

국가지하수 관측소 측정자료의 이상값 분석

이명재¹ · 이진용^{1*} · 김규범² · 원종호³

¹(주)지오그린21, ²한국수자원공사 국가지하수정보센터, ³한국수자원공사 지하수조사부

Analysis of Abnormal Values Obtained from National Groundwater Monitoring Stations

Myeong-Jae Yi¹ · Jin-Yong Lee^{1*} · Gyo-Bum Kim² · Jong-Ho Won³

¹GeoGreen21 Co., Ltd.

²National Groundwater Information Management & Service Center, KOWACO

³Groundwater Investigation Section, KOWACO

ABSTRACT

National groundwater monitoring stations have been managed throughout the country by Korea Water Resources Corporation (KOWACO) in order to monitor variations in quantity and quality of groundwater resources. A multi-sensor installed in each monitoring station well measures groundwater level, water temperature and electrical conductivity every six hours and the logged data are automatically transmitted to a host computer in KOWACO. Meanwhile despite regular station inspection and replacement of deteriorate or broken devices, abnormal values or outliers often occur due to intrinsic limitations of automatic monitoring and transmission. Thus prompt recognition and measures to these values are essentially required to reduce disturbance and missing period of the data. In this study, time and frequency of outlier occurrence were analyzed for the water level data obtained from national groundwater monitoring stations within the Han river basin in 2000. The analysis results indicated that the most prominent patterns of the outliers were rapid decline for water level, no variation for temperature and steep decline for electrical conductivity. This study provided a sample criterion for determining the outlier for each parameter.

Key words : National groundwater monitoring station, Water level, Water temperature, Electrical conductivity, Outlier

요 약 문

한국수자원공사는 전국을 대상으로 지하수의 수위 및 수질의 변동을 관측하고 국내 지하수자원의 보전 관리를 위한 기반 구축을 위하여 국가지하수 관측소를 설치·운영하고 있다. 자동 관측장비를 관측정 내에 설치하여 6시간 간격으로 지하수위, 온도, 전기전도도를 관측하고 있다. 주기적인 관측소 현장점검과 노후화된 혹은 고장난 관측장비 교체 및 수리 등의 작업에도 불구하고 자동 관측 및 전송의 내재적 한계로 인하여 부분적으로 이상값이 발생하고 있다. 그런데 이러한 이상값의 발생으로 인한 자료의 교란이나 결측 발생기간을 줄이기 위해서는 빠른 인지와 대처가 필요하다. 본 단보에서는 한강유역 국가 지하수 관측소 2000년도 지하수 관측 원자료에 대하여 이상값이 발생하는 시기와 빈도를 분석하였고 그 결과 지하수위는 급하강, 수온은 무변동 그리고 전기전도도는 급하강 이상이 가장 많이 발생하는 것으로 나타났다. 또한 각 해당항목에 대해 이상값 판단기준을 실례로 제시하였다.

주제어 : 국가 지하수관측소, 지하수위, 수온, 전기전도도, 이상값

*Corresponding author : hydrolee@empal.com

원고접수일 : 2004. 8. 5 게재승인일 : 2004. 1. 31

질의 및 토의 : 2005. 4. 30 까지

1. 서 론

국내 지하수자원의 보전 관리를 위하여 한국수자원공사는 1995년부터 국가지하수 관측소를 설치하여 운영하고 있다. 대부분의 관측소는 충적층 및 암반 관측정이 쌍으로 설치되어 있으며 지동 관측장비를 이용하여 일 4회 지하수위, 온도, 전기전도도 등을 관측하고 있다. 그리고 년 2회 지하수 수질분석을 수행하고 있다. 이후 지하수위 등 많은 관측자료들이 수집되었고 이들 자료들은 여러 연구에 활용되고 있다(김규범 등, 2000; 김규범 등, 2003; 문상기 등, 2002a, 문상기 등, 2002b; Moon *et al.*, 2004).

한국수자원공사는 국가지하수 관측소 현장 점검팀을 구성하여 주기적인 관측소 현장점검과 노후화된 혹은 고장난 관측장비 교체 및 수리 등의 작업을 하고 있으나, 자동 관측 및 전송의 내재적 한계로 인하여 부분적으로 이상값이나 결측값이 발생하고 있다. 그런데 이러한 이상값이나 결측값의 발생으로 인한 자료의 교란이나 결측 발생 기간을 줄이기 위해서는 빠른 인지와 대처가 필요하다. 이런 대책의 일환으로 한국수자원공사가 개발한 국가지하수 관측망 관리시스템에는 기존 관측자료에 대하여 이상값 및 결측값이 발생하는 시기와 빈도를 분석하여 경보를 생성하는 기능을 포함시켰다. 즉 실시간으로 전송되는 지하수 측정자료에 대해 자동으로 이상여부를 판단하게 된다.

관측자료에는 시스템 내·외적인 변화들에 의해서 야기된 비정상적인 관측값을 갖는데 이런 특이 관측값을 이상값(outlier)이라고 한다(Wei, 1990). 실제로 거의 모든 관측자료는 어떤 이유에서든지 이상값을 가지게 되는데, 이상값은 단순한 기록오류 이외에 시스템의 변화(자연적인 현상 또는 인위적인 변동)를 나타내는 인자로 사용할 수 있다. 이상값이 존재하게 되면 관측자료의 여러 가지 통계적 특징에 영향을 주며, 이를 해석하는데 있어 오류를 유발할 수 있으므로 이를 인지하는 것은 자료의 해석에 매우 중요하다. 특히 자동관측 시스템에서는 이상값에 대한 인위적인 판단 기회가 적으므로 관측기기에 문제가 생겼을 경우 장기적인 이상값이 발생할 수 있다.

이상값을 파악하는 방법으로 상자그림이나 분포곡선을 그려 특정한 범주에서 벗어나는 값을 찾아내거나 회귀분석을 통하여 이상값을 통계적으로 분류할 수 있다(Box *et al.*, 1994). 하지만 지하수위 관측자료와 같이 시간에 따라 연속적으로 측정된 시계열 자료의 경우는 상자그림을 통하여도 이상값이 나타나지 않을 수 있다. 예를 들어, 추세적인 특징을 갖거나 계절적으로 변동하는 시계열의 경우 전체 자료를 나타내었을 때 국지적인 이상값이 적절히 표

현되지 않을 수 있다. 이러한 시계열 자료에 대한 이상값을 찾기 위해서는 시계열 분석과 같은 적절한 방법을 이용해야 하는데, 관측값의 특징 및 관측기기의 특성 등과 연관하여 이상값 유무를 판단해야 한다.

본 기술자료에서는 국가지하수 관측망을 통해 획득한 수리자료들을 분석하여 이상값의 특징을 파악함으로써 향후 국가지하수관측소의 운영관리에 효율을 기하고자 하였다.

2. 이상값의 발생 요인

일반적인 이상값의 발생 요인으로는 잘못된 측정 및 오기 등이 있다. 관측소 주변에서의 지하수 이용과 같이 물리적으로 시스템에 변화가 생겼을 경우 지하수위에서 이상값이 발생할 수 있으며, 측정기기의 이상 및 자료 전송상의 오류 등도 원인으로 작용한다. 이러한 이상값의 발생 요인에 따라 나타나는 이상값의 형태에 차이-생기는데 예를 들어 어떤 시점에서 지하수위를 잘못 기재하였다면 그 시점에서만 이상값이 나타날 것이다. 반면 양수에 의하여 지하수 시스템이 교란되었다면 양수 시작부터 연속적으로 지하수위 측정값이 다른 형태로 나타나게 된다. 따라서 이상값이 발생하게 되는 과정을 분석하여 각각의 특성에 적합한 이상값을 분류할 필요가 있다. 하지만 이상값은 여러 가지 요인들이 혼재되어 나타나는 경우가 있으므로 각각의 요인을 정확히 분리하는 것은 용이하지 않다.

국가 지하수관측망 관측자료에서 나타나는 이상값의 주요 원인으로는 측정 센서부(지하수위, 온도, 전기전도도)의 오류 및 수질분석용 시료채취를 위한 양수작업 등이 있다. 측정 센서부의 감도가 나빠지는 것은 기기의 노후 이외에 센서부에 습기가 찼을 경우 발생하는데 현장점검팀이 기기보정을 시행하거나 센서를 예비품으로 교체하여 문제를 해결하고 있다. 양수작업과 같이 인위적인 행위로 인한 이상값은 발생하더라도 그 원인을 알 수 있으므로 정확한 작업기록만 가지고 있으면 문제는 쉽게 해결되지만, 센서의 문제로 인한 이상값의 경우 측정값이 서서히 변하는 경우가 많기 때문에 이상값 여부를 판단하기 어렵다. 이런 경우 이상값이 정상적인 관측값으로 기록·인식되기 때문에 지하수 자료를 이용한 각종 통계 및 해석에 오류를 발생시킬 수 있다.

3. 이상값의 이론적 분류 및 식별

이상값은 자료에 영향을 주는 특징에 따라서 여러 가지 형태로 구분되는데 다음과 같은 다섯 가지 형태로 나눌수

있다. 즉, Additive Outlier(AO), Innovative Outlier(IO), Outliers in long Patch(PO), Level-shift outlier(LS), Temporary(또는 Transient)-Change outlier(TC)가 그것이다 (Box and Jenkins, 1976; Chen and Liu, 1993). AO는 특정시점에서만 나타나는 이상값이며, 대부분 명확한 이유를 설명하기 어렵다. 외부적인 영향 즉 관측값을 잘못 읽는다거나 기록을 잘못하여 생기는 경우가 많아서 관측 이상값(observational outlier)이라고도 불린다. IO는 특정 기간동안 관측값이 영향을 받는 경우이고, TC는 IO의 특별한 예로 관측값의 영향이 지속적으로 감소하는 경우에 해당한다. IO나 TC의 경우, 어떤 물리적 또는 화학적인 기작에 의하여 시스템이 교란받았을 때 주로 발생한다. PO는 일정기간 동안 관측값이 일정 수준으로 변한 상태로 유지되다 정상으로 돌아오는 경우에 해당되는 이상값 들이고, LS는 관측값의 크기가 변한 상태로 지속되는 경우로 계측기기의 이상이 발생하였을 때 전형적으로 나타나는 이상값이다.

Fig. 1은 지하수 관측연보(건설교통부, 한국수자원공사, 2001)에 나타난 국가 지하수관측망 자료로 각 이상값의 종류에 따른 전형적인 패턴을 보여준다. Fig. 1(a)는 포항

신광관측소 암반층의 전기전도도로 전형적인 AO의 형태를 보여준다. Fig. 1(b)는 진천 진천관측소 암반층의 수온 그래프로 9월 이후 계단식으로 상승한 후 그 값이 유지되는 LS 이상값의 형태를 보여준다. Fig. 1(c)의 의령봉수 관측소의 지하수위 변화를 보면, 특정시점에서 지하수위가 감소하고 서서히 회복되는 형태를 반복하여 보여주고 있다. 이는 관측소 주변지역에서 인위적인 양수가 있었던 것으로 추정되는데, 이런 지하수위의 변화는 전형적인 TC 또는 IO의 예이다.

한편, Fig. 1(d)의 지하수위 변화 역시 TC 형태를 보여주는데, 이는 강우에 의한 자연적인 지하수위 변화 형태임을 알 수 있다. Figs. 1(c)와 1(d)에서 지하수 시스템에 어떤 기작이 작용하여 지하수위가 변동한 것에는 차이가 없지만 전자는 인위적인 양수에 의하여 수위가 하강하였고, 후자의 경우는 자연적인 강우에 의하여 지하수위가 상승하였다.

통계적인 의미에서 판단한다면, 두 경우 모두 이상값으로 취급해야 하지만 물리적인 의미에서 지하수위 변화를 평가할 때는 강우에 의한 변화를 이상값으로 취급하지는 않는다. 또한 주기적인 양수가 시행되고 양수에 의한 지

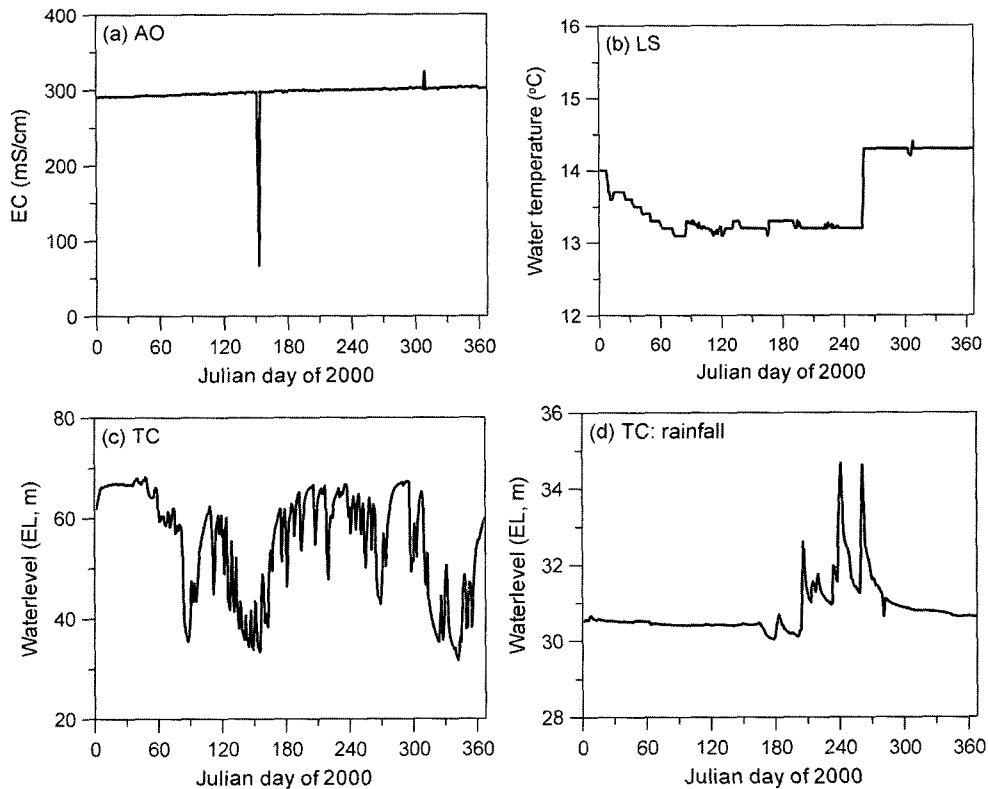


Fig. 1. Typical patterns of outliers identified from the monitoring data at the national groundwater monitoring stations: (a) Pohang-Singwang (bedrock) station, (b) Jincheon-Jincheon (bedrock) station, (c) Uiryeong-Bongsu (bedrock) station, (d) Yeosu-Yeosu (bedrock) station.

하수위 하강 폭을 알고, 그 폭 이내에서 지하수위가 변동할 경우 역시 정상적인 값으로 간주해야 한다. 즉 관측값에 대한 통계적 의미에서의 이상값 여부는 물론 각 국가 지하수관측소의 기존 변화 특성을 반영하여 이상값 여부를 결정할 필요가 있다.

4. 국가지하수 관측소 자료의 이상값 형태 분석

시계열의 통계적 특징을 이용하여 이상값을 찾기 위해서는 먼저 시계열 자료가 정상적(stationary)이어야 하며, 각 관측 시계열별로 시계열 모형이 설정되어야 한다. 또한 강우에 의한 수위변화와 같이 자연적인 현상 역시 이상값으로 분류되는 오류가 발생할 수 있으므로 시계열 분석이나 기타 통계적 특징을 이용하여 이상값을 찾는 것 또한 쉽지 않다. 각 관측소(관측점) 별로 측정된 시계열 자료를 개별적으로 분석하여 변동 경향 및 이상값 분포 형태를 경험적으로 판단하는 것이 필요하다. 이때 지하수위, 수온, 전기전도도의 변동 패턴이 각각 다르기 때문에 각 특징을 개별적으로 파악해야 한다.

이상값 여부를 판단하기 위해서는 현재 설치된 모든 관측소의 지하수위, 수온, 전기전도도 변동 경향을 기초로 각각 비교하는 것이 가장 좋은 방법이다. 우선 본 기술자료에서는 한강유역권에 속한 국가 지하수관측소 자료에 대하여 이상값으로 판단되는 경우를 분석하였다. 이상값으로 판단된 관측값들은 다시 몇 가지 경우로 나누어 공통된 판단근거를 마련하였다. 우선 이상값으로 판단할 수 있는 관측값으로 급격한 상승이나 하강을 보이는 경우를 선택하였다. 또한 일정기간 값이 변동하지 않는 경우나 극히 미약한 변화를 보이는 경우 역시 기기이상 등의 문제가 발생하였다고 판단하였다. 그러나 관측기기별 특성에 의해 변화가 서서히 일어나는 경우는 기준을 잡기 어렵고, 관측기기별로 이상발생 상황이 현재로서는 명확하지 않아 이상값 식별에서 제외하였다. 여기서는 수위, 수온, 전기전도도를 각각 따로 분석하였으며 다음에서 각 분석 항목별 이상값 판단기준의 사례를 제시하였다.

4.1. 지하수위의 이상값 식별

일반적으로 지하수위는 강우의 영향을 많이 받으며 강우에 따라 변동폭이 크게 다르다. 지하수위의 최고 및 최저치 등은 해마다 갱신되거나 월별 값들도 해마다 다르기 때문에 통계치를 이용한 이상값 판단은 신뢰도가 떨어지며 오히려 정상적인 수위를 이상값으로 진단할 가능성이 있다. 그러나 지하수위를 측정하는 센서는 수온이나 전기

전도도에 비해 안정되고 오류가 발생하는 경우가 적기 때문에 상대적으로 확실한 이상값이라고 판단할 수 있는 기준으로 관측값의 급하강, 급상승 및 무변동을 평가하였다.

4.1.1. 지하수위 급하강

앞서 언급하였듯이 지하수위는 주로 강우의 영향을 받는다. 지하수위를 강우와 비교해 보면 지하수위가 높은 때는 강우량이 많은 기간으로 강우량이 크면 지하수위가 급격히 상승하는 경우가 적지 않다. 강우영향으로 급격히 상승한 지하수위는 지수적으로 감소하기 때문에 초기에 비교적 급격한 강하곡선을 보이지만 이를 제외하고 지하수위가 자연적으로 급하강하는 경우는 드물다. 이외에 급하강하는 경우가 있는데 국가 지하수관측소 관리업 무중의 하나인 지하수 시료채취를 위한 양수작업이 진행되거나 관측소 주변지역에서 인위적인 양수가 발생한 경우이다. 연속적으로 기록되는 관측자료에 대하여 다음과 같은 기준으로 지하수위가 급하강한 경우의 이상값을 정의하였다.

① 급하강 이상값을 찾기 위해서는 해당시점 이전에 관측된 값이 10회 이상이어야 한다. 즉 10회의 관측시간은 다음과 같다(국가 지하수관측소는 매 6시간마다 수위, 수온, 전기전도도를 측정한다).

$$10\text{회 관측} = 10 \times 6(\text{시간/회}) = 60\text{시간} \quad (1a)$$

따라서 60시간 이상 연속적으로 관측된 이후에 이상값 여부를 판단한다.

② 선행 10개의 관측값을 이용하여 다음을 계산한다.

$$criteria_{down} = d_{max} - d_{min} + 1.0 \quad (1b)$$

여기서, d_{max} 는 $g_w(i) - g_w(i-1)$ 에서 최대값, d_{min} 는 $g_w(i) - g_w(i-1)$ 에서 최소값, $g_w(i)$ 는 i 번째 지하수위 그리고 $g_w(i-1)$ 는 $(i-1)$ 번째 지하수위이다.

즉 10개의 관측 지하수위에서 각각 바로 앞에 관측된 지하수위와의 변동폭을 계산하고 여기서 가장 높은 상승폭(+값)과 가장 작은 하강 폭(-값)과의 차이를 계산한다. 확실한 급하강을 찾는 기준을 위하여 두 값의 차이에 1.0 m를 더하였다. 이것이 급하강 여부를 판단하는 criteria down이다. 가장 높은 상승폭과 가장 낮은 하강폭과의 차이를 이용한 급하강 이상값 판단에서는 이전 60시간 동안 지하수위가 심하게 위아래로 변동하다가 수위가 크게 하강하는 경우는 제외된다. 즉 지하수위가 60시간 동안 소폭으로 변동하다가 갑자기 1.0 m 이상 하강하는 경우를 찾는 기준이다.

4.1.2. 지하수위 급상승

지하수위 급상승이 일어나는 경우는 강우에 의한 일시적인 상승이 대부분이다. 따라서 이러한 경우를 제외하고 지하수위가 급상승한 경우 이상값으로 판단해야 한다. 또한 센서이상으로 지하수위 상승이 나타날 수도 있으므로 보수적인 상황을 고려하여 이상값으로 판단하고 관측자료에 적용시켰다. 지하수위 상승으로 이상값이 발생할 경우는 해당 관측소 부근의 강수량을 검토하여 강수에 의한 지하수위 상승인지를 판단하여 불필요한 현장 점검을 줄일 필요가 있다. 지하수위 급하강의 경우, 10회 관측기간 동안(60시간) 지하수위가 소폭으로 진동하다 갑자기 1.0 m 이상으로 하강하였을 경우를 이상값으로 판단한 반면 급상승은 소폭으로 변동하다 급격히 상승하는 경우를 말한다. 지하수위 급하강의 경우와 같은 방법으로 다음과 같이 이상값을 식별한다.

- ① 선행된 관측값이 10회 이상이어야 한다.

$$10\text{회 관측} = 10 \times 6(\text{시간/회}) = 60\text{시간} \quad (2a)$$

- ② 선행 10개의 관측값을 이용하여 다음을 계산한다.

$$criteria_{up} = d_{max} - d_{min} + 1.0 \quad (2b)$$

10개의 관측 지하수위에서 각각 가장 높은 상승 폭(+값)과 가장 작은 하강폭(-값)과의 차이를 계산한다. 확실한 급상승을 찾기 위해 두 값의 차이에 1.0 m를 더한다. 이것이 급하강 여부를 판단하는 criteria up이 된다.

4.1.3. 지하수위 무변동

지하수위가 상승 혹은 하강이 없이 일정하게 유지된다면 수위센서 감도의 이상으로 수위 변동을 정확히 측정하지 못하는 경우일 수 있다. 일부 국가관측소에서는 월간 지하수위 변동폭이 1 cm가 되지 않는 곳이 있다. 따라서 자연변화 이외에 센서의 이상으로 수위변동이 없는 경우를 판단해야 한다. 지하수위 무변동의 이상값을 다음과 같이 정의하였다.

- ① 선행 지하수위 관측값이 40회 이상이어야 한다. 40회의 관측시간은 다음과 같다.

$$40\text{회 관측} = 40 \times 6(\text{시간/회}) = 240\text{시간} = 10\text{일} \quad (3a)$$

따라서 10일 동안 연속적으로 지하수위를 관측한 이후에 지하수위 무변동을 판단한다.

- ② 10일 동안 관측한 40회 관측된 지하수위 변화량을 구한다.

$$Level\ change = |g_w(i) - g_w(i-1)| \quad (3b)$$

여기서 $g_w(i) - g_w(i-1)$ 는 수위 급하강에서와 마찬가지로 이전 관측수위와의 변동폭을 나타낸다.

40개 변화량의 절대값을 계산하고 누적합을 구함으로써 10일 동안의 지하수위 변화량을 얻는다. 만약 10일 동안의 지하수위 변화량이 0.0이면 이상값으로 매우 명확하다.

이상과 같은 3가지 이상값 판단기준에 기초하여 2000년 한강유역 국가 지하수관측소 수위 원자료를 이용하여 수위 이상값을 분석하였다. Table 1은 이상값의 발생 횟수와 유형을 나타낸 것이다. 65개의 지하수위 이상값 중에서 가장 많이 나타나는 수위 이상값의 유형은 수위 급하강이다. 따라서 국가관측소 점검자 혹은 관리자는 관측소 주변의 양수에 의한 지하수의 급하강 여부를 판단하고 또한 적절한 시기에 수위 센서를 점검할 필요가 있다. 두 번째로 많이 나타나는 이상값의 유형은 수위 무변동이다. 그런데 수위 무변동은 영월상동 암반, 음성음성 암반, 제천고암 암반 관측소에서 집중적으로 발생하였다. 이것은 지하수위 센서에 이상이 나타나면 다시 교정되기 전까지는 계속 이상값을 관측하기 때문이다. 따라서 지하수위 무변동 이상값이 나타나면 해당관측소에 대한 신속한 점검이 필요하다.

4.2. 지하수 온도의 이상값 식별

수온은 변동폭이 가장 작고 안정된 항목이다. 그러나 센서 이상으로 이상값이 발생하는 경우가 있다. 수온의 이상값 유형을 지하수위와 유사하게 수온 급하강, 수온 무변동 및 수온 급변화 등 세 가지로 구분하였다. 수온 변동은 심하지 않고 완만하게 변동하는 특징이 있기 때문에 급하강이나 급변은 지하수위보다 비교적 쉽게 찾을 수 있지만 센서의 감도가 서서히 떨어져 이상값이 어느 시점부터 발생하는지는 판단하기 어렵다. 그리고 같은 이유에서 수온은 변동폭이 매우 작기 때문에 변동하지 않는 유형에 대해서 신중한 판단이 요구된다. 수온에 대한 이상값 정의는 다음과 같다.

4.2.1. 수온 급하강

한강 유역권에 설치된 관측소의 4~5년간 관측된 수온 자료를 살펴보면 수온의 변동폭이 크지 않으며 월별 변동폭은 각각 다르다. 따라서 수온의 변동폭을 구하기 위해서 한강유역에 있는 관측소의 수온 관측값을 월별로 구분하여 극값(최고값, 최저값)을 구하고, 최고값과 최저값의 차이를 계산하였다. 여기서 평상시의 수온은 최고값과 최저값의 사이에 있기 때문에 수온이 급변하는 변동폭은 최고값과 최저값 차이의 1/2로 삼았다. 따라서 한강유역 관

측소의 수온 극값 차이의 1/2을 월별로 계산하고 그 평균을 구하였다. Table 2는 한강유역 지하수 수온의 변동폭(극값 차이의 1/2)으로 7월에 가장 높은 수온 변동폭을 보여준다. 그리고 1월부터 12월까지 약 0.25°C 내외의 변동폭을 나타냈다. 따라서 수온이 급변할 수 있는 범위는 약 $\pm 0.25^\circ\text{C}$ 라고 판단되며 이를 기초로 수온 급하강을 다음과 같이 정의하였다.

① 수온 급하강 이상값을 찾기 위해서는 선행 관측값이 28회 이상이어야 한다. 28회의 관측기간은 다음과 같다.

$$28\text{회 관측} = 28\text{회} \times 6(\text{시간/회}) = 168\text{시간} = 7\text{일} \quad (4a)$$

따라서 7일 동안 6시간 간격으로 연속 관측한 이후에 이상값 여부를 판단한다.

② 선행 28개의 관측값에서 다음을 계산한다.

$$criteria_{down} = d_{max} - d_{min} + 0.25 \quad (4b)$$

여기서 d_{max} 는 $gw_{temp}(i) - gw_{temp}(i-1)$ 에서 최대값, d_{min} 는 $gw_{temp}(i) - gw_{temp}(i-1)$ 에서 최소값, $gw_{temp}(i)$ 는 i 번째 수온 그리고 $gw_{temp}(i-1)$ 는 $i-1$ 번째 수온을 나타낸다.

지하수위 급하강을 판단하는 것과 마찬가지로 28개의 관측된 수온에서 각각 선행 관측된 수온과의 변동폭을 계산한다. 여기서 가장 높은 상승 폭(+값)과 가장 작은 하강 폭(-값)과의 차이를 구한다. 그리고 보다 확실한 이상값을 판단하기 위해 앞서 언급한 $\pm 0.25^\circ\text{C}$ 를 반영하여 0.25를 더한다. 이것이 수온 급하강 여부를 판단하는 criteria down이다.

4.2.2. 수온의 무변동

수온은 많은 경우 연중내내 안정적이다. 그러나 수온이 1개월 이상 동안 0.01°C 의 변화도 없이 같은 값으로 관측되면 자연적인 현상으로 보기 어렵다. 더욱이 국가 지하수 관측소는 매 6시간마다 측정하기 때문에 새벽에 관측된 값과 한낮에 관측된 값에 차이가 있을 것으로 기대된다. 수온 무변동을 다음과 같이 판단한다.

① 선행 관측값이 180회 이상이어야 한다. 180회 관측 시간은 다음과 같다.

$$180\text{회 관측} = 180\text{회} \times 6(\text{시간/회}) = 1080\text{시간} = 45\text{일} \quad (5a)$$

즉 45일 동안 연속적으로 지하수온을 관측한 이후에 지하수온 무변동을 판단한다.

② 45일 동안 관측한 180회의 수온을 이용하여 변화량을 구한다.

$$Temp\ change = | gw_{temp}(i) - gw_{temp}(i-1) | \quad (5b)$$

여기서 $gw_{temp}(i) - gw_{temp}(i-1)$ 는 선행 관측 수온과의 변동폭을 나타낸다. 180개의 변화량의 절대값을 구하여 합산한다. 이렇게 구한 temp change가 0.0이면 수온 관측 센서의 이상으로 판단할 수 있다.

4.2.3. 지하수온 급변화

지하수온의 급변화는 앞서 언급한 지하수위의 급하강, 급상승 및 수온의 급하강과 비슷하다. 그러나 급하강 및 급상승은 선행 관측값이 소폭으로 변화하면서 급격히 하강 또는 상승하는 경우만을 의미한다. 수온의 급변화는 선행 관측값이 큰 폭이나 소폭으로 변동하는 것과 상관없이 일정량 이상으로 변화를 보이면 이상값으로 본다. 따라서 수위의 급하강이나 급상승보다 포괄적으로 이상값을 판단해야 한다. 특히 수온은 월별로 변동폭이 다르기 때문에 월별로 그리고 각 관측소마다 다른 기준을 설정할 필요가 있다.

① 기존의 수온 관측자료를 바탕으로 각 관측소마다 월별 수온 급변화의 기준을 정한다.

$$criteria\ change = \text{해당 월 최고값} - \text{해당 월 최저값} + 1.0 \quad (6)$$

여기서 기존 수년간의 관측자료를 이용하여 월별 최고값과 최저값의 차이를 계산하고 여기에 1.0을 더한 것이 criteria change가 된다. 1.0을 더한 것은 수온 급변화의 기준을 크게 설정하여 지나치게 많은 이상값 판정을 피하기 위함이다.

이상과 같은 3가지의 이상값 판단기준에 기초하여 2000년 수온 관측 원자료를 이용하여 수온 이상값의 발생 회수와 시기를 분석하였다. 수온의 급하강은 58개의 분석대상 관측소중 14개소에서 발생하였다. 그리고 이상값은 21번 발생하였다. 이것은 14개의 관측소에서 평균 1.5번 나타난다고 할 수 있다. 수온의 무변동 및 급변화는 각각 12개 관측소에서 59번 그리고 9개 관측소에서 19번의 이상값을 보였다. 반면에 3가지의 이상값이 한번도 발생하지 않은 관측소는 20개이다.

수온의 무변동은 수온 이상값 발생수의 60%를 차지하고 있다(Table 1). 여기서 수온 무변동의 이상값은 수위 무변동과 비슷하게 한번 발생하면 다시 교정되기 전까지 계속 이상값 발생하게 된다. 따라서 수온 무변동 이상값이 발생이 판단된 후 빠른 시간내에 점검이 실시된다면 이상값 발생수를 줄일 수 있다.

4.3. 전기전도도의 이상값 식별

전기전도도의 경우 지하수위나 수온과 달리 매우 큰 폭의 변화를 보여주며 변화유형도 관측소별로 크게 다르게

나타났다. 이러한 원인으로는 전기전도도를 측정하는 센서가 비교적 민감하고 다양한 주변 지하수 환경에 영향을 받기 때문으로 사료된다. 실제로 한국수자원공사가 발간하는 국가 지하수관측망 점검정비 보고서에서도 전기전도도 센서의 보정이 많은 부분을 차지하고 있다. 그러므로 전기전도도의 이상값 식별은 더욱 까다로운 문제이다. 또한 앞서 언급한 바와 같이 전기전도도의 물성값은 주변의 조건에 민감하게 반응하기 때문에 실제로 자연적인 전기전도도값에도 큰 차이를 보일 수 있다. 예를 들어 강우나 양수를 시행한 경우 관측점 내로 외부의 지하수가 유입됨으로써 전기전도도값이 상당히 변하는 양상을 보여준다. 전기전도도의 경우에도 이상값의 판단기준은 앞선 두 관측항목과 비슷하게 설정할 수 있다.

4.3.1. 전기전도도 급하강

① 선행 전기전도도 관측값이 28회 이상이어야 한다.

$$28\text{회 관측} = 28\text{회} \times 6(\text{시간/회}) = 168\text{시간} = 7\text{일} \quad (7a)$$

즉 168시간 이상 연속적으로 관측한 이후에 이상값 여부를 판단한다.

② 선행 28개의 전기전도도 관측값을 이용하여 다음을 계산한다.

$$criteria_{down} = 5 \times (d_{max} - d_{min} + 0.25) + 10.0 \quad (7b)$$

여기서 d_{max} 는 $gw_{EC}(i) - gw_{EC}(i-1)$ 에서 최고값, d_{min} 는 $gw_{EC}(i) - gw_{EC}(i-1)$ 에서 최저값, $gw_{EC}(i)$ 는 i 번째 전기전도도 그리고 $gw_{EC}(i-1)$ 는 $i-1$ 번째 전기전도도이다.

28개의 전기전도도에서 각각 바로 앞에 관측된 전기전도도와의 변동폭을 계산하고 여기서 가장 높은 상승 폭(+값)과 가장 작은 하강 폭(-값)과의 차이를 구한다. 전기전도도는 다소 무작위로 변하는 양상을 보일 수 있다. 따라서 지하수위 급하강의 이상값 판단 기준보다 더 크게 설정할 필요가 있다.

4.3.2. 전기전도도 무변동

전기전도도는 지하수위나 수온과는 다르게 변함이 없이 같은 값으로 유지되는 경우는 적으므로 이 경우 매우 확실한 이상값으로 판단할 수 있다. 수일 동안 1~3 $\mu\text{S/cm}$ 범위에서 변한 경우 이상값으로 판단할 수 있다. 이런 경우는 일주일 이상 같은 값으로 관측되면서 조금씩 변화하며, 대부분이 계단 모양의 변동형태를 보인다.

① 선행 전기전도도 관측값이 120회 이상이어야 한다.

$$120\text{회 관측} = 120\text{회} \times 6(\text{시간/회}) = 7200\text{시간} = 30\text{일} \quad (8a)$$

즉 30일 동안 연속적으로 관측한 후에 전기전도도 무변동을 판단한다.

② 30일 동안 관측한 120개의 전기전도도 변화량을 구한다.

$$EC_{change} = | gw_{EC}(i) - gw_{EC}(i-1) | \quad (8b)$$

여기서 $gw_{EC}(i) - gw_{EC}(i-1)$ 는 이전 관측 전기전도도와의 변동폭을 나타낸다. 그리고 120개 변화량의 절대값을 모두 합산한다. 만약에 30일 동안의 전기전도도 변화량이 5.0 $\mu\text{S/cm}$ 이하이면 이상값으로 판단한다.

4.3.3. 전기전도도 급상승

① 이전에 관측된 값이 28회 이상이어야 한다. 28회의 관측은 7일의 시간에 해당된다.

$$28\text{회 관측} = 28\text{회} \times 6(\text{시간/회}) = 168\text{시간} = 7\text{일} \quad (9a)$$

② 이전 28개의 관측값에서 다음을 계산한다.

$$Criteria_{up} = 5 \times (d_{max} - d_{min}) + 10.0 \quad (9b)$$

급하강처럼 전기전도도가 criteria up을 초과하게 되면 이상값으로 판단한다.

이상과 같은 세 가지의 이상값 판단기준에 기초하여 전기전도도 이상값 발생을 분석하였다. 전기전도도의 이상값은 89회가 발생하였는데 지하수위보다 많은 발생빈도를 보였다(Table 1). 전기전도도가 수위나 수온 이상값의 특성과 다른 점은 급하강 유형의 이상값이 가장 많이 발생하고 있다는 것이다. 지하수위 및 수온의 경우에는 무변동이 가장 많은 발생 빈도를 보인 반면 전기전도도는 무변동 유형이 가장 작은 부분을 차지한다. 그러나 지하수위와 수온처럼 전기전도도 무변동도 단지 4개의 관측소에서 19번의 무변동 이상값을 나타내었다. 무변동의 이상값을 보이는 경우에 센서의 점검이 조기에 이루어져야 한다.

Table 1. Numbers and types of outliers for water levels, temperatures, electrical conductivities obtained from national groundwater monitoring stations in the Han river basin in 2000

Parameters	Steep decline (SD)	No variation (NV)	Steep rise (SR)	Total
Water levels	27 (42%) ^a	23 (35%)	15 (23%)	65
Temperatures	21 (21%)	59 (60%)	19(19%) ^b	99
Electrical conductivity	42 (48%)	19 (21%)	28 (31%)	89

^aProportion

^bRapid variation

Table 2. Width of variation of monthly water temperature (half of difference between maximum and minimum values, unit: °C)

Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
0.20	0.16	0.19	0.19	0.24	0.23	0.27	0.23	0.24	0.20	0.20	0.17

5. 토 론

Table 3은 앞에서 설정한 수위, 수온, 전기전도도(EC) 이상값 판정 기준에 따라 발생하는 이상값의 유형을 관측소별로 정리하여 나타낸 것이다. 한강 유역에서 전체 이상값 발생 회수 253회에서 수위의 이상값을 나타내고

있는 것은 65회, 수온 및 전기전도도는 각각 99회 및 89회이다. 수온에서 가장 많은 이상값 발생빈도를 보이는 것은 수온의 무변동 유형 때문이다. 무변동의 이상값이 조기에 적절한 조치가 취해진다면 전반적인 이상값 발생회수는 줄어들 것으로 사료된다. 그러나 본 기술자료에서 이상값으로 판단된 253회 모두가 실제로 이상값으로 나타난 것은 아니다. 수위의 급하강은 주로 시료채취를 위해서 양수를 할 때 발생하였다. 전기전도도의 급하강이나 급상승도 장비 점검일에 큰 변화를 보였다. 실제로 국가 지하수관측망 관리자는 이상값 발생 판단을 할 때 해당 관측소의 점검일과 이상값 발생일을 비교하

Table 3. Summary of outliers obtained from monitoring data of national groundwater monitoring stations in the Han river basin for 2000. B: bedrock aquifer, A: alluvial aquifer

Monitoring stations		Water levels				Water temperature				Electrical conductivity				Total
		SD	NV	SR	Tot.	SD	NV	SR	Tot.	SD	NV	SR	Tot.	
Gapyung-Gapyung	B	0	0	0	0	0	0	1	1	1	0	0	1	2
Gapyung-Gapyung	A	0	0	0	0	0	0	1	1	1	0	2	3	4
Gapyung-Bukmyun	B	1	0	0	1	2	0	0	2	2	0	2	4	7
Kangnung-Hongje	B	0	0	0	0	0	3	0	3	0	0	0	0	3
Kangnung-Hongje	A	0	0	0	0	0	0	0	0	1	0	0	1	1
Gwangmyung-Cheolsan	B	0	0	0	0	0	0	0	0	0	0	0	0	0
Gwangmyung-Cheolsan	A	0	0	0	0	0	0	0	0	1	0	1	2	2
Gwangju-Gwangju	B	1	0	0	1	1	1	1	3	0	0	0	0	4
Goesan-Goesan	B	1	0	1	2	0	0	0	0	1	0	0	1	3
Gimpo-Gimpo	B	2	0	0	2	0	5	0	5	2	0	1	3	10
Gimpo-Gimpo	A	0	0	0	0	0	2	0	2	0	0	2	2	4
Danyang-Danyang	B	0	0	0	0	2	0	0	2	0	0	0	0	2
Dongducheon-Sangpae	B	2	0	0	2	0	3	0	3	0	0	0	0	5
Donghae-Gwiun	B	0	0	1	1	1	0	0	1	0	0	1	1	3
Donghae-Gwiun	A	0	0	0	0	0	0	0	0	0	0	4	4	4
Bucheon-Okgil	B	0	0	0	0	0	2	0	2	1	0	0	1	3
Bucheon-Okgil	A	0	0	0	0	0	0	0	0	0	0	1	1	1
Suwon-Omokcheon	B	0	0	0	0	0	0	1	1	1	0	2	3	4
Sihung-Gunja	B	2	0	0	2	0	0	0	0	0	0	1	1	3
Sihung-Gunja	A	0	0	0	0	0	1	0	1	0	0	0	0	1
Ansung-Samjuk	B	0	0	0	0	0	5	0	5	0	0	0	0	5
Ansung-Samjuk	A	0	0	0	0	0	2	0	2	0	0	0	0	2
Yanggu-Bangsang	B	0	0	0	0	0	0	0	0	2	0	0	2	2
Yangyang-Sonyang	B	2	0	0	2	0	0	0	0	0	0	0	0	2
Yangyang-Sonyang	A	0	0	0	0	0	0	0	0	0	0	1	1	1
Yangpyung-Gaegun	B	0	0	0	0	1	1	0	2	0	0	2	2	4
Yangpyung-Gaegun	A	0	0	1	1	0	0	0	0	1	0	0	1	2
Yangpyung-Yangdong	B	0	0	0	0	0	2	1	3	0	5	0	5	8
Yangpyung-Yangdong	A	0	0	0	0	1	0	0	1	0	0	1	1	2
Yangpyung-Yongmun	B	1	0	0	1	0	0	0	0	1	0	0	1	2

Table 3. Continued

Monitoring stations		Water levels				Water temperature				Electrical conductivity				Total
		SD	NV	SR	Tot.	SD	NV	SR	Tot.	SD	NV	SR	Tot.	
Yeosu-Yeosu	B	1	0	3	4	0	2	0	2	0	0	0	0	6
Youngwol-Sangdong	B	2	3	0	5	0	0	0	0	0	0	0	0	5
Yongin-Mapyung	B	0	0	1	1	2	1	1	4	2	1	0	3	8
Yongin-Mapyung	A	0	0	0	0	0	0	0	0	0	0	0	0	0
Wonju-Munmak	B	0	0	0	0	0	2	0	2	0	0	0	0	2
Wonju-Munmak	A	0	0	0	0	2	0	0	2	1	0	0	1	3
Eumsung-Saenguk	B	0	0	0	0	0	1	0	1	0	0	0	0	1
Eumsung-Saenguk	A	0	0	0	0	0	0	0	0	0	0	0	0	0
Eumsung-Eumsung	B	0	17	0	17	0	5	0	5	0	8	0	8	30
Icheon-Yulhyun	B	0	0	1	1	0	2	0	2	0	0	0	0	3
Icheon-Yulhyun	A	0	0	1	1	0	0	0	0	0	0	0	0	1
Jecheon-Goam	B	3	3	3	9	3	1	4	8	0	0	0	0	17
Jecheon-Goam	A	0	0	0	0	0	0	0	0	0	0	0	0	0
Cheonan-Sunger	B	2	0	0	2	0	4	0	4	0	0	0	0	6
Cheonan-Sunger	A	2	0	0	2	0	2	0	2	2	0	0	2	6
Cheongwon-Miwon	B	1	0	1	2	1	1	5	7	2	0	1	3	12
Chuncheon-Udoo	B	0	0	0	0	0	6	0	6	0	0	0	0	6
Chuncheon-Udoo	A	0	0	0	0	0	0	0	0	1	0	2	3	3
Chungju-Gagum	B	1	0	0	1	0	2	0	2	1	0	0	1	4
Chungju-Gagum	A	0	0	0	0	0	0	0	0	0	0	0	0	0
Chungju-Dongyang	B	0	0	0	0	0	0	0	0	6	0	1	7	7
Chungju-Dongyang	A	0	0	0	0	1	0	0	1	4	0	3	7	8
Paju-Beopwon	B	0	0	0	0	1	0	0	1	1	0	0	1	2
Pyungchang-Daehwa	B	0	0	1	1	0	1	0	1	1	0	0	1	3
Pyungtaek-Tongbok	B	0	0	0	0	0	2	0	2	1	0	0	1	3
Pyungtaek-Tongbok	A	0	0	0	0	1	0	0	1	1	0	0	1	2
Pocheon-Hwahyun	B	1	0	1	2	1	0	3	4	1	5	0	6	12
Hoengseong-Anhung	B	2	0	0	2	1	0	1	2	3	0	0	3	7
Incheon-Hajeom	B	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum					65				99				89	253

여 점검으로 인한 이상값 발생 여부를 평가해야 한다.

6. 결 론

한국수자원공사는 국가지하수 관측소를 통해 6시간 간격으로 지하수위, 온도, 전기전도도를 관측하고 있다. 주기적인 관측소 현장점검과 노후화된 혹은 고장난 관측장비 교체 및 수리 등의 작업에도 불구하고 자동 관측 및 전송의 내재적 한계로 인하여 부분적으로 이상값이 발생하고 있다. 그런데 이러한 이상값의 발생으로 인한 자료의 교란이나 결측 발생기간을 줄이기 위해서는 빠른 인지와 대처가 필요하다.

본 기술자료에서는 한강유역 국가 지하수 관측소 2000년도 지하수 관측 원자료에 대하여 이상값의 발생을 분석

하였다. 그 결과 지하수위는 급하강, 수온은 무변동 그리고 전기전도도는 급하강 이상이 가장 많이 발생하는 것으로 나타났다. 또한 각 해당항목에 대해 이상값 판단기준을 실효로 제시하였다. 실시간으로 관측되고 전송되는 원 자료들을 본 자료에서 제시한 방법으로 이상값 여부를 신속히 판단하여 조치를 취한다면 보다 신뢰성 있는 지하수 관측자료를 획득할 수 있을 것이다. 그러나 향후 실제 운용을 통해 여기서 제안된 이상값 판단기준의 적절성에 대한 지속적인 검증과 수정이 요구된다.

사 사

본 기술자료를 위해 좋은 수정의견을 주신 두 분의 심사위원님께 감사의 말씀을 드립니다.

참 고 문 헌

- 건설교통부, 한국수자원공사, 2001, 지하수관측연보, 한국수자원공사, 대전, p. 1073.
- 김규범, 최영진, 유영권, 류정아, 손영철, 2000, 국가 지하수 관측소의 장기관측자료에 의한 지하수 변동 특성, 2000년 한국지하수토양환경학회 추계학술발표회, 한국지하수토양환경학회, 포항공대, p. 36-39.
- 김규범, 김정우, 원종호, 이명재, 이진용, 이강근, 2003, 국가 지하수 관측소의 장기수위관측자료를 활용한 관측주기 결정 연구, 2003년 한국지하수토양환경학회 추계학술발표회, 한국지하수토양환경학회, 제주대, p. 199-201.
- 문상기, 우남철, 이광식, 2002a, 지하수위 유형과 유역별 지하수 함양률의 관련성 연구 : 국가 지하수 관측망 자료의 분석, 지하수토양환경, 7(3), p. 45-59.
- 문상기, 우남철, 한원식, 2002b, 국가 지하수 관측망 자료를 이용한 총적층 지하수 함양률의 공간적 변동성 연구, 2002년 한국지하수토양환경학회 총회 및 춘계학술발표회, 한국지하수토양환경학회, 서울시립대, p. 237-242.
- Box, G.E.P. and Jenkins, G.M., 1976, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, p. 531.
- Box, G.E.P., Jenkins, G.M., and Reinsel, G.C., 1994, *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, p. 598.
- Chen, C. and Liu, L., 1993, Joint estimation of model parameters and outlier effects in time series, *J. Amer. Statistical Assoc.*, **74**, p. 427-431.
- Moon, S.K., Woo, N.C., and Lee, K.S., 2004, Statistical analysis of hydrographs and water-table fluctuation to estimate groundwater recharge, *J. Hydrol.*, **292**, 198-209.
- Wei, W.W.S., 1990, *Time Series Analysis: Univariate and Multivariate Analysis*, Addison-Wesley Publishing Company, Inc., California, p. 478.