

통계분석을 이용한 지하수위 변동 특성 분류

문상기 · 우남철

연세대 지구시스템학과

e-mail : skmun@ysgeo.yonsei.ac.kr

요약문

A study on multivariate statistical classification of ground water hydrographs was conducted. The vast data of national ground water monitoring network (78 sites of alluvium) were used. 6 factors were selected to classify the ground water level change. Factor analysis was proved to be useful tool for classifying vast hydrogeological data.

Key word ; factor analysis, ground water hydrographs, alluvium, classification

1. 서론

본 연구의 목적은 국가 지하수 관측망 자료에 의한 전국 충적층의 지하수위 변동을 통계적으로 분류하는 데 있다. Winter(2000)는 다변량 통계 분석의 일종인 주성분분석(Principal Component Analysis, PCA)을 이용하여 미국에 있는 소규모의 호수 수역(Lake-watershed)에서 나타나는 지하수위 곡선의 유형을 분류한 바 있다. 이러한 분석을 통하여 (1) 지하수의 함양 특성을 파악할 수 있었고 (2) 지질학적 특성이 지하수위 변동에 미치는 영향도 파악할 수 있었다. 또한 PCA를 통하여 장기 모니터링을 위한 관정의 수를 줄이므로써 모니터링 프로그램의 비용을 경제적으로 절감할 수 있다. PCA에서는 관찰치들의 공분산을 설명하기 위하여 인수분해를 이용하였다. 그러나 이것은 하나의 단순한 변환에 불과한 것이며 공분산 구조에 대한 근본적인 모형추구는 아니다. 또한 PCA는 반응치의 관찰 척도에 따라 달라지기도 한다. 그리고 주성분에 의하여 전체변동을 충분히 설명하기 위한 주성분 선택기준이 명확하게 마련되어 있지 않다.

요인분석(Factor Analysis)은 이러한 주성분분석의 결점들을 보완하는 통계적 기법으로서 이 기법의 근본목적은 가능하면 많은 변수 사이의 공분산 관계를 기술하는 데 있다. 요인분석은 주성분분석의 연장이라고 할 수 있다(강병서,1996). 수문학 분야에서 Pierre Beaudoin 등(1982)은 요인분석을 사용하여 강수량의 공간적 변동에 관하여 연구하였다. 국내에서는 한국수자원공사에서 요인분석을 이용하여 충적층 지하수의 산출 특성을 규명한 바 있다(한국수자원공사, 1996). 본 연구에서는 요인분석을 이용하여 1997년부터 1999년까지의 방대한 충적층 지하수위 자료에 대한 통계적 분류를 실시하였다.

2. 연구방법 및 결과

- 분석에 사용된 자료

국가 지하수 관측망 자료는 총 77개 충적층 지하수위 자료를 포함하고 있다. 그러

나 1997년부터 모니터링이 실시된 것부터 1999년부터 실시된 것 등 그 측정 시기가 다양하다. 요인분석을 위하여 사용한 SPSS 통계 패키지에서 열(Column) 하나 하나는 변수 곧 지점을 의미하고, 행(Row) 하나 하나는 값(Case) 곧 지하수위의 일평균값을 의미한다(Table 1). 요인분석에서는 각 지점이 모두 가지고 있는 값에 대해서만 분석결과를 산출한다. 따라서 다른 지점들은 모두 값들을 가지고 있는데 어느 한 지점만 자료를 갖고 있지 않다면 이 지점과 함께 분석한 요인분석 결과는 그 지점이 갖지 않은 수만큼 분석값을 산출하지 않는다. 그러므로 본 연구에서는 자료를 충실하게 이용하기 위하여 년도별로 나누어 분석을 실시하였다. 즉, 지점이 가장 많은 1999년을 비롯하여 1998년과 1997년 한 해 동안의 자료에 대해서도 분석을 실시하였다. 다만, 각 경우에 있어서도 각 지점의 특수한 이유로 인해 많은 양의 자료가 측정되지 않았을 경우 그 지점은 그 경우의 분석에서 제외하였다. 78개의 총적층 지점에서 지하수위의 1년 자료 중 자료가 거의 유실되지 않은 지점 수는 1999년의 경우 51개 지점이다.

Table 1. An example of data preparation in spreadsheet for factor analysis

| 충주가금 | 청원가덕 | 부여부여 | 남원도통 | 순천풍덕 | 포항연일 | 예천예천 | ... |
|--------|-------|-------|--------|------|-------|-------|-----|
| 52.243 | 76.31 | -3.76 | 104.08 | 2.89 | -0.12 | 85.97 | ... |
| 52.24 | 76.32 | -3.69 | 104.08 | 2.89 | -0.12 | 85.99 | ... |
| 52.233 | 76.28 | -3.77 | 104.08 | 2.88 | -0.09 | 85.97 | ... |
| 52.223 | 76.25 | -3.75 | 104.08 | 2.87 | -0.07 | 85.95 | ... |
| 52.21 | 76.26 | -3.74 | 104.08 | 2.87 | 0.03 | 85.97 | ... |
| 52.203 | 76.26 | -3.8 | 104.08 | 2.87 | 0.01 | 85.99 | ... |
| 52.195 | 76.23 | -3.68 | 104.08 | 2.86 | 0.13 | 85.99 | ... |
| 52.188 | 76.22 | -3.72 | 104.08 | 2.85 | -0.03 | 85.98 | ... |
| 52.175 | 76.18 | -3.74 | 104.08 | 2.85 | -0.01 | 85.94 | ... |
| 52.165 | 76.18 | -3.93 | 104.08 | 2.84 | 0.03 | 85.92 | ... |
| 52.16 | 76.17 | -2.73 | 104.08 | 2.84 | 0.09 | 85.92 | ... |
| 52.15 | 76.15 | -1.85 | 104.09 | 2.83 | 0.13 | 85.91 | ... |
| 52.14 | 76.14 | -2.27 | 104.08 | 2.83 | 0.21 | 85.9 | ... |
| 52.13 | 76.14 | -1.6 | 104.08 | 2.83 | 0.23 | 85.9 | ... |
| ... | ... | ... | ... | ... | ... | ... | ... |

- 요인 분석

요인 분석을 통하여 유사한 수위 변동 특성을 보이는 지점들을 분류하였다. 각 요인들은 유사한 변동 특성을 보인다. 요인 분석을 통하여 '설명된 총분산(Total Variance Explained)'은 다음과 같다(Table 2).

Table 2. Total Variance Explained

| Component | Initial Eigenvalues | | |
|-----------|---------------------|---------------|--------------|
| | Total | % of Variance | Cumulative % |
| 1 | 28.479 | 55.841 | 55.841 |
| 2 | 6.025 | 11.814 | 67.654 |
| 3 | 3.905 | 7.656 | 75.310 |
| 4 | 2.103 | 4.124 | 79.434 |
| 5 | 1.425 | 2.794 | 82.228 |
| 6 | 1.312 | 2.572 | 84.800 |
| 7 | 0.887 | 1.739 | 86.539 |
| 8 | 0.804 | 1.576 | 88.115 |

Extraction Method: Principal Component Analysis.

Table 2에서 Initial Eigenvalues 는 요인변량의 크기를 나타내는 수치로서 이것이 크면 집단 중에서 차지하고 있는 비중이 상대적으로 높음을 의미한다. % of Variance 는 각 Eigenvalues 가 설명한 분산의 양으로서 해당 변수의 공분산에서 차지하는 비율이 된다. 따라서 요인으로 분류될 경우 이 값이 높을수록 설명력이 있는 독립적인 요인값이 된다. 또 Cumulative % 는 % of Variance 의 누적값을 산출한 것이다. 일반적으로 요인은 고유값이 1 이상인 것을 선택한다(최기창 외, 1999). 따라서 Table 2 에서는 6개의 요인이 추출될 수 있다. 이 경우 요인 중 성분 1 이 설명할 수 있는 양은 55.841% 이고 성분 2 는 11.814% 이다. 성분 1~6 이 설명할 수 있는 양은 84.8 % 이다.

Table 3. Rotated Component Matrix

| | Component | | | | | |
|------|-----------|----------|---------|----------|---------|----------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 충주동량 | 0.957 | 0.005008 | -0.0771 | -0.01280 | 0.07149 | -0.154 |
| 진해자은 | 0.892 | 0.246 | 0.185 | 0.06504 | 0.104 | 0.05152 |
| 진주초전 | 0.891 | -0.07846 | 0.132 | 0.137 | 0.260 | 0.156 |
| 밀양가곡 | 0.303 | 0.710 | 0.184 | 0.143 | 0.268 | 0.108 |
| 울산온양 | -0.09478 | 0.155 | 0.819 | 0.02919 | 0.214 | 0.09101 |
| 예천예천 | 0.07771 | -0.145 | 0.335 | 0.800 | 0.146 | 0.176 |
| 강릉홍제 | 0.448 | 0.135 | 0.310 | 0.205 | 0.716 | -0.07384 |
| ... | ... | ... | ... | ... | ... | ... |

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Table 2 와 Table 3 모두 요인 추출 방법은 주성분분석을 사용하였다. Table 3 은 회전된 성분 행렬(Rotated Component Matrix) 이다. 여기서 사용한 회전방법은 Kaiser 정규화가 있는 Varimax 방법이며 12 번의 반복계산 끝에 요인회전이 수렴되었다. 이것은 회전한 후

요인을 6 개로 분류하고 이 분류된 요인과 각 변수(지점)간의 상관관계를 나타낸 것이다. 그 결과 충주동량, 진해자은, 진주초전은 요인 1에 부하량이 가장 크게 걸려 있고, 밀양가곡은 요인 2에, 울산온양은 요인 3에, 예천예천은 요인 4에, 강릉홍제는 요인 5에 부하량이 각각 가장 크게 걸려 있다. 이와 같은 결과로부터 51개 지점의 층적층 지하수위가 각각 어떤 요인에 가장 큰 부하량이 걸렸는지를 파악하므로써 요인별로 분류하였다. 이와 같이 파악된 요인들이 대표하는 지하수위 변동 곡선은 Figure 1 과 같다. 각 그림에서 알 수 있듯이 요인별로 지하수위 변동이 분류되었다. 이 때 Component scores 그래프에 가까운 정도는 'Rotated Component Matrix'에서 확인할 수 있다(충주동량:0.957 ; 진해자은:0.892 ; 밀양가곡:0.710 ; 남원도통: 0.651).

3. 토의 및 결론

본 연구에서는 방대한 국가 지하수위 관측망 자료를 통계적으로 분류하였다. 기존에는 지하수위의 유형을 단순히 연구자의 주관에 따라 분류하였으나 이러한 방법은 객관성이 결여된 방법이라고 할 수 밖에 없다. 본 연구에서 사용한 요인분석은 검증된 통계패키지를 이용하여 손쉽게 분류를 실시할 수 있다. 이 분류 결과는 앞에서 열거한 것과 같이 다양한 수리지질학적 연구에 사용할 수 있을 뿐만 아니라 본 연구에서 계속 진행 예정인 우리나라 함양량 평가를 위한 부지 선정에도 사용될 수 있다. 단, 향후 분류된 수위곡선에 대한 지질, 토양, 토지사용 그리고 층적층의 다양한 자료들과의 상관관계 분석이 실시되어야 할 것으로 사료된다.

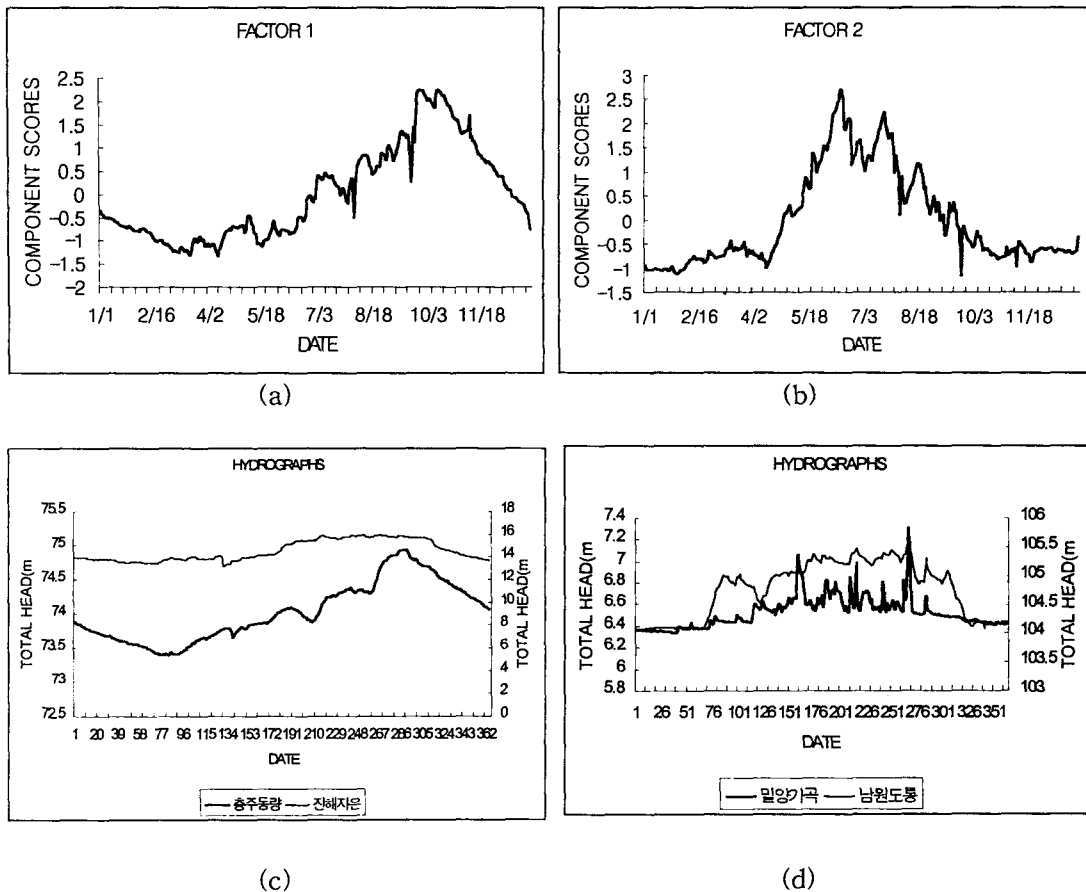


Figure 1. Component scores and Ground water hydrographs ; (a) component scores for factor 1 ; (b) component scores for factor 2 ; (c) water level in well Chungju and Jinhae ; (d) water level in well Milyang and Namwon.

4. 참고문헌

- 최기창, 박혜련, 오세열, 1999, 실무자를 위한 통계학 -SPSS/win 의 활용-, pp. 234-247, 한울출판사.
- 강병서, 행렬과 SPSS/PC' 이용 다변량통계분석 [2판], 1996, p. 413, 학현사.
- 한국수자원공사, 1996, 전국 층적층 지하수 조사 보고서, pp. 46-47.
- Winter, T.C., Mallory, S.E., Allen, T.R., and Rosenberry, D.O., The use of principal component analysis for interpreting ground water hydrographs, GROUND WATER, 38(2), pp. 234-246.
- Pierre Beaudoin and Jean Rousselle, 1982, A study of space variations of precipitation by factor analysis, Journal of Hydrology, 59, pp. 123-138.