이질적 지반의 확률적 지층구분을 위한 이상치 분석의 적용

Probabilistic stratification method for heterogeneous soils using outlier analysis

김정열[1], Jeong-Yul Kim, 김현기[2], Hyun-Ki Kim, 조남준[3], Nam-Jun Cho

[1] 국민대학교 건설시스템공학부 석사과정, Master Student, School of Civil and Environmental Engineering, Kookmin Univ.
[2] 국민대학교 건설시스템공학부 조교수, Assistant Professor, Dept. of Civil and Environmental Engineering, Kookmin Univ.
[3] 국민대학교 건설시스템공학부 교수, Professor, Dept. of Civil and Environmental Engineering, Kookmin Univ.

SYNOPSIS : Subsurface investigation results obtained from soils with strong heterogeneity sometimes show a high level of uncertainty about stratigraphy and mechanical characteristics. In these cases, important engineering judgments are dependent on the personal experiences of engineers. However, many of such problems can be solved by applying appropriate statistical approaches. This study introduces the outlier analysis as a one of the statistical solutions and a simple case study is presented as an example.

Keywords : Heterogeneity, outlier analysis, stratification

1. 서 론

지반구조물의 설계변수를 결정하기 위해 현장이나 실내에서 이루어진 지반조사의 결과를 살펴보게 되면 본질적으로 토양의 형성과정에서 발생하는 이질적인 성질을 관찰하게 된다. 이러한 이질적인 성질로 인해 관측된 지반조사 결과의 이질적인 성질은 설계 목적이나 용도에 맞게 통계적인 기법을 통해 처리하는 것은 이제 설계과정에서의 선택적인 사항이 아니라 필수적인 사항이 되었다. 이에 본 연구에서는 일정한 구역 내에서 행해진 지반 조사 결과에 이상치 분석을 실시하여 동일지층구조로 판단할 수 있는 기준을 정량적으로 제시하는 방법을 제안하고자 한다.

2. 통계적 분석 기법

2.1 이상치 지수에 의한 검토

지반 조사 결과의 통계적 처리가 이루어진 후 자료를 통해 동일층으로 구분하기 위해서는 이에 대한 적합성을 검토하여야 하는데 이 때 이상치 지수가 사용되는 이상치 분석을 통한 방법이 사용된다. 이상치 지수의 통계적 모델은 지반 조사 자료를 이용하게 되는데 새로 얻어진 자료와 기존의 자료를 비교함으로써 알 수 있다(Oh C. K. et al, 2009) 이상치 지수는 다음 식 (1)과 같은 자료 관계에 따라 나타내어진다(Sohn H. et al, 2001).

$$Novelty\ Index(y) = \| x - y \| \qquad (1)$$

여기서, x는 기존 자료값, y는 새로 얻어진 자료값을 나타낸다. 이상치 지수는 새로 얻어진 자료가 이상상태에서 얻어진 자료일수록 크기가 커진다.

이상치 지수를 통한 통계적 분석은 자료의 평균과 표준편차를 사용하여 나타내는 Gaussian방법에 적용하기는 다소 무리가 있다(Castillo E., 1988). 이상치들은 대부분 한계값 이상을 가지게 되는데 이러한 값들은 확률 분포의 꼬리 부분에 위치하기 때문이다. 분포의 꼬리부분에 관련된 통계학 분석방법은 각 부분집합에서의 최대값 또는 최소값을 결정하는 최대 연산 및 최소 연산방법이다. 이상치 지수의 간단한 모델을 들어보면 다음 식(2), (3), (4)와 같다(Castillo E., 1988).

$$Gumbel : F_G(x|\kappa,\delta) = \exp\left[-\left(\frac{x-\kappa}{\delta}\right)\right] \qquad (2)$$

$$Frechet : F_F(x|\kappa,\delta,\beta) = \begin{cases} \exp\left[-\left(\frac{\delta}{x-\kappa}\right)^{\beta}\right] & \text{if } x \geq \kappa \\ 0 & otherwise \end{cases} \qquad (3)$$

$$Weibull : F_W(x|\kappa,\delta,\beta) = \begin{cases} 1 & \text{if } x \geq \kappa \\ \exp\left[-\left(\frac{\kappa-x}{\delta}\right)^{\beta}\right] & otherwise \end{cases} \qquad (4)$$

여기서 κ는 분포의 위치계수, δ는 분포의 크기계수, β는 분포의 형상계수이다. 이 세가지 분포함수를 통해 최소값과 관련된 좌측 꼬리 부분을 모델화 할 수 있다. 이 세가지 분포를 좀 더 균일하게 나타내는 GEV(Generalized extreme value)함수는 식(5)와 같이 나타낼 수 있다(Park H.W. et al, 2006; Jenkinson A.F., 1955).

$$F_{GEV}(x|\mu,\sigma,\gamma) = \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\delta}\right)^{-1/\gamma}\right]\right\}, \quad -\sigma-\gamma(x-\mu) \leq 0,\ \sigma \geq 0 \qquad (5)$$

여기서 μ는 GEV함수의 위치계수, σ는 GEV함수의 크기계수, γ는 GEV함수의 형상계수이다. GEV의 계수들은 반복 계수 추정 방법을 통해 확인할 수 있다(Jenkinson A.F., 1955). 형상계수는 GEV함수의 분포 형상을 통해 추정하게 되고 형상계수가 결정되면 나머지 계수들이 계산된다. 끝으로 특정 신뢰수준에 일치하는 한계값을 설정하여 이상치들을 분석하게 된다.

2.2 상관거리에 의한 검토

위의 이상치 검정을 통해 한계 범위 이내에 들어오는 분포를 보이는 지반을 동일지층으로 구분하게 되면 지층내의 물성 분포에 대해 통계적 분석을 실시하게 된다. 이 때, 각각의 물리적/역학적 성질들이 갖는 위치적인 상관성을 정량적으로 평가하기 위해 Variogram을 통한 상관거리 correlation length를 산정한다. 일반적으로 Variogram 기법은 주어진 자료가 등간격인 경우에 적용하나, 지반 조사가 항상 등간격으로 이루어지기는 어렵기 때문에 불균등한 간격으로 얻어진 자료에 대해서는 일정한 허용한계를 이용하여 완화된 분리거리 내에 존재하는 자료의 조합들을 동일한 간격내 자료로 간주하는 방법을 사용하게 된다. 다음 식 (6)과 식 (7)과 같이 각 지점별 분리거리 h가 분리거리 허용한계 $\triangle L_{\max}$와 완화된 분리거리 L에 의해 정의되는 범위 내에 있게 되면 완화된 분리거리 L 간격으로 분포하여 존재하는 자료로 간주하게 된다 (최종근, 2007).

$$\triangle L = \min(0.5L, \triangle L_{\max}) \qquad (6)$$

$$L - \triangle L < h < L + \triangle L \qquad (7)$$

이렇게 정의된 h 범위에 존재하는 모든 자료들에 대해 다음 식 (8)을 적용하여 Variogram을 구하게 된다.

$$\gamma(h) = \frac{1}{2N}\sum_{i=1}^{N}[(Z(x_i) - Z(x_{i+h})]^2 \qquad (8)$$

여기서, h는 분리거리, N은 분리거리별 자료쌍의 개수, $Z(x_i)$는 자료값, $Z(x_{i+h})$는 x_i로부터 완화된 분리거리 L로 정의된 h의 범위 내에서 떨어진 자료의 값을 나타낸다. 위의 과정을 모든 자료에 적용하여 분리거리를 변화시켜가며 계산하고 variogram으로 나타낸다. 문턱값을 벗어난 자료들은 서로간의 상관관계가 약하다는 것을 의미하므로 지반 조사 자료에서 동일층으로 구분할 때에 이러한 자료들을 고려해 줄 수 있게 된다.(최종근, 2007)

3. 여의도 지역 지반 조사 자료의 적용

여의도 지역내에서 실시한 70여 개소의 표준관입시험 결과에 대해서 위에서 소개한 이상치 분석을 실시하고 자료간의 상관거리 분석을 였다. 먼저 각 시추공에서 깊이별로 얻어진 표준관입시험 N치를 표준화하여 표현하는 수정 N값 $(N_1)_{60}$를 다음과 같이 구하였다.

$$(N_1)_{60} = N \bullet \eta \bullet C_n \qquad (9)$$

여기서 N=SPT N-value, η=해머 효율 60%에 대한 보정계수(로드길이, 보어홀 직경, 샘플러 종류, 해머 종류), C_n=유효응력 보정계수이다. 이렇게 보정을 하기 전 N값과 보정을 마친 후의 N값의 확률분포를 살펴보면 다음 그림 1과 같다.

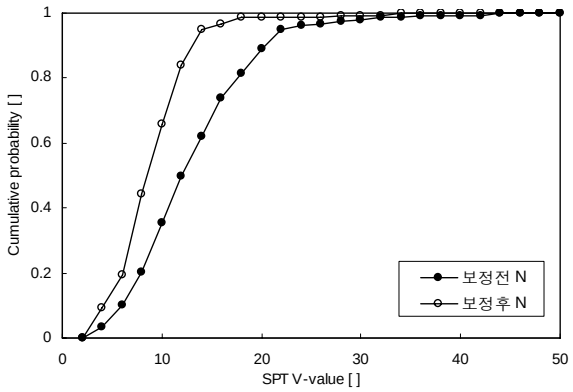
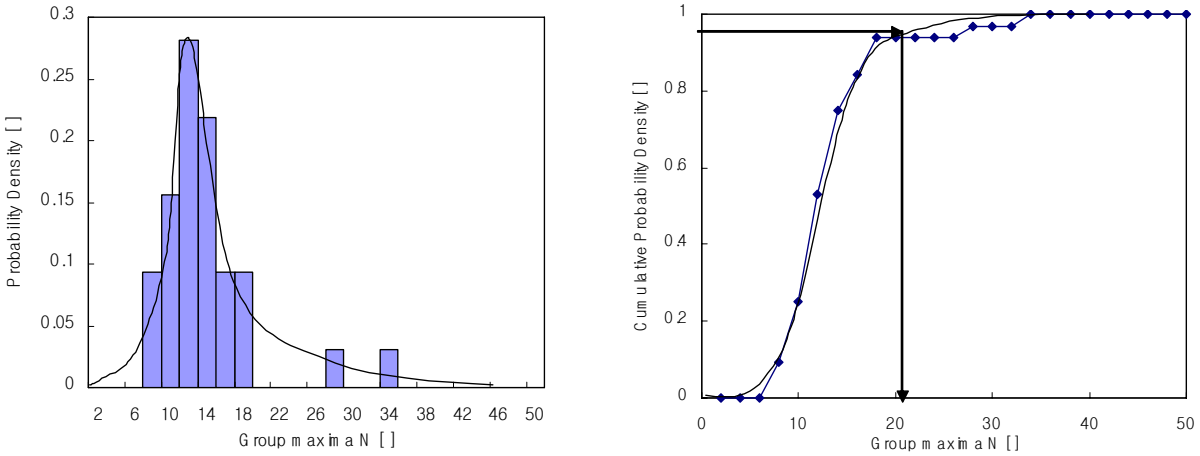


그림 1. 여의도 지역내 퇴적층에서 얻어진 N-value 누적확률분포 곡선

이렇게 구한 $(N_1)_{60}$에 대하여 이상치 분석을 적용하였더니 5% 신뢰수준에 대해 다음의 그림 2와 같이 22이상의 $(N_1)_{60}$이 나타나게 되면 동일한 퇴적층이라 보기 어려운 것으로 나타났다..



(a) 하위 그룹에 대한 최대값의 분포도와 그에 따른 극한값 확률분포의 형태

(b) 극한값 확률분포형태에 대해 신뢰수준 5%에 해당하는 이상치 검정 기준 설정

그림 2. 여의도 지역내 퇴적층에서 얻어진 $(N_1)_{60}$에 대한 이상치 검정 기준 설정에 대한 예

그리고 보정된 N값 $(N_1)_{60}$가 50이상이 되어 견고한 층으로 판단되는 지층의 분포가 위치에 따라 어느 정도의 상관성을 갖고 있는지를 살펴보기 위해 위의 시추공의 위치에 따라 $(N_1)_{60}$가 50 이상이 되는 깊이에 대한 자료를 이용하여 Variowin 3.2를 이용하여 Variogram을 작성하였다. 그림 3은 본 자료를 이용하여 단위 분리거리를 각각 5m, 10m, 15m, 20m로 정의하였을 때의 Variogram이다. 그림에서 보이다시피, 위치적 상관성을 정의하기에는 자료의 산도가 매우 높게 나타나 본 자료를 통해서는 상관거리 산정이 거의 불가능한 것으로 보인다.

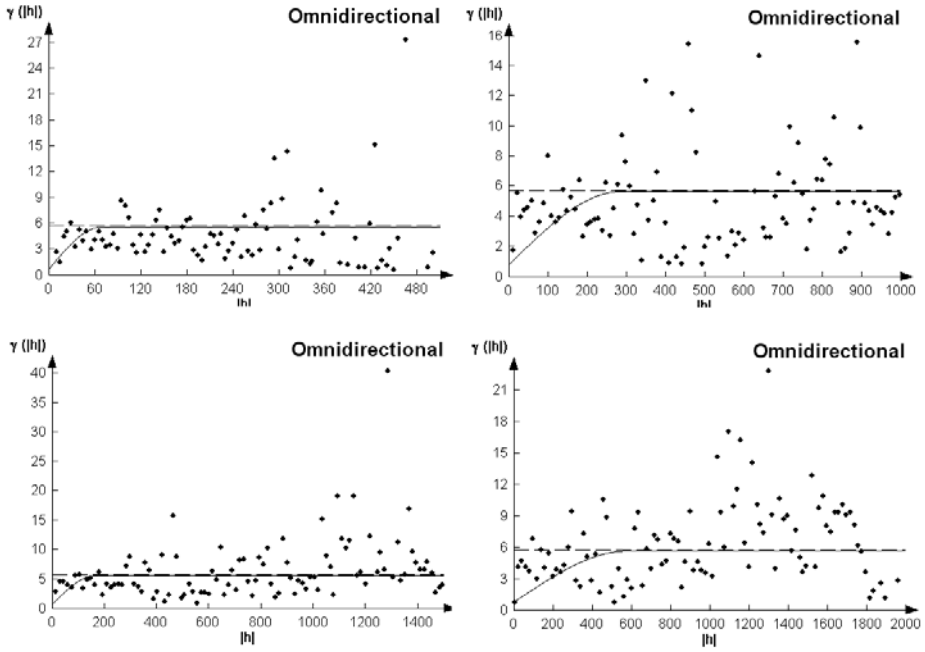


그림 3. 각기 다른 단위 분리거리(*dh*)에 따른 variogram (상하좌우 순으로 *dh*=5m, 10m, 15m, 20m)

4. 요 약

이질적인 지층의 성질을 정량적으로 분석하기 위해 적용할 수 있는 여러 가지 통계적 기법 가운데 이상치 분석과 Variogram 기법을 적용하여 여의도 지역의 퇴적층에 대한 성질을 분석해 본 결과 다음과 같은 결론을 내릴 수 있었다.

- 여의도 지역의 상부 퇴적층은 5% 신뢰수준에 대해 22이상의 $(N_1)_{60}$이 나타나게 되면 동일한 퇴적층이라 보기 어려운 것으로 나타났다.
- $(N_1)_{60}$가 50이상이 되어 견고한 층으로 판단되는 지층의 분포가 위치에 따라 어느 정도의 상관성을 갖고 있는지를 살펴보기 위해 위의 시추공의 위치에 따라 $(N_1)_{60}$가 50 이상이 되는 깊이에 대한 자료를 이용하여 지지층 깊이에 대한 상관거리를 구해보려 하였으나 Variogram을 통해 본 결과 위치적 상관성이 거의 나타나지 않는 것으로 분석되었다.

감사의 글

본 연구는 국토해양부가 주관하고 한국건설교통기술평가원이 시행하는 2006년도 첨단도시개발사업(07도시재생B03)에 의해 수행되었습니다.

참고문헌

1. 최종근, 2007, 지구통계학, 시그마프레스, pp136~154
2. Castillo E., (1998), "Extreme value therory in engineering", Academic press series in statistical modeling and decision science.
3. Oh, C.K., Sohn, H. and Bae, I.-H., (2009), "Statistical novelty detection within the Yeongjong suspension bridge under environmental and operational variations", Smar Mater. Struct. vol.18, 125022, pp. 9~18
4. Jenkinson A.F., (1955), "The frequency distribution of the annual maximum of meteorological elements", Q.J.R. Meteorol. Soc., Vol.81, pp.158~171
5. Park, H. W. and Sohn, H., (2006), "Parameter estimation of the generalized extreme value distribution for structural health monitoring Probab." Eng. Mech., Vol.21 pp.366~376
6. Sohn, H., Worden K. and Farrar, C. R., (2001), "Novelty detection under changing environmental conditions", SPIE's 8th Annual Int. Symp. on Smart Structures and Materials.