

실시간 물사용량 자료의 특성 분석 및 보정

An Analysis of the Characteristics and Data Correction of Real-Time Water Use Data

윤정환* · 이준형** · 유영훈*** · 왕원준**** · 김형수*****

Jung-Hwan Yun · Junhyeong Lee · Younghoon Yoo · Won-joon Wang · Hung Soo Kim

요약: 수자원 관리 분야의 디지털 전환 기술로서 스마트워터그리드(smart water grid, SWG) 기술이 도입되고 있으며, 수도계량기를 통해 수집되는 실시간 물사용량 자료는 SWG 기술의 기초 자료로 활용되고 있다. 하지만 통신오류 등의 문제로 다수의 이상치와 결측치가 포함된 채 기록되며, 이를 실시간으로 보정하기 위한 기존의 방법론은 온전히 기록된 과거시간대의 값을 그대로 입력하는 등 정확도가 낮은 보정 방법임으로 개선이 필요한 실정이다. 따라서 본 연구에서는 과거 누적 자료의 이상치 처리와 개선된 결측 처리 방법을 제안하고, 실시간 물사용량 자료의 요일별 정보를 이용한 실시간 물사용량 보정 기법을 제안하고자 한다. 기존 결측 처리 방법과 새로 제안하는 결측 처리 방법을 각각 적용하여 보정 결과를 비교 및 평가한 결과, 새로 제안하는 결측 처리 방법의 오차지표(RMSE)는 0.002로 기존 방법론의 오차지표 (RMSE) 0.079 보다 낮게 산정되어 정확성이 개선되었음을 확인하였다. 이후 요일별-시간별 물사용량의 평균을 산정하여 실시간 결측 발생시 대체값으로 사용하는 방법을 제시하였다. 본 연구결과는 효율적인 스마트 물관리의 근거자료로 사용될 수 있으며, 개발된 방법론을 사용하여 결측 보정 오류로 인한 의사결정의 불확실성을 줄일 수 있을 것으로 판단된다.

핵심주제어: 수자원 관리, 디지털 전환, 스마트워터그리드, 물사용량 보정

Abstract: Smart Water Grid (SWG) technology is an alternative to digital transformation in water resource management. SWG technology uses real-time water use data collected through a “smart water meter”. However, these data often include several outliers and missing values due to problems such as communication errors. The existing method for correcting problems in real-time includes some errors and needs to be improved. Therefore, we propose a real-time water use data correction technique using daily information to correct for outliers and missing values. In this study, results from the existing and the proposed correction methods are compared and evaluated using the root mean square error (RMSE) index. The newly proposed method has an RMSE = 0.002, whereas the existing method has an RMSE = 0.079. These results suggest that the proposed method is more accurate than the existing method and improves the correction results. When real-time data on daily water use are missing, the average hourly water use by day is used to calculate daily water use. The results of this study indicate that the proposed method is an improvement over the existing method for addressing correction errors and can be used as the basis for more effective smart water management. This will reduce the uncertainty of decision-making due to correction errors.

Key Words: Water Resource Management, Digital Transformation, Smart Water Grid, Water Use Data Correction

* 주저자, 주식회사 에스제이엠엔씨 대표이사

** 교신저자, 인하대학교 스마트시티공학전공 박사과정

*** 공동저자, 인하대학교 스마트시티공학전공 박사과정

**** 공동저자, 인하대학교 스마트시티공학전공 박사과정

***** 공동저자, 인하대학교 사회인프라공학과 교수

I. 서론

2020년 7월 정부는 코로나19 사태로 인한 유례없는 경기침체 극복 및 구조적 전환 대응이라는 목표하에 국민보고대회를 개최하여 「한국판 뉴딜 종합계획」을 확정·발표하였다. 최근 코로나19로 인한 세계적인 봉쇄 및 비대면화의 확산으로 전례없는 경기침체에 직면하였으며, 이를 완화하기 위해 기존의 사회 구조에 대한 혁신이 요구되고 있다. 이에 따라 원격통신, 네트워크, 인공지능 등 디지털 기술을 바탕으로 기존 전통적인 운영방식과 산업 서비스구조를 원격화 및 자동화로 전환하는 것을 의미하는 디지털 전환(digital transformation)은 산업 혁신 및 국가경쟁력의 주요 요소로 떠올랐다. 이러한 측면에서 정부의 디지털 뉴딜 정책은 경제·사회구조의 디지털 전환 가속화를 위한 대규모 투자 및 추진으로, 우리나라의 디지털 역량의 중요성을 강조하는 것임을 알 수 있다.

스마트워터그리드(smart water grid, SWG)는 수자원 관리 분야의 디지털 전환을 위한 기술로서 매우 적합하며, 효율적인 수자원 관리를 위해 필요한 기술이다. SWG 기술은 지능형 물관리체계 즉, ICT기술을 융합·활용한 고효율 물관리 인프라 시스템을 말하며, 물의 효율적인 배분, 관리, 운송 등을 통해 지역적·시간적 수자원의 불균형을 해소하는 기술이다(최계운 등, 2013a). SWG 기술을 포함하는 ‘블루골드(blue gold)’ 시장은 ‘블랙골드(black gold)’로 불리는 1차 에너지원의 주요 요소인 석유에 빗대어 평가하는 용어이며, 물의 가치가 증가함에 따라 물과 관련된 산업을 통칭하는 용어로 쓰이고 있다. 우리나라의 경우 블루골드 시장의 원천 기술을 확보하고, 적극적인 상용화 및 해외 기술 진출의 기반을 조성하는 목적으로 지난 2010년 범 부처 물산업 육성정책이 시행되었다. 국토부의 ‘Water Grid 지능화 기술 기획보고서’를 시작으로 SWG 기술개발을 위해 SWG 연구단이 구성되어, 기존 물관리 산업이 갖고있는 문제점과 한계점을 개선하기 위한 정보통신기술 기반의 SWG기술이 본격적으로 개발되었다.

우리나라의 경우 비계획적으로 물을 이용하던 과거부터 물을 자원으로

서 인식하고 관리하는 물관리 선진화가 된 지금까지 인구수 증가, 특정 지역으로의 인구 및 자본 밀집, 하천변 활용 목적의 하천 직선화로 인한 하류부 하천수 집중 등을 겪으면서 용수부족 문제가 발생하였고, 미래의 수자원 부족 문제가 우려되고 있다. 이러한 수자원 부족 문제를 해결하기 위해 관리 조직과 법, 인프라시설 등이 구축되었다. 1960년대에 국가 수자원 관리 기관으로서 수자원국이 신설되고, 댐법, 수도법, 하천법 등 수자원관련 상위법이 시행되었다. 또한 ‘수자원 종합개발 10개년 계획’을 통해 최초의 수자원 계획이 수립되었고, 1990년대 까지 수자원 관리를 위한 수도시설 및 댐시설이 활발하게 확장되었다. 이러한 노력들은 ‘자원’의 개념에서 물을 관리하게 되는 기반이 되었고, 이용가능한 물을 공급할 수 있는 인프라시설로서 국토의 거의 모든 지역에 걸쳐 물 공급이 가능하게 되었다. 하지만 시설 확장 위주의 용수확보 방안은 예산 부족 및 관리 인력 부족 등의 한계가 나타났고, 단일화된 물공급 처리시스템만 고수하게 되었다. 이러한 공급자 중심의 단일형 물공급체계는 기존 수자원의 분배·운영·관리의 비효율성을 초래하고 저에너지 및 고효율 물공급 처리가 어려운 실정이다.

SWG기술은 첨단 정보통신 기술을 기존 전통적인 수자원 관리 체계에 활용하여 한계를 극복하고 효율성을 증대시키기 위한 기술로 말할 수 있다. SWG기술의 목적은 크게 수원 발굴 및 활용을 통한 다양한 수자원 확보, 물사용 유형에 따른 맞춤형 용수공급, 모니터링 및 평가를 통한 물공급 안정성 확보, ICT기술을 활용한 공급·소비 간 양방향 운영 등이 있다.

이와 같이 SWG기술은 수요자 중심의 물공급을 실현하는데 있으며, 다수의 기술들에서 수요자의 물사용량 정보를 실시간으로 입력하여 분석의 기반으로 활용하고 있다. 이때 실시간으로 물사용량 정보를 SWG기술에 사용하기 위해서는 결측없이 물사용량 정보가 온전히 수신되는 것과 수신된 정보에 오류가 포함되진 않았는지, 만일 결측 혹은 이상치 기록 등의 문제가 발생했다면 어떻게 보정할지에 대한 문제에 직면하게 된다. 실시간으로 수집되는 물사용량의 판단 및 보정 기법이 필요하며, 기법 활용을

위한 정확도 평가가 필요하다.

따라서 본 연구에서는 실시간으로 수집되는 수용가 물사용량의 이상치 판단, 최적의 결측치 보정 기법 선정을 통해 보정 방법론을 제시하고자 한다. 또한 기법의 적용 및 분석을 통해 개발된 기법의 활용성을 평가하고자 한다.

II. 선행연구

실시간 물사용량과 관련된 국내·외 연구는 크게 인프라 구축에 대한 연구와 기록된 데이터를 분석하여 사용량 변화를 파악하는 연구가 진행되었다. 인프라 구축에 대한 연구로는 개발된 스마트 수도계량기 모델을 실증 지역에 설치하고, 오류 분석과 기록된 정보의 정확도 분석을 통해 적용성을 평가하여 계량기 모델의 보완사항을 도출하는 연구가 진행되었다. 또한 기록된 데이터를 토대로 물사용량의 주기성 분석, 세대 구성원에 따른 물사용량 비교 등 물사용량 패턴을 분석하는 연구가 진행되었다.

Doolan(2011)은 스마트 수도계량기를 설치한 가정에게 물 소비에 대한 분석 결과를 실시간으로 제공할 수 있는 디스플레이를 옥내 설치하여 물 사용자에게 물사용량 정보를 주었을 때 물 소비량이 줄어드는 결과를 토대로 실시간 물사용량 자료의 물 절약 효과를 분석하였다.

주진철 등(2012)은 자기저항 방식의 스마트 수도계량기를 Test bed에 설치하고 통신 두절, 통신 음영지역 파악, 유·무선 공유기 오작동 등의 수신 오류를 파악하고 수신율 분석을 통해 현장 적용성 평가를 수행하였다. 또한 수집된 자료 중 수신율이 양호한 계량기의 수신 자료를 이용해 용수 사용량 지표를 산정하여 구성원에 따른 사용량 변화 분석을 수행하여 수도 공동 활용에 따른 물사용량 절감 효과를 분석하였다.

Britton et al.(2013)은 누수율이 높은 가정을 대상으로 스마트 수도계량기를 설치하고, 잠재적인 물 손실량을 포함한 누수정보를 사용자에게 알

려주는 기술을 개발하고, 이 기술로 인해 물사용량이 감소하였음을 밝혔다.

Sonderlund et al.(2014), Liu et al.(2016)은 스마트 수도계량기 수용가에게 물사용 패턴을 분석하여 피드백 해주고, 피드백을 받기 이전과 이후의 물사용 패턴 변화를 분석하여 물사용량이 감소하였음을 제시하였다.

홍공현 등(2017)은 스마트 수도계량기와 연동된 물사용량 분석 및 물 정보 제공 툴을 개발하여 Test bed에 설치하고, 수집된 데이터와 툴에서 도출된 결과를 토대로 개발된 툴의 기능평가를 수행하였다.

유재현 등(2020)은 실시간 물 사용량 정보를 분석하여 물 사용의 패턴에 벗어나는 기준에 해당하면 위급상황으로 정의하고, 위급상황을 모니터링하고 관리할 수 있는 시스템을 개발하였다. 이 시스템을 이용하여 독거노인의 실시간 물사용량 변화로부터 위급상황을 감지하고, 노인복지에 기여할 수 있는 방안을 제시하였다.

Roberta et al.(2020)은 도시지역의 효율적인 상수공급 시스템을 위해 SWG의 도입이 확대되고 있으며, SWG 기술의 기본 데이터인 물사용량 시계열의 신뢰성 문제를 해결하기 위한 방안으로 이상치 감지 및 처리 방법을 제안하였다. 수집된 물사용량 시계열의 전체 데이터 대비 정상데이터와 이상데이터의 비율을 이용하여 이상치를 감지하고 분석 목적에 따른 처리방안을 제시하여 신뢰성 있는 처리방안을 제안하였다.

Koo et al.(2021)은 SWG 기술이 적용된 지역의 실시간 물사용량 자료를 이용한 단기 물수요량 예측을 위해 물사용처별, 모형별 예측 및 평가를 수행하였다. 이때 신뢰성있는 예측을 위해 물사용량 수집지점 중 결측률이 적은 지점을 선별하였고, 물사용처는 가정, 식당, 종교시설, 세탁시설, 점포, 복지회관 등으로 구성하였다. 예측모형은 통계학적 예측모형과 머신러닝 기반 모형을 사용하였으며, 물 소비량 시계열만을 이용한 물사용량 예측의 한계를 밝혔다. 또한 물사용량은 소비자의 생활 특성을 반영하기 때문에, 보다 정확한 예측을 위해서는 날씨, 계절 등의 영향을 고려하는 방안을 제시하였다.

선행 연구를 살펴본 결과 주로 스마트 수도계량기로부터 얻은 실시간 물사용량 정보를 이용해 물사용 패턴을 분석하고, 패턴변화 감지, 누수 감지 등 이로부터 얻을 수 있는 효과를 분석하는 연구가 진행되었다. 또한 원격통신을 이용한 물사용량 자료 수집시 발생하는 결측 및 이상치의 문제점이 제시되었고, 이를 완화하기 위해 전체 자료 중 이상치 및 결측치의 비율 분석, 적절한 값으로의 반환 등이 수행되었다. 이러한 분석은 물사용량 자료를 기초자료로서 사용하는 SWG기술의 정확도와 신뢰성을 확보하기 위한 연구이며, 실시간으로 얻어지는 물사용량 정보에 대한 정확성 확보의 중요성을 강조한다. 하지만 실시간 물사용량 정보의 이상치 및 결측치를 파악하고 이를 적절한 값으로 대체하는 등 보정방법에 대한 연구는 미비한 실정이다. 따라서 본 연구에서 실시간 물사용량 자료의 특성을 분석하고 적합한 보정 방법론을 제시 및 적용성을 평가하였다.

III. 연구방법

본 연구에서는 다수의 이상치와 결측이 포함된 실시간 물사용량 자료를 대상으로 이상치 및 결측치 처리, 요일별 특성분석을 통해 실시간 물사용량 자료의 보정 기법을 제시하고자 한다. 이를 위해 기존 SWG기술에서 사용하고 있는 결측치 처리방법과 본 연구에서 제시한 방법을 각각 적용하고, 결측치 대체값을 실제값과의 오차지표 비교를 통해 본 연구에서 제안하는 방법의 결측 처리 결과가 더 타당함을 밝히고, 최종적으로 최적의 실시간 물사용량 보정 기법을 제시하고자 한다.

1. 분석 자료

본 연구의 분석대상 자료는 인천광역시 영종도 112블록의 SWG 사업단 상수도전의 스마트 수도계량기로 수집된 '수용가 시간별 물사용량'이며, 해당 블록은 신공항도시로 계획된 도심지의 대규모 물수용가 지역이다.

자료의 시간적 범위는 수용가를 대상으로 약 2016년 5월 2019년 12월 까지 실시간으로 수집한 시단위 물사용량(m^3/h) 자료이다. 여기서 수용가(customer)는 ‘자신이 사용할 목적으로 물을 구입하는 고객을 의미’하며, 지능형 원격 검침 인프라(advanced metering infrastructure, AMI)를 이용하여 측정된 상수도를 통해 수용가에게 시간별로 공급되는 물의 양을 ‘실시간 물사용량’으로 정의하였다(최계운, 2013b). 한시간 마다 기록되는 실시간 물사용량은 1시간 이전에 기록된 값에 기록이후 1시간 동안 공급된 물의 양을 더한 값이 기록되는 누가 자료(accumulative data)이다. 즉, 수용가에서 기록시간 사이에 물사용량이 없다면 이전값과 동일한 값이 기록되고, 물사용량이 발생하면 사용량 만큼 증가된 값이 기록되는 일종의 단조증가 시계열이다.

전체 자료 중 결측비율이 작고, 비교적 자료 수집기간이 긴 10개 지점의 누가 물사용량 자료를 연구대상 자료로 선정하였다. 이후 4장(적용 및 결과)에서는 연구대상 자료 중 대표적인 1개 지점의 자료를 기준으로 연구과정의 적용 및 결과를 서술하였으며, 나머지 9개 지점의 경우 동일한 과정으로 적용하였으나 결과만 요약하여 제시하였다. 선정된 10개 지점 자료의 수집기간, 기본 통계량을 <표 1>에 나타내었다. 자료의 수집기간의 경우 자료 수집이 시작된 시기는 각각 다르며, 자료 수집이 종료된 시기는 2019년 12월 31일로 동일하게 수집되었다. 그 중 이상치 및 결측치 없이 연속으로 자료가 수집된 기간이 가장 길었던 1번 자료를 기준으로 적용 결과를 제시하였고, 자료의 시간적 범위는 2016년 7월 1일 0시부터 2019년 12월 31일 23시까지이다.

〈표 1〉 분석 대상 실시간 물사용량 자료의 기본통계량

[단위: m^3/h]

구분	수집기간	Min	Q_1	Mean	Q_3	Max	결측비율
1	2016.07.01~	0	44.49	109.59	178.09	12427.43	8.45%
2	2016.05.25~	0	348.69	27×10^6	1179.94	84×10^{10}	9.24%
3	2016.05.24~	-3122.72	270.54	565.32	907.30	1046.05	10.82%
4	2016.05.25~	-52×10^6	21657.78	48950.36	83522.45	109036	10.38%
5	2016.05.18~	0	42024.67	69×10^8	146710.4	22×10^{13}	14.39%
6	2016.05.25~	0	278.58	601.97	963.92	1304.28	8.49%
7	2016.07.15~	0	783.75	1634.90	2485.24	54224.03	10.94%
8	2016.05.25~	-21×10^{10}	422.40	-66×10^5	3149.80	12570.4	19.38%
9	2016.05.25~	-0.05	566.23	1208.87	1714.62	32×10^5	7.52%
10	2016.06.15~	-13×10^6	3342.20	6411.19	10564.44	15453.97	8.11%

2. 이상치 판단 방법론

이상치란 과거 관측값들과는 전혀 다른 매커니즘에 의해 기록되어 그 값이 너무 동떨어져 있는 관측값을 의미하며, 통계학적으로 이상치는 관측치들이 주로 모여있는 곳에서 멀리 떨어져 있는 관측치로 정의된다. 이상치 판단 기법은 실시간 시계열 기록 분야에서 다양한 형태로 요구 및 연구되어져 왔으며, 본 연구의 분석에 이용한 자료는 수용가 단위의 물사용량 단변량 자료이기 때문에, 단변량 분석 기반의 방법론을 사용하였다.

사분위수는 관측치를 정렬하여 나열하였을 때 4등분위 점에 위치한 값이 갖는 정량적인 수를 의미하며, 이러한 사분위수를 이용하여 이상치 범위를 판단하기 위해 사분위수범위(IQR) 개념을 사용한다. 사분위수범위는 제 1사분위수와 제 3사분위수의 차이로 정의되며, 일반적으로 사분위수범위의 1.5배 혹은 3배를 초과하는 관측치는 이상치로 정의한다.

3. 결측치 처리 방법론

실시간 수집 자료의 결측이 발생하는 체계(mechanism)는 계측기의 미설치 구간(혹은 미설치 시기)이나 신호 송출 불량, 신호 수신 불량 등 복잡·다양한 경로로 발생한다. 이러한 결측을 처리하는 방법으로는 온전히

관측된 자료를 이용해 결측된 자료를 대체할 수 있는 값을 새로 부여하는 방법이 주로 사용되어 왔다. 대체방법은 대체값을 적용하는 방법에 따라 이월대체, 집단내 평균대체, 집단내 확률대체 등으로 구분된다(건강보험심사평가원, 2019). 이월대체는 결측발생 직전 혹은 특정기간 전 시점의 온전히 조사된 자료를 이용하여 대체값으로 사용하는 방법으로, 시간이 경과함에 따라 변수들의 변동은 발생하더라도 전체적인 변수들의 특징이 변함없거나 시간경과에 따른 변수들 간의 상관성이 높은 경우에 효과적으로 적용할 수 있는 방법이다. 집단내 평균대체는 전체 관측자료를 몇 개의 대체층(imputation class)으로 나누어 구분하고, 각 층마다 평균을 구하여 그 층의 모든 결측에 대한 대체값으로 사용하는 방법이다. 집단내 확률대체는 전체 관측자료를 몇 개의 대체층으로 나누어 구분하고, 각 층마다 관측값 중 무작위로 하나를 선정하여 그 층의 모든 결측에 대한 대체값으로 이 무작위값을 사용하는 것으로, 이 방법은 대체이후 표본의 분포가 어느정도 유지되는 특징이 있다(Little and Rubin, 2002).

지능형 원격 검침인프라(AMI)를 통해 수집되는 실시간 물사용량 정보는 통신환경상의 오류 등의 이유로 인해 수집 주기에 맞춰 수집되지 않는 경우가 발생한다. 이 경우에 물사용량 정보의 결측이 발생하며, 물사용량 정보 기반의 기술 분석을 위해서는 결측이 존재할 경우 분석 결과의 신뢰성이 떨어지는 문제가 있기 때문에 결측치 처리가 필요하다. 물사용량 정보의 결측을 처리하기 위한 기존 방법은 결측 데이터 발생시각과 동일한 시간의 전일 데이터를 입력하여 대체하는 방법이다. 이 방법을 누적치 자료로 수집되는 실시간 물사용량 자료에 대해 적용할 경우, 결측 데이터 발생 시각으로부터 전일 동일한 시각사이에 사용된 물사용량이 합계되지 않은 값으로 대체하는 것이기 때문에 정보가 손실된 값으로 대체될 우려가 있다. 따라서 이러한 기존 결측 처리방법을 대체할 새로운 방법을 제시하고자 하며, 제시하는 방법과 기존 방법의 비교를 통해 개선된 방법을 제시하고자 한다.

본 연구에서는 실시간으로 수집되는 물사용량 결측이 확인될 경우, 대

제하여 입력할 값을 결정하기 위해 물사용량 자료의 특성을 반영한 방법을 고안하였다. 선행연구 및 물사용량 자료 특성을 고려하여 이월대체법을 응용하여 아래와 같은 결측치 처리방법을 고안하였고 방법별로 오차평가를 통해 최적의 결측치 처리 방법을 선정한다. 또한 기존 결측치 처리방법의 결과와 비교하여 본 연구에서 제시하는 방법이 개선된 방법임을 평가한다.

1. 일별 산술평균 방법: 결측치가 발생한 시점의 직전 4개의 일별 동시간대 값의 산술평균값 입력
2. 요일별 산술평균 방법: 결측치가 발생한 시점의 직전 4개의 요일별 동시간대 값의 산술평균값 입력
3. 요일별 가중평균 방법: 결측치가 발생한 시점의 직전 4개의 요일별 동시간대 값의 가중평균값 입력(가까운 시간대 값에 더 큰 가중치 부여)
4. 최근린 산술평균 방법: 결측치가 발생한 시점 전방의 최근린값과 후방의 최근린값의 산술평균값 입력

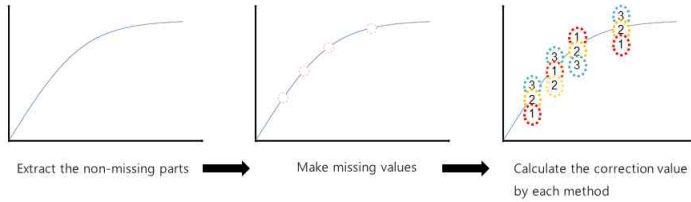
4. 결측치 처리 방법별 결과 평가방법

본 연구에서 사용한 자료는 원시오류 등의 원인으로 참값(true value)을 알 수 없는 결측치가 포함된 자료이기 때문에, 결측치 처리 방법에 의해 대체된 값을 평가할 때 자료 전체를 대상으로 하는것은 어려움이 있다. 따라서 본 연구에서는 전체 자료 중 결측없이 연속으로 정보가 장기간 수집된 구간을 부분 추출하여 결측치 처리 방법론을 평가하였다.

우선 결측없이 연속적으로 정보가 수집된 구간 중 가장 긴 구간을 추출하여 부분 시계열로 설정하였다. 부분 시계열의 값 중 무작위에 근거한 임의의 결측치를 부여하였으며, 이때 원 자료가 가지고 있는 결측 비율과 유사한 비율의 결측치를 부여하였다. 결측치가 부여된 부분 시계열 자료를 이용하여 본 연구에서 제안한 결측치 처리방법들을 적용하여 대체값을 산정하였다. 결측치 부여 전의 부분 시계열 값을 실제값으로 고려하고, 각

결측치 처리 방법에 의해 대체된 값을 실제값과 오차 평가하여 오차가 작게 발생한 방법론을 선정하였다(〈그림 1〉).

〈그림 1〉 결측치 처리 방법별 보정결과 평가방법 모식도



각 방법별 결측치 처리결과의 정확성을 상호평가하기 위해 오차평가 지표로서 RMSE(root mean square error)와 MARE(mean absolute relative error)를 사용하였다. 평균제곱오차제곱근(RMSE)은 대체값과 실제값의 오차가 크게 발생하는 순간에 더 가중치를 주어 오차정도를 평가하는 지표이며, 평균절대상대오차(MARE)는 대체값과 실제값의 차이를 실제값으로 나눠 줌으로써 보정값에 대한 오차 정도를 평가하는 지표이다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$$

5. 실시간 물사용량 보정 기법 제안

물사용량 정보의 실시간 보정을 위해서는 오측 및 결측 발생이 감지되면 해당 값을 보정값으로 즉시 대체가 가능해야한다. 이를 위해서는 결측이 발생하면 즉각적으로 대체될 값을 사전에 결정해둘 필요가 있다. 본 연구에서는 결측시 즉시 대체가능한 보정값을 결정하기 위해 앞서 제안된 이상치 및 결측치 처리과정을 통해 보정처리된 물사용량 자료를 이용해

특정 주기별 시간별 정보를 기반으로한 결측 대체용 보정값을 산정하고 이를 이용한 실시간 물사용량 보정 기법을 제안한다.

본 연구의 자료는 시점별 유동인구 차이가 큰 대규모 도심지에서 수집된 자료이므로, 시점에 따라 물사용 인구수의 차이가 크게 발생할 수 있고, 이는 실제 물사용량에 영향을 줄 수 있다. 또한 물사용량 자료의 경우 값 변화의 주체가 사람이며, 실시간으로 시간당 수집된 자료이기 때문에 생활양상의 변화에 영향을 크게 받을 것이다. 따라서 물사용량 주체의 생활양상을 대표할 수 있는 계절, 월, 요일의 주기를 고려해볼 수 있으며, 대상지역의 물사용 인구수 변화에 가장 직접적인 연관이 있는 시간적 주기로 요일을 선정하였다.

이상치 제거 및 결측치 처리된 누가자료 형태의 물사용량 자료를 시간당 물사용량으로 변환한다. 이후 전체 자료 중 동일한 요일 및 시간에 해당하는 시간당 물사용량 자료를 평균하여, 모든 요일, 모든 시간별 평균값을 미리 산정하여 실시간 결측 대체용 보정값으로 사용한다. 예를 들어, 시간당 물사용량 자료를 이용해 산정한 요일별-시간별 평균값을 저장해 두고, 실시간 자료 수집 중 월요일 3시에 결측이 발생할 경우, 미리 산정해둔 평균값 중 ‘월요일’, ‘3시’의 값을 불러와서 즉시 보정하는 방법이다.

IV. 적용 및 결과

1. 이상치 판단 및 처리

분석 자료를 대상으로 방법론에서 제시한 이상치 판단 방법을 모두 적용하였다. 적용한 결과 이상치로 분류된 지점이 7지점으로 산정되었고, 이는 누가자료 형태에서 발생할 수 없는 극단적인 이상치인 것으로 판단하였다. 10개 지점 전체 자료에 대하여 같은 방식으로 이상치 처리한 결과를 <표 2>에 나타내었다.

〈표 2〉 분석 대상 실시간 물사용량 자료의 이상치 처리 후 기본통계량

[단위: m^3/h]

구분	수집기간	Min	Q_1	Mean	Q_3	Max	결측비율
1	2016.07.01~	0	44.39	109.08	178.01	226.98	8.49%
2	2016.05.25~	0	348.68	776.45	1179.87	1610.70	9.26%
3	2016.05.24~	0	270.43	592.76	909.49	1046.05	10.90%
4	2016.05.25~	0	21638.92	53794.98	83490.19	109036	11.91%
5	2016.05.18~	0	42011.73	94081.22	146710.4	193215.58	14.62%
6	2016.05.25~	0	278.58	601.96	963.92	1304.28	8.54%
7	2016.07.15~	0	783.02	1613.46	2484.66	3221.83	10.99%
8	2016.05.25~	0	422.37	1970.22	3149.63	4904.15	19.43%
9	2016.05.25~	0	566.18	1107.02	1714.54	2101.00	7.55%
10	2016.06.15~	0	3342.20	6825.92	10564.44	15453.97	8.12%

2. 최적 결측치 처리기법 선정 및 처리

본 연구에서 제시한 결측치 보정 방법은 과거 정보를 이용해 결측보정을 수행하는 4가지 방법이며, 본 연구에서 제시한 결측보정 방법은 최대 1개월치의 온전한 과거정보를 이용하는 방법이기 때문에, 원시 오류에 의한 이상치로 원래 값을 알 수 없는 기간이 섞인 원시 시계열 자료를 그대로 이용하기에는 무리가 있다. 따라서 원시 시계열 중 오류 및 이상치 없이 연속적으로 누가 물사용량이 기록되었던 2018년 5월부터 2019년 1월까지의 정보를 부분 추출하여, 결측치 처리 방법을 평가하기 위한 평가 자료로 사용하였다. 평가 방법은 2장에서 설명한 것과 같이 부분 추출한 평가 자료에 원시 시계열이 갖고있는 결측치 비율 만큼 임의의 결측치를 부여하고, 본 연구에서 제시하는 4가지 결측 처리 방법과 기존 결측치 처리 방법을 모두 적용하여 결측 처리를 수행한다. 이후 각 방법에 의해 보정된 값과 결측치 부여 전의 온전한 값의 오차지표를 산정하여 각 방법의 성능을 비교 및 평가하였다. 각 결측치 처리 방법별 오차지표 산정 결과는 〈표 3〉에 나타내었으며, 기존 결측치 처리 방법을 적용한 결과와 비교하였을때 본 연구에서 제시한 결측치 처리 방법의 오차지표가 더 낮게 산정되었기 때문에 기존 보다 개선되었음을 확인하였다.

〈표 3〉 결측치 처리 방법별 RMSE, MARE 비교

오차지표	일별 산술평균 방법	요일별 산술평균 방법	요일별 가중평균 방법	최근린 산술평균 방법	기존 결측치 처리 방법
RMSE(m^3/h)	0.004339	0.004684	0.004602	0.002022	0.079284
MARE(%)	0.000435	0.000400	0.000399	0.000172	0.009925

결측치 처리방법 적용 결과 최근린 산술평균 방법에서 오차지표가 가장 낮게 산정되었으며, 기존 결측치 보정 방법의 RMSE 보다 약 40배 낮은 RMSE가 산정되었다. 같은 방식으로 10개 지점 전체 자료에 대하여 결측치 처리한 결과를 〈표 4〉에 나타내었다.

〈표 4〉 분석 대상 실시간 물사용량 자료의 결측치 처리 방법별 RMSE, MARE 비교

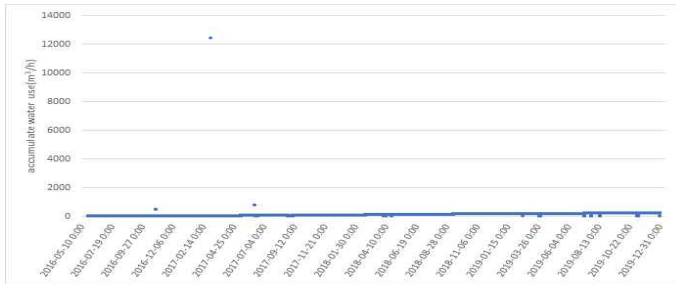
구분	오차 지표	일별 산술평균 방법	요일별 산술평균 방법	요일별 가중평균 방법	최근린 산술평균 방법	기존 결측치 처리 방법
1	RMSE(m^3/h)	0.004339	0.004684	0.004602	0.002022	0.079284
	MARE(%)	0.000435	0.000400	0.000399	0.000172	0.009925
2	RMSE(m^3/h)	0.053793	0.055410	0.055485	0.025384	0.575263
	MARE(%)	0.000612	0.000637	0.000653	0.000333	0.016409
3	RMSE(m^3/h)	0.011019	0.010601	0.010533	0.007351	0.429256
	MARE(%)	0.000250	0.000243	0.000238	0.000153	0.012149
4	RMSE(m^3/h)	3.070447	3.059232	3.138497	2.036917	28.619391
	MARE(%)	0.000804	0.000769	0.000775	0.000432	0.011449
5	RMSE(m^3/h)	8.441212	8.581389	8.853424	3.660181	61.883887
	MARE(%)	0.001021	0.001018	0.001045	0.000283	0.013406
6	RMSE(m^3/h)	0.020869	0.021126	0.022575	0.015684	0.435232
	MARE(%)	0.000823	0.000797	0.000832	0.000501	0.024886
7	RMSE(m^3/h)	0.066754	0.059734	0.061401	0.026885	1.066978
	MARE(%)	0.000410	0.000309	0.000308	0.000134	0.010488
8	RMSE(m^3/h)	0.071962	0.055030	0.053217	0.041259	1.666297
	MARE(%)	0.000383	0.000269	0.000259	0.000164	0.013869
9	RMSE(m^3/h)	0.035144	0.033329	0.033754	0.017064	0.692216
	MARE(%)	0.000446	0.000441	0.000448	0.000233	0.013338
10	RMSE(m^3/h)	0.100117	0.116942	0.114667	0.090195	5.610276
	MARE(%)	0.000235	0.000282	0.000277	0.000178	0.018036

분석 자료를 대상으로 결측치 처리 방법을 적용한 결과 10개 지점 모두에서 최근린 산술평균 방법에 의한 결측치 처리시 가장 낮은 오차지표가 산정된 것을 파악하였다. 또한 기존 결측치 보정 방법에 비해 큰 차이로 오차지표가 작게 산정되었으므로 본 연구의 최적 결측치 처리 방법은 최근린 산술평균 방법으로 선정하였다.

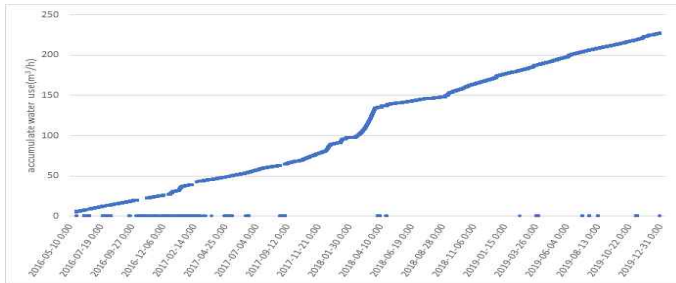
3. 실시간 물사용량 보정 기법 적용을 위한 시간당 물사용량 시계열 산정

이상치 처리와 결측치 처리까지 수행한 최종 누가 실시간 물사용량 자료의 그래프를 도시하면 <그림 2>와 같다.

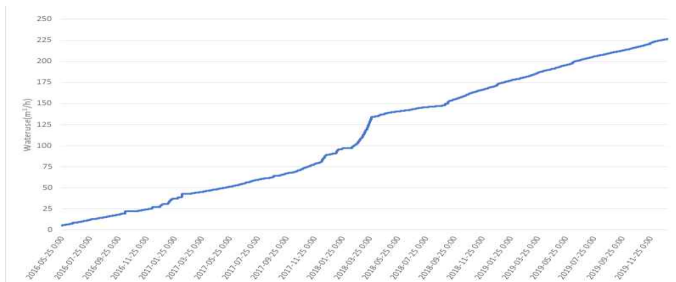
<그림 2> 분석 자료의 이상치 및 결측치 처리 결과 시계열



(a) 분석 자료의 원시 시계열



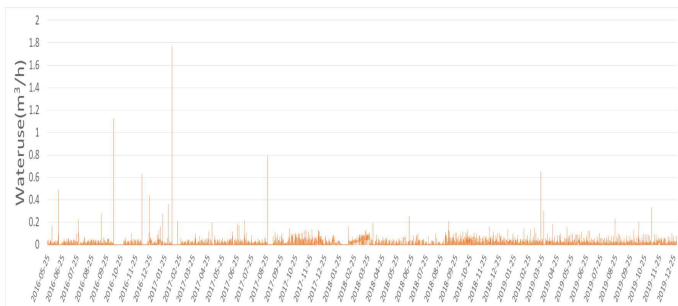
(b) 분석 자료의 이상치 처리 시계열



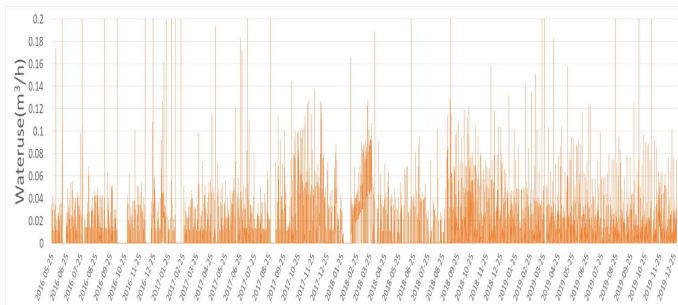
(c) 분석 자료의 결측치 처리 시계열

이를 이용하여 실시간 물사용량 보정 기법 적용을 위해 시간당 실시간 물사용량 자료로 변환하였다. 시간당 물사용량 자료는 시간에 따라 연속적으로 기록된 누가 물사용량 자료의 시간별 차이값을 산정함으로써 시간당 자료로 변환하였으며, 변환된 시계열을 연속된 시간에 따른 시간당 물사용량 그래프로 도시하면 <그림 3>과 같다.

〈그림 3〉 시간당 실시간 물사용량 시계열



(a) 세로축 범위 : $0 \sim 2 \text{ m}^3/\text{h}$



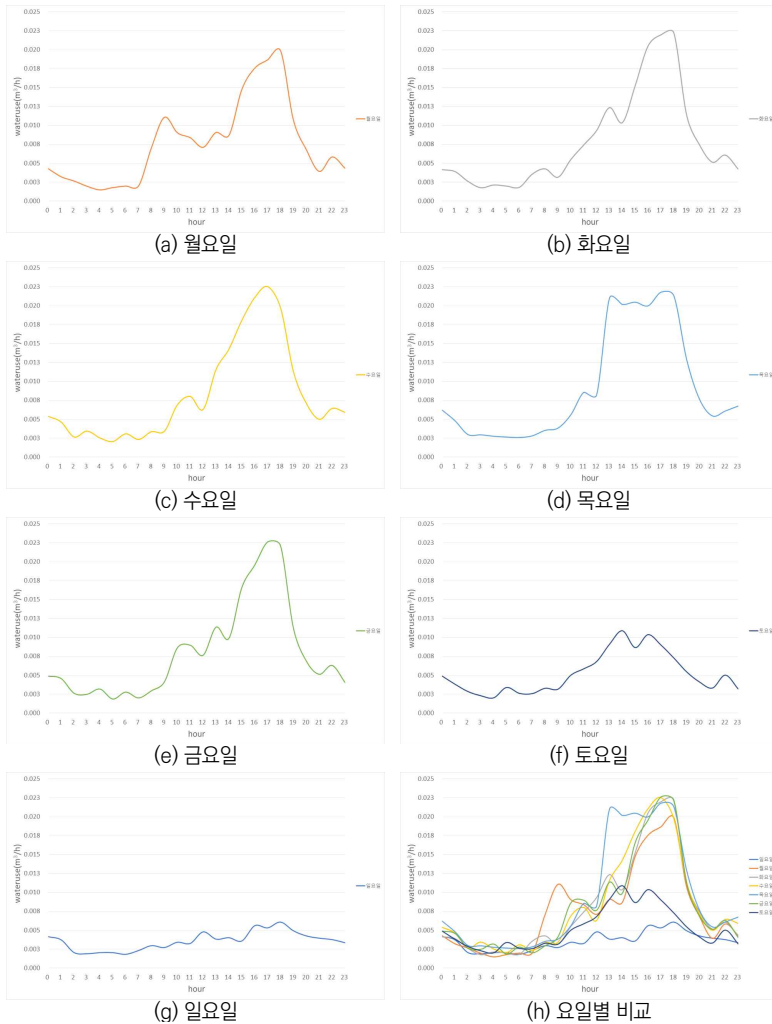
(b) 세로축 범위 : $0 \sim 0.2 \text{ m}^3/\text{h}$

4. 실시간 물사용량 보정 기법 적용

실시간 자료 수집 중 결측이 발생하면 즉각 대체될 값이 필요하며, 이를 산정하기 위한 방법을 3.4절에서 제시하였다. 최종 보정된 시간당 실시간 물사용량 시계열을 이용하여 요일별-시간별 물사용량 정보를 산정하였고 이를 기반으로 결측 발생시 대체용 보정값을 산정하였다.

시간당 물사용량 자료를 이용하여 요일별-시간별 평균값을 산정한 결과를 <그림 4>와 같이 나타내었다.

<그림 4> 요일별-시간별 물사용량 그래프



요일별-시간별 물사용량 그래프를 산정한 결과, 요일별로 차이는 있지만 대체로 저녁시간(16시부터 19시까지)에 침두 물사용량이 발생하는 것

을 알 수 있다. 또한 하루 중 가장 낮은 물사용량이 기록되는 새벽시간(0시부터 7시까지)의 물사용량에 비해 6배 이상의 물사용량이 발생하는 것을 알 수 있다. 월요일의 시간별 물사용량 그래프의 경우 아침시간(8시부터 10시까지)에 침두 물사용량이 발생하는 특징이 있으며, 목요일의 시간별 물사용량 그래프의 경우 점심시간(12시부터 14시까지)에 침두 물사용량이 발생하는 특징이 있다. 토요일과 일요일의 시간별 물사용량 그래프는 평일에 비해 절반 수준의 물사용량이 발생하는 특징이 있다.

하루(0시부터 23시까지)의 물사용량 변화 특징을 요일별 그래프를 통해 살펴보면 해당 수용가는 평일(월, 화, 수, 목, 금요일)의 물사용량은 평균 0.00817이고, 주말(토, 일요일)의 물사용량은 평균 0.00439으로 시간당 물사용량 차이가 뚜렷하게 나타난다. 또한 저녁시간(16시부터 19시까지)에 물사용량이 가장 많고 17시와 18시 사이에서 침두 물사용량이 기록되었다. 해당 수용가가 도심지에 위치한 점을 고려할 때 수용주체의 생활패턴이 반영된 결과로 판단할 수 있다. 따라서 요일별 정보를 기반으로 한 실시간 물사용량 보정시에 16시부터 19시까지 물사용량이 많은 시간 또는 평일과 다르게 주말에는 물사용량이 적은 특징 등을 고려하여 비교적 물사용량이 클 것으로 예상되는 평일 저녁시간 등에 활용성이 클 것으로 판단된다.

VI. 결론

본 연구에서는 실시간으로 수집되는 수용가의 물사용량 자료에 포함된 다수의 이상치와 결측치를 처리하고 최적의 보정 기법을 제시하고자, 이상치 탐색 및 처리, 결측치 처리를 수행하였다. 또한 실시간으로 수집되는 자료의 특성을 고려하여 결측 발생시, 요일별 물사용량 정보의 특성을 반영하여 즉각 보정이 이루어질 수 있는 방법론을 제안하였다.

시계열 자료의 이상치 판단 결과 이상치와 정상 자료의 차이가 비교적

뚜렷한 것으로 확인되었고, 이상치로 판단된 값을 결측치로 처리하였다. 기존의 결측치 보정 방법론의 개선을 위해 선행연구 및 통계적 이론을 토대로 새로운 결측치 보정 방법 4가지를 제안하였다. 새로운 방법 4가지와 기존 결측 보정 방법을 연구 자료에 대하여 모두 적용하고, 각 결측치 보정을 수행한 결과를 2가지 오차지표로 평가하였다. 평가결과, 본 연구에서 제시한 방법 중 ‘최근린 산술평균 방법’에서 오차지표가 가장 작게 산정되었으며, 기존 결측치 보정 방법에 비해서도 오차지표가 큰 차이로 작게 산정되었다. 따라서 기존 결측치 보정 방법에 대한 개선으로 본 연구에서 제시한 ‘최근린 산술평균 방법’에 의한 결측 보정을 제안하였다.

결측 보정까지 수행한 실시간 물사용량 자료를 시간단위의 물사용량 자료로 변환하고, 이를 이용하여 요일별-시간별 특징을 분석하였다. 시간당 물사용량 자료의 요일별 정보특징은 평일과 주말의 물사용량 차이가 비교적 뚜렷하게 나타났고, 저녁시간대에 침두 물사용량이 기록되는 양상이 나타났다. 결측 보정시 이러한 요일별 특징을 함께 고려하여 실시간 보정 값으로 활용한다면 보다 합리적인 보정에 활용 가능할 것으로 판단된다.

이때, 요일별-시간별 정보의 특징이 평일과 주말의 물사용량 차이가 뚜렷하게 나타난 것은 본 연구 대상 수용가가 대규모 도심지역이기 때문에 평일과 주말의 대상지역 내 물사용 인구수 차이의 영향이 큰 것으로 판단된다. 하지만 법정공휴일, 대체공휴일 등으로 평일이지만 공휴일로 지정된 날의 물사용량 특징은 주말의 특징과 유사하게 기록되었지만, 물사용량 정보의 특징을 산정할때는 요일을 기준으로 산정되어 일반적인 평일의 값으로 포함되었다. 1번 자료를 기준으로 이러한 주말이 아닌 공휴일은 전체 1278일 중 48일로 약 3.76%를 차지하고, 다른 자료에서도 유사한 비중을 차지한다. 이러한 주말이 아닌 공휴일에 대해서는 별도로 값을 관리하거나, 공휴일만의 시간별 물사용량 특성을 반영하는 등의 분석을 추가하여 보완할 필요가 있다.

본 연구에서 제안한 결측 보정 방법론은 결측이 발생한 구간을 선형적으로 보정하는 방법이기 때문에 시간당 물사용량의 변화가 크거나 돌발적

인 상황에 결측이 발생한 경우 반영하기 어렵다. 또한 수집된 자료의 원시 오류의 종류가 다양하고, 자료 중 원시 오류가 다수 포함되어 있기 때문에 원시 오류를 판단하고 처리하는 기법을 추가로 개발할 필요가 있다.

지리·지형적 특성이 반영되는 실시간 유량, 실시간 수위, 실시간 강수량 등의 자료의 경우, 결측발생 지점 인근의 관측값을 평균하거나, 결측발생 지점과 유사한 지점의 값을 대체하여 사용하는 등 지형 정보를 활용한 다양한 보정기법이 연구되어있다. 그러나 수용가의 실시간 물사용량 자료의 경우 수용주체인 사람의 인위적인 혹은 개인적인 생활패턴에 의해 물사용량이 달라지므로, 인근 수용가의 값을 대체값으로 활용하는 등의 지형적 정보를 활용한 보정은 정확도를 담보하기 어렵다. 보다 장기간의 실시간 물사용량 자료가 축적된다면 생활패턴에 의해 달라지는 물사용량의 양상이 유사한 수용가를 군집하여 대체값을 생성하는 등의 연구가 가능할 것으로 판단된다.

본 연구에서 제안한 방법론과 그 결과는 실시간 물사용량 정보를 이용한 스마트 물관리의 기초 자료를 관리하는데 활용할 수 있을 것으로 판단된다.

■ 참고문헌 ■

- 건강보험심사평가원, 2019, 『이상치 탐색을 위한 통계적 방법과 활용 방안』, 강원도: 건강보험심사평가원 심사평가연구소.
- 유재현·김계현·박용길·염경택, 2020, “스마트폰 앱 기반의 물 사용량을 활용한 실시간 독거노인 관리 시스템 개발,” 『한국디지털콘텐츠학회 논문지』, 21(6), pp.1017-1024, DOI: 10.9728/dcs.2020.21.6.1017.
- 주진철·안호상·안창혁·고경록·오현제, 2012, “상수도 원격검침시스템의 현장 적용성 평가 및 가정용수 사용량 분석,” 『대한환경공학회지』, 34(10), pp.656-663.
- 최계운·이호선·박미형·윤근호·윤지민, 2013a, “SWG 연구단 소개 및 개요,” 『물과 미래』, 46(7), pp.8-14.
- 최계운·김주환·박수완·이호선·최진탁, 2013b, 『스마트 워터 그리드 기초 용어사전』, 서울: 양서각.

- 홍공현·김주환·이두진·최태호, 2017, “물 사용량 분석 및 물 정보 제공을 위한 스마트워터미터링 시스템 구축,” 『대한상하수도학회·한국물환경학회 2017 공동학술발표회』, 광주 김대중컨벤션센터, pp.700-701.
- Britton, T. C., R. A. Stewart, and K. R. O'Halloran, 2013, “Smart metering: Enabler for rapid and effective post meter leakage identification and water loss management,” *Journal of Cleaner Production*, 54, pp.166-176, DOI: 10.1016/j.jclepro.2013.05.018.
- Doolan, C., 2011, “Sydney water’s smart metering residential study,” *Proceedings of 4th AWA national water efficiency conference*, Melbourne Sydney.
- Koo, K. M., K. H. Han, K. S. Jun, G. Lee, J. S. Kim, and K. T. Yum, 2021, “Performance assessment for short-Term water demand forecasting models on distinctive water uses in Korea,” *Sustainability*, 13(11), 6056, pp.1-18, DOI: 10.3390/su13116056.
- Little, R. J. A. and D. B. Rubin, 2002, *Statistical analysis with missing data*, Hoboken, NJ: John Wiley & Sons.
- Liu, B., W. Zhou, T. Zhu, H. Zhou, and X. Lin, 2016, “Invisible hand: A privacy preserving mobile crowd sensing framework based on economic models,” *IEEE Transaction on Vehicular Technology*, 66(5), pp.4410-4423, DOI: 10.1109/TVT.2016.2611761.
- Roberta, P. and D. G. Giuseppe, 2020, “A nonparametric framework for water consumption data cleansing: An application to a smart water network in Naples (Italy),” *Journal of Hydroinformatics*, 22(4), pp.666-680, DOI: 10.2166/hydro.2020.133.
- Sonderlund, A. L., J. R. Smith, C. J. Hutton, and Z. Kapelan, 2014, “Using smart meters for household water consumption feedback: Knowns and unknowns,” *Procedia Engineering*, 89, pp.990-997, DOI: 10.1016/j.proeng.2014.11.216.

윤정환: 인하대학교 대학원 수자원분야 박사과정으로 재학 중이며, 스마트 물관리 솔루션을 제공하는 소프트웨어회사를 운영하고 있다. 물, 기상, 환경정보를 이용하여 GIS, AI, 머신러닝, 빅데이터 분석 등의 연구를 수행하고 있다(lenablue12@gmail.com).

이준형: 인하대학교 대학원 수자원분야 박사과정으로 재학 중이며, 동 대학원 토목공학 석사학위를 취득하였다. 기후변화, 수자원 이용, 수문 빅데이터 분석이 주요 관심 분야이며, 통계기반 시계열 자료처리, 머신러닝 등을 이용한 분석 연구를 수행하고 있다(lee_junhyeong@naver.com).

유영훈: 인하대학교 대학원 수자원분야 박사과정으로 재학 중이며, 동 대학원 토목공학 석사학위를 취득하였다. 기후변화, 지속가능한 수자원 관리, 최적 관리 방안이 주요 관심 분야이며, 홍수예경보, 단기 및 장기유출 등의 연구를 수행하고 있다(dudgns5971@naver.com).

왕원준: 인하대학교 대학원 수자원분야 박사과정으로 재학 중이며, 동 대학원 토목공학 석사학위를 취득하였다. 치수사업 경제성 분석, 정성적 위험도 평가, 공간통계 분석 등이 주요 관심 분야이며, 격자기반 공간분석자료, 빅데이터 등을 이용한 분석 연구를 수행하고 있다(makelest@naver.com).

김형수: 미국 Colorado State University에서 공학 박사학위를 취득하고 현재 인하대학교 사회인프라공학과에서 재직 중이다. 주요 관심분야는 수문분석, 카오스, 수재해, 습지 등이다(sookim@inha.ac.kr).

투 고 일: 2021년 06월 25일
심 사 일: 2021년 06월 30일
게재확정일: 2021년 07월 18일