



스캔통계량 분석을 통한 상수도 누수 및 수질 민원 발생 클러스터 탐색

Cluster exploration of water pipe leak and complaints surveillance using a spatio-temporal statistical analysis

이주원 · 김은주 · 남숙현 · 황태문*

Juwon Lee · Eunju Kim · Sookhyun Nam · Tae-Mun Hwang*

한국건설기술연구원 환경연구본부

Korea Institute of Civil Engineering and Building Technology, The Department of Environmental Research

ABSTRACT

In light of recent social concerns related to issues such as water supply pipe deterioration leading to problems like leaks and degraded water quality, the significance of maintenance efforts to enhance water source quality and ensure a stable water supply has grown substantially. In this study, scan statistic was applied to analyze water quality complaints and water leakage accidents from 2015 to 2021 to present a reasonable method to identify areas requiring improvement in water management. SaTScan, a spatio-temporal statistical analysis program, and ArcGIS were used for spatial information analysis, and clusters with high relative risk (RR) were determined using the maximum log-likelihood ratio, relative risk, and Monte Carlo hypothesis test for I city, the target area. Specifically, in the case of water quality complaints, the analysis results were compared by distinguishing cases occurring before and after the onset of "red water." The period between 2015 and 2019 revealed that preceding the occurrence of red water, the leak cluster at location L2 posed a significantly higher risk (RR: 2.45) than other regions. As for water quality complaints, cluster C2 exhibited a notably elevated RR (RR: 2.21) and appeared concentrated in areas D and S, respectively. On the other hand, post-red water incidents of water quality complaints were predominantly concentrated in area S. The analysis found that the locations

Received 14 August 2023, revised 19 September 2023, accepted 19 September 2023.

*Corresponding author: Tae-Mun Hwang (E-mail: taemun@kict.re.kr; Fax: 82-31-910-0291, Tel. 82-31-910-0741)

이주원 (UST 학생연구원) / Juwon Lee (UST Student Researcher)

경기도 고양시 일산서구 고양대로 283, 10223
283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 10223, Republic of Korea

김은주 (선임연구위원) / Eunju Kim (Senior Research Fellow)

경기도 고양시 일산서구 고양대로 283, 10223
283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 10223, Republic of Korea

남숙현 (선임연구위원) / Sookhyun Nam (Senior Research Fellow)

경기도 고양시 일산서구 고양대로 283, 10223
283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 10223, Republic of Korea

황태문 (선임연구위원) / Tae-Mun Hwang (Senior Research Fellow)

경기도 고양시 일산서구 고양대로 283, 10223
283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 10223, Republic of Korea

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

pp. 253-259

pp. 261-269

pp. 271-279

pp. 281-288

of complaint clusters were similar to those of red water incidents. Of these, cluster C7 exhibited a substantial RR of 4.58, signifying more than a twofold increase compared to pre-incident levels. A kernel density map analysis was performed using GIS to identify priority areas for waterworks management based on the central location of clusters and complaint cluster RR data.

Key words: Leakage, Clustering, GIS, Space-time scan statistic, Water quality complaints

주제어: 누수, 밀도맵, 수질 민원, 스캔 통계학, 위험반경

1. 서 론

최근 몇 년간 정수장에서 생산한 물이 소비자에게 도달하는 과정에서 가압장 운영자의 오조작, 배급수 계통 관로 노후화 및 부식에 따른 누수 및 오염물 침투 등 다양한 수질 사고가 발생하고 있으며, 이에 따른 시민들의 불안과 민원이 지속적으로 발생하고 있다 (Lalle et al., 2021; Kang et al., 2013). 최근 I시의 상수도에서 적수 및 유출 발견 등 수질 사고가 발생한 바 있는데, 이러한 사회적 이슈를 사전 예방하기 위한 방법은 스마트 상수도 관리를 통한 상수도 서비스 개선을 위해 다양한 정책 사업이 추진되고 있다. 상수도 수질 사고는 사전예방 기술이 무엇보다 중요하다. 이를 위해서는 상수 공급과정에서 발생하는 수질 관련 각종 사고 및 민원을 통계적으로 분석하고 이력 관리를 체계화 및 표준화하는 것이 필요하다. 따라서, 사고 발생에 영향을 줄 수 있는 요인을 미리 파악하여, 사고우려 지역에 대한 사전예방 차원의 중점관리 선정 및 관리가 중요하므로, 이를 위해 지리정보시스템 기반의 클러스터 분석 기법 적용이 요구된다. 최근 스캔 통계학은 지역별 비교를 통해 사건 발생 빈도가 높은 고위험 지역을 선정하여 위험반경을 추정하는 방법으로 널리 사용되고 있다.

스캔 통계학은 대상 공간의 클러스터를 탐색하여 핫스팟 지역을 찾는 통계 기법이다 (Han and Lee, 2016). 관측된 영역의 원을 스캐닝 윈도우(Window)로 가정하고 윈도우의 크기와 위치를 바꿔가며 대상 지역을 탐색 후, 원의 내부 및 외부 위험도의 상대적 비율(Relative Risk; RR)을 계산하여 최대우도비가 발생하는 군집을 찾는 방법이다 (Kulldorff, 1997). 위 과정을 통해 다른 지역에 비해서 비교적 높은 위험률을 보이는 클러스터를 찾아낼 수 있다. 시공간 자료를 SaTScan v10.1.2를 활용하여 분석함으로써 위험 지역의 탐지 및 반경 추정이 가능하다. SaTScan은 공간 분

포의 상관관계를 시공간적으로 분석하여 위험 클러스터를 결정할 수 있는 프로그램이며(<http://www.satscan.org/>), Kulldorff(2010)에 의해 개발되었다. 이 프로그램은 클러스터를 통계적으로 정량화할 수 있기 때문에 일반적으로 질병 발생 위험이 큰 지역의 조기 감지를 위해 널리 사용된다 (Desjardins et al., 2018; Whiteman et al., 2019). 또한, GIS(Geographic information system)에서 데이터를 규칙적이거나 무작위적인 패턴으로 공간상에 나타낼 수 있다. 이렇게 표현된 패턴은 시각적으로 판단하기 어려울 수 있는데, 본 연구에서는 공간 패턴을 빠르고 쉽게 파악할 수 있는 커널밀도추정(kernel density estimation; KDE) 방법을 적용한다 (Milic et al., 2019). KDE는 입력된 현상의 양을 공간에 퍼뜨리며 이때 확산은 반지름의 매개 변수에 의해 제어된다. 커널이 각 포인트에 배치되고 추가되어 영향 영역의 중첩을 기반으로 현상의 발생 밀도를 추정한다. 커널의 크기는 포인트의 가중치, 커널의 모양 및 반경에 따라 변화하며 값이 클수록 부드럽고 일반화된 밀도맵(Density map)이 생성된다 (Herzog and Yopez, 2010; Oliveira et al., 2015). 커널밀도추정 방법은 공간 데이터의 패턴을 시각화하고 사고지역의 핫스팟을 정의하는데 이용된다 (Harirforoush and ellalite, 2019; Milic et al., 2019; Srokanth, 2019).

본 연구에서는 2015년부터 2021년까지 국내 I시에서 발생한 상수도 수질 사고 및 민원 데이터 정보를 스캔 통계학 방법론으로 분석하여 상수도 수질 중점관리 지역을 추정하고자 하였다. 해당 과정에서 적수 사태 전후 수질 민원 발생 위치의 변화를 비교하기 위하여 사고 이전(2015-2019년) 발생된 민원 데이터와 이후 발생된 민원 데이터를 나누어 분석하여 통계적으로 유의한 상대적 리스크(Relative risk)가 높은 클러스터의 위치 변화를 비교하였다. 이후 분석된 클러스터 중심지점의 위치와 상대 위험도 값을 데이터의 공간 분포패턴을 이용한 분석결과를 지도위에 시각화하



고 관리가 우선시되는 지역을 제시하였다.

2. 연구방법

2.1 데이터 수집 및 처리

본 연구의 대상지역은 최근 적수 등 사고 사례가 발생했던 I시로 하였다. I시의 상수관망 시스템의 경우 총 7개의 정수장을 중심으로 6개 계통으로 분류되며 지역 전체를 대상으로 하루 약 108만 톤의 용수를 생산 및 공급하고 있다. 일반적으로 상수도 안전에서, 물리역학적 측면에서의 수리 안전사고는 파손된 관로에서 발생하는 누수를 고려한다. 따라서 2015년부터 2019년까지 대상 지역에서 발생한 누수 데이터와 수질 민원 데이터를 분석에 적용하였으며 사용된 모든 데이터는 행정동을 기준으로 블록코드 기반으로 정리하였다. Table 1은 해당 데이터의 발생 건수와 위·경도를 포함한 자료이며 지역의 위치정보는 주소를 지리적 좌표로 변환하는 오픈소스 웹인 지오서비스웹(Geoservice-web)을 통해 변환하였다. SaTScan 분석시 필요한 I시의 인구데이터는 웹사이트(DATA.co.kr)에서 크롤링하여 사용하였다.

2.2 푸아송 시공간 스캔 통계

본 연구에서는 SaTScan에서 구현된 푸아송 시공간 스캔 통계(Space-time scan statistic(STSS))를 사용하여 수질 사고 사례 클러스터를 탐색하였다. 클러스터는 모든 가능한 지점으로부터 시작되는 소구간 내의 최

대 관측치와 구간의 관측치 수를 비교하여 결정된다. 위 과정에서 다른 지역에 비해서 비교적 높은 위험률을 보이는 클러스터를 찾아내며, 관측치는 일반적으로 공변량을 고려하여 푸아송 분포를 따른다고 가정된다 (Kulldorff, 1997). STSS는 가설검정을 기본 원리로 이루어지며 귀무가설과 대립가설 사이에서 우도(Likelihood)를 계산하고 이 둘을 비교하는 우도비 검정(Likelihood ratio; LR)을 이용하여 통계적으로 유의한 클러스터를 찾는다. 우도비는 어떤 사건이 특정 속성에서 일어나지 않을 확률에 대한 일어날 확률의 비로, 정량적으로 그 차이를 부각한다는 장점이 있다 (Chung, 1998). 본 연구에서는 윈도우 내의 관측 케이스 수보다 위험도가 높은 예측(Expected case) 케이스 수가 더 많은 클러스터를 찾는 것이 목적이며($I(q_z > p_z)$), SaTScan의 경우에는 우도비에 로그를 취한 로그우도비(Log likelihood ratio; LLR)가 적용된다. 모든 발생 가능한 윈도우의 집합을 Z , Z 의 임의 원소를 z (행정구역), q_z 와 p_z 를 윈도우 z 내의 실제 관측된 케이스 수와 예측 케이스 수라고 하고, 연구 영역내의 모든 관측된 케이스의 합과 모든 예측 케이스의 합은 식 (1)과 같다.

$$q = \sum_z q_z, p = \sum_z p_z \quad (1)$$

푸아송 모델을 가정할 경우 z 에 대한 우도비 LR는 식 (2) 및 식 (3)과 같다. Z 내의 모든 윈도우에 대해 로그우도비의 최대값 $T(\text{Max}_z \text{ LLR}(z))$ 를 구하면, 그 때의 윈도우가 가장 유의한 주요 클러스터(Primary

Table 1. Number of cases in regions with latitude and longitude (Study period: 2015-2019)

District	Population	Leaks	Complaints	Longitude	Latitude
B	488,579	7,842	501	126.72191	37.50784
D	60,023	1,660	85	126.64324	37.47402
G	59,968	3,474	50	126.48784	37.74693
K	223,999	4,368	243	126.73774	37.53771
J	92,773	3,692	197	129.33282	35.56971
NU	443,626	6,634	507	126.73097	37.44971
N	330,375	7,549	54	126.65030	37.46370
O	6,608	333	3	126.46085	37.25569
S	401,134	7,195	580	126.67596	37.54546
Y	377,669	2,248	291	126.67827	37.41038

cluster)가 된다 (Han and Lee, 2016).

$$LR(z) = \left\{ \left(\frac{q_z}{p_z} \right)^{q_z} - \left(\frac{q - q_z}{p - p_z} \right)^{q - q_z} \right\} I(q_z > p_z) \quad (2)$$

$$LLR(z) = \log(LR(z)) \quad (3)$$

여기서, LR : 우도비(Likelihood)

LLR : 로그 우도비(Log likelihood ratio)

q_z, p_z : Z 지역내 실제 관측된 케이스 수 및 예측된 케이스 수

이후 몬테카를로(Monte Carlo) 접근법으로 유의성을 검정한다, 몬테카를로법은 임의 지점을 원의 중심으로 고정 후 반경의 크기를 변화시켜가며 사고 발생률이 통계적으로 유의하게 높게 계산된 원의 반경을 위험 반경으로 결정하는 방식이다. 몬테카를로 결과의 p-value는 모든 연구영역에서 위험 정도가 같다는 가정 하에 수많은 데이터 셋을 임의로 생성 후, 각 데이터 셋에 대한 T를 계산하여 실제 데이터로부터 계산된 검정 통계량 값의 순위를 확인하여 유의성을 판별한다(식 (4)). SaTScan은 몬테카를로 시뮬레이션(Monte Carlo simulation)을 통해 가설의 유의성 검정이 이루어진다. 실제 데이터와 같은 전체 관측치의 개수 및 사건으로부터 발생된 관측치의 개수를 가지는 공간맵을 랜덤하게 생성하여 T 통계량을 가지는 클러스터가 분석된다. 몬테카를로 시뮬레이션을 통해 생성된 데이터셋에서 계산된 검정통계량 값과 비교했을 때 실제 데이터에서 계산된 값이 상위 5%에 든다면 이는 유의하며, SaTScan은 보통 999번의 시뮬레이션 과정을 거치기 때문에 계산된 값이 상위 50% 안에 든다면 해당 클러스터가 유의하다고 판단한다 (Hohl et al., 2020).

$$p = \frac{r}{1+m} \quad (4)$$

여기서, r 은 계산된 테스트 통계 순위이며, m 은 시뮬레이션 수를 나타낸다.

2.3 공간패턴 시각화

KDE 방법은 오랫동안 사고 핫스팟을 식별하기 위해 널리 적용되었다. 식 (5)에 의해 공간상 지점에 의한 강도가 계산된다 (Xia and Yan, 2008).

$$f(s) = \sum_{i=1}^n \frac{1}{\pi r^2} k\left(\frac{d_{is}}{r}\right) \quad (5)$$

여기서, $f(s)$ 는 s 지점에 대한 강도이며, r 은 검색반경이며, k 는 커널 함수, d_{is} 는 s 와 i^{th} 사이의 거리이다. KDE는 ArcGIS Pro(version 3.1)(ESRI, Redlands, CA, USA)를 사용하여 표현하였다.

3. 결 과

3.1 시공간 클러스터

Table 2는 2015년부터 2019년까지 5년간 발생한 누수와 수질 민원의 전체 데이터를 SaTScan을 통해 분석하여 도출된 상위 5개 고위험 클러스터의 속성값이며, Table 3은 2019년도 발생한 적수 사태 이후 발생한 수질 민원 데이터를 SaTScan으로 분석한 결과이다. 실제 자료로부터 계산된 로그우도비(LLR), 실제 관측된 사고건수(Observed), 각 지역의 사고 발생율이 동일하다는 가정에서 계산된 사고건수(Expected)를 나타낸다. 상대 위험도(RR)는 관측치를 클러스터 내부의 예측값(Expected value)로 나눈 값과 클러스터 외부의 예측값의 비율로 나타낼 수 있으며(식 (6)), 각 클러스터에 대한 상대 위험도를 나타낸다 (Xu and Beard, 2021).

$$RR = \frac{c/e}{(C-c)/(C-e)} \quad (6)$$

여기서, RR은 상대 위험도, c 는 데이터 세트의 총 케이스 수, C 는 클러스터에서 관찰된 총 사례 수, e 는 대상 지역의 인구로 계산된 예상 사례 수× C 를 총 인구수로 나눈 값이다.

Fig. 1 및 2는 각각 SaTScan으로 년도별 누수 데이터와 2015년부터 2019년까지 5년 동안 발생한 누수 데이터와 민원 데이터의 위성 클러스터 맵(Cluster



map) 분석 결과를 나타낸 것이다. 분석된 클러스터가 클수록 클러스터 자체의 상대 위험도는 낮으며, 크기가 작아질수록 상대 위험도가 증가한다고 볼 수 있다. 클러스터의 크기의 경우 윈도우 내의 인구수가 연구 영역의 전체 인구수에서 차지하는 비율로 정의되며, 이때 윈도우의 최대 허용 크기는 50% 보다 작아야 한다. 본 연구에서는 사고발생 확률이 높은 중점지역을 찾는 것이 목표이므로, 최대 클러스터 반경 허용 범위를 5 km로 제한하였다. 몬테카를로 검정법 결과, 모든 p-value의 분석값이 0.05 보다 작은 값을 나타내었으며, 해당 클러스터의 발생률은 통계적으로 유의한 것으로 판단되었다. Table 2에 정리된 누수클러스터(L1~L5)와 민원클러스터(C1~C5) 5개를 대상으로 고위험 클러스터를 선택하였다. 누수클러스터 중 가장 위험한 클러스터는 D 지역에 포함되는 클러스터 L2로 LLR 값이 1618.89, 상대 위험도가 2.45이며 두 번째로 위험한 클러스터는 L1으로 D 지역에 포함되며, LLR 값이 663.77, 상대 위험도가 2.25이다. 민원클러스터의 경우 가장 위험한 클러스터는 S 지역에 위치하는 클러스터 C2로 LLR 값이 88.97, 상대 위험도가 2.21이었

으며, 두 번째로 위험한 클러스터는 J 지역에 위치한 C3로 LLR 값은 69.97, 상대 위험도는 4.64이다.

I시의 적수 사태는 2019년도에 발생하였으며, 따라서 사고 이후 민원클러스터 발생 위치의 추이를 비교하였다. Fig. 3은 사고 이후 발생한 수질 민원 데이터의 위성 클러스터 맵 분석 결과이다. 적수 피해 지역은 I시 S 지역, J 지역 Y 지역, K 지역 일대였으며, 주요 사고 발생지역인 S지역에 클러스터가 집중된 것을 확인할 수 있다. 발생된 고위험 클러스터 C6~C10의 경우 사고 이전과 비교했을 때(Table 2, 3), LLR과 상대 위험도 값이 증가하는 것을 확인하였다. 특히 C7의 경우 LLR이 252.52이며, 상대 위험도는 4.57로, 사고 이전과 비교하여 두 배 이상 증가하는 것으로 분석되었으며, 해당 지역의 경우 다른 지역보다 집중적인 감시가 이루어져야 할 것으로 판단된다.

3.2 커널밀도추정

본 연구에서는 다양한 방법론 중 실제 값과 예측 값 사이의 편차가 크지 않은 안정적인 방법을 적용하

Table 2. Attributes of prospective space-time clusters for total water leak and complaints at the district level

Accident	Cluster	Radius (km)	LLR	Observed/Expected	p-value	RR	Population at risk	Longitude	Latitude
Complaints	C1	3.10	21.48	1.35	<0.00000011	1.42	327790	126.73373	37.45323
	C2	5.35	88.97	2.01	<1E-17	2.21	197242	126.69093	37.58733
	C3	0.45	69.43	4.5	<1E-17	4.64	20401	126.63544	37.46577
	C4	0.00	42.78	8.83	<3.3E-16	8.93	3654	126.37645	37.45530
	C5	0.00	14.52	2.03	<0.000067	2.06	34182	126.73059	37.53019
Water leak	L1	0.00	663.77	2.18	<1E-17	2.25	67623	126.68532	37.45839
	L2	3.98	1618.89	2.27	<1E-17	2.45	138385	126.62308	37.48586
	L3	0.00	100.02	15.71	<1E-17	15.73	191	126.70371	37.57343
	L4	0.00	47.68	1.37	<1E-17	1.38	41657	126.71246	37.51363
	L5	0.00	35.09	1.30	<8.2E-13	1.30	46904	126.66485	37.50639

Table 3. Attributes of prospective space-time clusters of 2021 water quality complaints at the district level

Accident	Cluster	Radius (km)	LLR	Observed/Expected	p-value	RR	Population at risk	Longitude	Latitude
Complaints	C6	0.00	17.48	2.18	<3.6E-7	2.22	41657	126.71246	37.51363
	C7	5.35	252.52	3.87	<1E-17	4.57	122429	126.64195	37.61584
	C8	1.49	43.14	2.37	<2.2E-16	2.48	77114	126.66485	37.50639
	C9	0.00	37.12	2.87	<4.6E-15	2.96	39515	126.67925	37.54136

여 측정되지 않은 미지 공간에서 발생 가능한 누수와 수질 민원 발생 빈도를 추정하고 등고선이 표출된 레

스터(Raster) 결과를 도출하였다. Fig. 4는 2015년부터 2019년까지 발생한 수질 민원과 누수데이터의 전체

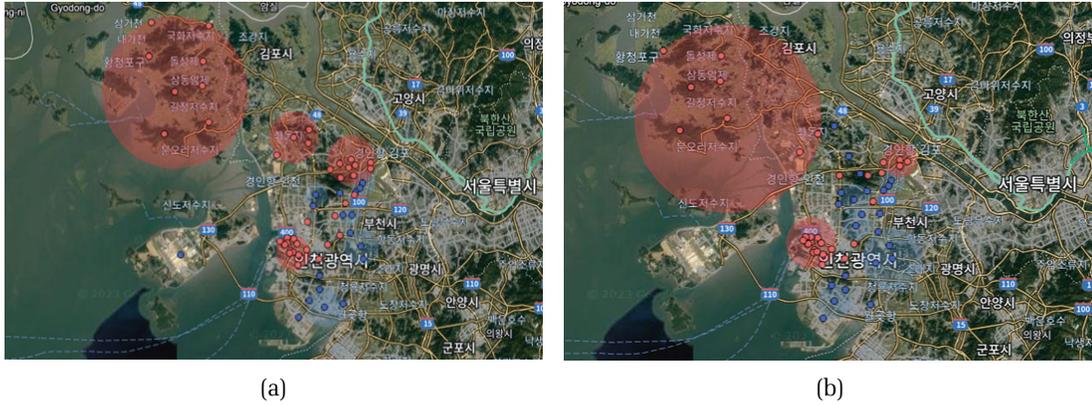


Fig. 1. Location of significant clusters by SaTScan analysis; (a) 2015, (b) 2019.

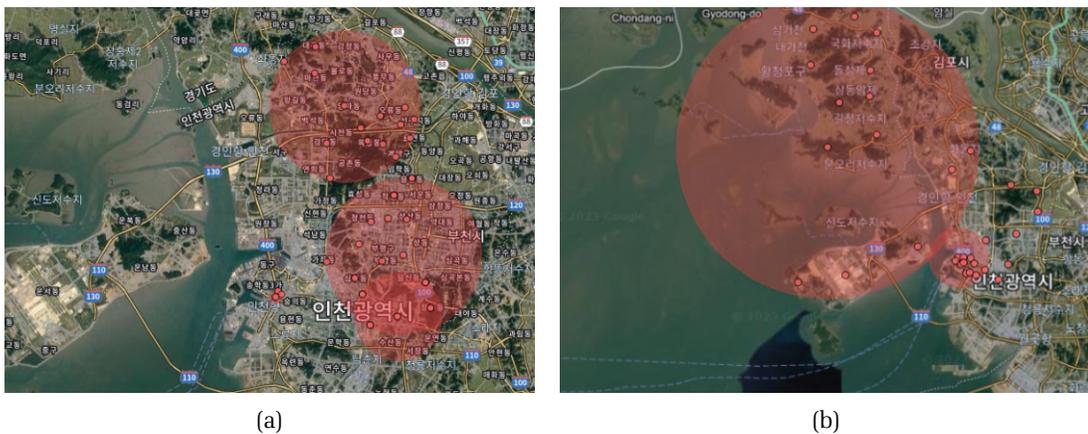


Fig. 2. Location of significant clusters by SaTScan analysis; (a) Complaints, (b) Leaks.

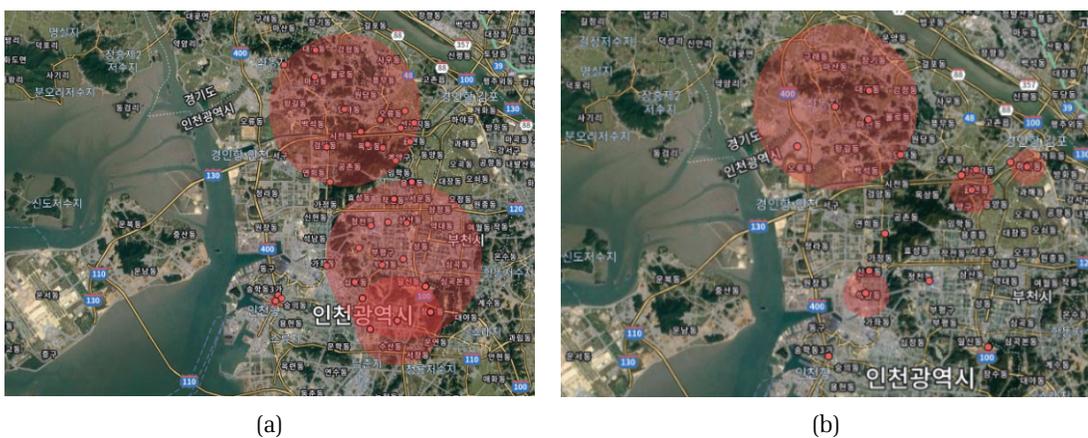


Fig. 3. Comparison of significant water quality complaints clusters location changes through SaTScan analysis; (a) 2015-2019, (b) 2021.



값을 적용하여 도출된 클러스터의 중심위치를 기반으로 수행된 커널밀도추정 분석 결과이며, I시 전역에서 발생 가능한 민원과 누수 사고의 상대 위험도 예측값을 나타낸다. 적색에 가까울수록 상대적으로 상대 위험도가 높은 지역을 의미한다. 여기서, 상대 위험도는 대상지역의 실제 발생한 수질 민원 및 누수 사고에 대한 실제 건수를 푸아송 통계분포로 추정된 추정값

에 대한 상대적 비율을 의미하는데, 상대 위험도가 수치로 표출됨으로 상수도 수질 민원 및 누수 사고에 대한 우선관리 위치를 결정하는데 중요한 지표로 활용할 수 있다. 누수의 경우 G지역에서 가장 높은 상대 위험도 값을 나타냈으며 J 지역, S 지역, D 지역 일부 지역에서 높은 값을 보이는 것을 확인하였다. 수질 민원의 경우 J 지역의 서쪽 지역에서 가장 높은 값을 나

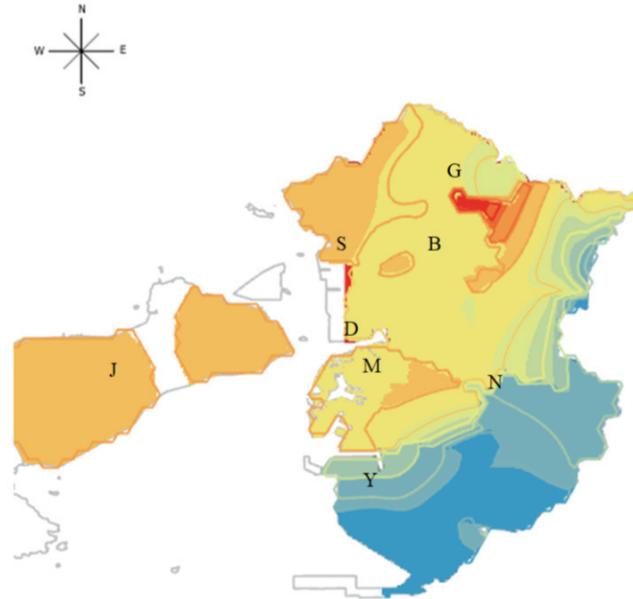


Fig. 4. Map based on water leak accidents hotspot.

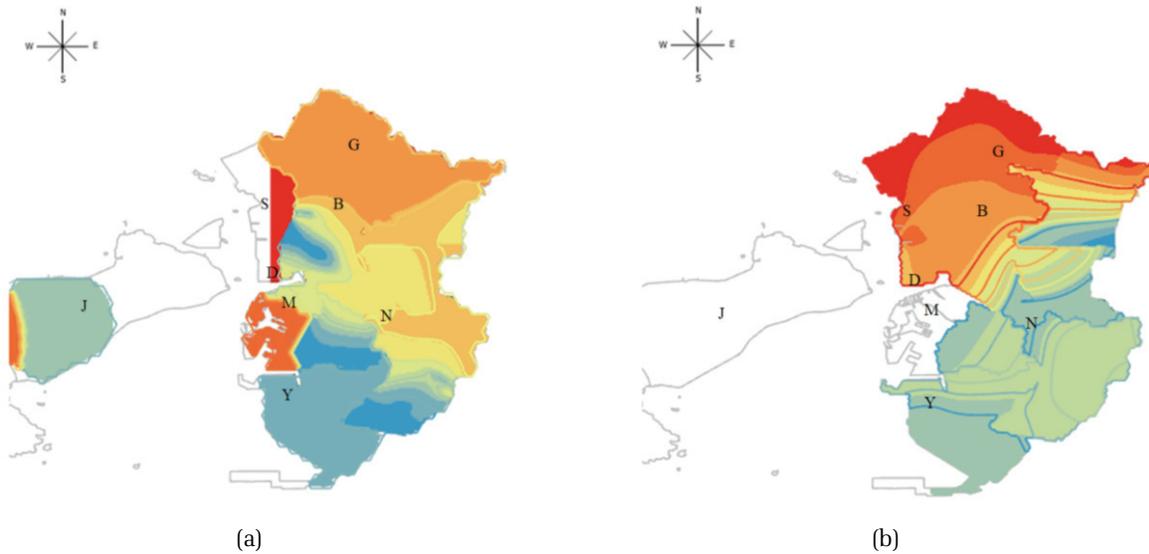


Fig. 5. Map of changes in the relative risk of water quality complaints before and after the red water accidents (a) 2015-2019, (b) 2019-2021.

pp. 253-259

pp. 261-269

pp. 271-279

pp. 281-288

타냈으며 D지역, S 지역과 G 지역 북쪽 일부 지역 순으로 높게 분석되었다. 예측값의 유의성 순위는 상이 하나 누수와 민원데이터 분석 결과 G 지역, S 지역, J 지역, D 지역이 공통 지역으로 나타나 해당 지역을 수질 사고의 주요 감시 대상으로 선택하였다. Fig. 5는 적수 사태 이전과 이후 발생한 수질 민원클러스터의 중심위치를 사용하여 보간하였을 때 상대 위험도의 변화를 비교한 것이다. 사고 이후, 주요 적수 피해 지역인 S 지역에 높은 상대 위험도 값이 밀집된 것으로 분석되었다. SaTScan는 지역 크기와 인구수에 따라 통계 분석을 통해 유의한 시공간 클러스터를 도출하며 이는 GIS의 클러스터링 방법의 수치적 한계를 보완하는 방법으로 적용가능하다. 수도법 제21조에 근거하여 수도사업자는 상수도관망 중점관리지역을 선정하고 정기적인 수질측정 등을 실시하고 관망관리 개선계획을 수립하여야 한다. 본 연구에서 제안한 SaTScan-GIS을 활용한 지리적 통계 분석방법은 상수도관망의 노후 등으로 수질오염이 발생하거나 발생할 우려가 있는 지역을 선정함에 있어 보다 객관적인 평가도구로 활용 가능할 것으로 판단된다.

4. 결 론

본 연구에서는 시공간데이터에서 비교적 위험률이 높은 클러스터를 찾는 방법인 스캔 통계량을 먹는물 공급과정에서 발생하는 수질 사고 분석에 적용하였다. 2015년부터 2021년까지 I시에서 발생한 누수 사고와 수질 민원 발생 수를 분석하여 주요 감시 지역을 제안하였으며 2019년 적수 사고 전후의 민원발생 추이를 비교하여 고위험 클러스터의 발생 분포를 비교하였다.

SaTScan에서 구현된 푸아송 시공간 스캔 통계를 통해 우도비를 이용한 몬테카를로법으로 검정 결과, 유의한 고위험 클러스터는 2015~2019년에는 D 지역, S 지역, J 지역에 위치하는 것으로 분석되었으며 적수 사태 이후에는 사고 주요 발생지인 S 지역에 밀집된 것으로 분석되었다. 특히 사고 이후 분석된 클러스터의 경우 사고 이전보다 높은 상대 위험도 값을 나타내어, 해당 지역의 경우 사고 예방과 대응을 위한 감시체계를 보다 완벽하게 수립해야 할 필요가 있을 것으로 판단하였다.

분석된 누수와 민원클러스터의 중심위치와 상대 위험도 값을 ArcGIS에 적용하였으며 커널밀도추정을 통해 사고발생 위험도를 지도에 가시화하였다. 적수 사태 이전에는 G 지역, S 지역, J 지역, D 지역에 높은 상대 위험도 값이 분포하였으나, 사태 이후 주요 사고 발생지인 S 지역이 가장 위험도가 높은 것으로 분석되었다.

본 연구에서 적용한 상수관망에 대한 스캔 통계학을 활용한 연구사례는 많지 않으며, 추후 지역 범위와 데이터의 확장을 통해 다양한 적용이 가능할 것으로 예상된다. SaTScan의 장점은 인구수를 고려하여 통계적으로 유의한 시공간 클러스터를 도출하여 해당 지역에 대한 로그우도비와 상대 위험도 등을 산출하여 이를 구글어스, GIS 등과 같은 지리정보 시스템에 해당 정보를 표출해주는 것이다.

특히 상대 위험도가 수치로 표출됨으로 상수도 수질 민원 및 누수사고에 대한 상수도 중점관리지역을 결정하는데 객관적 지표로 활용할 수 있으며, 향후 상수도 수질 사고 예방을 위한 감시 방법으로 유용할 것으로 기대된다.

사 사

본 연구는 환경부의 재원으로 한국환경산업기술원의 상수도 혁신 기술개발사업의 지원을 받아 연구되었습니다(2020002700004).

References

Chung, C.F. (1998). "Three Bayesian prediction models for landslide hazard", in: *Proceedings of International Association for Mathematical Geology 1998 Annual Meeting*.

Desjardins, M.R., Whiteman, A., Casas, I., and Delmelle, E. (2018). Space-time clusters and co-occurrence of chikungunya and dengue fever in Colombia from 2015 to 2016, *Acta Trop.*, 185, 77 - 85.

Han, J., and Lee, M. (2016). Cancer cluster detection using scan statistic, *J. Korean Data Inf. Sci. Soc.*, 27, 1193 - 1201.

Harirforoush, H., Bellalite, L. (2019). A new integrated GIS-based ana-lysis to detect hotspots: A case study of the city of Sherbrooke, *Acc. Anal. Prev.*, 130:62 - 74.



- Herzog, I., Yépez, A. (2010). Least-cost kernel density estimation and interpolation-based density analysis applied to survey data, in: Fusion of Cultures, *CAA 2010 -Proceedings of the 38th Conference*, 6-9 April, 2010, Granada, Spain, 367 - 374.
- Hohl, A., Delmelle, E.M., Desjardins, M.R., and Lan, Y. (2020). Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States, *Spat.*, 34, 100354.
- Ji, B., and Chung, D. (2017). Inference about Hazardous Clusters of Varicella Incidence in Korea: Using SaTScan and ArcGIS, *J. Korean Data Inf. Sci. Soc.*, 19, 229 - 238.
- Kang, D., Kim, T.W., and Ahn, J.H. (2013). Water Resources Infrastructure: Sustainability and Resilience, *J. Korean Soc. Hazard Mitig.*, 2, 13, 309 - 316.
- Khouni, I., Louhichi, and G., Ghrabi, A. (2021). Use of GIS based Inverse Distance Weighted interpolation to assess surface water quality: Case of Wadi El Bey, Tunisia, *Environ. Technol.*, 24, 101892.
- Kulldorff, M. (1997). A spatial scan statistic, *Commun. Stat. Theory Methods.*, 26, 1481 - 1496.
- Lalle, Y., Fourati, M., Fourati, L.C., and Barraca, J.P. (2021). Communication technologies for Smart Water Grid applications: Overview, opportunities, and research directions, *Comput. Netw.*, 190, 107940.
- Li, J., and Heap, A.D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors, *Ecol. Inform.*, 6, 228 - 241.
- Mühlenstädt, T., and Kuhnt, S. (2011). Kernel interpolation, *Comput. Stat. Data Anal.*, 55, 2962 - 2974.
- Oliveira, U., Brescovit, A.D., and Santos, A.J. (2015). Delimiting Areas of Endemism through Kernel Interpolation, *PLOS ONE*, 10, e0116673.
- Milic, N., Popovic, B., Mijalkovic, S., Marinkovic, D. (2019). The influence of data classification methods on predictive accuracy of kernel density estimation hotspot maps, *Int. Arab. J. Inf. Technol.*, 16(6), 1053-1062.
- Srokanth, L., and Srokanth, I. (2020). A case study on kernel density estimation and hotspot analysis methods in traffic safety management, *IEEE Xplore*, 19432010.
- Whiteman, A., Desjardins, M.R., Eskildsen, G.A., and Loaiza, J.R. (2019). Detecting space-time clusters of dengue fever in Panama after adjusting for vector surveillance data, *PLoS Negl Trop Dis.*, 13, e0007266.
- Xia, Z., and Yan, J. (2008). Kernel density estimation of traffic accidents in a network space, *Comput. Environ, Urban, Syst.*, 32(5):396 - 406.
- Zhang, H., Pu, R., and Liu, X. (2016). A new image processing procedure integrating PCI-RPC and ArcGIS-spline tools to improve the orthorectification accuracy of high-resolution satellite imagery, *Remote. Sens.*, 8, 827.

pp. 253-259

pp. 261-269

pp. 271-279

pp. 281-288