



# 회귀나무(Regression tree)를 이용한 생활용(가정용 및 일반용) 지하수 이용량 추정 방법

김규범<sup>1,‡</sup> · 김지욱<sup>2</sup> · 신현정<sup>3</sup>

<sup>1</sup>대전대학교 건설안전방재공학과

<sup>2</sup>한국수자원공사

<sup>3</sup>한국농어촌공사

## 요 약

지하수 개발가능량과 이용량의 정확한 산정은 합리적인 지하수 이용관리 정책을 위하여 필요하다. 본 연구에서는 행정구역(읍면)별로 가정용 및 일반용 관정의 지하수 이용량에 대한 보다 정확한 통계를 산정하기 위하여 회귀나무 분류 기법을 사용하였다. 모델에 사용된 표본수는 가정용 관정 502개, 일반용 관정 771개로서, 모델 입력 인자는 가정용은 7개, 일반용은 4개를 적용하였고 출력인자는 실측된 일 이용량을 사용하였다. 가정용 관정의 회귀나무 모델에 의한 일 이용량은 실측 이용량과 매우 유사한 것으로 평가되었으며, 일반용 관정의 회귀나무 모델에 의한 일 이용량도 기존의 국토교통부에서 제시한 추정 식 보다 실측 이용량과 보다 유사한 것으로 평가되었다. 향후 추가적인 실측 자료를 확보하여 본 연구의 방법을 적용, 보완한다면 행정구역별 이용량 통계의 신뢰도가 크게 향상될 것으로 기대된다.

**주요어:** 지하수 이용량, 회귀나무, CHAID, 대표단위수

**Gyoo-Bum Kim, Ji-Wook Kim and Hyeon-Jeong Shin, 2019, Estimation of groundwater usage for the living (domestic and business) purpose wells by using a regression tree method. Journal of the Geological Society of Korea. v. 55, no. 6, p. 683-691**

**ABSTRACT:** Accurate estimation of groundwater sustainable yield and usage amount is required to make reasonable groundwater management policies. In this study, a regression tree classification technique is used to estimate the accurate values on the amount of groundwater use of administrative (Eup and Myeon) districts for domestic and business wells. The number of samples used in this model is 502 for domestic wells and 771 for business wells. The number of input factors for tree models were 7 and 4, respectively, and the measured daily usage was used as output value. The amount of daily use by the regression tree model of the domestic wells is estimated to be very similar to the actual usage, and the amount by the model of the business wells is more similar to the actual value than the estimated value by the MOCT's equation. In the future, it is expected that the reliability of the usage statistics of administrative districts will be greatly improved if additional data is secured and this classification method is applied and modified.

**Key words:** groundwater use, regression tree, CHAID, representative elementary counts

(Gyoo-Bum Kim, Daejeon University, Daejeon 34520, Republic of Korea; Ji-Wook Kim, Korea Water Resources Corporation, Daejeon 34350, Republic of Korea; Hyeon-Jeong Shin, Korea Rural Community Corporation, Daejeon 35209, Republic of Korea)

## 1. 서 언

우리나라는 1993년 지하수법 제정 이후 매년 지하수 통계를 작성해 오고 있다. 지하수 개발가능량

과 이용량에 대한 국가 통계는 지하수 관리의 기초 자료로서, 지하수의 과다개발 관리 및 적정 활용 등을 위한 기본 자료로 사용된다. 즉, 특정 지역에서 지하수를 개발하고자 할 경우 추가적으로 취수가 가능한

<sup>‡</sup> Corresponding author: +82-42-280-4804, E-mail: [geowater@dju.kr](mailto:geowater@dju.kr)

지하수량과 현재 사용되고 있는 지하수량의 파악은 주변에 영향을 최소화하면서 신규 개발 규모를 결정하기 위한 필수적인 요소가 된다. 또한, 가뭄 발생 등 기후변화 시대에 지하수에 대한 개발 욕구가 증대되고 있는 만큼, 지방자치단체 등 행정기관의 지속가능한 지하수 자원의 확보 및 보전을 위하여 과다개발을 방지하고 오염 확산을 차단하기 위한 기본 통계로서 의미가 있다.

최근 정부 통계에 의하면, 지난 20여 년 간 지하수 이용량은 꾸준히 증가되어 왔으며, 특히 생활용과 농업용 지하수 이용량이 지속 증가되고 있는 특성을 보인다(MOCT, 2018). 2014년 말 우리나라의 지하수 이용량은 연간 41억  $m^3$ 으로서, 이는 하천유지용수를 제외한 물 이용량 251억  $m^3$ 의 16.3%를 차지하고 있다(MOCT, 2016). 유럽의 경우를 보면, 지표수 지하수의 전체 이용량 중에서 지하수의 비율이 약 48%로서 국가별로는 덴마크가 약 99%, 아이슬랜드가 약 99%, 말타가 약 94%, 코소보 약 93% 등으로 매우 높고, 그리스, 싸이프러스, 라트비아, 룩셈부르크, 슬로바키아, 보스니아 헤르체고비나 등은 50%를 넘어 우리나라 보다 지하수 이용 비율이 높을 뿐 아니라, 그 비율도 2005년에 비하여 전반적으로 증가되고 있다(European Commission, 2019, [https://ec.europa.eu/eurostat/statistics-explained/index.php/Water\\_statistics#Water\\_uses](https://ec.europa.eu/eurostat/statistics-explained/index.php/Water_statistics#Water_uses)). 최근 들어 낙동강 주변의 강변여과수 개발, 가뭄시 지하수 활용, 기존 관정의 개보수 사업 등과 같은 지하수 이용에 대한 관심이 제고되고 있으나, 유럽에 비하여 여전히 지하수 이용률이 낮은 상태로서 우리나라의 지하수 개발가능량 129.9억  $m^3$ /년을 고려할 때 지하수의 추가적인 개발 가능성은 충분히 존재하는 편이다(MOCT, 2016). 그러나, 최근 들어 가뭄 등 기후변화에 대비하여 지하수 이용을 확대하고자 하는 정부의 노력이 제고되고 있으나 지하수 이용량에 대한 정확한 통계가 부족한 상태에서 이와 같은 지하수 활용 확대 정책의 추진에는 한계가 있을 것으로 보인다. 또한, 지하수의 고갈을 방지하고 주변의 환경적 피해를 최소화하면서 지하수를 개발하기 위해서는 현재 사용하고 있는 지하수량에 대한 정확한 진단과 추가 개발 가능량에 대한 평가가 선행되어야 한다. 일부 연구자들에 의하여 지하수 이용량 추정 기법들이 검토된 바 있는데, Siebert *et al.* (2010)은 국가 또는 주와 같은

광역적 단위의 농업용 지하수 이용량에 대하여 수인 현황에 대한 센서스 기반의 추정 기법을 적용하여 전세계 농경지 면적 38%가 545  $km^3/y$ 의 지하수를 사용하는 것으로 평가한 바 있으며, Martínez-Santos and Martínez-Alfaro (2010)은 스페인의 농업 지역에서의 지하수 이용량을 지하수위를 활용한 지하수의 물수지 평가를 통하여 추정한 바 있다. 또한, 최근 연구에서는 인공신경망을 활용한 물 수요를 추정 기법이 개발된 바 있다(Huang *et al.*, 2017).

그 동안 우리나라 지하수 이용량 통계의 정확성에 대한 전문가들의 논쟁과 더불어 이용량 추정을 위한 연구가 수행된 바 있다(Woo *et al.*, 1994; Park *et al.*, 2004; Kim *et al.*, 2013). 지하수 이용량은 유량계, 시간계 등을 이용하여 직접 조사하여야 하나 실질적으로는 행정 인력의 부족 등으로 인하여 전수조사가 불가능할 뿐 아니라 전체 시설의 약 70%가 넘는 소형시설(약 104만공)에는 유량계, 시간계 등이 설치되어 있지 않은 것이 현실이다(Kim *et al.*, 2013). 따라서 일부 지방자치단체에서는 지하수 개발이용허가 또는 신고시 제출한 취수계획량을 이용량으로 적용하는 경우가 많아 과다 산정에 대한 우려가 제기되기도 하였다. 이를 해결하고자 국토교통부에서는 2006년부터 2012년까지 전국 32개 시군의 7,693개 지하수 관정을 대상으로 월별 지하수 이용량에 대한 실측 조사 및 연간 이용량 조사를 실시한 바 있다. 이 결과를 토대로 국토교통부는 2013년에 지하수 이용량 산정 기준을 수립하여 현재까지 활용되고 있다(MOCT, 2013).

본 연구에서는 현장에서 실측된 월간 지하수 이용량 자료를 활용하여 생활용(가정용과 일반용) 지하수 관정의 연간 이용량 추정 기법을 도출함으로써 보다 신뢰성 높은 행정구역(읍면)별 지하수 이용량 산정 방법을 제시하고자 한다.

## 2. 지하수 이용량 실측 현황

본 연구의 기초자료로 사용된 지하수 이용량 실측 데이터는 국토교통부에서 2006~2012년까지 32개 시군을 대상으로 실시한 7,693개 관정의 월 및 연간 지하수 이용량 자료와 충청남도 및 한국농어촌공사에서 2016~2018년에 실시한 충청남도 지역내 362개 관정의 지하수 이용량 실측 자료를 포함하였다.

**Table 1.** Statistics for original samples of living purpose wells; domestic and business.

Sub-use	Before removing the outliers			Data used for a model construction		
	Number of wells	Average daily use (m <sup>3</sup> /d)	Standard deviation (m <sup>3</sup> /d)	Number of wells	Average daily use (m <sup>3</sup> /d)	Standard deviation (m <sup>3</sup> /d)
Domestic use	524	1.3	2.4	502	0.9	0.9
Business use	822	17.2	36.0	771	8.7	13.1

지하수 이용량 실측은 지하수업무수행지침(MOCT, 2015)에서 규정하고 있는 주용도 및 세부용도별로 구분하여 매월 동일 날짜에 현장을 방문하여 유량계, 시간계 또는 전력계를 조사하는 방식으로 이루어졌다. 현장 조사 항목은 월간 지하수 이용량, 급수인구, 급수기구, 관정구경, 관정심도, 토출관구경, 펌프 마력, 양수능력, 가축사육두수, 가축사육종, 측정방법 등으로서, 지하수 개발이용 신고서의 항목이 포함되도록 하였다.

이상 조사된 자료의 기본 현황을 보면, 생활용이 4,959 개소, 공업용 31 개소, 농업용 3,065 개소이며, 단순 산술평균하면 생활용 관정의 일평균 이용량은 약 24.5 m<sup>3</sup>/일, 공업용 관정은 약 35.5 m<sup>3</sup>/일, 농업용 관정은 약 40.1 m<sup>3</sup>/일로 나타났다. 이와 같은 일 이용량은 실제 사용량보다 큰 것으로 파악되는데, 생활용 관정의 이용량이 높게 나타난 것은 상수도, 학교 등 공공용으로 공급되는 대용량 시설이 포함된 경우이거나, 유량계와 달리 이용량이 과다 계산되는 시간계와 전력계에 의한 측정이 많이 포함되어 있기 때문이다. 따라서, 본 연구에서는 유량계와 비교하여 약 2~4 배 정도 과다 산정된 시간계와 전력계의 조사 자료는 분석용 자료에서 제외하도록 하였다.

### 3. 표본 추출

본 연구에서는 생활용 지하수 관정의 대부분을 차지하는 가정용과 일반용 지하수 이용량 추정 방법을 개발하고자 하였으며, 추후 추정 모델의 적용성을 위하여 새울정보시스템(지자체의 지하수 행정관리 업무용 시스템)내에 구축되어 있는 DB 항목을 사용하여야 하는 조건하에서 수행하였다. 위에서 언급한 바와 같이 측정된 지하수 이용량 자료 중에는 신뢰성이 낮은 값이나 모델에 적합하지 않은 값들이 존재할 수 있으므로 분석 결과의 신뢰성을 제고하기 위

하여 표본의 선별 과정을 다음과 같이 수행하였다.

첫째, 이용량 추정 모델이 가능하도록 입력인자를 충실히 갖추고 있는 자료를 표본으로 사용하였다. 즉, 일부 조사자료 중에는 심도, 양수능력, 펌프 마력 등 시설 제원이 누락된 경우가 있으므로 가용한 모든 입력 인자가 확보되어 있는 관정들을 표본으로 선정하였다. 따라서, 급수인구, 관정구경, 관정심도, 토출관구경, 펌프 마력, 양수능력 등의 정보를 모두 갖고 있는 관정을 표본으로 사용하였다.

둘째, 위에서 언급한 바와 같이 시간계와 전력계로 측정된 자료는 과다 산정되는 경향이 있으므로 제외하였다.

셋째, 상자 도표를 사용하여 일 이용량을 도시한 후 산술평균의 약  $\pm 1.5\sigma$  (여기에서  $\sigma$ 는 표준편차)를 벗어나는 자료는 이상값으로 보고 제거하였다(Kim *et al.*, 2013). 가정용 이용량의 평균은 0.9 m<sup>3</sup>/일, 일반용 이용량은 8.7 m<sup>3</sup>/일이며, 이상값의 범위는 가정용의 경우는 약 5.0 m<sup>3</sup>/일, 일반용의 경우는 70.0 m<sup>3</sup>/일 이상으로서 분석에서 제외하였다(표 1). 이상의 과정을 통하여 최종적으로 모델에 사용된 자료는 가정용 502개, 일반용 771개에 해당한다.

### 4. 회귀나무 분석법(Regression tree)

Kim *et al.* (2013)은 단순 선형 회귀식에 의한 지하수 이용량 추정식을 제시한 바 있으나 회귀식에 의한 이용량 추정이 지나치게 단순화되어 있어 읍면동 단위 등 좁은 지역에 대한 이용량 통계를 도출하는데는 한계를 갖고 있다. 회귀나무 분석 결과는 나무구조로 표현되어 있어 이해가 쉽고 여러 개의 변수가 결합하여 목표 변수에 어떻게 영향을 주는지 파악이 용이한 장점이 있다. 아울러, 선형 회귀식과 달리 데이터의 선형성, 정규성 또는 등분산성 등의 가정을 필요로 하지 않으므로 적용에 제한점이 적다. 그

**Table 2.** Input and output variables of regression tree models for two purposes of wells: domestic and business.

Content	Input variables	Output variable
Domestic use	7 variables (Eup-Myeon-Dong code <sup>1)</sup> , Person, Well depth, Well diameter, HP of pump, Pipe diameter, Pumping capacity)	Daily groundwater use
Business use	4 variables (Eup-Myeon-Dong code, Well depth, HP of pump, Pumping capacity)	Daily groundwater use

Remark: 1) Dong code: 1, Eup code: 2, Myeon code: 3

러나, 회귀나무 분석에서는 자료의 비연속성 및 구획화가 존재하게 되며, 이와 같은 분리의 경계점에서는 예측 오류의 가능성이 존재하므로 유의해야 한다.

의사결정나무(Decision tree) 분석 방법은 나무 기반의 분류 모형을 만드는 것으로서 독립 변수 값을 기반으로 종속 변수 값을 예측하거나 케이스를 집단으로 분류하기 위하여 사용되는 방법이다. 데이터를 분석하여 이들 사이에 존재하는 패턴을 예측 가능한 규칙들의 조합으로 나타내게 되며 그 모양이 나무와 유사하다고 하여 의사결정나무라고 하며, 의사결정나무에는 분류나무(Classification tree)와 회귀나무(Regression tree)로 구분이 된다.

본 연구에서는 회귀나무의 적용을 위하여 CHAID (Chi-squared automatic interaction detection) 알고리즘을 적용하였다. CHAID는 카이제곱 검정 또는 F-검정을 이용하여 다지 분리(Multiway split)을 수행하는 알고리즘으로서 1963년 J.A. Morgan과 J.N. Sonquist가 발표한 AID (Automatic interaction detection) 시스템에서 유래되었는데, 입력 변수들간의 통계적 관계를 찾는 목적으로 개발되었으며, 이들 통계적 관계를 의사결정나무에 의하여 표현할 수 있도록 개발되었다(Thearling, 1995; IBM, 2012).

회귀나무 분석법은 종속변수가 연속형 자료일 때 적용 가능한 의사결정나무로서, 변수의 분류와 분류 기준값의 선택 방법으로 F-통계량을 사용하였다. F-통계량은 집단 간 분산과 집단 내 분산의 비를 나타내는 것으로서 일원배치법에서의 검정 통계량과 동일하며 그 값이 클수록 오차의 변동에 비해 처리의 변동이 크다는 것을 의미하며, 이는 자식 노드들이 서로 이질적임을 의미하므로 이 값이 커지는 방향으로 가지 분할을 수행하는 과정으로 이루어진다. F-통계량은 다음과 같이 표현된다.

$$F = \frac{MST}{MSE} = \frac{SST/(k-1)}{SSE/(n-k)}$$

여기에서, *MST*는 총제곱평균(Mean squares for treatments), *MSE*는 잔차제곱평균(Mean squares for error), *SST*는 총제곱합(Sum of squares for treatments)으로서  $\sum(\hat{y} - \bar{y})^2 + \sum(y - \hat{y})^2$ 이며, *SSE*는 잔차제곱합(Sum of squares for error)으로서  $\sum(y - \hat{y})^2$ 을 의미하며, *k*는 입력변수의 갯수, *n*은 데이터의 갯수,  $\hat{y}$ 는 예측값,  $\bar{y}$ 는 관측값 *y*의 평균을 의미한다.

## 5. 가정용 지하수 이용량 추정 기법

가정용 지하수 관정에 대한 일 이용량 추정을 위하여 입력 변수로서 위에서 언급한 현장 조사 항목을 활용하였으며, 여러 번의 시행 착오를 거친 후 회귀나무 모델의 입력 자료로 사용할 변수를 결정하였다(표 2).

나무 구조의 분할 및 병합을 위한 유의수준은 0.05를 적용하였으며, 최종 구축된 모델의 각 입력 변수의 중요도를 보면, 가장 중요한 입력 변수는 급수인구(예측변수 중요도 0.52)이며, 다음으로는 양수능력(0.29), 펌프마력(0.14) 및 관정 심도(0.05) 등의 순으로 나타나 일반 가정에서 사용되는 용도인 만큼 급수인구에 의한 영향이 잘 반영되어 있다. 이로부터 개발된 회귀나무 모델은 그림 1과 같다. 최상위 나무 구조를 보면, 급수인구가 3명 보다 큰 경우의 일 이용량은 약 1.430 m<sup>3</sup>/d로 분류되며, 3명 보다 작은 경우에는 하부 구조의 조건에 따라서 일 이용량이 분류됨을 알 수 있다. 한편, 실측된 일 이용량과 추정된 일 이용량의 상관계수는 약 0.49 정도로서 낮아 개별 관정의 이용량을 추정하는 것은 한계가 존재한다. 따라서, 본 연구에서는 개별 관정의 이용량 추정 모델보다는, 일정 규모의 면적내에서 다수의 관정이 존재하는 경우인 행정구역별 지하수 이용량 추정의 적합성을 보기 위하여 본 회귀나무 모델을 검토하였다.

이를 위하여, 모델 구축에 사용된 표본으로부터 10개, 30개, 50개, 100개, 150개, 200개, 250개 등으로 새로운 표본을 각각 10가지 경우를 추출한 후 일 이용량 평균을 도시해 보았다(그림 2). 이로부터 파악된 관정의 REC (Representative elementary counts, 대표단위수)가 약 150개 이상이 되면 일 이용량 평

균에 큰 변화가 나타나지 않음을 알 수 있다. 이와 같은 결과는 일정 규모 이상의 관정수를 갖는 경우에는 가정용 관정의 지하수 이용량 추정에 본 회귀나무 모델이 적용 가능함을 보여준다. 우리나라 전체의 가정용 지하수 관정수는 607,930개로서 국토 면적으로 나누면 1 km<sup>2</sup> 당 약 6개가 존재하며, 150개

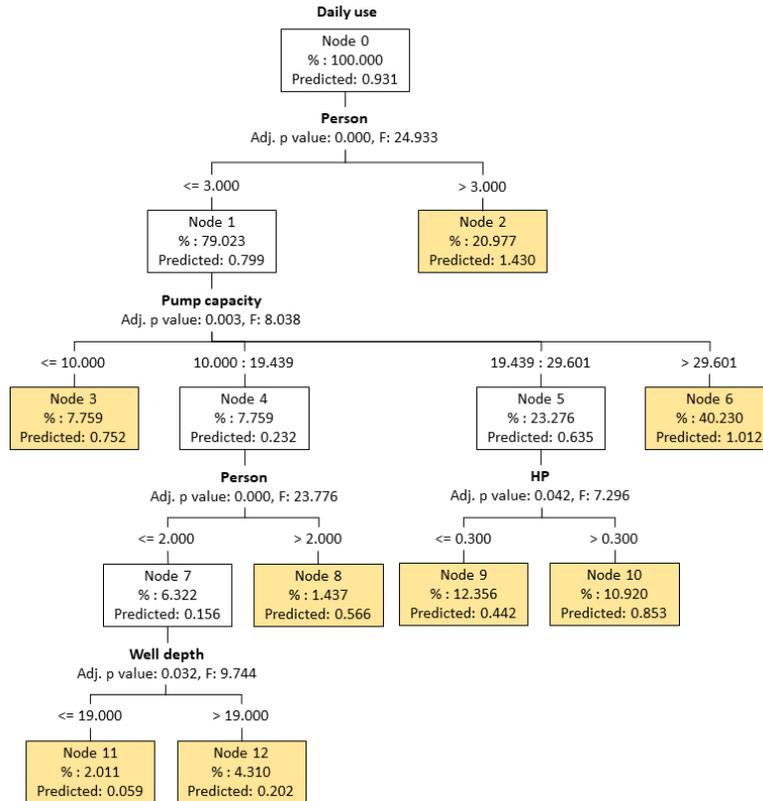


Fig. 1. Regression tree model for domestic groundwater use estimation.

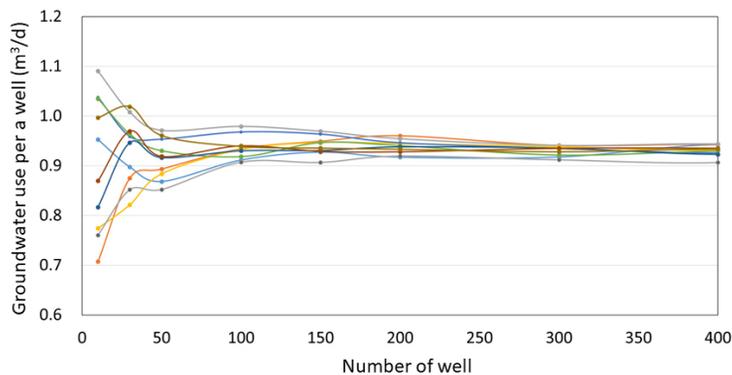


Fig. 2. Representative elementary counts distribution for groundwater use estimation of domestic purpose wells.

관정에 대한 면적은 약 24 km<sup>2</sup>에 해당한다(MOCT, 2018). 이 면적은 일반적인 읍면 면적보다 작으므로 읍면 단위로 가정용 관정의 지하수 이용량을 추정하는 수단으로서 본 회귀나무 모델은 적용 가능할 것으로 평가된다.

### 6. 일반용 지하수 이용량 추정 기법

일반용으로 사용되는 생활용 지하수 관정의 이용량에 대해서도 4개의 입력 변수를 토대로 회귀나무 모델을 구축하였다(표 2; 그림 3). 나무 구조의 분할 및 병합을 위한 유의수준은 0.05를 적용하였으며, 최종 구축된 모델의 각 입력 변수의 중요도를 보면, 가장 중요한 입력 변수는 양수능력(예측변수 중요도 0.65)이며, 다음으로는 관정 심도(0.18), 읍면동의 구분(0.17) 등의 순으로 나타났다. 일반용의 경우에는 지하수의 사용처가 다양하고 가정용에 비하여 많은 양을 취수하게 되므로 목표 취수량에 부합하는 규모의 펌프를 설치하게 되고 이 펌프의 양정고를 고려한 양수능력에 의하여 실제 이용량이 영향을 받는

것으로 파악된다. 한편, 실측된 일 이용량과 예측된 일 이용량의 상관계수는 약 0.47 정도로서 낮으므로 가정용과 마찬가지로 개별 관정의 이용량을 추정하는 것은 한계가 존재한다. 따라서, 일반용 관정에 대해서도 개별 관정의 이용량 추정 모델로서는 한계가 있는 것으로 보고 읍면 단위의 행정구역별 지하수 이용량 추정의 적용성을 검토하였다.

이를 위하여, 모델 구축에 사용된 표본으로부터 10개, 30개, 50개, 100개, 150개, 200개, 300개, 400개 등으로 새로운 표본을 각각 10가지 경우를 추출한 후 일 이용량 평균을 도기한 결과, 관정 수가 약 150개 이상이 되면 일 이용량 평균에 큰 변화가 나타나지 않음을 알 수 있다(그림 4). 우리나라 전체의 일반용 지하수 관정수는 166,165개로서 국토 면적으로 나누면 1 km<sup>2</sup> 당 약 1.7개가 존재하며, 150개 관정에 대한 면적은 약 90 km<sup>2</sup>에 해당한다(MOCT, 2018). 일반적으로 읍면의 면적은 개략적으로 60~130 km<sup>2</sup> 정도이므로 본 회귀나무 모델을 이용하여 일반용 관정의 지하수 이용량 추정 수단으로 활용 가능할 것으로 보였다.

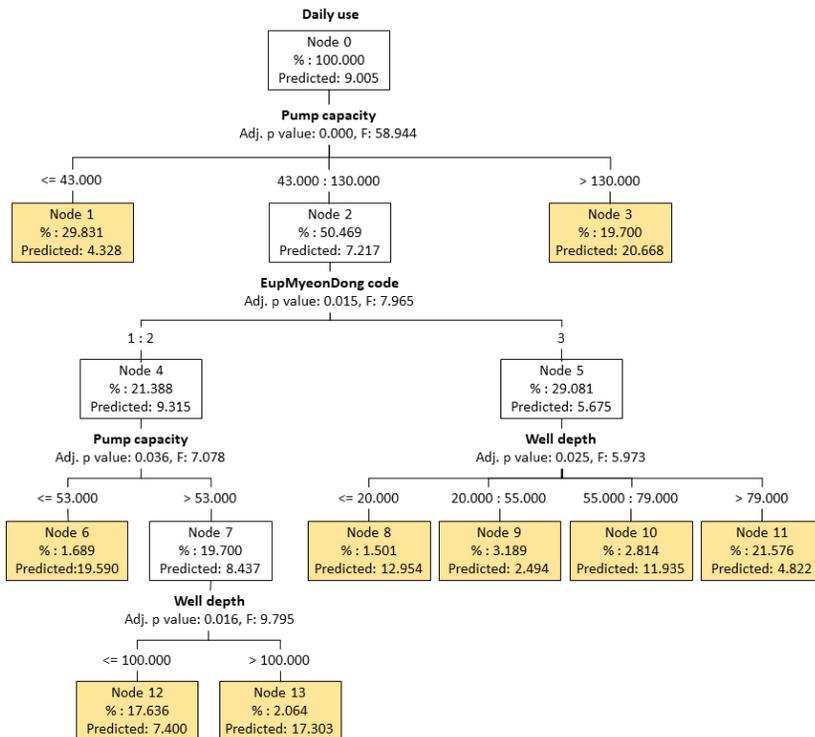


Fig. 3. Regression tree model for business groundwater use estimation.

## 7. 토 론

본 연구에서 도출된 회귀나무 모델과 기존의 국토교통부(현 환경부)에서 제시한 이용량 추정 방법과의 비교를 통하여 본 방법의 적용성을 평가하였다(표 3).

분석에 사용된 표본을 대상으로 도시지역, 도농복합지역 및 농촌지역으로 구분한 후 국토교통부에서 제시한 가정용 및 일반용의 지하수 이용량 추정식을 적용한 후, 본 연구에서 개발된 회귀나무 모델의 결과와 비교해 보았다(표 4). 가정용의 경우에는 국토교통부의 방법에 의한 값(관정당 일 이용량은  $0.78 \text{ m}^3$ , 1인당 일 이용량은  $0.27 \text{ m}^3$ )이 실측값(관정당 일 이용량은  $0.88 \text{ m}^3$ , 1인당 일 이용량은  $0.31 \text{ m}^3$ )이나 회귀나무 모델에 의한 값(관정당 일 이용량은  $0.92 \text{ m}^3$ , 1인당 일 이용량은  $0.32 \text{ m}^3$ ) 보다 약간 작게 나타났다. 도시지역의 경우는 표본의 수가 적어 실측값과 차이가 다소 존재하나 도농복합지역과 농촌지역은 회귀 모델에 의한 값과 실측 값이 잘 일치함을 알 수 있다. 기존 국토교통부의 방법에 의한 이용량 추정은 실제 이용량과 약 11%의 차이가 존재하였으나, 금회 모델에 의해서는 5% 이내로 줄어드는 것으로 나타났다.

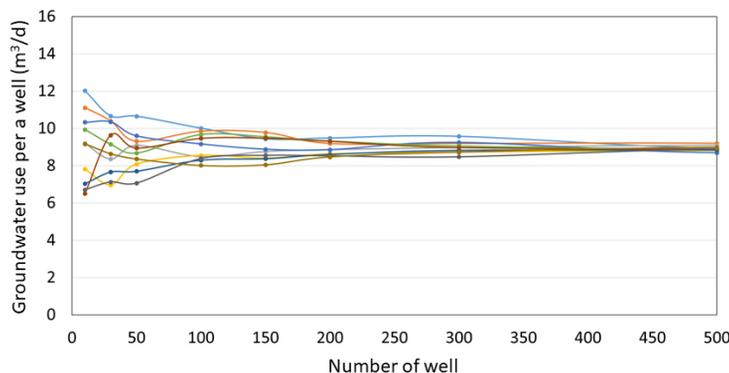
일반용의 경우에는 국토교통부의 방법에 의한 추정 값의 오차가 큰 것으로 나타났는데, 추정 일 이용량이 실측 값의 약 0.57배에 불과한 것으로 나타났다. 도시지역의 경우에는 표본의 수가 적어 각 방법별 편차가 크나, 도농복합지역과 농촌지역의 경우에는 회귀나무 모델에 의한 추정 결과가 실제 일 이용량과 유사함을 알 수 있다. 일반용의 경우에 국토교통부의 방법이 실측 값과 큰 차이가 존재하는 것은 기존 추정 식이 양수능력을 입력인자로 하는 단순 회귀모형을 적용하였고 모델의 설명력( $R^2$ )도 0.014 (도시지역), 0.062 (도농복합지역), 0.011 (농촌지역) 등으로 매우 낮기 때문으로 보인다(Kim *et al.*, 2013). 기존 국토교통부의 방법에 의한 이용량 추정은 실제 이용량과 약 43%의 차이가 존재하였으나, 금회 모델에 의해서는 3% 이내로 줄어드는 것으로 나타났다.

가정용 관정의 일 이용량은 일반 가정에서의 생활 용수로 사용되기 때문에 관정에 의하여 공급받는 급수 인구에 의하여 주로 결정되게 되는데, 일반적으로 가구 당 인구 수의 편차가 크지 않으므로 가정용 관정의 지하수 일 이용량의 편차도 크지 않을 것으로 예상 된다(그림 5). 이와 같은 특성은 표 4에서 보는 바와 같이 2가지 추정 방법과 실측 값의 차이가 크지 않다는 점에 의하여 설명된다. 반면에, 일반용 관

**Table 3.** Estimation equation of groundwater use for domestic and business purpose wells proposed by the MOCT (2013).

Purposes	Urban area	Urban-rural complex area	Rural area
Domestic use	$0.173+(\text{Person} \times 0.210)$	$0.173+(\text{Person} \times 0.210)$	$0.265+(\text{Person} \times 0.181)$
Business use	$2.199+(\text{PC} \times 0.021)$	$3.154+(\text{PC} \times 0.023)$	$3.521+(\text{PC} \times 0.011)$

Remark) PC : Pumping capacity ( $\text{m}^3/\text{d}$ )



**Fig. 4.** Representative elementary counts distribution for groundwater use estimation of business purpose wells.

**Table 4.** Comparison of daily and yearly groundwater use estimations and actual measurements for two purposes: domestic and business (unit: m<sup>3</sup>/d).

Purposes	Methods	Contents	Urban	Urban-rural complex	Rural	Total		
Domestic use		No. of well	22	255	225	502		
		Person (average)	2.50	3.29	2.40	2.86		
	MOCT (2013)		Average daily use per single well	0.70	0.86	0.70	0.78	
			Average daily use per person	0.28	0.26	0.29	0.27	
	Regression tree		Average daily use per single well	0.83	0.98	0.86	0.92	
			Average daily use per person	0.33	0.30	0.36	0.32	
		Actual use		Average daily use per single well	0.62	0.94	0.85	0.88
				Average daily use per person	0.25	0.29	0.35	0.31
	Business use		No. of well	34	547	190	771	
		MOCT (2013)	Average daily use per single well	4.47	5.12	4.41	4.91	
Regression tree		Average daily use per single well	11.23	8.98	8.44	8.95		
Actual use			Average daily use per single well	8.99	9.05	7.59	8.69	

정의 경우에는 사용 목적과 사용 방법 등이 매우 다양하여 관정별 일 이용량이 큰 편차를 보일 것으로 예상된다. 일반용으로 분류되는 관정은 주로 식당, 여관, 목욕탕, 세차장, 수영장, 빌딩 및 소규모 개인사업체 등에서 영업용 목적으로 사용되거나 공원, 박물관, 도서관 등에서 비영업용으로 사용되는 것이기 때문에 관정당 이용량의 편차가 큰 편이다(그림 5). 따라서, 일반용 지하수 관정의 일 이용량 추정의 정확도를 보다 높이기 위해서는 위에서 언급한 용도별 이용량에 영향을 미치는 인자를 추가로 확보하여 조사하는 것이 필요할 것이다. 일반용 관정의 이와 같은 이용량 편차의 특성은 표 4에서 보는 바와 같이 회귀나무 모델에 의한 일반용의 관정당 지하수 일 이용량 추정 값과 실측 값의 차이가 가정용 보다 크게 나타나는 점을 설명해 준다.

상기 모델을 실제 충청남도 각 시군에 대하여 적용하여 관정당 연간 지하수 이용량을 추정한 결과, 기존의 통계상의 추정 이용량과 비교해 보면 시군별 편차가 줄어들었음을 알 수 있다(그림 6). 충청남도

의 시군 지역은 대부분 농업 또는 도농 복합형으로 유사한 생활용수 이용 패턴을 가질 것으로 추정되는 바 시군별 관정당 지하수 이용량의 편차도 크지 않아야 할 것이므로 본 추정 기법이 기존에 비하여 개선된 것으로 보았다. 지하수 조사연보(MOCT, 2018) 자료에 의하면, 아산시와 예산군의 가정용 지하수 이용량이 각각 공당 2,181 m<sup>3</sup>/y 및 1,424 m<sup>3</sup>/y로 일반적인 이용량보다 훨씬 높게 산정되는 등 통계의 신뢰도가 낮은 것으로 보인다. 반면에, 본 연구의 회귀나무 모델로 추정된 가정용 지하수 이용량은 시군별 편차가 크지 않으며, 이는 충청남도 각 도시의 유형과 물 소비 패턴이 유사한 점을 고려할 때 보다 합리적이라고 판단된다. 일반용의 경우에도 각 시군별 편차가 크게 개선되어 나타남을 알 수 있다.

본 연구에서 개발된 회귀나무 모델에 의한 가정용 및 일반용 지하수 일 이용량 추정 방법은 읍면과 같이 일정 면적 내에서 백여 개 이상의 관정을 보유한 지역에서의 이용량 추정을 위한 적용시 신뢰성이 높은 것으로 판단된다. 다만, 지역내 일부 대용량 일

반용 관정이 존재하는 경우에는 이들 관정에서의 일 이용량에 대한 실측을 통하여 읍면별 총이용량의 오차를 줄여나가도록 해야 할 것이다.

## 8. 결 언

효율적인 지하수 개발 이용 및 보전 관리 정책을 수립, 추진하기 위해서는 지하수의 개발가능량과 이용량에 대한 정확한 산정이 선행되어야 한다. 우리나라의 지하수 개발가능량 산정 방법에 대한 연구는 지난 수십 년간 활발히 이루어져 왔으나, 지하수 이용량에 대해서는 체계적인 연구가 미흡한 실정이다. 또한, 정부에서 매년 발간하고 있는 지하수 이용량 통계 자료에 대해서도 정확성에 대한 논쟁이 지속되고 있다. 본 연구에서는 국토교통부 및 충청남도에서 지하수 이용량 실측 조사를 수행하였던 8,055개 지하수 관정 중에서 생활용의 가정용과 일반용으로 사용되는 1,273개 관정 자료를 활용하여 회귀나무 모델 기반의 지하수 이용량 추정 기법을 도출하였다.

가정용과 일반용 관정의 실측 일 이용량과 회귀나무 모델에 의한 추정 값은 상관계수가 0.5 미만으로 높지 않아 개별 관정의 이용량 추정에는 적절하지 않으나, 관정의 대표단위수(REC)를 평가한 결과 약 150개 이상의 관정에 대한 일 이용량 산정시에는 적용 가능한 것으로 나타났다. 또한, 본 회귀나무 모델에 의한 일 이용량 추정값은 기존의 국토교통부에서 제시한 추정식에 의한 값보다 개선된 것으로 나타났다. 한편, 일부 대용량 관정에 의하여 행정구역별 총 이용량이 영향을 받을 수 있으므로 본 연구에서 개발된 회귀나무 모델에 의한 방법과 병행하여 대용량 관정의 일 사용량에 대해서는 실측을 병행하여 행정구역별 이용량을 결정하는 것도 필요해 보인다.

아울러, 지하수의 이용 패턴은 경제 발전과 생활 환경의 변화 등으로 바뀌게 되므로, 주기적인 지하수 이용량 실측 조사를 병행하여 본 연구에서 제시된 모델을 개선해나가는 노력이 필요할 것이다.

## 감사의 글

본 연구는 환경부(한국환경산업기술원)의 수요대응형 물공급서비스 과제(상시 가뭄지역의 지하수 최적공급 관리를 위한 IoT 기반 인공함양 및 Well net-

work 기술 개발, #146523)의 지원 및 충청남도·한국농어촌공사의 “충청남도 지하수 실이용량 등 분석연구”의 일환으로 수행되었습니다.

## REFERENCES

- European Commission, 2019, [https://ec.europa.eu/eurostat/statistics-explained/index.php/Water\\_statistics#Water\\_uses](https://ec.europa.eu/eurostat/statistics-explained/index.php/Water_statistics#Water_uses) (July 10, 2019).
- Huang, A.C., Lee, T.Y., Lin, Y.C., Huang, C.F. and Shu, C.M., 2017, Factor analysis and estimation model of water consumption of government institutions in Taiwan. *Water*, 9, 492; doi:10.3390/w9070492.
- IBM, 2012, IBM SPSS Modeler 15; Modeling Nodes. IBM Corporation, 496 p.
- Kim, J.W., Jun, H.P., Lee, C.J., Kim, N.J. and Kim, G.B., 2013, Groundwater-use estimation method based on field monitoring data in South Korea. *The Journal of Engineering Geology*, 23, 1-10 (in Korean with English abstract).
- Martinez-Santos, P. and Martínez-Alfaro, P.E., 2010, Estimating groundwater withdrawal in areas of intensive agricultural pumping in central Spain. *Agricultural Water Management*, 98, 172-181.
- MOCT, 2013, Manual for Groundwater Use Estimation. Seoul, 36 p (in Korean).
- MOCT, 2015, Manual for Groundwater Development and Management Affairs. Seoul, 344 p (in Korean).
- MOCT, 2016, Long Term Master Plan of Water Resources (2001~2020), 3rd Revision, Sejong, 131 p (in Korean).
- MOCT, 2018, Annual Report of Groundwater. Sejong, 662 p (in Korean).
- MOCT and K-water, 2006, Report on the Long Term Master Plan of Water Resources (2006~2020). Seoul, 892 p (in Korean).
- Park, W.B., Koh, G.W., Kim, B.S., Moon, D.C. and Yang, S.K., 2004, Estimation of agricultural groundwater withdrawal and characteristics of groundwater use in Jeju island. *Proceedings of KoSSGE 2004 Spring Conference*, 110-113 (in Korean).
- Siebert, S. Burke, J., Faures, J.M., Frenken, K., Hoogeveen, J., Döll, P. and Portmann, F.T., 2010, Groundwater use for irrigation - a global inventory. *Hydrology and Earth System Sciences*, 14, 1863-1880.
- Thearling, K., 1995, An Overview of Data Mining at Dun and Bradstreet. DIG White Paper 95/01, Data Intelligence Group, Pilot Software, Cambridge, MA 02141.
- Woo, N.C., Joe, M.J. and Kim, N.J., 1994, Development of a predictive model for groundwater use. *The Journal of Engineering Geology*, 4, 297-309 (in Korean with English abstract).

Received : August 9, 2019

Revised : September 3, 2019

Accepted : September 5, 2019