

# 공간 통계 분석을 이용한 DEM 오차 패턴 연구

안 은 자

경희대학교 지리학과

## 1. 서론

지리학적 정보는 지구의 표면이나 가까이에 나타나는 현상과 사상에 대한 정보로서 정의된다 (Goodchild et al., 1999). 지리학에서, 이러한 지리학적 정보는 특정한 현상을 연구하기 위한 공간자료로 이용되는데, 이는 공간적 패턴을 통해 유형화된다. 이러한 공간자료는 현지답사를 통해 수집·분석되며, 관찰자의 주관적 판단, 기술적인 오류로 인해, 오차의 필연적 발생 가능성을 안고 있다 (Maffini, 1989; Bolstad, 1990; Dunn, 1990; Keefer, 1991). 따라서 오차가 포함되어 있는 공간자료를 대상으로 한 GIS의 공간분석 결과는 분석의 단계가 복잡할수록, 오차가 확산될 수 있는 개연성이 있다. 이러한 맥락에서 결과에 대한 신뢰도 평가에 있어 공간자료 오차는 심각하게 고려되어야 한다 (Muller, 1987; Goodchild 등, 1992).

이에, 본 연구는 지리학 연구에서 유용하게 활용되는 공간 자료인 DEM<sup>1)</sup>을 대상으로 새로운 오차 분석을 시도하고자 한다. 현재 DEM의 오차에 관한 연구는 자료의 오차 발생원인을 확인하거나 전체적인 오차 값을 추정하는데 초점이 맞추어져 있다 (Burrough, 1992b; Lanter and Veregin, 1992; Klinkenberg and Joy, 1994; Burrough and McDonnell, 1998). 그런데 이러한 유형의 오차평가는 오차의 정량적인 합계만을 고려한 것으로 오차 자체의 공간적 패턴을 해석할 수 없다는 한계점을 드러내고 있다. 결국, DEM의 오차를 감소시켰다는 총량의 수치적 평가는 부정확할 수 있기 때문에, 그에 따른 공간 자료 또한 신뢰할 수 없게 된다. 따라서, DEM의 정확성을 높이고 궁극적으로 자료의 신뢰도를 높이기 위해, 오차의 공간적 특성을 반영할 수 있는 탐색적 공간 자료 분석 (exploratory spatial data analysis; ESDA)이 이루어져야 할 필요성이 대두된다.

## 2. 이론적 배경과 기술적 연구 동향

### 1) DEM 오차

DEM은 “공간상에 나타나는 지형기록의 연속적 변동의 수치적 표현”으로 정의된다. 고도 자료는 고도가 규칙적인 격자행렬 (gridded matrix), TIN, 그리고 등고선의 각 셀에 대해 추정되는 그리드 모델을 포함하는 많은 방법으로 수치화된다. 이러한 DEM의 오차는 자료 수집·입력·조작의 실수 또는 결합의 의미 뿐만 아니라, 변동 (variation)의 의미인 오차의 통계적 개념도 나타낸다 (Burrough, 1986). 한편, DEM 오차에 대한 연구의 대부분은 DEM 생성 방법인 내삽 기법과 관련되어 진행되었다. 이는 DEM을 만드는 내삽 기법, 즉 TIN, Spline, Kriging, IDW 등을 이용하여 고도 점을 연속면으로 만든 다음, 원래의 고도 점과 만들어진 DEM 간의 차이를 RMSE 값으로 나타내어 오차를 평가하는 것이다. 최근 DEM 오차 연구는 공간 통계 기법을 적용하는 연구로 차츰 이행되고 있다.

### 2) 공간 통계 분석

공간 통계 분석은 자료의 다양한 공간적 속성을 하나의 변수로만 취급하던 기존 통계 분석의 일

1) DEM (Digital Elevation Model; 수치고도모형)은 경사도, 사면방향, 지형단면 등과 같이 한 지역의 기본적인 지형 특성을 분석하거나 (Lee et al., 1992; Wood), 지형과 관련된 통계적 분석 등 (Fisher, 1993), 지리학 연구에 유용하게 활용된다.

부분으로, 공간적으로 상호 연관성이 있는 자료에 통계적 기법을 적용하여 실세계의 다양한 문제들을 해결하는 방법이다. GIS 공간 분석은 Bailey의 공간분석 정의에 따라, 자료의 선택, 자료의 처리, 자료의 탐색적 분석과 자료의 확정적 분석으로 나뉜다. 이러한 ESDA는 공간자료의 공간 분포를 묘사하고 시각화하며, 불규칙적인 위치나 공간적인 이례지점(outliers)을 확인하고, 공간적 체계 등의 공간 자료의 패턴을 발견하기 위하여 공간정보들을 효율적으로 제시하고 요약하는 방법이다 (Anselin, 1998). 이는 탐색적 자료 분석(Exploratory Data Analysis, EDA)<sup>2)</sup>의 확장된 분야로 특히, 자료의 공간적 특성인 공간적 자기상관성과 공간적 이질성을 분석하는데 중점을 둔다. 공간자료의 두 가지 속성, 즉 공간적 의존성과 이질성은 DEM의 오차에도 그대로 반영될 가능성이 높다. 따라서 공간상의 한 위치에서 발생하는 오차와 그 외 지역에서 발생하는 오차와의 기능적 상관관계를 면밀하게 분석할 필요가 있다. 이는 공간 통계적 DEM 오차 연구를 통해 오차의 공간적 패턴에 따라 오차 수정 전략과 구체적 오차 수정 모형을 가장 적절하게 제시할 수 있기 때문이다. 본 연구에서는 'ESDA'의 범주에 속하는 몇 가지 방법들을 통해, 그 가능성을 실험하고 DEM의 오차 혹은 정확성 지표로서 이용가능성을 확인하려 한다. 먼저 DEM 오차의 상관성 분석을 위해, Spatial neighbors를 이용해 이웃한 폴리곤 간에 가중치를 부여하고 전역적 공간 자기상관(Moran's I와 Geary's C)과 국지적 공간적 자기상관을 분석한 후, 이를 구조적으로 설명하기 위해 변동도 분석을 수행한다.

### 3. 분석 자료 구축

#### 1) DEM 자료 구축과 오차 확인

본 연구는 1: 25,000 축척의 수치지형도를 사용하였으며, 연구 대상 지역은 도엽 명 하봉(충청남도 공주시)에 해당하는 곳으로 평지와 산지, 그리고 계곡이 함께 분포하는 지형적 특성을 갖고 있다. 레이어(layer) 7111(10m 간격의 주곡선)과 7114(50m 간격의 계곡선)를 내삽에 이용하였으며, DEM의 그리드 공간 해상력은 지상 5m, 10m, 15m로 선정하였다(평지 표본 지역 기준). 본 연구의 전체 대상 지역에서 지형 특성에 따라 평지, 산지, 계곡의 3가지 지역을 표본연구지역으로 분리 선정하였다. DEM에서 오차를 추출하여 이들의 공간 분포 패턴을 확인하기 위해서, 먼저 DEM을 지형별 표본 지역(평지, 산지, 계곡), 내삽 방법(IDW, TIN, 크리깅), 해상도(5, 10, 15m)에 따라 구분하여 구축하였다.

DEM 오차를 비롯한 공간 자료의 오차는 실제와 실제의 표현 간 차이라고 할 수 있다. 일반적으로,  $x$ 라는 공간 속성의 실제 값을  $a(x)$ 라 하고, 이의 표시를  $b(x)$ 라고 하면, 위의 정의에 따른 오차  $v(x)$ 는  $a(x)$ 에서  $b(x)$ 를 뺀 단순한 수치적 차이로 표현된다. 이를 모델화 한 표준 자료 오차 모델은 실제 면과 왜곡 요소(오차)를 더하여 고도 자료 면  $Z$ 를 구하는 것이다. 여기에서 오차는 연속된 공간 현상으로 다루지기 때문에, 스칼라(scalar)가 아니라 필드(field)의 개념이다. 또한 오차에 대한 정보는 가능한 값의 범위와 분포를 지정하기 위해 한정되는데, 이런 유형의 정보는 확률변수(random variable)로 표현된다. 따라서, 확률변수로 표현되는 DEM 오차는 공간 변동을 포함하고, 이는 다음 공간 변화의 일반 모델(GMSV: General Model of Spatial Variation)식으로 설명할 수 있다. 이에 따라, 본 연구에서 지형별로 구축된 DEM과 해상도 별로 구축된 DEM을 대상으로 추출한 오차는 그리드 자료 형태로 처리한다.

$$Z(x) = \mu(x) + \epsilon(x) \quad (x \in D) \quad \text{수식 1}$$

$\mu(x)$ 는 경향으로서 알려진 실제 값, 대규모의 변동

$\epsilon(x)$ 는 공간적으로 자기상관된 오차, 소규모의 변동(Cressie, 1991)

$D$  :  $n$  차원의 공간  $R^n$  의 영역(대부분의 경우 2차나 3차이지만 실제의 제한은 없다.)

2) EDA는 자료의 패턴을 발견하기 위한 기술적 또는 그래픽적 통계분석 방법들의 집합체를 말하며, 정보의 제시 및 요약이 주요 목적이다.

#### 4. DEM 오차의 공간적 특성

##### 1) 비공간 탐색 기법을 이용한 오차 분석

자료의 전반적인 특징을 알아보기 위해서 기술 통계 분석을 실시하였다. 분석 결과 중 특이할만한 것은 산지 지형의 IDW DEM만이 해상도에 따른 분산과 표준편차 값에 차이를 나타내지 않았다는 것이다. 이는 산지 지형의 특성상 고도점이 밀집해있고 가까이 위치해 있는 점에 가중치를 주는 IDW 내삽 기법의 계산과정의 결과라 할 수 있다. 또한 RMSE를 이용한 분석에서 크리깅을 이용한 DEM의 정확도가 상대적으로 우수한 것으로 나타났다. 하지만, RMSE를 이용한 오차표준을 만족시키는 DEM이라고 할지라도, 그 오차가 체계적인 분포특성을 가지고 있다면 그 DEM은 정확성이 높은 자료라고 할 수 없다.

표 1. 지형별, 해상도별, 내삽방법별 RMSE

| 지형 | 내삽  | 해상도 | RMSE | 지형 | 내삽  | 해상도 | RMSE | 지형 | 내삽  | 해상도 | RMSE |
|----|-----|-----|------|----|-----|-----|------|----|-----|-----|------|
| 평지 | IDW | 5m  | 5.91 | 산지 | IDW | 5m  | 4.64 | 계곡 | IDW | 5m  | 6.54 |
|    |     | 10m | 6.12 |    |     | 10m | 4.32 |    |     | 10m | 6.2  |
|    |     | 15m | 9.02 |    |     | 15m | 5.32 |    |     | 15m | 7    |
|    | 크리깅 | 5m  | 3.61 |    | 크리깅 | 5m  | 2.71 |    | 크리깅 | 5m  | 3.75 |
|    |     | 10m | 4.56 |    |     | 10m | 3.16 |    |     | 10m | 5.76 |
|    |     | 15m | 6.12 |    |     | 15m | 3.08 |    |     | 15m | 6.83 |
|    | TIN | 5m  | 4.33 |    | TIN | 5m  | 5.94 |    | TIN | 5m  | 7.06 |

비공간적 탐색 기법(aspacial exploration methods)은 자료의 비공간적 분포를 분석하여 속성값 자체가 가지는 분포특성을 밝히는 방법이다. 도수분포도(histogram)는 자료의 분포 특성을 표현하는 가장 단순한 기법으로, 속성값에 대한 최대값·최소값·범위, 속성값의 계급분류 가능성, 속성값의 비대칭성 여부, 이례지점(outlier)의 존재 여부 등이 이에 나타난다. 지형별 DEM 오차는 평지-계곡-산지의 순으로 낮게 나타나는데, 이는 측정 자료 수의 차이라고 할 수 있다. 한편, 평지 표본 지역의 경우 전체 RMSE 값은 크리깅 내삽 보다 IDW 내삽을 사용한 DEM이 더 크게 계산되었지만, 도수분포도를 보면 오차의 크기 면에서 IDW 내삽을 이용한 DEM의 오차가 더 작게 분포함을 알 수 있다. 이는 RMSE 값이 단순히 오차의 가감을 합하여 이를 상쇄해버리는 오류를 포함하고 있음을 알려준다. 즉 실제로는 절대값 만큼 오차가 발생한 것이지만, 양과 음의 값이 양적으로만 계산되어 이를 정확히 반영할 수 없다는 것이다. 한편, 해상도 별 도수분포도는 오차 측정 점의 수에 따른 결과이므로, 유의하지 않게 판단된다. 이처럼 도수분포도를 통해 단일 값으로는 알 수 없는 분포(경향)를 살펴봄으로써, 오차를 분석하는데 보다 자세한 정보를 얻을 수 있다. 하지만, 이러한 도수분포도는 경향만을 나타 낼 뿐, 공간적 위치는 파악할 수 없는 단점을 지니고 있다. 따라서, 보다 공간자료의 분석에 적합한 공간적 탐색 기법을 이용해 DEM 오차를 분석해야 하겠다.

##### 2) 공간적 탐색 기법(ESDA)를 이용한 오차 분석

DEM 오차의 군집의 정도를 확인함에 있어서, 오차의 공간적 패턴을 어느 정도까지 반영하는 오차의 확률적 모델을 만드는 것이 필요하다 (Monckton, 1994). 공간적 자기상관은 일반적으로 비슷한 오차 값 간의 공간적 연관성의 정도의 편리한 추정을 제공하는 Moran'I 통계(Cliff and Ord, 1981)를 이용하여 추정한다. DEM 오차는 공간적인 자료의 내부에서 발생하는 것으로, 그 발생과

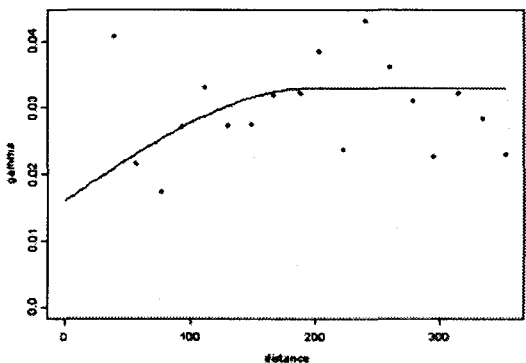


그림 1 DEM 오차의 변동도

위치에 있어서 서로 연관성을 가지고 있으리라 추측된다. 이를 통계적으로 검증하기 위해서 DEM 오차의 공간적 자기상관성의 존재 유무를 분석하였다. 본 연구에서 Moran'I 통계치는 1000m 범위에서 2.057(p=0.02)로 오차간 자기 상관성이 존재한다고 할 수 있다. 자료의 공간적 자기상관의 단일 추정치는 오직 공간적 연관성의 국지적인 정보만을 제공하기 때문에 다소 한계가 있다. 따라서, 공간적 자기상관 분석과 더불어 공간 패턴을 보다 유용하게 묘사하며 자기상관성의 변화를 설명하는 변동도(variogram)분석을 수행해야 한다. 경험적 변동도 패턴에 따라 계산한 DEM 오차의 변동도는 그림 1과 같다.

## 5. 결론

본 연구에서는 수치지도의 등고선으로부터 생성된 DEM 오차를 평가하는데 있어서 지금까지 대표적인 척도로 사용되고 있는 RMSE를 보완할 수 있는 방안을 탐색적 공간자료 분석기법(ESDA)을 통해 모색하였다. 이를 위해 3가지 대표적인 내삽기법으로 DEM을 구축하고, 탐색적 공간자료 분석 기법을 통해서 RMSE와 같은 단일 오차 척도에서 발견할 수 없는 DEM 오차의 공간적 분포 특성을 파악하였다. 비공간 통계기법인 도수분포도에서 확인한 내삽별 오차 분석은 RMSE값과는 다른 오차의 패턴을 보여주었다. 특히, 평지 표본 지역의 오차 분석에서 전체 RMSE 값은 크리깅 내삽 보다 IDW 내삽을 사용한 DEM이 더 크게 계산되었지만, 도수분포도에선 이와는 반대의 결과가 나타났다. 즉, IDW 내삽을 이용한 DEM의 오차가 크리깅 내삽 DEM 오차 보다 그 분포면에서 더 적은 경향을 보이는 것이다. 이는 RMSE 값이 단순히 오차의 가감을 합하여 이를 상쇄해버리는 오류를 포함하고 있음을 알려주며, 이로써, DEM 오차를 RMSE 단일 값으로는 정확히 반영할 수 없음을 분명히 밝혀준다. 또한, DEM 오차의 공간적 자기 상관성을 밝히고, 이를 구조적으로 설명하기 위해 변동도 분석을 수행하였다. 이는 오차의 공간적 유사성을 밝히는 것으로 지도상에서 오차의 범위를 정의할 수 있게 해줄 것이다. RMSE는 공간적으로 체계적 분포패턴을 갖는 오차를 밝히기 어렵다는 한계를 갖지만 공간적 탐색 기법은 내삽기법에 따라 발생하는 체계적 오차의 분포 특성을 분석함으로써 궁극적으로 DEM의 자료 질을 향상시킬 수 있도록 내삽기법을 조정하는데 유용한 수단으로 이용될 수 있다.

## 참고문헌

- 황철수, 1999, "DEM 오차 평가 방법에 관한 연구", 한국지형공간정보학회논문집, 7(2), pp. 23-34.
- Burrough, P.A. and McDonnell, R., 1998, Principles of Geographic Information Systems, Oxford University Press, New York, pp.39.
- Charlisle, B. H., 2000. "The Highs and Lows of Digital Elevation Model(DEM) Error: Developing a Spatially Distributed DEM Error Model", Proceedings of 5th International Conference on GeoComputation, University of Greenwich, United Kingdom 23-25 August.
- Ehlschlaeger, C. R., 2002, "Representing Multiple Spatial Statistics in General Elevation Uncertainty Models: Moving Beyond the Variogram", Geographical Information Science, 16(3), pp. 259-285.
- Ehlschlaeger, C. R. & A. Shortridge, 1996, Modeling Elevation Uncertainty in Geographical Analyses, Proceeding of the International Symposium on Spatial Data Handling, Delft, Netherlands, 9B. 15-9B. 25.
- Wechsler, S. P., 1999, "Digital Elevation Model (DEM) Uncertainty: Evaluation and Effect on Topographic Parameters", ESRI User Conference, San Diego, CA, July.