

상수관로 누수위치 자료를 이용한 계층적 군집분석

Hierarchical Clustering Analysis of Water Main Leak Location Data

박수완* / 임광채** / 최창록*** / 김규리****
Park, Suwan / Im, Gwangchae / Choi, Changlok / Kim, Kyulee

Abstract

Rehabilitation projects for old water mains typically require considerable capital investments. One of the economical ways of pursuing the rehabilitation projects is to focus on a specific area within the entire region under management. In this paper the hierarchical clustering methods that analyze spatial inter-relationship of location data are applied to about 8,000 water leak location data recorded in a case study area from 1992 to 1997. Among the hierarchical clustering methods Single, Complete, and Average Linkage Methods are used to identify clusters of the water leak locations and to divide the area according to the defined clusters. By comparing the clusters identified by the clustering methods, the best clustering method for the case study area is suggested. Prioritization of the area for maintenance is obtained based on the water leak incident intensity for the clustered area using the suggested best clustering method.

keywords : old water main, rehabilitation, prioritization, water leak location, clustering, maintenance

요 지

노후수도관 개량사업은 예산상, 시공상 등의 여러 제약조건에 의해서 장기적인 계획 하에 시행되게 된다. 본 연구에서는 연구대상지역에서 1992년부터 1997년 사이에 기록된 누수 위치좌표 약 8,000개를 이용하여 누수 위치들 간의 공간적 상관관계에 대한 계층적 군집분석을 수행한다. 계층적 군집분석방법 중 최단 연결법, 최장 연결법 및 평균 연결법을 적용하여 연구대상지역을 누수위치의 공간적 상관관계에 따라 분할하였으며, 각 군집 방법 별로 분할된 구역들을 비교하여 연구대상지역에 가장 적절한 군집 분석방법을 제시한다. 제시된 최적의 군집분석 방법을 이용하여 연구대상지역을 누수 위치들을 군집으로 분할한 후 군집으로 분할된 각 구역의 단위면적당 누수건수를 산정하고 이에 따라서 분할된 구역들에 대한 상수관망 유지관리 우선순위를 결정한다.

핵심용어 : 노후수도관, 개량, 누수위치, 군집분석, 유지관리, 우선순위

* 교신저자, 부산대학교 사회환경시스템공학부 조교수
Corresponding Author, Assistance Prof., Dept. of Civil Eng, Pusan National Univ, Busan 609-735, Korea
(e-mail: swanpark@pusan.ac.kr)
** 부산대학교 사회환경시스템공학부 · 석사과정
Master's Course, Dept. of Civil Eng., Pusan National Univ, Busan 609-735, Korea
*** 부산대학교 사회환경시스템공학부 · 석사과정
Master's Course, Dept. of Civil Eng., Pusan National Univ, Busan 609-735, Korea
**** 부산대학교 사회환경시스템공학부 · 석사과정
Master's Course, Dept. of Civil Eng., Pusan National Univ, Busan 609-735, Korea

1. 서론

1.1 연구배경 및 필요성

현대적인 상수도 시설이 도입된 이래 약 100년의 역사를 가진 한국의 상수도는 2006년 12월말 276개 기초자치단체에서 전체인구의 91.3%인 45,303천명에게 상수도를 공급하고 있으며, 특히 7개 특별시와 광역시의 수도 보급률이 평균 98%를 나타내며, 1975년 전체 급수율 72.9%(14,089천명)에 비해서 많은 양적인 성장을 하였다(환경부, 2006). 그러나 환경부의 2006년 상수도통계에 따르면 2006년도 1년간 우리나라에서 생산 및 공급한 수도물 총량은 5,753백만 m^3 으로 이 가운데 누수량 등을 제외한 실제 유효수량은 4,933백만 m^3 이며, 수도요금이 부과된 양(유수수량)은 4,613백만 m^3 이며, 유수율은 80.2%로 나타나고 있다. 또한 실제 총 누수량으로 추정되는 14.2%(819백만 m^3)는 전체 부산광역시민이 1년에 사용하는 유효수량이 약 363백만 m^3 이라고 본다면, 총 누수량은 부산광역시민이 약 2.3(전국 총 누수량/부산광역시 유효수량)년간 사용할 수 있는 양과 동일하며, 이를 금액으로 환산할 경우 연간 5,769억 원(2006년 평균 생산원가 기준, 704.4원/ m^3)의 막대한 수도물 생산 비용 및 수자원 손실이 일어나고 있다. 그 대부분은 배, 급수관망의 누수에서 기인한 것으로 보고되고 있다. 상수관망 유지관리 사업은 여러 제약조건에 때문에 장기적인 계획 하에서 시행되어 진다. 이러한 사업의 효율적인 시행을 위해서는 상수관망의 누수사고 등의 실제자료를 통해서 상수관망 우선관리 대상지역을 결정하는 것이 필요하다. 현재 우리나라의 노후관 개량사업은 정량적인 방법이 아닌 근무자의 과거 경험을 바탕으로 상수관망 교체구역이 결정되어 지고 있는 것이 현실이다. 상수관로의 노후도를 밝히는 연구는 20세기 중반부터 꾸준히 진행되어 왔다. 그러나 상수관망 유지관리지역을 선별하는 연구는 많지 않은 실정이다. 따라서 본 연구는 상수관망의 누수 위치좌표를 이용해서 전체 누수발생지역중 상수관망 누수위치가 타 구역과는 다르게 근접해있는 구역을 선별하고 각 구역별 상대적인 누수발생우열을 가려서 상수관망 우선관리대상지역을 결정하는 방법론을 제시하고자 한다.

1.2 연구내용 및 구성

본 연구에서는 미국의 연구대상지역에서 1992년부터 1997년 사이에 기록된 상수관망 누수위치좌표 8,000개를 대상으로 계층적 군집분석(Hierarchical Cluster Analysis, HCA)을 실시하고 도출된 누수위치군집을 이

용해서 군집들 간의 경계를 설정하였다. 그리고 분할된 구역면적(Partitioned Region Area, PRA)과 분할된 구역면적 내의 총 누수건수(Total Number of Leak, TNL)를 이용해서 각 구역들의 단위면적당 발생한 누수건수인 총 누수강도(Total Leak Intensity, TLI)를 분석하였다. 마지막으로 총 누수강도를 이용해서 상수관망의 효율적 유지관리를 위한 우선관리대상지역을 결정하였다.

Bailey and Gatrell (1995)에 의해서 제안된 강도(Event Intensity)분석법을 이용하면 연구대상지역을 위도 0.5°와 경도 0.5°로 이루어진 정사각형지역으로 정의하고, 이를 여러 개의 동일한 크기의 셀(Cell)로 분할하여 단위 셀 면적당 존재하는 누수사건의 수를 이용해서 누수강도(Leak Intensity)를 측정할 수 있다. Fig. 1은 연구대상지역을 Bailey and Gatrell (1995)의 사건 강도 분석법에 따라서 2,500 개의 셀로 분할하고, 각 셀 내에서 발생한 누수강도 크기를 색깔로 나타내고 있다.

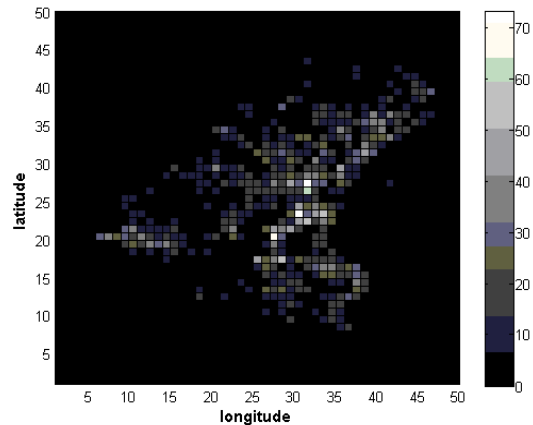


Fig. 1. Water Leak Intensity of the Case Study Region

Fig. 1의 셀 색깔은 강도가 클수록 흰색에 가깝고, 누수건수가 작을수록 검정색에 가깝다. 유기적으로 연결된 상수관망매설지역을 사실적으로 분할하기 위해서는 셀과 주변의 셀이 병합되어야 한다. 그런데 셀의 누수강도 정보만으로는 병합할 셀의 지목이 어렵다. 또한 셀의 크기가 상대적으로 커지게 되면 사각형 모양의 셀이 누수가 없는 구역까지 포함할 수 있게 되고, 그 셀을 통해서 분석된 누수강도의 신뢰성은 떨어지게 된다. 그러나 HCA(Hierarchical Cluster Analysis)을 이용할 경우에는 누수위치좌표를 유사성 또는 비유사성 분석을 통해서 군집으로 분할할 수 있고 분할된 누수위치좌표를 이용해서 실제누수위치를 반영한 지역분할을 할 수 있게 된다. Fig. 2는 연구과정을 나타내는 연구 흐름도이다. 본 연구의 구성은 크게 세 가지 과정으로 나눌 수

있다. 첫 번째, HCA를 이용해서 유사성 또는 비유사성을 가지는 누수위치좌표들을 군집으로 분할한다. 두 번째, 통계적으로 적절한 군집개수(Number of Cluster, NCL)를 찾는다. 세 번째, 군집 분할 프로그램인 HCE (Hierarchical Clustering Explorer)와 선정된 군집개수를 이용해서 누수위치좌표를 그림으로 나타낸다. 그리고 분할된 각 누수위치군집에 적절한 경계선을 설정해서 군집이 차지하는 구역의 면적을 산정한다. 마지막으로 분할된 구역의 총 누수강도에 대한 분석을 통해서 누수가 상대적으로 많이 발생하는 지역을 파악하고 각 구역별로 상수관망 유지관리 우선순위를 결정한다.

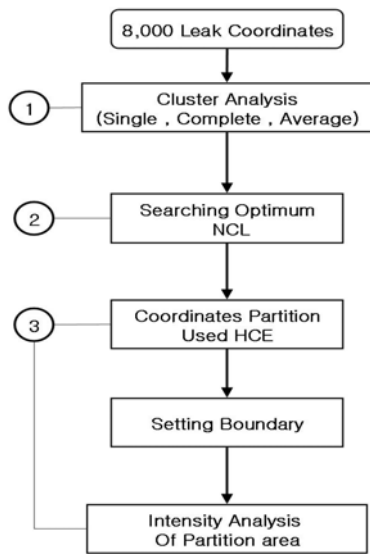


Fig. 2. Flow Chart of the Methodology

1.3 연구 동향

상수관망의 누수 방지 또는 저감을 포함한 상수관망의 효율적 유지관리를 위한 방법론에 대한 연구는 물리적, 통계적 그리고 지리정보시스템(Geographic Information System, GIS)을 이용한 방법론으로 구분 지을 수 있다. 상수관망의 유지관리에 관망의 물리적 특성을 주로 고려한 대표적인 연구는 다음과 같다. Rajani and McDonald (1993)는 결빙지역에 매설되어 있는 상수관망이 받는 압력을 연구하였고, Spangler (1941)는 상수관망과 토질의 관계를 연구하였으며 토양내부에 있는 상수관망이 받는 압력을 분석하는 방법을 개발하였다. Kumar *et al.* (1984)은 토질에 대한 주철, 덕타일, 아연, 관의 부식 지수를 제안하였고, Doleac (1979)은 지하에 있는 상수관망의 부식도를 상수관망의 매설기간과 토질의 성질을 이용해서 지수 식을 제안하였다. Rossman and Grayman (1998)은 수리학적 요소들이 상수도물 저장조의 수질에 미치는 영향을 연구하였다. 상수관망의

유지관리에 통계학적 방법을 적용한 대표적인 연구는 다음과 같다. Kettler and Goulter (1985)은 상수관망의 누수건수와 관망의 수명 사이에 선형 회귀 식을 제안하였고, Andreou *et al.* (1987)은 보다 발전된 비례위험모델을 연구해서 파손건수가 추가 될수록 다음파손의 시점을 앞당긴다는 것을 발견하였다. Randall *et al.*(1992)은 상수관망의 부식도는 일정하게 증가한다는 가정 하에 선형회귀식을 제안하였고, Kleiner and Rajani (2001)은 관의 개량기준에 있어서 최소비용으로 수행 가능한 개량시기를 제시하였다. Rajani and Marker (2000)은 부식된 관의 강도를 이용해서 상수관의 나머지 사용 가능한 기간을 제안하였고, 박수완 (2007)은 상수관망 최적 교체시기를 결정하기 위한 방법론을 바탕으로 관망 교체 우선순위와 관망 교체 정책이 상수관망 시스템에 미치는 영향을 분석하였다. 상수관망의 유지관리에 GIS를 적용한 대표적인 연구는 다음과 같다. 최성규와 안병찬 (2001)은 지형공간정보와 EPANET을 연계해서 상수도 데이터베이스와 상수관망 모델의 입, 출력 관계를 구현 하였고, 문 호(2001)는 상수관망 사고 지점의 시설물 정보를 체계화해서 화재 또는 재해발생에 신속히 대처하는 방법을 제시하였다. 전환돈 등 (2007)은 도시 배수시스템에 필요한 펌프의 효율적인 운영에 관한 연구를 하였고, 박종일 (2002)은 복잡하고 방대한 실제 관망에서의 발생하는 누수의 실제 위치를 밝혀내고자 하였다. 오창주 등(2001)은 상수관망 파손이 일어날 경우 해당 관과 함께 격리되는 주변 관들을 밝히고 이에 따른 연구대상지역의 제수 밸브의 위치와 개수를 구하였고, Jun and Loganathan (2007)은 상수관망이 파괴되었을 경우 상수관망 시스템의 효율적인 관망 격리를 통해서 경제적 손실을 최소화하는 제수밸브 탐색 알고리즘을 제안하였다.

2. 이론 및 연구방법

2.1 HCA(Hierarchical Clustering Analysis) 이론

군집분석(Cluster Analysis)이란 각 사건(대상)간의 유사성을 측정해서 유사성이 높은 집단을 분류하고, 군집에 속한 사건들의 유사성과 서로 다른 군집에 속한 사건간의 상이성을 규명하는 통계분석방법이다. 그리고 군집이란 모임을 구성하고 있는 구성원간의 상호 유사성 또는 관련성을 중요시하는 개념이다. 이러한 군집분석에 관한 연구로서 Mojena (1977)는 군집화 과정에서 유도되는 상사성계수를 이용한 두 가지의 정지규칙(Stopping Rule)을 제안하였고, Beale (1969)는 모집단 분포가 다변량 정규분포를 이룬다는 가정 하에서 F통

계량을 제안하였다. Calinski and Harabasz (1970)는 공분산행렬에 대한 트레이스의 비를 위하여 유사 F통계량을 제시하였고, Wolfe (1970)는 혼합정규분포의 가정 하에 보완된 우도비 검정을 제안하였다. Marriot (1971)는 균등분포를 이룬다는 가정 하에 공분산행렬에 대한 행렬식 값의 비에 대한 표본분포를 몬테칼로 방법으로 얻었고, Sarle (1983)는 초직사각형 상에 균등 분포하는 초입방체와 같은 형태의 군집으로 분할될 것이라는 가정 하에 군집개수를 결정하였다. Wong and Schaack (1982)은 k-최인접 방법에 있어서 군집의 개수를 결정하기 위한 방대한 시뮬레이션 연구를 수행하였고, Seo and Shneiderman (2006)은 시각적 군집분석 프로그램을 개발하였다.

본 연구에서 이용하는 계층적 군집(Hierarchical Cluster)이란 한 군집이 다른 군집의 내부에 포함되지만 군집간의 중복이 허용되지 않고 나무모양(Dendrogram)의 형식을 취함을 뜻한다. HCA(Hierarchical Cluster Analysis)방법에는 최단연결법(Single Linkage Method, SLM), 최장연결법(Complete Linkage Method, CLM), 평균연결법(Average Linkage Method, ALM), 중위수연결법(Median Linkage Method, MLM), 중심연결법(Centroid Linkage Method, CELM) 그리고 Ward법 등이 있으며, 이들 중 SLM, CLM, ALM이 일반적으로 잘 사용되므로, 본 연구에서는 HCA방법으로 앞서 말한 3가지 방법을 적용한다. SLM은 군집과 군집간의 사건들의 최소거리를 유사성의 척도로 사용하며 ALM은 군집과 군집간의 사건들의 모든 거리의 평균거리를 유사성의 척도로 사용한다. 그리고 CLM은 군집과 군집간의 사건들의 최장거리를 유사성의 척도로 사용한다. 8,000개의 군집은 우선적으로 유사성이 높은 군집끼리 병합되어 지며 군집의 개수는 점차 줄어든다.

누수위치군집간의 거리를 정의하는 방법에 따라 여러 가지 방법이 존재하며 임의의 누수위치 $x_i'=(x_{i1}, \dots, x_{ip})$ 와 $x_j'=(x_{j1}, \dots, x_{jp})$ 사이의 거리는 Eq. (1)과 같은 유클리드(Euclid)거리가 대표적으로 사용된다.

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (1)$$

2.1.1 SLM(Single Linkage Method)

SLM에서 두 군집 C_1 과 C_2 사이의 거리는 각 군집에 속하는 임의의 두 사건들 사이의 거리 중 최단거리로서 Eq. (2)로 정의된다.

$$d[(C_1)(C_2)] = \min[\min(d(x, y)) | x \in C_1, y \in C_2] \quad (2)$$

SLM은 Eq. (2)로 정의되는 임의의 두 군집간의 거리가 다른 두 군집간의 거리보다 작을 때 그 두 군집을 병합시킨다.

2.1.2 CLM(Complete Linkage Method)

SLM과는 정반대로 CLM에서는 두 군집 C_1 과 C_2 사이의 거리를 두 군집에 속하는 사건들 사이의 거리 중 가장 먼 거리를 사용하며 Eq. (3)과 같이 정의된다.

$$d[(C_1)(C_2)] = \min[\max(d(x, y)) | x \in C_1, y \in C_2] \quad (3)$$

군집의 수가 정해졌을 때는 같은 군집 내에 있는 임의의 두 사건 사이의 최대거리를 최소화하는 분리군집 방법과 근사적으로 동일한 결과를 제공하는 성질을 가진다.

2.1.3 ALM(Average Linkage Method)

ALM은 군집 내의 사건의 개수로 정의되는 군집의 크기가 각각 N_1, N_2 인 두 군집 C_1 과 C_2 간의 거리를 각 군집에서 하나의 사건을 선정하여 생성한 모든 가능한 $N_1 \times N_2$ 쌍의 거리의 평균으로 아래와 같이 Eq. (4)로 정의한다.

$$d[(C_1)(C_2)] = (1/(N_1 N_2)) \sum_r \sum_s d_{rs} \quad (4)$$

이 방법을 따르면 군집 내의 사건의 개수로 정의되는 군집의 분산이 작은 분산을 가지는 군집을 형성시키는 방향으로 군집절차가 편의(bias)되는 경향이 있다.

2.1.4 군집개수 판정

본 연구에서는 RSQ(Square of Multiple Correlation Coefficient, R^2)와 CCC(Cubic Clustering Criterion)를 군집개수(Number of Cluster, NCL)결정의 척도로 이용한다. 여기서 RSQ는 다중상관계수의 제곱을 나타내며 군집의 분리 정도에 관한 하나의 지표이고, 그 구조는 Eq. (5) ~ Eq. (8) 과 같다(Liu *et al.*, 2006).

$$W_k = \sum_{i \in C_k} (X_i - \bar{X})^2 = \sum_{i \in C_k} (X_i - \bar{X}_k)' (X_i - \bar{X}_k) \quad (5)$$

X_i : k번째 Cluster의 i 번째 좌표

\bar{X}_k : k번째 Cluster의 평균 좌표

$$P_G = \sum_{J=1}^C W_J \quad (6)$$

C : 총 군집 개수

$$T = \sum_{i=1}^n \|X_i - \bar{X}\|^2 \quad (7)$$

n : 전체 좌표 개수

X_i : i 번째 좌표

\bar{X} : 전체 좌표의 평균

$$R^2 = 1 - [(P_C)/T] \quad (8)$$

군집의 병합으로 인해서 RSQ의 수치가 작아지는데, 일반적으로 RSQ는 최적의 군집개수에서 RSQ 감소 폭인 SPRSQ(Semipartial RSQ)가 상대적으로 크다고 알려져 있다. 그리고 CCC는 최적의 군집개수에서 CCC 증가 폭인 SPCCC(Semipartial CCC)이 상대적으로 크다고 알려져 있다(김기영과 전명식, 1989).

3. 연구대상지역 상수관로 누수좌표에 대한 군집분석 적용

3.1 연구대상지역의 상수관로 누수자료

연구대상지역은 2000년 당시 인구가 348,189명인 미국의 소도시이다. 이 지역에서는 1992년부터 1997년 사이에 6인치 상수관망의 누수기록은 총 8,000 여 개가 기록되어 있으며, 누수발생건수는 해마다 증가는 경향을 보이고 있다. 년 간 평균 누수건수는 약 1,600 여 개로

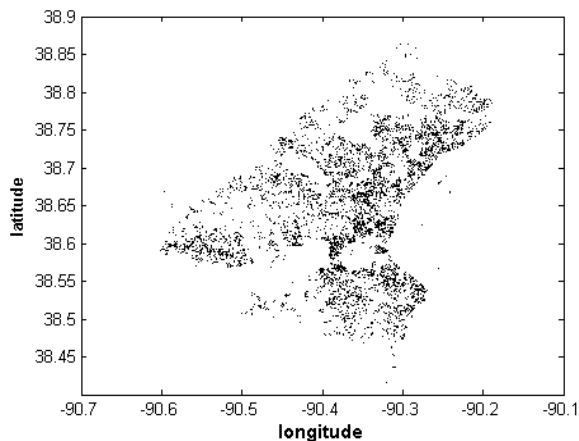


Fig. 3. Water Leak Location

상당한 수의 누수가 일어나고 있는 것으로 판단된다. Fig. 3은 상수관망 누수위치를 위도, 경도 상에 나타낸 것이며, 그림 중앙에 퍼져있는 검은색 점이 누수위치에 해당한다.

3.2 SLM을 이용한 지역 분할

Fig. 4부터 Fig. 6은 SLM으로 분할된 각 누수위치군집이 이루는 구역을 나타내고 있으며 각 구역상의 번호는 군집번호(Cluster Number, CN)이다. Table 1부터 Table 3은 분할된 구역들의 구역면적(Partitioned Region Area, PRA), 총 누수건수(Total Number of Leak, TNL), 총 누수강도(Total Leak Intensity, TLI) 그리고 총 누수강도의 총 누수강도서열(Total Intensity Order, TIO)을 나타내고 있다.

Table 1부터 Table 3을 통해서 군집번호 1의 면적이 타 구역에 비해서 약 150배 정도 크고 총 누수건수 또한 타 구역에 비해서 약 900배 정도 많은 것을 확인할 수 있다. 따라서 SLM은 고립된 소규모 누수위치군집탐색에 적합하다고 판단된다.

3.3 CLM을 이용한 지역 분할

Fig. 7부터 Fig. 9는 CLM으로 분할된 각 군집이 이루는 구역을 나타내고 있다. Table 4부터 Table 6은 분할된 구역들의 구역면적(PRA), 총 누수건수(TNL), 총 누수강도(TLI), 총 누수강도서열(TIO)을 나타내고 있다.

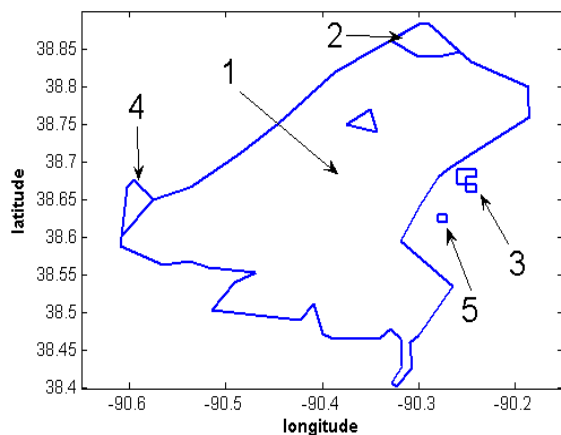


Fig. 4. Partitioned Region by SLM, NCL=5

Table 1. TLI and TIO of Partitioned Region by SLM, NCL=5

CN	1	2	3	4	5
PRA(km ²)	838.63	17.40	3.86	11.00	0.97
TNL	7,957	11	18	5	6
TLI	9.49	0.63	4.66	0.45	6.21
TIO	1	4	3	5	2

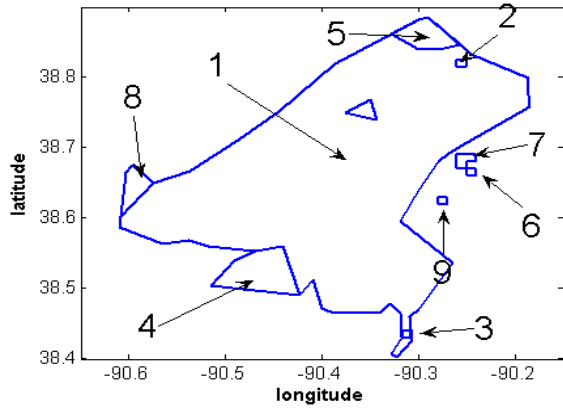


Fig. 5. Partitioned Region by the SLM, NCL9

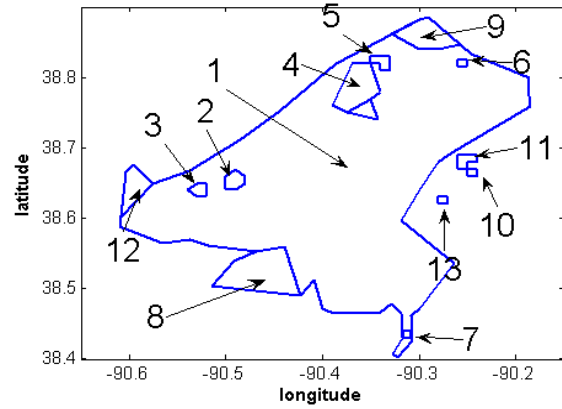


Fig. 6. Partitioned Region by the SLM, NCL=13

Table 2. TLI and TIO of partitioned region by SLM, NCL9

CN	1	2	3	4	5	6	7	8	9
PRA	799.16	0.96	0.85	37.66	17.40	0.97	2.90	11.00	0.97
TNL	7,868	2	4	82	11	9	9	5	6
TLI	9.85	2.07	4.69	2.18	0.63	9.32	3.11	0.45	6.21
TIO	1	7	4	6	8	2	5	9	3

Table 3. TLI and TIO of partitioned region, NCL13

CN	1	2	3	4	5	6	7	8	9	10	11	12	13
PRA	768.35	4.35	2.90	20.67	2.89	0.96	0.85	37.66	17.40	0.97	2.90	11.00	0.97
TNL	7732	11	4	108	11	2	4	81	11	9	9	5	6
TLI	10.06	2.53	1.38	5.23	3.80	2.07	4.69	2.15	0.63	9.32	3.11	0.45	6.21
TIO	1	8	11	4	6	10	5	9	12	2	7	13	3

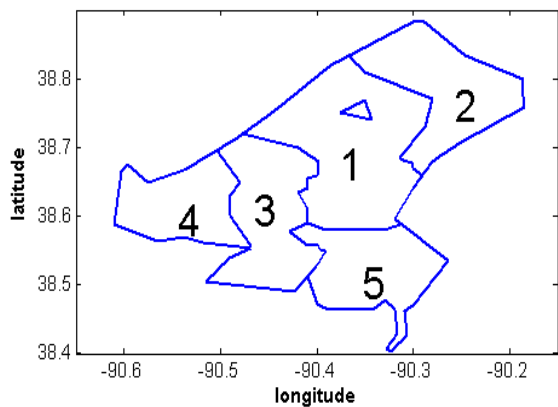


Fig. 7. Partitioned Region by CLM, NCL5

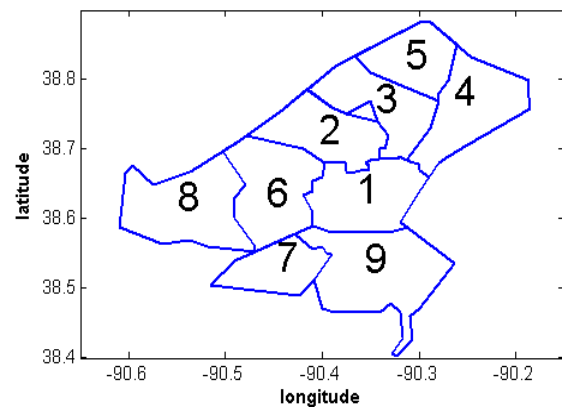


Fig. 8. Partitioned Region by CLM, NCL9

Table 4. TLI and TIO of Partitioned Region by CLM, NCL5

CN	1	2	3	4	5
PRA(km ²)	266.44	169.97	166.93	122.18	141.06
TNL	3,345	1,357	884	844	1,570
TLI	12.55	7.98	5.30	6.91	11.13
TIO	1	3	5	4	2

Table 5. TLI and TIO of Partitioned Region by CLM, NCL9

CN	1	2	3	4	5	6	7	8	9
PRA	98.98	83.22	84.24	103.92	66.05	106.47	60.47	122.18	141.06
TNL	1,995	750	600	1,254	103	745	139	844	1,570
TLI	20.16	9.01	7.12	12.07	1.56	7.00	2.30	6.91	11.13
TIO	1	4	5	2	9	6	8	7	3

Table 6. TLI and TIO of Partitioned Region by CLM, NCL13

CN	1	2	3	4	5	6	7	8	9	10	11	12	13
PRA	98.98	83.22	84.24	87.25	16.67	66.05	106.4	60.47	87.83	34.35	47.73	17.06	76.2
TNL	1,995	750	600	934	320	103	745	139	648	196	731	304	535
TLI	20.16	9.01	7.12	10.71	19.19	1.56	7.00	2.30	7.38	5.71	15.31	17.82	7.01
TIO	1	6	8	5	2	13	10	12	7	11	4	3	9

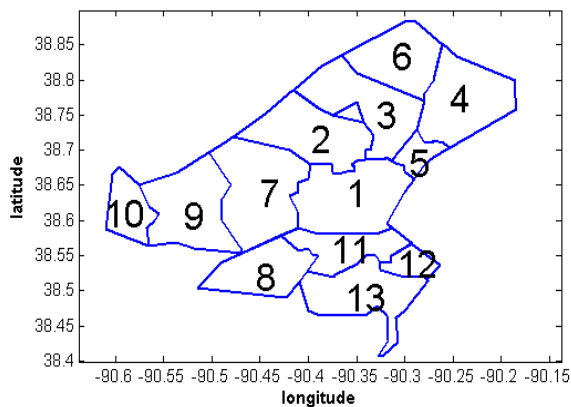


Fig. 9. Partitioned Region by CLM, NCL13

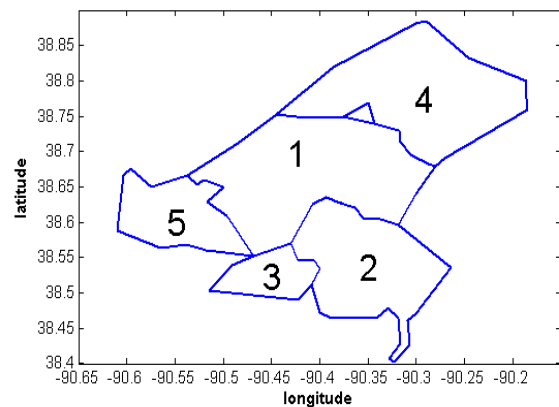


Fig. 10. Partitioned Region by ALM, NCL5

Table 4부터 Table 6을 통해서 CLM으로 분할된 구역면적들은 전반적으로 균등한 것을 확인 할 수 있으며, 군집개수가 늘어나면서 기존의 구역이 다시 분할되는 것을 확인 할 수 있다. 군집개수 9의 군집번호 8을 바탕으로 다시 분할된 군집개수 13의 군집번호 9와 군집번호 10의 총 누수강도는 기존의 군집번호 8의 총 누수강도보다 크거나 작은 지역으로 분할되어 있다. 그리고 군집개수 9의 군집번호 9를 통해서 다시 분할된 군집개수 13의 군집번호 11, 군집번호 12, 군집번호 13 중에서 2개의 구역은 기존의 군집번호 9의 총 누수강도보다 크고 나머지 한 구역은 작은 것을 확인 할 수 있

다. 따라서 기존의 구역이 새로운 분할을 통해서 기존 구역의 총 누수강도 보다 크거나 작은 새로운 구역으로 분할되는 것을 확인 할 수 있다.

3.4 ALM을 이용한 지역 분할

Fig. 10부터 Fig. 12는 ALM으로 분할된 각 군집이 이루는 구역을 나타내고 있다. Table 4부터 Table 6은 분할된 구역들의 구역면적(PRA), 총 누수건수(TNL), 총 누수강도(TLI), 총 누수강도서열(TIO)을 나타내고 있다.

Table 7부터 Table 9를 통해서 ALM으로 분할된 구

Table 7. TLI and TIO of Partitioned Region by ALM, NCL5

CN	1	2	3	4	5
PRA	281.33	181.20	52.76	251.66	99.87
TNL	3,293	2,062	107	1,763	775
TLI	11.70	11.38	2.03	7.01	7.76
TIO	1	2	5	4	3

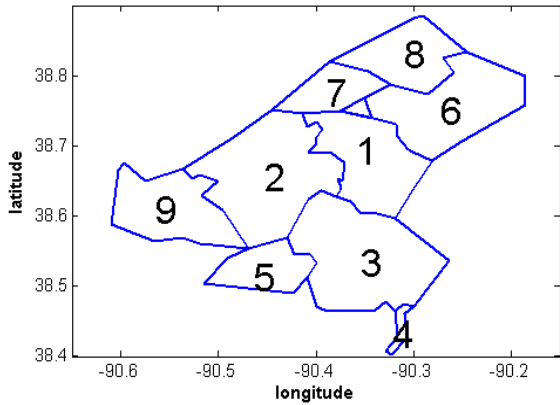


Fig. 11. Partitioned Region by ALM, NCL9

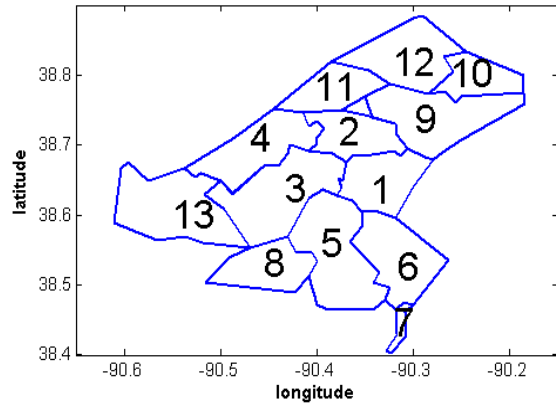


Fig. 12. Partitioned Region by ALM, NCL13

Table 8. TLI and TIO of Partitioned Region by ALM, NCL9

CN	1	2	3	4	5	6	7	8	9
PRA	107.86	173.47	174.48	6.72	52.76	129.88	46.35	75.42	99.87
TNL	2,044	1,249	2,051	11	107	1,494	157	112	775
TWI	18.95	7.20	11.75	1.64	2.03	11.50	3.39	1.48	7.76
TIO	1	5	2	8	7	3	6	9	4

Table 9. TLI and TIO of Partitioned Region by ALM, NCL13

CN	1	2	3	4	5	6	7	8	9	10	11	12	13
PRA	59.39	48.47	100.7	72.77	101.8	72.64	6.72	52.76	91.32	38.56	46.35	75.42	99.8
TNL	1,451	593	801	448	1,088	963	11	107	1,223	271	157	112	775
TLI	24.43	12.23	7.95	6.16	10.68	13.26	1.64	2.03	13.39	7.03	3.39	1.48	7.76
TIO	1	4	6	9	5	3	12	11	2	8	10	13	7

역면적들은 전반적으로 균등한 것을 확인 할 수 있으며, 군집개수 9에서 군집개수 13으로 다시 분할될 경우 기존의 구역에 비해서 총 누수강도가 상대적으로 큰 구역과 작은 구역으로 분할되는 현상은 CLM과 유사하다.

3.5 종합 분석

3.5.1 분할된 구역의 면적 비교

3가지 HCA의 지역 분할 성향을 3.2절에서부터 3.4절을 통해서 분석하였다. Fig. 13은 3가지 HCA에 의해서 연구대상지역을 9개의 구역으로 분할하였을 경우 각 군집의 구역면적을 나타내고 있다.

ALM으로 분할된 각 구역의 면적크기는 대체로 균등하지만 SLM에서 볼 수 있는 소규모 고립구역도 확인할 수 있다. 예를 들어서 Fig. 13에서 SLM의 3번 구역과 ALM의 4번 구역은 거의 동일한 위치에 존재하는 소규모 고립구역이다. 따라서 ALM은 누수위치좌표를 균등하게 분할하면서 동시에 고립된 누수위치좌표를 군집분할하는 성향도 있는 것으로 사료된다.

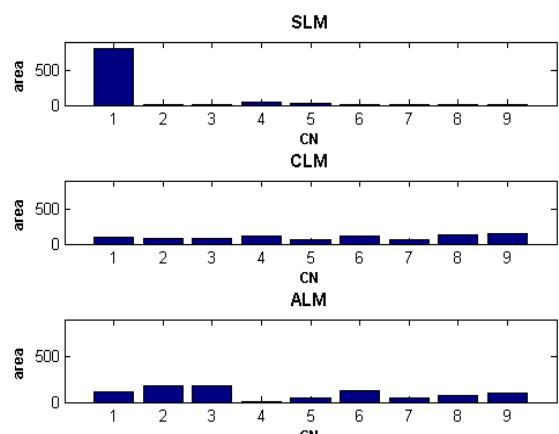


Fig. 13. Partitioned Region Areas by SLM, CLM and ALM, NCL=9

Table 10은 3가지 HCA로 분할된 구역들의 면적표준편차이다. CLM으로 분할된 군집들이 차지하는 구역면적은 3가지 HCA방법 중 면적표준편차가 상대적으로 작은 것을 알 수 있다. 그러므로 연구대상 지역을 균등한 크기의 면적으로 분할하는 방법으로 적당하다고 사

료된다. 그리고 ALM으로 분할된 군집들이 차지하는 구역면적은 CLM에 비해서 표준편차가 다소 크지만 전반적으로 각 구역의 면적은 균등하다.

Table 10. Standard Deviation of Partitioned Region Areas

NCL	5	9	13
SLM	371.4	263.6	211.0
CLM	55.6	25.9	29.7
ALM	97.2	57.2	28.3

3.5.2 분할된 구역들의 연도별 서열변동(Order Change)

Table 11부터 Table 13은 CLM을 이용해서 분할된 각 구역들의 연도별 누적누수강도(Cumulative Leak Intensity, CLI)와 총 누수강도서열(Total Intensity Order, TIO)의 서열변동(Order Change, OC)을 나타내고 있다. 누적누수강도는 1993년도를 시점으로 각 연도별 총 누수강도를 평가시점까지 합산한 것이다. 서열변동은 각 연도별 누적누수강도의 서열과 총 누수강도의 총 누수강도서열과의 차이 값이다. 즉 평가 년도 누적누수강도의 서열이 총 누수강도서열보다 한 단계 떨어

지면 -1의 값이 되고 한 단계 올라서면 +1의 값이 된다. 서열변동 분석을 위한 평가연도기간은 1993년부터 1997년까지 이다.

Table 11부터 Table 13을 통해서, 군집개수 5 일 경우에는 과거 5년 동안 서열변동이 발생하지 않았음을 확인 할 수 있다. 그러나 군집개수 9, 13 일 경우에는 총 누수강도서열의 중간층에서 서열변동이 발생한 것을 확인 할 수 있고, 총 누수강도서열이 상대적으로 높을 경우에는 서열변동 빈도가 낮다. 서열변동이 발생했을 경우에는 총 누수강도서열의 정확도가 떨어진다는 의미이지만, 군집개수 9, 13에서 발생한 서열변동의 절대값이 대부분 1인 것을 감안하면 큰 폭은 아닌 것으로 판단된다.

Table 14부터 Table 16은 ALM을 이용해서 분할된 각 구역들의 연도별 누적누수강도(CLI)와 총 누수강도서열(TIO)의 서열변동(OC)을 나타내고 있다.

Table 14부터 Table 16을 통해서, 군집개수 5 일 경우에는 과거 5년 동안 서열변동이 일어나지 않았음을 확인 할 수 있다. 그러나 군집개수 9, 13의 경우에는 총 누수강도서열의 상, 하층에서 주로 서열변동이 발생한 것을 확인 할 수 있다.

Table 11. CLM OC and CLI by CLM, NCL=5

CLM		OC					CLI				
TIO	CN	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
1	1	0	0	0	0	0	2.14	4.84	7.08	9.21	11.50
2	5	0	0	0	0	0	1.61	3.92	6.19	8.46	10.26
3	2	0	0	0	0	0	1.20	2.85	4.38	5.90	7.37
4	4	0	0	0	0	0	0.92	2.58	3.75	4.89	6.36
5	3	0	0	0	0	0	0.74	1.87	2.83	3.89	4.89

Table 12. CLM OC and CLI by CLM, NCL=9

CLM		OC					CLI				
TIO	CN	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
1	1	0	0	0	0	0	3.62	7.83	11.40	14.71	18.37
2	4	0	0	0	0	0	1.83	4.32	6.57	8.88	11.10
3	9	0	0	0	0	0	1.61	3.92	6.19	8.46	10.26
4	2	0	0	0	0	0	1.60	3.63	5.15	6.68	8.39
5	3	-1	-1	0	0	0	0.95	2.53	3.92	5.24	6.52
6	6	1	-1	-1	0	0	1.08	2.51	3.71	5.22	6.49
7	8	0	2	1	0	0	0.92	2.58	3.75	4.89	6.36
8	7	-1	0	0	0	0	0.15	0.76	1.27	1.55	2.08
9	5	1	0	0	0	0	0.21	0.55	0.94	1.21	1.50

Table 13. CLM OC and CLI by CLM, NCL=13

CLM		OC					CLI				
TIO	CN	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
1	1	0	0	0	0	0	3.62	7.83	11.40	14.71	18.37
2	5	-1	-1	0	0	0	3.12	6.84	11.22	14.21	17.93
3	12	1	1	0	0	0	3.22	7.68	10.61	14.18	16.94
4	11	0	0	0	0	0	1.95	4.99	8.28	11.48	14.06
5	4	-1	0	0	0	0	1.58	3.84	5.69	7.86	9.80
6	2	1	0	0	0	0	1.60	3.63	5.15	6.68	8.39
7	9	-2	0	0	-3	0	0.99	2.77	3.93	5.06	6.76
8	3	-2	0	0	0	0	0.95	2.53	3.92	5.24	6.52
9	13	1	-1	0	2	0	1.04	2.41	3.89	5.28	6.49
10	7	3	1	0	1	0	1.08	2.51	3.71	5.22	6.39
11	10	0	0	0	0	0	0.73	2.10	3.29	4.45	5.33
12	8	-1	0	0	0	0	0.15	0.76	1.27	1.55	2.08
13	6	1	0	0	0	0	0.21	0.55	0.94	1.21	1.50

Table 14. ALM OC and CLI by ALM, NCL=5

ALM		OC					CLI				
TIO	CN	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
1	1	0	0	0	0	0	2.02	4.65	6.71	8.77	10.82
2	2	0	0	0	0	0	1.66	3.86	6.13	8.33	10.40
3	5	0	0	0	0	0	1.05	2.91	4.25	5.49	7.10
4	4	0	0	0	0	0	1.02	2.49	3.81	5.17	6.46
5	3	0	0	0	0	0	0.11	0.61	1.02	1.33	1.82

Table 15. ALM OC and CLI by ALM, NCL=9

ALM		OC					CLI				
TIO	CN	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
1	1	0	0	0	0	0	3.61	7.90	11.2	14.1	17.4
2	3	0	-1	0	0	0	1.72	3.99	6.33	8.60	10.7
3	6	0	1	0	0	0	1.72	4.13	6.23	8.51	10.5
4	9	0	0	0	0	0	1.05	2.91	4.25	5.49	7.10
5	2	0	0	0	0	0	1.04	2.62	3.89	5.41	6.72
6	7	0	0	0	0	0	0.41	1.10	1.81	2.39	3.19
7	5	-2	0	0	0	0	0.11	0.61	1.02	1.33	1.82
8	4	0	-1	-1	0	0	0.15	0.30	0.74	1.19	1.49
9	8	2	1	1	0	0	0.20	0.53	0.89	1.14	1.41

Table 17은 1993년부터 1996년까지의 CLM과 ALM으로 분할된 구역들의 서열변동을 대상으로 서열변동의 절대 값이 0보다 클 빈도(Frequency of order change)와 서열변동 절대 값의 평균(Mean of order change)을 나타내고 있다. 분할된 구역의 서열변동 빈도와 평균크기가 클 경우 상수관망 유지관리를 위한 각 구역의 누수강도서열이 시간에 따라서 변화되므로 지역분할의 실

효성이 떨어진다. 따라서 서열변동 빈도와 절대 값의 평균크기는 작을수록 좋다. Table 17을 통해서 ALM으로 분할된 구역이 서열변동 빈도와 평균이 상대적으로 작고 구역의 누수발생서열이 일관된 것을 확인 할 수 있다. 그러므로 ALM으로 분할된 구역들의 누수강도서열은 1993부터 1997년까지 일정하게 유지된 것으로 사료된다. SLM으로 분할된 누수위치좌표의 경우 1번 구

Table 16. ALM OC and CLI by ALM, NCL=13

ALM		OC					CLI				
TIO	CN	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
1	1	0	0	0	0	0	4.77	10.2	14.6	18.4	22.4
2	9	-1	-1	0	0	0	2.01	4.95	7.29	9.98	12.2
3	6	-1	-1	-1	0	0	1.87	4.64	7.10	9.84	12.1
4	2	2	2	1	0	0	2.19	5.01	7.14	8.95	11.2
5	5	0	0	0	0	0	1.61	3.53	5.78	7.72	9.72
6	3	0	0	0	0	0	1.24	2.96	4.29	6.07	7.41
7	13	0	0	0	0	0	1.05	2.91	4.25	5.49	7.10
8	10	0	0	0	0	0	1.01	2.18	3.71	5.03	6.48
9	4	0	0	0	0	0	0.76	2.16	3.33	4.51	5.77
10	11	0	0	0	0	0	0.41	1.10	1.81	2.39	3.19
11	8	-2	0	0	0	0	0.11	0.61	1.02	1.33	1.82
12	7	0	-1	-1	0	0	0.15	0.30	0.74	1.19	1.49
13	12	2	1	1	0	0	0.20	0.53	0.89	1.14	1.41

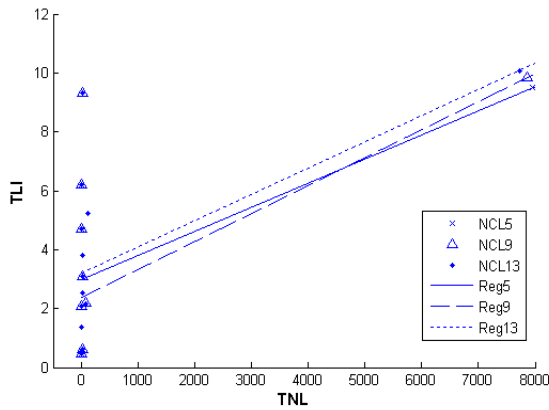


Fig. 14. Relation between TLI and TNL of SLM, NCL[5 9 13]

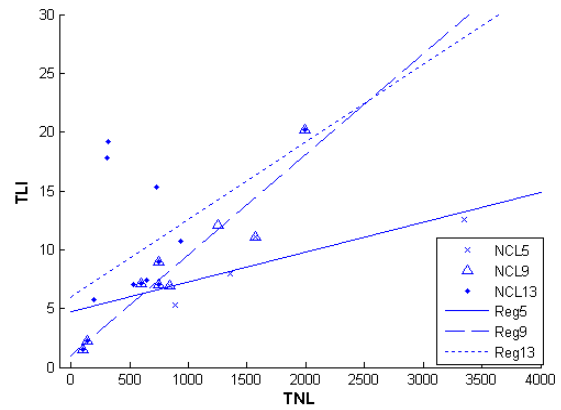


Fig. 15. Relation between TLI and TNL of CLM, NCL[5 9 13]

Table 17. Probability of Order Change

NCL	CLM			ALM		
	5	9	13	5	9	13
Frequency of OC(%)	0	25	31	0	22	27
Mean of OC	0	0.28	0.46	0	0.28	0.35

역에 전체 누수건수의 95% 이상이 편중되어 있어서 각 구역의 서열변동 폭이 상대적으로 크다.

3.5.3 총 누수건수와 총 누수강도의 관계

Fig. 14부터 Fig. 16은 3가지 HCA로 분할된 각 구역의 총 누수건수와 총 누수강도의 관계를 최소자승법을 이용해서 회귀식으로 나타내고 있다.

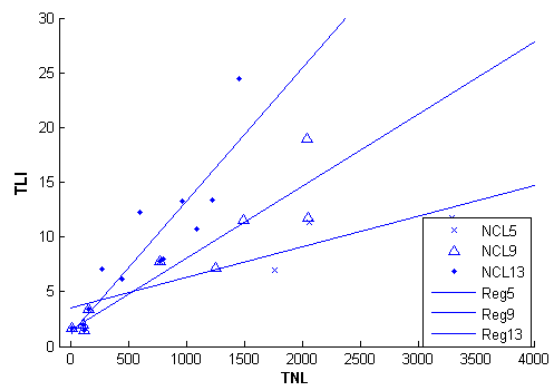


Fig. 16. Relation between TLI and TNL of ALM, NCL[5 9 13]

Table 18은 3가지 HCA로 분할된 각 구역의 총 누수건수와 총 누수강도의 결정계수(Coefficient of determination)를 나타내고 있다. ALM으로 분할된 구역의

결정계수가 상대적으로 큰 것을 확인 할 수 있다. SLM과 CLM의 총 누수강도와 총 누수건수의 관계에서는 상대적으로 큰 불확실성이 존재하는 것으로 판단된다. 따라서 ALM으로 연구대상지역을 분할 할 경우에는 분할된 총 누수건수만으로 각 구역의 총 누수강도 서열을 가늠할 수 있을 것으로 사료된다.

Table 18. Coefficient of determination between TNL and TLI

NCL	SLM	CLM	ALM
5	0.574	0.747	0.757
9	0.351	0.916	0.874
13	0.358	0.274	0.812

3.5.4 상수관망 우선관리대상지역 선정

3.5.1항부터 3.5.3항을 통해서 3가지 HCA의 연구대상 지역 분할 경향과 각 구역의 누수강도서열의 일관성을 확인 할 수 있었다. 결과적으로 ALM(Average Linkage Method)이 지역 분할 성향과 일관된 누수강도서열의 측면에서 연구대상지역 분할을 위한 가장 적절한 HCA로 판단된다. Fig. 17부터 Fig. 19는 ALM으로 연구대상지역을 5개, 9개, 13개 구역으로 분할하고 각 구역의 총 누수강도서열을 분할된 구역위에 표기한 것이다. Fig. 17부터 Fig. 19를 통해서 상수관망 유지관리 1순위 지역부터 마지막 순위지역까지 결정지을 수 있다.

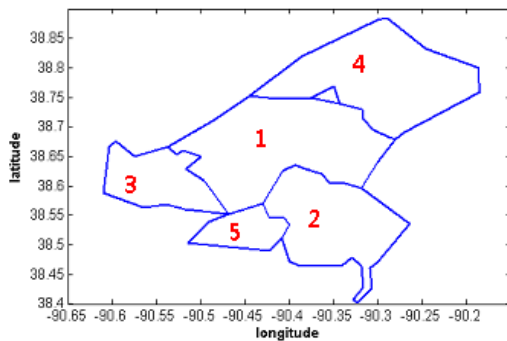


Fig. 17. Prioritization of the area for maintenance, NCL5

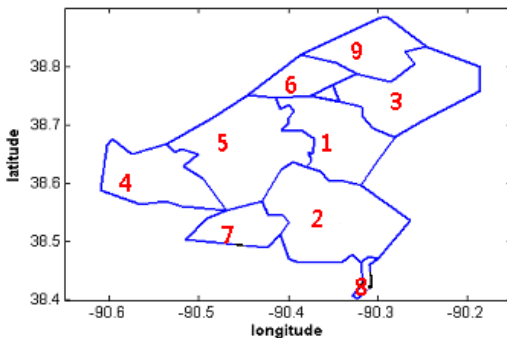


Fig. 18. Prioritization of the area for maintenance, NCL9

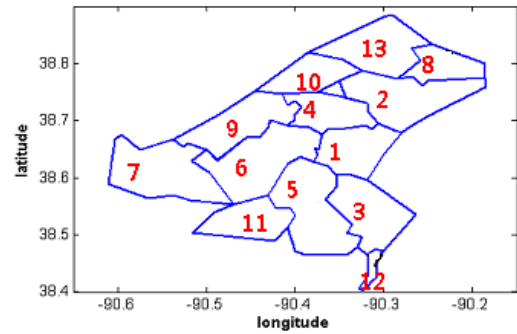


Fig. 19. Prioritization of the area for maintenance, NCL13

4. 결론

본 연구에서는 연구대상지역 내 8,000개의 상수관망 누수위치좌표들을 대상으로 HCA(Hierarchical Cluster Analysis)를 적용해서 누수위치좌표들을 적절한 수의 군집으로 분할한다. 그리고 분할된 누수위치좌표들이 속하는 구역의 지역적 경계선을 설정해서 각 구역에서 발생한 총 누수강도를 산출하고 총 누수강도를 이용하여 총 누수서열을 각 구역의 상수관망 유지관리 우선순위로 사용한다. 연구대상지역의 상수관망 누수위치좌표를 이용해서 상수관망 유지관리 차원의 지역분할과 우선관리대상지역의 결정에 HCA방법론을 적용한 결론은 다음과 같다.

- 1) Bailey Gatrell (1995)의 강도분석법은 일정한 크기의 셀로 연구대상지역을 분할해서 각 셀의 총 누수강도가 높은 구역을 개략적으로 산정할 수 있다. 그러나 적당한 셀의 개수와 누수위치좌표를 바탕으로 하는 사실적인 지역분할에는 부족한 점이 있다. HCA는 누수위치좌표들의 유사성 또는 비유사성의 직접적인 특성을 이용해서 누수위치좌표를 적당한 개수의 군집으로 병합 또는 분할하고, 분할된 군집을 바탕으로 연구대상지역의 사실적인 분할을 가능하게 한다. 따라서 누수위치좌표를 대상으로 하는 HCA 강도측정방법은 기존의 사각형 셀을 이용한 강도측정방법에 비해서 합리적 이라고 사료된다.
- 2) HCA로 분할된 연구대상지역의 구역들은 비교적 그 편차가 작은 것이 바람직한데, 그 편차가 매우 클 경우에는 연구대상지역의 군집면적이 매우 큰 군집 하나와 매우 작은 군집 하나로 산정될 수도 있다. 따라서 이러한 경우 분할된 구역에 대한 유지관리 우선순위를 부여하는 의미 자체가 퇴색되며, 면적이 매우 큰 군집 구역에 소속된 상수관망의 효율적인 유지관리를 위해서 구역을 다시 세

부 구역으로 구분해야 할 필요가 생긴다. 3가지 HCA방법의 탐색 경향을 감안해 볼 때 ALM(Average Linkage Method)이 고립지역을 탐색하면서 분할된 구역면적의 표준편차가 비교적 적은 경향으로 연구대상지역을 분할하였다. 또한 연구대상 기간 동안에 각 구역의 서열변동 크기가 상대적으로 작았다. 이러한 측면에서 연구대상지역의 상수관망 우선관리대상지역 선정을 위한 HCA방법으로는 ALM이 가장 효과적이라고 사료된다.

- 본 연구에 보완되어야 될 사항으로는 누수위치군 집 구역의 경계설정을 위해서 상수관망지도를 이용하는 것이다. 상수관망지도를 이용해서 각 누수 위치좌표가 차지하는 연구대상지역의 명확한 범위가 파악 된다면 보다 정확한 누수강도를 측정할 수 있을 것이다.

참 고 문 헌

- 김기영, 전명식 (1989). **SAS 군집분석**, 자유아카데미, pp. 31, 55-60.
- 문 호 (2001). **GIS를 기반으로 한 지하시설물(상수관로)네트워크 모델링 및 구현**, 석사학위논문, 서울산업대학교대학원.
- 박수완 (2007). “상수도 배수관로 시스템의 장기적 유지관리를 위한 방법론과 컴퓨터 알고리즘의 개발 및 적용.” **한국물환경학회지**, 한국물환경학회, 제23권, 제3호, pp. 356-366.
- 박풍일 (2002). **누수탐지를 위한 최적화 기법의 비교**, 석사학위논문, 동아대학교대학원.
- 오창주, 이경훈, 강용덕 (2001). “관망제어를 위한 벨브 탐색 알고리즘.” **대한상수도학회지**, 대한상수도학회, 제15권, 제3호, pp. 222-228.
- 전환돈, 이양재, 이정호, 김중훈 (2007). “도시 내배수시스템 실시간 운영모형의 개발.” **한국물환경학회지**, 한국물환경학회, 제23권, 제5호, pp. 748-755.
- 최성규, 안병찬 (2001). “GIS와 EPANET 상수관망모델의 연계 방안 연구.” **한국GIS학회지**, 한국GIS학회, 제9권, 제3호, pp. 477-491.
- 환경부 (2006). **2006년 상수도통계연보**.
- Andreou, S. A., Marks, D. H., and Clark, R. M. (1987). “A new methodology for modelling break failure patterns in deteriorating water distribution systems.” *Theory. Advance in Water Resources* 10, 2-10.
- Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Longman Scientific & Technical, London.
- Beale, E. M. L. (1969). *Cluster Analysis*, Scientific Control Systems, London.
- Calinski, T. and Harabasz, J. (1974). “A Dendrite Method for Cluster Analysis.” *Communication in Statistic*, 3, 1-27.
- Doleac, M. L.(1979). *Time-to failure analysis of cast iron water main*. Report the City of Vancouver by CH2M HILL, BC, Canada.
- Jun, H. and Loganathan, G. V. (2007). “Valve-controlled segments in water distribution systems.” *J. Water Resour. Plng. and Mgmt*, Vol. 133, Issue 2, pp. 145-155.
- Kettler, A. J. and Goulter, I. C. (1985). “An analysis of pipe breakage in urban water distribution network.” *Canadian Journal of Civil Engineering*, 12, pp 286-293.
- Kleiner, Y. and Rajani, B. B. (2001). *Comprehensive review of structural deterioration of water mains, statistical model*, Urban Water.
- Kumar, A., Meronyk, E. and Segan, E. (1984). *Development of concepts for cohesion assessment and evaluation of underground pipeline*. Technical Report CERL-TR-M-337. US Army Corps of Engineers, Construction Engineering Research Laboratory, IL.
- Liu, J. L., Bai, Y., Kang, J. and An, N. (2006). “A new approach to hierarchical clustering using partial least squares.” *Proceeding of the Fifth International Conference on Machine Learning and Cybernetics*, IEEE, Dalian, Vol. 1, pp 1125-1131.
- Marriott, R. (1977). “Practical problems in a method of cluster analysis.” *Biometrics*, 27, pp. 501-514.
- Mojena, R. (1977). “Hierarchical grouping method and stopping rules: an evaluation.” *The Computer Journal*, 20, pp 359-363.
- Randall, M., Russell, A. and Oliphant, R. (1992). *Guidance manual for the structural condition assessment of the trunk mains*. Water Research Centre, Swindon, UK.
- Rajani, B. B. and Makar, J. M. (2000). “A methodology to estimate remaining service life of

- grey cast iron water mains." *Canadian Journal of Civil Engineering*, 27, pp 1259-1272.
- Rajani, B. and McDonald, S. (1993). "Water main break data on different pipe materials for 1992 and 1993." *National Research Council of Canada*, Ottawa, .No. A-7019.
- Rossman, L. A. and Grayman, W. M. (1998) "The effect of hydraulic factors on quality in drinking water storage tanks." *National Conference on Environmental Engineering*, Vol. 6, pp. 627-632.
- Sarle, W. S. (1983). *Cubic Clustering Criterion*, SAS Technical Report A-108, SAS Institute Inc., Cary, NC.
- Seo, J. and Shneiderman, B. (2006). "Knowledge discovery in high dimensional data: Case studies and a user survey for the rank-by-feature framework." *Transactions on Visualization and Computer Graphics*, IEEE, Vol. 12, No. 3, pp. 311-322.
- Spangler, M. G. (1941). "The structural design of flexible pipe culverts." *Iowa Engineering Experimental station*, Bulletin No. 53, Ames, Iowa.
- Wolfe, J. H. (1970). "Pattern clustering by multi-variate mixture analysis." *Multivariate Behavioral Research*, 5, pp 329-350.
- Wong, M. A. and Schaack, C. (1982). "Using the k-th nearest neighbor clustering procedure to determine the number of subpopulations." *American Statistical Association 1982 Proceedings of the Statistical Computing Section*, pp 40-48.
- (논문번호:08-99/접수:2008.09.25/심사완료:2008.12.05)