜, 욕실 내 용수의 평균사용량을 예측하고자 한다.

이를 위하여 가변수를 이용한 시계열 회귀모형 및 자기회귀오차모형(ARE: autoregressive error model)을 고려하여 두 모형에 대한 적합도를 비교하여 적절한 모형을 선택하고 이에 대한 결과를 제시한다. 구체적으로, 2장에서는 가변수를 이용한 시계열 회귀모형과 자기회귀오차모형에 대하여 정의하며, 3장에서는 욕실 내 용수자료에 대한 원격측정시스템을 이용한 수집기법 및 특성에 대하여 설명한다. 4장에서는 수집된 욕실 내 용수예측을 위한 시계열 모형의 구축과정과 그 결과를 비교하여, 마지막 5장에서는 결론을 통해 본 연구의 시사점에 대해 논의한다.

II. 시계열 회귀모형과 적합성

1. 가변수를 이용한 시계열 회귀모형

일반적으로 경제 및 사회분야에서 발생된 시계열 자료들은 대부분 계절변동을 포함하고 있어 이들의 추세를 파악하기에 어려움이 따른다. 또한 주어진 데이터가 월별 혹은 분기별 자

료인 경우에 계절변동을 뚜렷하게 확인할 수 있다. 계절요인을 분석모형에 반영하는 방법은 고정 계절변동과 확산계절변동의 두 가지 형태로 고려할 수 있는데, 이러한 시계열 회귀 분석 방법으로 가변수(dummy variable)를 이용한다. 시계열 y_t 에 대하여 다음의 모형을 가정한다[13].

$$y_t = T_t + S_t + E_t \tag{1}$$

T_t 는 t 시점의 추세성분을 의미하며, S_t 는 t 시점의 계절 성분, E_t 는 t 시점의 확률오차를 의미한다. 시계열 y_t 의 평균은 시간에 의존하여 $\mu_t = T_t + S_t$ 이며, 시간에 따른 시계열의 변동은 오차항 E_t 에 의하여 나타난다. 오차항에 대하여 일반 회귀분석과 동일하게 가정하고 가변수를 이용하여 계절요인을 반영한 시계열 회귀모형은 다음과 같다.

$$y_t = \alpha_0 + \alpha t + \sum_{j=1}^{m-1} \beta_j d_{tj} + \epsilon_t \tag{2}$$

여기서 d_{tj} 는 가변수로서, t 시점 관측치가 계절 j 에 속하는 경우 $d_{tj} = 1$, 그렇지 않았을 때 $d_{tj} = 0$ 로 정의된다. 예를 들어, 관측된 시계열 자료가 월별 측정자료 이고($m = 12$), 시점 t 에서 1월인 경우 모형은 다음과 같다.

$$\begin{aligned} y_t &= \alpha_0 + \alpha t + \beta_1 d_{t1}(1) + \beta_2 d_{t2}(0) + \dots + \beta_{11} d_{t11}(0) + \epsilon_t \\ &= \alpha_0 + \alpha t + \beta_1 d_{t1}(1) + \epsilon_t \end{aligned} \tag{3}$$

위의 모형에서 이용한 방법인 가변수를 이용한 시계열 회귀모형은 시계열자료가 고정 계절변동을 갖는 경우 적합한 방법이다. 만약 시계열 자료가 확산 계절변동을 갖는 경우 가변수를 이용한 시계열 회귀모형을 적용하려면 시계열 자료를 적절한 변환(제곱근변환, 로그변환)을 통하여 확산 계절변동을 고정 계절변동으로 변환한 후 모형을 적합하여야 하며, 적합한 모형에 의한 예측치 들은 원래 시계열의 값과 동일한 단위로 나타내기 위하여 역변환을 적용하여야 한다[13,14].

2. 자기회귀오차모형

시계열 자료는 시간의 흐름에 따라 계속적으로 측정된 자료이기 때문에 항상 자기상관이 존재한다고 생각할 수 있다. 자기상관이란 현재 측정된 값이 과거의 측정된 값의 영향을 받는다는 것을 의미한다.

자기회귀오차모형은 시계열 자료를 회귀모형에 적합시킬 때, 오차항 들이 자기상관관계를 가지는 경우에 적합한 모형이다. 일반적인 p 차 자기회귀오차모형은 다음과 같다.

$$\begin{aligned} y_t &= \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \epsilon_t \\ \epsilon_t &= \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_p \epsilon_{t-p} + a_t \end{aligned} \tag{4}$$

여기서 a_t 는 서로 독립이고, 평균이 0, 분산이 σ_a^2 이다.

자기회귀오차모형과 일반회귀모형의 차이점은 오차항이 서로 독립이 아니라 시차에 따른 자기상관을 가지고 있으며, p 차 자기회귀과정 AR(p)를 따른다는 것이다. 이 자기상관에 의해서 발생하는 통계 추론의 문제점들은 등분산성 가정을 만족하지 않는 경우의 문제와 유사하게 나타나는데, 회귀계수들의 최소제곱추정량이 비효율적으로 나타난다는 것이다[14].

모형추정을 위하여 사용되는 방법은 조건부 최소제곱추정법(conditional least squares estimation), 비조건부 최소제곱추정법(unconditional least square estimation), 최대우도추정법(maximum likelihood estimation), Yule-Walker 추정법 등이 있다[15].

모형의 적합성을 측정하기 위하여 결정계수와 RMSE (root mean square error)를 이용한다. 결정계수는 종속변수가 독립변수에 의해 어느 정도 설명되는지를 나타내는 지표이며, RMSE는 예측값과 실제값의 분산정도를 의미하며 값이 작을수록 좋은 모형을 의미한다. 일반적인 RMSE의 식은 다음과 같다[14].

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \tag{5}$$

III. 옥실 내 용수자료 수집 및 특성

옥실 내 용수 사용량의 조사 및 분석을 위해서 한국수자원공사에서 취득한 자료를 이용하였으며, 기간은 2002년 10월~2006년 2월까지 3년 6개월간 각 시·군에 위치한 총 140 가구들로부터 용수 사용량을 조사하였다. 측정방법은 그림 1과 같이 각 가정 내 수도꼭지에 데이터 저장과 무선전송이 가능하도록 전자식 유량계에 로거를 부착한 유량모니터링 시스템을 제작, 설치하였다.

이를 통하여 표본가구별로 실시간 자료취득이 가능하도록 시스템화 하였는데, 시스템 구축 방법은 인터넷 데이터센터(internet data center)를 통하여 각 유량계에서 CDMA 방

식의 무선신호를 송출하고 중계기를 거쳐 자료를 수집, 관리

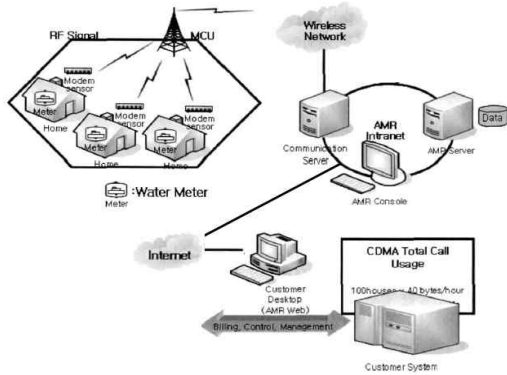


그림 1. 원격측정시스템을 이용한 가정용수 취득 과정
Figure 1. Data collection process of end-use water in household using radio telemetry system

하는 방식이다[2].

이와 같이 수집한 유량데이터를 이상치 (outlier)의 판별 및 제거 등의 자료처리과정 (data processing)을 거쳐서 각 용도별 용수를 도출하였다. 본 연구에서는 용도별 용수들 중 욕실 내 용수만 고려하고, 이를 다시 욕조용수와 세면용수로 세분화하였다.

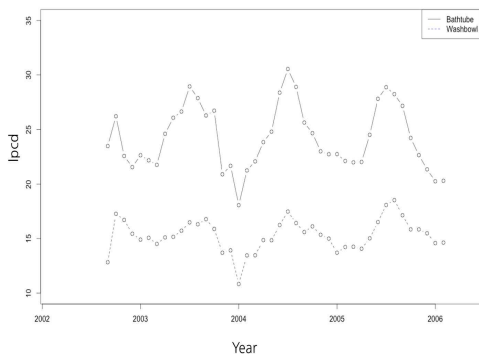


그림 2. 욕실 내 용수에 대한 시계열 도표
Figure 2. Time series plot of bathtub and washbowl water-uses

가구별 욕조용수와 세면용수에 대하여 가구별 구성원으로 나눈 값인 1인 1일 용수 사용량 (lpcd: liter per capita day)을 산출하였는데, 이는 일반적인 평균값의 개념과 일치하여 용수 사용량의 객관적인 지표로서 이용하며[2], 본 연구에서는 시계열 분석을 위하여 월별자료로 전환하였다.

욕실 내 용수에 대한 시도표 (time series plot)를 확인해

보면, 욕조용수의 경우 선형추세는 없지만, 월별 사용량이 겨울철에 비해 여름철에 많은 뚜렷한 패턴을 가짐을 확인하였다. 세면용수의 경우에도 욕조용수와 같이 시간에 따른 추세는 존재하지 않는 것으로 판단된다 (그림 2).

IV. 욕실 내 용수에 대한 시계열 분석

1. 모형의 설정 및 적합도 검증

본 절에서는 욕실 내 용수인 욕조용수, 세면용수의 원단위 사용량에 대한 분석결과를 살펴본다. 먼저 각 용수별 원단위 사용량에 대하여 월별 가변수와 선형추세만 반영된 일반회귀 모형과 추세효과와 자기상관, 월별 가변수를 고려한 자기회귀 오차모형을 비교하고 최종적인 모형에 대한 시계열 산포도를 제시한다. 자료에 대한 분석은 SAS V9.2를 이용하였으며, 시계열 예측 도표는 R-package 3.0.2을 이용하였다.

본 자료에서 자기회귀오차모형의 설정은 다음과 같다. 먼저 시계열 변동성분들을 고려해야 하는데, 첫 번째로 추세성분의 경우 용수사용량을 시간의 흐름에 따라 증가 혹은 감소하는 경향이 있을 수 있으므로 시간에 대하여 선형추세를 적용하였다. 선형추세를 적용한 이유는 확보된 자료가 3년 6개월로 단기간이기 때문에, 2차 혹은 3차를 적용하기에는 어렵다고 판단하였기 때문이다. 두 번째로 계절성분의 경우, 용수 사용의 특성상 겨울철 (12월~1월)보다 여름철 (6월~8월) 사용량이 많으리라는 월별 효과가 예상되기 때문에, 가변수를 이용하여 월별효과를 모형화 하였으며, 참조변수는 12월을 기준으로 가변수를 정의하였다.

위와 같이 욕실 내 용수에 영향을 미치는 여러 요인들을 적절히 반영하여 본 연구에서 가정하는 자기회귀오차모형은 다음과 같다.

$$y_t = \beta_0 + \alpha t + \sum_{j=1}^{11} \beta_j d_{tj} + \epsilon_t, \epsilon_t = \sum_{k=1}^{12} \phi_k \epsilon_{t-k} + a_t \quad (6)$$

여기서, 반응변수 y_t 는 욕조/세면용수에 대한 1인 1일 용수 사용량을 나타낸다. α 는 시점 t 에서 반응변수에 영향을 주는 추세요인을 나타낸다. d_{tj} 는 월별 가변수로서, t 시점 관측치가 j 월에 해당되는 경우 $d_{tj} = 1$, 그렇지 않을 때 $d_{tj} = 0$ 로 정의된다. 그러므로 β_j 는 j 월과 12월의 조정된 평균차이를 나타내는 모수가 된다. 마지막으로 ϵ_t 는 회귀모형의 오차항을 나타내는데 시계열 자료가 가지고 있는 자기상관성을 반영하

기 위하여 오차항에 대한 자기회귀모형을 고려하였다. 최종 잔차인 a_t 는 백색잡음항을 나타내며, 평균이 0, 분산이 σ^2 인 정규분포를 따르는 확률변수로 정의된다. 일반 회귀모형은 위의 모형과 동일하지만, 오차항 ϵ_t 가 자기회귀오차모형의 최종 잔차 a_t 와 동일하게 나타난다.

욕실 내 용수의 시계열 모형을 설정하기 위하여 원단위를 종속변수로 하고 설명변수로는 일반회귀모형을 적용하는 경우 12월을 기준으로 한 월별 가변수만을, 자기회귀오차모형의 경우 월별 가변수와 자기상관오차를 고려하였다. 욕조용수와 세면용수 모두 유의한 선행추세가 존재하지 않았기 때문에 선행 추세항은 모형에서 제거하였다. 비교를 위한 통계량은 AIC, RMSE, 전체 결정계수 (total R^2)를 이용하였으며, 선택된 모형에 대한 잔차의 정규성 검정을 Shapiro-Wilks 방법을 이용하여 표 1을 통하여 확인하였다.

표 1. OLS와 ARE의 적합도 검증
Table 1. Goodness of fit criteria for OLS and ARE

water-uses	OLS			ARE			Normality Test (p-value)
	AIC	RMSE	Total R^2	AIC	RMSE	Total R^2	
bath tub	142.78	1.18	0.88	134.51	1.03	0.91	0.2850
wash bowl	137.90	1.11	0.60				0.9465

세면용수의 경우 오차에 대한 유의한 자기상관관계가 존재하지 않았으므로 월별 가변수만이 반영된 일반회귀모형을 적합하였다. 전체 결정계수 R^2 가 60%이며, 잔차가 정규분포를 따른다는 귀무가설이 채택되었다 (p-value=0.9465).

욕조용수의 경우 자기회귀오차모형의 전체 결정계수 R^2 가 91%로서 일반회귀모형에 비해 향상되었으며, AIC도 일반회귀모형 142.78에 비해 134.51로서 현저히 감소하였기 때문에 자기회귀오차모형이 더 적절하다고 판단하였다. 선택된 자기회귀오차모형의 잔차에 대한 정규성 검정 (Shapiro-Wilks test)결과 정규성 가정이 충족되었다 (p-value=0.2850).

2. 욕실 내 용수에 대한 시계열 회귀모형

욕실 내 용수 원단위에 대한 시계열 모형결과는 표 2에 제시하였으며, 각 용수에 대한 시계열 예측도표는 그림 3,4에 제시하였다.

위의 표 1에서 설명하였듯이, 욕조용수의 경우 자기회귀오차모형을 적용하였으며, 세면용수의 경우 월별 가변수만이 반영된 일반회귀모형이 적용되었다.

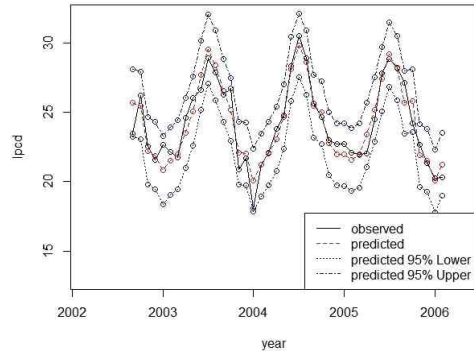


그림 3. 욕조 용수에 대한 시계열 예측도표
Figure 3. Plot of forecasting result(Bath tub)

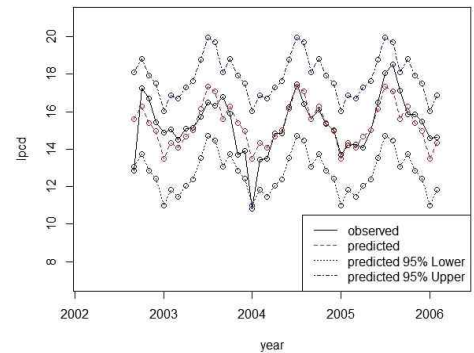


그림 4. 세면 용수에 대한 시계열 예측도표
Figure 4. Plot of forecasting result(Wash bowl)

표 2의 각 모수들의 추정치를 확인해 보면, 욕조 용수의 경우 4월의 경우 계수 추정치가 1.6430으로서, 이는 12월에 비해서 욕조용수를 1.6430 만큼 더 많이 사용했다고 해석할 수 있다. 동일한 방법으로 5월~10월 사용량이 12월에 비하여 많이 나타남을 알 수 있었으며, 특히 7월, 8월의 경우 다른 월에 비해 많이 사용하였다는 점(7월 7.6961, 8월 6.5190)을 확인할 수 있었다. 나머지 11월, 1월~4월까지의 욕조용수 사용량 추정치가 12월에 비해서 차이가 없었다는 것을 확인하였다. 이는 겨울에 비해서 초여름~초가을에는 욕조 용수사용량이 많다고 해석할 수 있으며, 봄, 늦가을, 겨울에는 욕조용수사용량의 차이는 없다는 것을 알 수 있다. 후방 제거법(backstep elimination) 적용결과 최종적으로 12차 자기상관오차 중에서 12차 자기상관만 유의한 것으로 판정되었다.

적합한 후의 잔차가 자기상관의 존재여부를 판단하기 위한 Durbin-Watson값은 1.6679로서 더 이상의 자기상관이 존재하지 않았다. 세면용수의 경우, 12월에 비하여 7월~8월

사용량이 상대적으로 많이 사용하였음을 확인하였다. 7월의 세면 용수 사용량의 회귀계수 추정치가 2.3854로서 12월에 비해 세면 용수 사용량이 2.3854정도 차이가 나며, 이는 통계적으로 유의하게 나타났다고 할 수 있다. 8월 경우도 마찬가지로 통계적으로 유의하며, 회귀계수 추정치는 2.1257로서 12월에 비하여 상대적으로 세면 용수가 많이 사용되었음을 확인할 수 있다. 이는 여름에 세면용수를 더 많이 사용한다는 것을 뒷받침하며, Durbin-Watson값은 1.141로서 자기상관이 존재하지 않는다고 판단하였다.

표 2. OLS/ARE 모형에 따른 시계열 모형 결과
Table 2. Coefficient estimates of OLS/ARE model

Variable	Bathtub			Washbowl		
	estimate	std. error	p-value	estimate	std. error	p-value
intercept	21.8874	0.3972	<.0001	14.9607	0.5535	<.0001
Jan	-1.0508	0.5617	0.0715	-1.4518	0.7828	0.0735
Feb	-0.3930	0.5617	0.4897	-0.6159	0.7828	0.4375
Mar	0.0690	0.6170	0.9117	-0.8848	0.8455	0.3037
Apr	1.6430	0.6170	0.0125	-0.2825	0.8455	0.7407
May	3.2023	0.6170	<.0001	0.0471	0.8455	0.9559
Jun	5.8142	0.6170	<.0001	1.1969	0.8455	0.1672
Jul	7.6961	0.6170	<.0001	2.3854	0.8455	0.0084
Aug	6.5190	0.6170	<.0001	2.1257	0.8455	0.0175
Sep	3.8074	0.5617	<.0001	0.6258	0.7828	0.4303
Oct	3.6073	0.5617	<.0001	1.3188	0.7828	0.1024
Nov	0.3367	0.5617	0.5536	0.4391	0.7828	0.5790
Dec
autoregressive parameter(nlag =12)	0.4056	0.1697

실제값과 예측값의 차이를 나타내는 RMSE값은 욕조용수 1.1034, 세면용수 1.0329 로 나타났다.

시계열 자료의 모든 시차간의 자기 상관이 0으로 자기 계열간의 서로 독립인 성질을 백색잡음이라고 하는데, 자기회귀오차 모형에서는 잔차들이 기본적으로 백색잡음과정이라고 가정한다. 이에 대한 판정방법은 Ljung-Box 검정을 이용하는데, 욕조용수의 경우 자기회귀오차모형을 적용하였으므로 이를 표 3에 제시하였다. 그 결과, 6차 자기상관계수까지 모두 0이라는 가설과 12차 자기상관계수가 모두 0라는 가설이 모두 채택되었다. 이는 잔차에 남아있는 자기상관관계는 존재하지 않는다고 할 수 있다.

표 3. 욕조용수의 자기회귀모형에 대한 Ljung-Box 검정결과
Table 3. Ljung-Box test for autoregressive error model in Bathtub

To lag	χ^2	DF	Pr>Chisq	Autocorrelations				
6	3.18	6	0.7853	-0.094	0.038	-0.064	-0.146	-0.090
12	7.74	12	0.8052	-0.087	-0.165	0.079	-0.069	-0.054

3. 시계열 모형의 실제값과 추정값의 산점도

욕실 내 용수들의 실제값과 회귀모형 및 자기회귀오차모형 추정치의 산점도를 그림 5에 제시하였다.

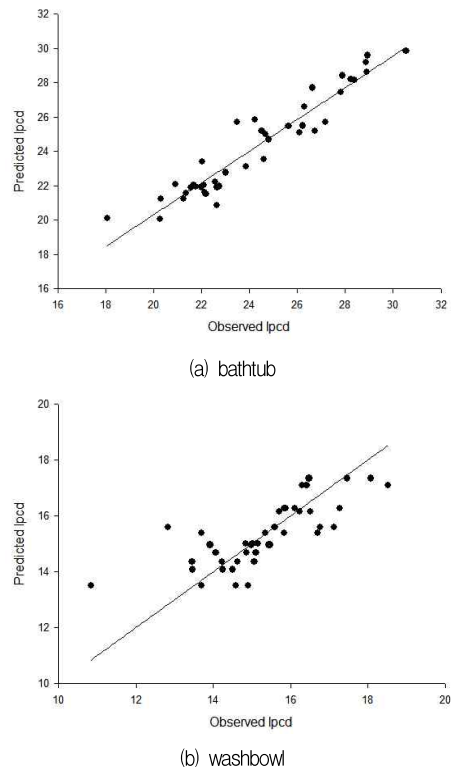


그림 5. 실제값과 추정값의 산점도
Figure 5. Scatterplot between actual value and predicted value of lpcd

그림 5에서 x 축은 욕조용수/세면용수 원단위의 실제값이고, y 축은 가변수를 이용한 시계열 회귀모형 또는 자기회귀오차모형을 이용한 추정치를 나타낸다. 일반적으로, 원점을 지나는 대각선 주변에 관측값들이 많을수록 시계열 회귀모형

에 잘 적합된다고 볼 수 있는데, 본 자료의 경우 욕조용수/세면용수 모두 비교적 적합이 잘 되었다고 판단되며, 상대적으로 세면용수가 대가전에서 벗어나는 값들이 좀 더 많이 나타났다.

V. 결론

신뢰성 있는 물 수요예측을 위해서는 실측자료를 이용하여 다양한 수요구조의 변화를 반영할 수 있어야 한다.

본 연구에서는 전국 140여 가구를 대상으로 취득한 용수 자료들 중 욕실 내 용수에 대한 수요예측을 실시하는 것을 목적으로 하였다. 이를 위하여 몇 가지 시계열 회귀모형을 소개하고 실제 욕실 내 용수를 욕조용수와 세면용수로 세분화하여 각 시계열 회귀모형들에 적용하였다. 각 용수에 대한 시도표를 통하여, 욕조용수의 경우 선형추세는 없지만, 월별 사용량이 겨울철에 비해 여름철에 많은 뚜렷한 패턴을 가짐을 확인하였으며, 세면용수의 경우 시간에 따른 추세는 존재하지 않는 것으로 판단하였다. 모형의 적절성을 확인하기 위하여 결정계수 및 AIC, RMSE를 도출하여 가변수를 이용한 시계열 회귀모형과 자기회귀오차모형에 대하여 비교, 최종모형을 선정하였다. 그 결과, 욕조용수의 경우 자기회귀오차모형이 더 적합하다고 판단되었으며, 세면용수의 경우 가변수를 이용한 시계열 회귀모형이 적합하다는 것을 확인 할 수 있었다.

선정된 모형에 대하여 시계열 모형을 적용한 결과는 다음과 같다. 첫째, 욕조, 세면 용수 모두 선형추세가 존재하지 않았다. 둘째, 욕조용수의 경우 용수사용량이 12월에 비하여 4월~10월 사용량이 많게 나타남을 확인하였으며, 세면용수의 경우 7~8월에 사용량이 상대적으로 많이 나타났다. 셋째, 적합도 척도인 전체 결정계수 R^2 는 욕조용수 91.28%, 세면용수 59.94%로서 비교적 높게 나타났으며, RMSE값은 세면용수 1.0329, 욕조용수 1.1070으로 나타났다. 각 용수별 실제값과 추정값의 산점도를 통하여 적합이 잘 되어있음을 확인하였다.

본 연구에서는 시계열 요인을 모형화 하는 방법으로 자기회귀오차 모형과 가변수를 이용한 시계열 모형을 이용한 방법을 살펴보고, 이를 수집된 욕실 내 용수모형에 적용하여 그 특성을 분석하였다는 것에 의의가 있다.

본 연구로부터 도출된 결과는 가정용수의 용도별 수요예측, 수요관리정책수립, 수도관련 기자재 및 시설의 규격결정 등에 기초자료로 활용될 수 있을 것으로 기대된다.

향후 연구방향으로서, 본 연구에서 이용된 자료는 3년 동

안의 단기자료이기 때문에 선형추세만을 고려하였으나, 앞으로 그 이상의 장기적인 자료를 확보한다면 이차, 삼차 또는 비선형 추세를 적용하여 다양한 수요 예측을 수행할 수 있을 것이라 판단된다.

참고문헌

- [1] Ministry of Environment, "Statistics of waterworks: 2011", The Ministry of Environment, Sejong, pp. 1-10, 2013.
- [2] Kim, H., Lee, D., Park, N., Jung, K., "Analysis of statistical characteristics of household water end-uses", Journal of the Korean Society of Civil Engineers, Vol.28, pp. 603-614, 2008.
- [3] Lee, D., Kim, J., Kim, H., "Statistical analysis of non-household unit water use for business categories", Journal of the Korean Society of Civil Engineers, Vol.29, pp. 385-396, 2009.
- [4] Cochran, R., Cotton, A. W., "Municipal water demand study, oklahoma city and tulsa, oklahoma", Water Resources Research, Vol. 21, pp. 941-943, 1985.
- [5] Ministry of Environment, "Residential end uses of water: a pilot study", The Ministry of Environment, Sejong, pp. 15-20, 2003.
- [6] AWWA Research Foundation, "Residential end uses of water", American Water Works Association Research Foundation, Denver, pp. 70-80, 1999.
- [7] AWWA Research Foundation, "Commerical and institutional end uses of water", American Water Works Association Research Foundation, Denver, pp. 1-10, 2000.
- [8] Zhou, S. L., McMahon, T. A., Walton, A., Lewis, J., "Forecasting daily water demand: A case study of Melbourne", Journal of Hydrology, Vol. 236, pp. 153-164, 2000.
- [9] Kim, J., Kim, H., Lee, D., Kim, G., "Analysis of water use characteristics by household demand monitoring", Journal of Korean Society of Environmental Engineering, Vol.29, pp. 864-869, 2007.

- [10] Myoung, S., Lee, D., Kim, H., Jo, J., "A comparison of statistical prediction models in household water end-uses", The Korean Journal of Applied Statistics, Vol.24, pp. 567-573, 2011.
- [11] Ghiassi, M., David K. Zimbra, H. Saidane. "Urban water demand forecasting with a dynamic artificial neural network model.", Journal of Water Resources Planning and Management, Vol.134, pp.138-146, 2008.
- [12] Wong, J. S., Zhang, Q., Chen, Y. D., "Statistical modelling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast", Water Resources Research, Vol.46, W03506, 2010.
- [13] Kim, Y., Park W., "Estimating automobile insurance premiums based on time series regression", The Korean Journal of Applied Statistics, Vol. 26, pp. 237-252, 2013.
- [14] Choi, B., "Univariate time series analysis", Sekyungsa, Seoul, 1992.
- [15] Lee, H., "Analysis of time series models for PM10 concentrations at the Suwon city in Korea", Journal of the Korean Data & Information Science Society, Vol. 23, pp. 535-542, 2010.
- [16] Kim, H., "The application of time series analysis under R environment", Journal of the Korean Data Analysis Society, Vol.13, pp. 331-341, 2011.
- [17] Jeong, H., A study on Korea domain registration forecasting, Journal of the Korean Data Analysis Society, Vol. 14, pp. 1889-1898, 2012.
- [18] Choi, B., Kang, H., Lee, K., Han, S., "A development of time-series model for city gas demand forecasting", The Korean Journal of Applied Statistics, Vol. 22, pp. 1019-1032, 2009.
- [19] Cho, Y., Lee, G., "Modelling and prediction of time series data based on markov model", Journal of the Korea Society of Computer and Information, Vol. 16, pp. 225-233, 2011.

저 자 소 개



명 성 민

2000: 연세대학교 통계학과 이학사.
 2002: 연세대학교
 의학전산통계학과 보건학 석사.
 2006: 연세대학교
 의학전산통계학과 의학통계학 박사
 현재: 중원대학교
 의료정보행정학과 조교수
 관심분야: 보건통계학, 데이터마이닝,
 통계계산
 Email: smmyoung@jwu.ac.kr



김 동 건

1986: 연세대학교 경영학과 학사.
 1990: Virginia Tech 통계학 석사.
 1995: Virginia Tech 통계학 박사.
 현재: 동덕여자대학교
 정보통계학과 교수
 관심분야: 데이터마이닝, 통계계산
 Email: dongg@dongduk.ac.kr



조 진 남

1980: 연세대학교
 응용통계학과 경제학사.
 1982: 연세대학교
 응용통계학과 경제학 석사.
 1992: Virginia Tech 통계학 박사.
 현재: 동덕여자대학교
 정보통계학과 교수
 관심분야: 실험계획법, 표본조사
 Email: jinnam@dongduk.ac.kr