

# 무작위합역 절차의 다양성에 대한 시뮬레이션 연구

이상일\* · 이몽현\*\*

## The Diversity of Random Spatial Aggregation Procedures: A Simulation Study

Sang-Il Lee\* · Monghyeon Lee\*\*

**요약 :** 본 논문의 주된 연구목적은 무작위합역 절차의 다양성을 검토하고 서로 다른 무작위합역 절차의 특성을 비교 분석함으로써 상대적으로 적절성이 높은 무작위합역 절차를 선정하여 제시하는 것이다. 기본적인 무작위합역 절차가 정식화되고, 효율적인 실행 알고리즘으로서 공간근접성행렬에 기반한 방법이 소개된다. 무작위합역 절차의 다양성을 제공해주는 두 가지 차원을 바탕으로 모두 여섯 가지의 서로 다른 무작위합역 유형이 도출되었다. 이 유형의 서로 다른 특성을 파악하기 위해 시뮬레이션 실험이 이루어졌고, 그 결과가 연결도와 원형도의 두 가지 기준에 의거해 평가되었다. 시뮬레이션 실험의 주요 결과를 요약하면 다음과 같다. 첫째, 방식 B가 방식 A와 방식 C에 비해 연결도 측면에서 상대적으로 일관성 있는 결과를 산출해 준다. 둘째, 방식 2가 방식 1에 비해 최종구역의 형태를 보다 원형에 가까워지도록 만들어준다. 셋째, 방식 B, C(특히 B)가 원형도 측면에서 상대적으로 일관성 있는 결과를 산출해 준다. 시뮬레이션 연구 결과, B2 유형이 가장 적절한 무작위합역 절차인 것으로 드러났다. B2 유형은 상대적으로 원형도가 높은 최종구역을 산출할 뿐만 아니라, 연결도와 원형도 두 가지 측면 모두에서 상대적으로 일관성 있는 결과를 산출하는 것으로 평가되었다. 본 연구는 무작위합역 절차가 고정불변의 것이 아니라 다양하게 정의될 수 있으며, 그 다양성 속에서도 무작위성과 현실유관성의 관점에서 상대적으로 더 적절한 선택이 가능할 수 있다는 점을 실험 연구를 통해 보여주었다는 점에서 그 학문적인 의의가 있다고 평가할 수 있다.

주요어 : 합역, 무작위합역, 공간근접성행렬, 연결도, 원형도, 시뮬레이션

**Abstract :** The main objectives of this study are to discuss on the diversity of random spatial aggregation, to compares pros and cons of the different random spatial aggregation procedures, and to recommend a procedure which is believed to be superior in validity. A basic procedure for random spatial aggregation is proposed and an implementation algorithm based on the spatial proximity matrix is introduced. All six distinctive algorithms are derived on the basis of two dimensions which facilitate the diversity of random spatial aggregation. A simulation experiment is carried out to compare pros and cons of the six procedures and its results are evaluated based on the contiguousness and compactness criteria. Main findings are as follows. First, Category B seems to be more consistent in terms of contiguousness in comparison to Category A and C. Second, Category 2 appears to generate more compact final zones in comparison to Category 1. Third, Category B and C (especially B) seem to be more consistent in terms of compactness. Based on all the experiment results, Type B2 seems to be the superior algorithm for random spatial aggregation. It tends to generate more compact final zones and more consistent results in terms of both contiguousness and compactness. This study can be seen as a significant

\*서울대학교 사범대학 지리교육과 교수(Professor, Department of Geography Education, College of Education, Seoul National University, si\_lee@snu.ac.kr)

\*\*삼성전자 메모리사업부 책임연구원(Staff Engineer, Memory Business, Samsung Electronics Co. Ltd., monghlee@gmail.com)

achievement in the sense that it shows that the concept of random spatial aggregation can be considered in a diversified manner, and one out of a certain group of candidates can be chosen as superior on the basis of the degree of randomness and relevancy.

Key Words : Spatial aggregation, Random spatial aggregation, Spatial proximity matrix, Contiguousness, Compactness, Simulation

## I. 서론

공간적 애그리게이션(spatial aggregation)은, GIS 맥락의 가장 단순한 정의를 따르면, 인접한 두 개 이상의 공간적 개체를 묶어 하나의 집합적 개체를 만드는 과정으로 볼 수 있다(Wade and Sommer, 2006, 4). 좀 더 구체적으로 살펴보면, 벡터 데이터의 맥락에서는 디졸브(dissolve) 오퍼레이션의 적용과 그에 따른 속성합산(attribute aggregation)을 의미하는 일종의 재분류(reclassification) 기법으로 이해되기도 하고(Bolstad, 2019, 396), 래스터 데이터의 맥락에서는 공간 해상도를 낮추는 일종의 일반화(generalization) 기법으로 이해되기도 한다(Chang, 2019, 270). 그러나 공간데이터분석(spatial data analysis)의 보다 폭넓은 맥락으로 보면, 공간적 애그리게이션은 공간단위 체제의 전환 과정으로 이해되어야 한다(Sun and Wong, 2017). 특히, 해당 공간데이터가 2차원 역형(域形) 개체로 구성되어 있다면, 이는 “보다 면적이 작고 다수인 공간단위로 구성된 체제를 보다 면적이 크고 소수인 공간단위로 구성된 체제로 전환하는 과정”으로 정의될 수 있고, 이에 대한 번역어로 합역(合域)을 사용하는 것에 큰 무리는 없을 것으로 생각된다.

공간데이터분석에서의 합역 개념을 보다 명확히 하기 위해서는 다음의 두가지 개념적 논의가 필요하다. 첫째, 투입 공간단위 체제나 합역 과정을 통해 생성된 산출 공간단위 체제나 모두 MECE(mutually exclusive and collectively exhaustive, 상호배제와 전체포괄)의 조건을 만족시켜야 한다. 즉, 공간단위들은 서로 중첩되지 말아야 하며, 모두 합쳤을 때 전체 지역을 완전히 구성할 수 있어야 한다(김감영, 2014; Bailey and Gatrell, 1995). 둘째, 합역은 오로지 인접한, 즉 경계를 공유하고 있는 공간단위 혹은 구역 간에서만 발생한다. 이를 연결성 요구(contiguity requirement) 혹은 제약(constraint)이라고 하는데, 최초의 합역 과정에서는 인접한 공간단위 사이에서, 합역이 진행되는 과정에서는 중간에 생성된 합역체와 공간단위 사이에서, 오로지 연결성이 존재하는 경우에만 병합이 이루어진다. 따라서 합역은, 연결성 요구를 만족시키는 병합 과정을 통해, 투

입 공간단위 체제(기본 공간단위로 구성)를 산출 공간단위 체제(최종구역으로 구성)로 전환하는 과정을 의미하며, 이 과정에서 모든 기본 공간단위는 예외없이 특정 최종구역의 한 요소가 되어야 한다.

이렇게 정의된 합역 개념은 지리학 혹은 지리정보과학의 보다 광범위한 세 가지 연구 영역과 연결되어 있다. 첫째, 지역구분(regionalization) 연구이다. 지역구분은 “전체 지역을 구성하는, 상대적으로 독립적인 하위지역(sub-region)을 확인하는 과정”으로 정의할 수 있다. 그런데 데이터-기반의 지역구분은 결국 “기본 공간단위를 병합하여 하위지역을 구성하는 과정”을 의미하게 되기 때문에 지역구분은 합역 개념과 필연적으로 연결될 수밖에 없다(이상일, 1999; 이상일 등, 2012; Koo, 2012; Jaegal, 2013; Adams *et al.*, 2016). 둘째, 재구획(redistricting) 연구이다. 재구획은 “특정한 기본 공간단위를 바탕으로, 특정한 목적함수를 만족시키는 최적의 구획 체제를 형성 혹은 재형성하는 과정”으로 정의될 수 있는 데, 주로 선거구나 학구의 경계 획정/재획정과 관련되어 있다. 간단히 말해, 재구획은 합역 개념을 공간최적화(spatial optimization)와 연결시킨 것이다(김감영 등, 2009; 김영훈, 2009; 김감영 등, 2010; 강영욱 · 조선희, 2012; 김명진, 2012; 김감영, 2014; Cope, 1971; Macmillan, 2001; Kim *et al.*, 2016). 셋째, MAUP(modifiable areal unit problem, 공간단위 임의성의 문제) 연구이다. MAUP 연구는 공간단위 체제가 변화함에 따라 통계적 결과가 어떠한 방식으로 달라지는지, 그리고 공간단위 체제와 통계적 결과 사이에 어떠한 체계적인 관련성이 있는지를 탐구한다(이상일, 1999; 김감영, 2011; Fotheringham and Wong, 1991; Amrhein, 1995; Green and Flowerdew, 1996; Wong, 1996; Flowerdew *et al.*, 2001; Stillwell *et al.*, 2018; Lee *et al.*, 2019). MAUP 효과는 스케일 효과(scale effect)와 구획 효과(zoning effect)에 의해 발생하는데, 이러한 두 가지 효과를 체계적으로 분석하기 위해서는 합역에 바탕을 둔 시뮬레이션 연구가 필수적이다. 본 연구는 앞의 세 가지 연구 주제 중 MAUP과 관련된 합역 절차에 초점을 맞추고자 한다. 지역구분과 재구획을 위

한 것이 일종의 목적함수 기반 합역 절차라면(Kim *et al.*, 2016), MAUP 연구를 위한 것은 무작위합역(random spatial aggregation) 절차이다. 이러한 무작위합역 절차를 요약하면 다음과 같다. 초기 공간단위 체계를 선정하고, 여러 개의 서로 다른 합역 수준(혹은 스케일 수준)을 설정한다. 각 합역 수준에 대해, 구역 방식이 서로 다른 합역체를 가능한 많이 생성한다. 이렇게 하면 다수의 합역 수준에 대해, 구역 방식을 달리하는 매우 많은 수의 합역체가 만들어진다. 이 합역체별로 공간통계분석의 결과를 산출하고, 그것을 서로 비교하면, MAUP의 스케일 효과와 구역 효과를 동시에 검토해 볼 수 있는 시뮬레이션의 프레임이 완성되는 것이다. 그런데, 무작위합역 절차에 기반한 MAUP 연구가 매우 다양하게 진행되어 왔음에도 불구하고(Fotheringham and Wong, 1991; Amrhein, 1995; Green and Flowerdew, 1996; Wong, 1996), 무작위합역의 내부 알고리즘은 선명하게 드러나 있지 않거나, 크게 중요하지 않는 것으로 취급되어 왔다. 그러나 최근의 몇몇 연구는 무작위합역 절차에도 다양성이 존재할 수 있으며, 어떤 무작위합역 절차를 선택하느냐에 따라 MAUP의 연구 결과가 달라질 수도 있음을 지적하고 있다(Flowerdew *et al.*, 2001; Lee *et al.*, 2019).

따라서 본 논문의 주된 연구목적은 무작위합역 절차의 다양성을 검토하고, 서로 다른 무작위합역 절차의 상대적 특성을 비교 분석함으로써, 상대적으로 적절성이 높은 무작위합역 절차를 선정하여 제시하는 것이다. 우선, 기본적인 무작위합역 절차를 정식화하고, 효율적인 실행 알고리즘으로서 이항 공간근접성행렬에 기반한 방법을 제안한다. 무작위합역 절차의 다양성을 제공해주는 두 가지 차원에 대한 논의를 바탕으로 다수의 무작위합역 절차를 제시한다. 무작위합역 절차의 상이한 특성을 검토하기 위해 시뮬레이션 실험을 실시한다. 서로 다른 무작위합역 절차를 두 가지 기준에 근거하여 상호 비교함으로써 상대적으로 적절성이 높은 한 가지 절차의 사용을 제안한다.

## II. 무작위합역 절차의 기본 원리와 다양성

### 1. 무작위합역의 기본 원리

무작위합역의 기본 상황은  $n$  개 기본 공간단위(basic spatial units)를 합역하여  $m$  개의 최종구역(final zones)을

형성하는 것이다. 무작위합역은 기본적으로 병합 발생에서의 무작위성이 담보되어야 하며, 합역의 매 라운드마다 이전 라운드에서 만들어진 합역체가 연접한 기본 공간단위 하나씩을 병합해나가 최종 라운드에서 최종구역 설정을 완수하는 반복적, 점진적, 팽창적 과정을 통해 이루어진다. 우선, 전체 기본 공간단위 중  $m$  개의 공간단위를 무작위로 선정하여 시드 공간단위(seed spatial unit)로 사용한다. 모든 합역은 이 시드 공간단위를 중심으로 이루어지며 최종적으로 이 시드 공간단위를 하나씩 포함하는 최종 구역을 확정한다. 매 라운드마다 합역 과정을 통해 중간 단계에 해당하는 라운드구역(round zone)이 생성될 수 있으며, 차후 라운드의 합역 과정의 기본 단위가 된다. 또한 매 라운드마다 더 이상 병합이 불가능한 최종구역이 확정될 수 있으며, 한번 확정된 최종구역은 차후 라운드의 합역 과정에 가담하지 않는다. 라운드 1은 시드 공간단위에 대한 것이기 때문에 특징적이지만, 라운드 2부터는 이전 라운드에서 생성된 라운드구역의 병합 과정을 정의하는 부분이라 최종 라운드까지 동일한 과정이 반복된다. 이 과정을 정리하면 다음과 같다.

#### [라운드 1]

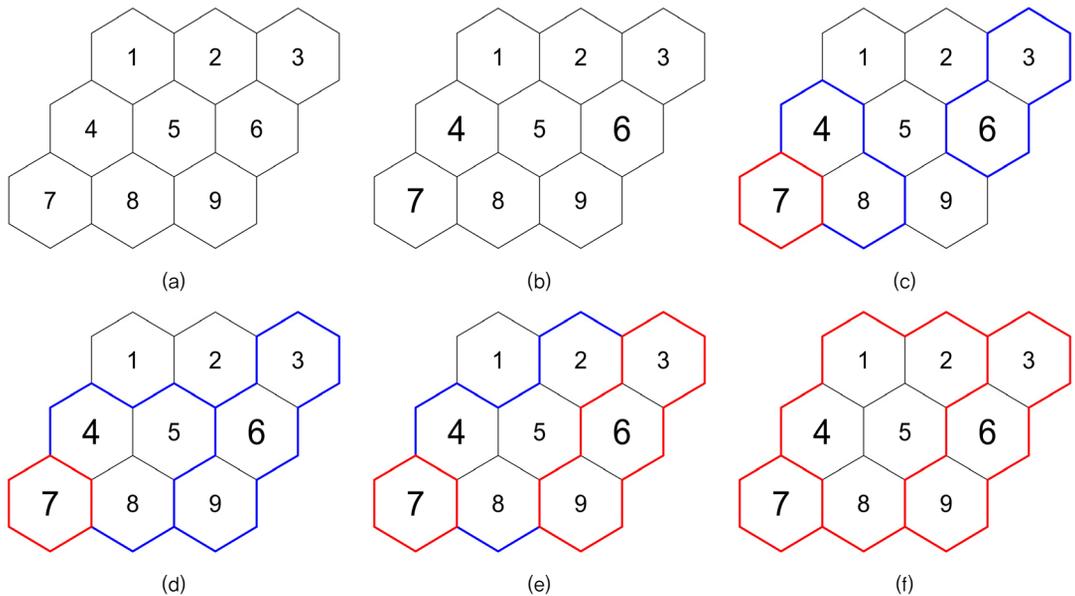
- R1-1: 시드 공간단위에 대해 무작위 순위를 부여한다. 순서에 따라 각 시드 공간단위를 고려한다.
- R1-2: 각 시드 공간단위에 대해, 연접한 기본 공간단위의 개수를 구하고 다음의 규칙에 따라 합역 절차를 진행한다.
  - R1-2-1: 0개인 경우, 시드 공간단위를 최종구역으로 확정한다.
  - R1-2-2: 1개인 경우, 시드 공간단위에 해당 기본 공간단위를 병합한다. 이 결과 생성된 구역을 라운드 1의 라운드구역이라고 부른다. 단, 해당 기본 공간단위가 남아있는 유일한 기본 공간단위일 경우 생성된 구역을 최종구역으로 확정한다.
  - R1-2-3: 2개 이상인 경우, 무작위로 한 개의 공간단위를 선택한 후, 시드 공간단위에 병합한다. 이 결과 생성된 구역을 라운드 1의 라운드구역이라고 부른다.
- R1-3: 확정된 최종구역은 다음 라운드의 합역 과정에 개입하지 않는다. 생성된 라운드구역을 검토한다.
  - R1-3-1: 라운드구역의 개수가 0이면 합역 절차를 중단하고 최종 결과를 산출한다.
  - R1-3-2: 그렇지 않으면 다음 라운드로 진전한다.

**[라운드 2 ~ 최종 라운드]**

- R2-1: 이전 라운드에서 생성된 라운드구역에 대해 무작위 순위를 부여한다. 순서에 따라 각 라운드구역을 고려한다.
- R2-2: 각 라운드구역에 대해, 연접한 기본 공간단위의 개수를 구하고 다음의 규칙에 따라 합역 절차를 진행한다.
  - R2-2-1: 0개인 경우, 라운드구역을 최종구역으로 확정한다.
  - R2-2-2: 1개인 경우, 라운드구역에 해당 기본 공간단위를 병합한다. 이 결과 생성된 구역을 해당 라운드의 라운드구역이라고 부른다. 단, 해당 기본 공간단위가 남아있는 유일한 기본 공간단위일 경우 생성된 구역을 최종구역으로 확정한다.
  - R2-2-3: 2개 이상인 경우, 무작위로 한 개의 공간단위를 선택한 후, 라운드구역에 병합한다. 이 결과 생성된 구역을 해당 라운드의 라운드구역이라고 부른다.
- R2-3: 확정된 최종구역은 다음 라운드의 합역 과정에 개입하지 않는다. 생성된 라운드구역을 검토한다.
  - R2-3-1: 라운드구역의 개수가 0이면 합역 절차를 중단하고 최종 결과를 산출한다.
  - R2-3-2: 그렇지 않으면 다음 라운드로 진전한다. 라운드 2와 동일한 과정을 반복한다.

그림 1은 이러한 과정을 도식적으로 보여주고 있다. 그림 1(a)는 9개의 육각형으로 이루어진 가상의 데이터 셋을 보여주고 있다. 이 9개의 기본 공간단위를 합역해 3개의 최종구역을 생성하는 것이 목적이다. 그림 1(b)는 이들 중 4, 6, 7번 기본 공간단위가 시드 공간단위로 선정된 상황을 보여주고 있다. 이 세 개의 시드 공간단위로부터 합역이 외연적으로 이루어지며, 결국 세 개의 시드 공간단위를 포함하는 세 개의 최종구역이 생성될 것이다. 우선, 이 세 개의 시드 공간단위에 대해 무작위 순위를 부여한다. 편의상 4, 6, 7의 순으로 결정된 것으로 한다.

그림 1(c)는 개별 시드 공간단위로부터 R1-2 과정이 이루어진 것을 보여주고 있다. 우선 4번 시드 공간단위에 대해, 연접한 기본 공간단위의 개수를 구하면 세 개(1, 5, 8)이다. 따라서 이것은 R1-2-3 상황에 해당하는 것인데, 1, 5, 8 중 8번 기본 공간단위가 무작위로 선정되었다고 할 때, 그 공간단위가 4번 시드 공간단위에 병합된다. 이때 4와 8의 결합을 통해 새로이 생성된 구역이 라운드 1의 라운드 구역이 된다( $RZ_{4,8}$ ). 다음으로 6번 시드 공간단위에 대해 동일한 과정을 진행한다. 2, 3, 5, 9가 연접하고 있어 마찬가지로 R1-2-3 상황이며, 이 중 3번이 무작위로 선정되었다고 할 때, 그 공간단위가 6번 시드 공간단위에 병합된다. 이때 6과 3의 결합을 통해 새로이 생성된 구역 역시 라운



**그림 1.** 무작위합역의 기본 절차 도해

드 1의 라운드구역이 된다( $RZ_{6-3}$ ). 마지막으로 6번 시드 공간단위에 대해서도 동일한 과정을 진행한다. 그런데, 4번 중심의 라운드구역( $RZ_{4-8}$ ) 생성으로 말미암아 7번 시드 공간단위는 연결한 기본 공간단위를 가지지 못하게 됨에 따라(R1-2-1 상황), 그 자체로 최종구역으로 확정된다( $FZ_7$ ). 결국 라운드 1이 진행되고 나면 한 개의 최종구역과 두 개의 라운드구역이 생성된다. 라운드구역의 개수가 0이 아니므로(R1-3-2 상황), 합역 과정은 다음 라운드로 이어진다.

라운드 2는 라운드 1에서 생성된 라운드구역에 대해서만 이루어지는데, 4번 중심의 라운드구역( $RZ_{4-8}$ )과 6번 중심의 라운드구역( $RZ_{6-3}$ )에 대해 합역 절차가 진행된다. R2-1에 의거해, 이 두 개의 라운드구역에 대해 무작위 순위를 부여한다. 편의상 무작위순위가 4번 중심의 라운드구역, 6번 중심의 라운드구역으로 주어진 것으로 가정한다. 그림 1(d)는 개별 라운드구역에 대해 R2-2 과정이 이루어진 것을 보여주고 있다. 우선 4번 중심의 라운드구역( $RZ_{4-8}$ )은 1, 5, 9라는 세 개의 연결 기본 공간단위를 가진다. 따라서 R2-2-3 상황이 되는데, 5번이 임의적으로 선정되었다고 가정할 때 새로운 2차 라운드구역( $RZ_{4-8-5}$ )이 생성된다. 동일한 방식으로, 6번 중심의 라운드구역( $RZ_{6-3}$ )도 연결한 2와 9번 기본 공간단위 중 임의적으로 선정된 9번 공간단위를 병합해 새로운 2차 라운드구역( $RZ_{6-3-9}$ )이 생성된다. 결국 라운드 2가 진행되고 나면, 두 개의 라운드구역이 생성된다. 라운드구역의 개수가 0이 아니므로(R2-3-2 상황), 합역 과정은 다음 라운드로 이어진다.

라운드 3은 라운드 2에서 생성된 라운드구역에 대해서만 이루어지는데, 4번 중심의 라운드구역( $RZ_{4-8-5}$ )과 6번 중심의 라운드구역( $RZ_{6-3-9}$ )에 대해 합역 절차가 진행된다. 이 두 개의 라운드구역에 대해 무작위 순위를 부여한다. 편의상 무작위순위가 4번 중심의 라운드구역, 6번 중심의 라운드구역으로 주어진 것으로 가정한다. 그림 1(e)는 개별 라운드구역에 대해 합역이 이루어지는 과정을 보여주고 있다. 우선 4번 중심의 라운드구역( $RZ_{4-8-5}$ )은 1과 2라는 두 개의 연결 기본 공간단위를 가진다. 이 중 2번이 임의적으로 선정되어 새로운 3차 라운드구역( $RZ_{4-8-5-2}$ )이 생성된다. 동일한 방식을 두 번째 라운드 구역에 적용한다. 그런데, 4번 중심의 라운드구역( $RZ_{4-8-5-2}$ )의 생성으로 말미암아 6번 중심의 라운드구역( $RZ_{6-3-9}$ )은 연결한 기본 공간단위를 가지지 못하게 됨에 따라, 그 자체로 최종구역으로 확정된다( $FZ_{6-3-9}$ ). 결국 라운드 3이 진행되고 나

면, 두 개의 최종구역과 한 개의 라운드구역이 남게 된다. 라운드구역이 여전히 존재하므로, 합역 과정은 다음 라운드로 이어진다.

라운드 4는 최종 라운드가 되는데, 그림 1(f)는 합역 절차가 종료된 최종 결과를 보여주고 있다. 라운드 3에 의해 생성된 라운드구역( $RZ_{4-8-5-2}$ )과 연결한 기본 공간단위가 1번 하나 밖에 없으므로, 전자에 후자를 병합하여 4번 중심의 최종구역( $FZ_{4-8-5-2-1}$ )이 생성된다. 결국 라운드 4가 진행되고 나면, 더 이상 라운드구역이 존재하지 않게 되므로 모든 합역 절차를 중단한다. 1~4 라운드를 거치는 동안 확정된 최종구역, 즉 라운드 1에서 생성된 7번 중심의 최종구역( $FZ_7$ ), 라운드 3에서 생성된 6번 중심의 최종구역( $FZ_{6-3-9}$ ), 마지막 라운드 4에서 생성된 4번 중심의 최종구역( $FZ_{4-8-5-2-1}$ )으로 구성된 최종 합역 결과물이 산출되는 것이다.

## 2. 이항 공간근접성행렬에 기반한 실행 알고리즘

앞에서 설명한 기본 알고리즘을 실행하는 방법은 다양할 수 있다. 여기서의 연결성-기반 이항 공간근접성행렬(contiguity-based binary spatial proximity matrix, 이하 CB\_SPM)에 기반한 방법을 소개하고자 한다. CB\_SPM은 에어리어 데이터의 기본 공간단위 간의 공간근접성을 정의하는데 주로 사용된다. 즉, 경계를 공유한 두 공간단위 간에는 1의 값을 그렇지 않은 두 공간단위 간에는 0의 값을 부여하는, 1차 근린(first-order neighbors) 규정 방식에 기반한다(Bailey and Gatrell, 1995; Tiefelsdorf *et al.*, 1999). 무작위합역 알고리즘의 실행에 있어서 CB\_SPM에 기반한 방법을 개발하고자 하는 데는 두 가지 이유가 있다. 첫째, 기본 공간단위의 위상 관계를 가장 단순하고 간명하게 저장하는 방법일 뿐만 아니라 본질적으로 합역 개념과 연결된다. 왜냐하면 합역이란 연결한 두 역형 개체의 경계를 허무는 것이므로, 기본 공간단위 간의 경계 공유 정보는 합역의 전 과정이 요구하는 유일한 정보이다. 둘째, 1차 근린 공간근접성행렬로부터 고차 근린(higher-order neighbors) 관계에 대한 정보를 추출할 수 있다(Anselin and Smirnov, 1996). 합역의 진전은 결국 고차 근린 관계에 있는 공간단위의 병합을 의미하는 것이기 때문에 효율적인 무작위합역 알고리즘 개발에 단초를 제공할 수 있다. 하지만 본 연구에서는 첫 번째 이유에 집중하여 실행 알고리즘을 개발하고자 한다.

우선, CB\_SPM을 활용한 무작위합역 알고리즘은 CB\_SPM을 긴 서식(long fom)으로 변환하는 것으로부터 시작한다. 이것을 3CSR(three-column sparse representation, 3 열희소재현)이라고 하는데(Bivand *et al.*, 2013, 274), 행 ID, 열ID, 근접도(혹은 가중치)의 세 컬럼으로 구성된 긴 서식의 데이터이다. CB\_SPM을 3CSR로 전환하는 과정을 설명하기 위해 그림 1에서 사용한 정육각형 테셀레이션 예로 들고자 한다. 이것의 CB\_SPM은 9×9 정방 행렬(square matrix)로 주어지고(식 (1) 참조), 이것의 3CSR은 32×3의 비정방 행렬(non-square matrix)의 형태를 취한다(표 1의 첫 번째 IN3CSR 참조). 행의 개수가 32인 것은 CB\_SPM의 모든 요소를 합산한 것으로, 기본 공간단위 간 공유 경계의 총수(실질적으로는 그것의 2배)를 의미한다.

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} \quad (1)$$

CB\_SPM 기반 무작위합역 실행 알고리즘은 이렇게 규정된 3CSR로부터 합역의 결과를 나타내는 새로운 3CSR을 도출하는 과정으로 정의할 수 있다. 두 개의 3CSR을 구분하기 위해 전자를 IN3CSR(Input 3CSR, 투입 3CSR)로 후자를 OUT3CSR(output 3CSR, 산출 3CSR)로 명명하고자 한다. 그리고 실행 알고리즘은 매 라운드마다 IN3CSR에서는 삭감, OUT3CSR에서는 첨가가 이루어지는 일종의 마이너스-플러스 기법(minus-plus technique)이 된다. 보다 구체적으로 말해 IN3CSR의 모든 요소가 삭제되면(혹은, OUT3CSR의 행의 길이가 기본 공간단위 수와 같아지면) 무작위합역 과정은 종결된다. IN3CSR의 모든 요소가 삭제된다는 것은, 더 이상 병합할 기본 공간단위가 남아 있지 않다는 것을 의미하며, OUT3CSR의 행의 길이가 기본 공간단위 수와 같아진다는 것은 모든 기본 공간단위가 특정한 최종구역에 병합되었다는 것을 의미한다. 마이너스의 과정은 굉장히 단순하다. 시드 공간단위로 선정되거나 라운드구역에 병합되는 순간 해당 공간단위를 열ID로 갖고 있는 IN3CSR의 모든 요소는 삭제된다. 왜냐하면 한번 병합된 공간단위는 차후의 라운드에서 병합의 후보가 될 수 없기 때문이다. 플러스의 과정도 매우 단순하다. 라운드구역에

병합이 이루어지는 순간 해당 라운드구역이 어느 시드 공간단위 중심인지를 OUT3CSR의 행ID로 하고, 병합되는 공간단위를 열ID로 하여 첨가하면 된다. 라운드가 진행됨에 따라 IN3CSR의 행은 점점 줄어들고, OUT3CSR의 행은 점점 늘어, 결국 IN3CSR의 행의 개수는 0에 도달하고, OUT3CSR의 행의 개수는 기본 공간단위의 개수와 같아지게 된다.

이 과정을 그림 1에 나타나 있는 6단계를 통해 도해하면 표 1과 같다. 그림 1(a)는 초기 상태로 IN3CSR은 32×3의 비정방 행렬의 형태를 띠고 있고, OUT3CSR의 요소는 존재하지 않는다. 그림 1(b)에서는 4, 6, 7이 시드 공간단위로 선정되는데, 이 시드 공간단위는 합역 과정의 중심이 될 뿐 다른 구역에 합역되지 않기 때문에 이 세 공간단위를 열ID로 가지는 요소, 즉 해당 행을 모두 삭제해야 한다. 그렇게 하면 표 1에서 보듯이 IN3CSR의 행 개수는 22개로 줄어든다. 반면에 OUT3CSR은 이 세 시드 공간단위가 최종구역의 한 요소가 된다는 것을 저장해야 하므로 세 개의 행을 갖게 된다. 그림 1(c)에서는 8번 공간단위가 4번 시드 공간단위에, 3번 공간단위가 6번 시드 공간단위에 병합된다. 따라서 8번과 3번을 열ID로 갖는 모든 요소, 즉 해당 행을 IN3CSR로부터 삭제해야 한다. 최종구역으로 확정되는 7번 구역은 이후의 라운드에서 병합의 대상이 될 수 없기 때문에 그것을 열ID로 갖는 요소를 IN3CSR로부터 제거해야 하지만, 유일한 병합 가능 공간단위였던 8번이 제거될 때 이미 제거되었다. 결국 IN3CSR의 행의 개수는 16개로 줄어든다. 반면에 병합 사항이 첨가된 OUT3CSR의 행의 수는 5개로 늘어난다.

그림 1(d)에서는 4번 중심 라운드구역에 5번 공간단위가, 6번 중심 라운드구역에 9번 공간단위가 병합되는데, 앞의 방식과 동일하게 5번과 9번을 열ID로 갖는 모든 요소를 IN3CSR로부터 삭제한다. 동시에 이 병합 상황을 OUT3CSR에 첨가한다. 이렇게 하면 IN3CSR의 행 수는 7개로 줄어들고, OUT3CSR의 행 수는 7개로 늘어난다. 그림 1(e)에서는 4번 중심 라운드구역에 2번 공간단위가 병합되고 동시에 6번 중심 라운드구역이 최종구역으로 확정된다. 따라서 2번을 열ID로 갖는 모든 요소를 IN3CSR로부터 제거하고, 병합 상황을 OUT3CSR에 업데이트한다. 이렇게 하면 IN3CSR의 행 수는 3개로 줄어들고, OUT3CSR의 행 수는 8개로 늘어난다. 마지막인 그림 1(f)에서는 4번 중심 라운드구역이 1번 공간단위를 병합함으로써 최종구역으로 확정되고, 모든 합역 과정이 종결된다. 1번을 열ID로 갖는 모든 요소를 IN3CSR에서 제거하면 IN3CSR의 모든 요소는 사라지게

표 1. 이항 공간근접성행렬을 활용한 무작위합역 절차의 실행 알고리즘 도해

		그림 1에 나타나 있는 무작위합역의 단계					
		(a)	(b)	(c)	(d)	(e)	(f)
IN3CSR		$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 1 \\ 1 & 5 & 1 \\ 2 & 1 & 1 \\ 2 & 3 & 1 \\ 2 & 5 & 1 \\ 2 & 6 & 1 \\ 3 & 2 & 1 \\ 3 & 6 & 1 \\ 4 & 1 & 1 \\ 4 & 5 & 1 \\ 4 & 7 & 1 \\ 4 & 8 & 1 \\ 5 & 1 & 1 \\ 5 & 2 & 1 \\ 5 & 4 & 1 \\ 5 & 6 & 1 \\ 5 & 8 & 1 \\ 5 & 9 & 1 \\ 6 & 2 & 1 \\ 6 & 3 & 1 \\ 6 & 5 & 1 \\ 6 & 9 & 1 \\ 7 & 4 & 1 \\ 7 & 8 & 1 \\ 8 & 4 & 1 \\ 8 & 5 & 1 \\ 8 & 7 & 1 \\ 8 & 9 & 1 \\ 9 & 5 & 1 \\ 9 & 6 & 1 \\ 9 & 8 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 5 & 1 \\ 2 & 1 & 1 \\ 2 & 3 & 1 \\ 2 & 5 & 1 \\ 3 & 2 & 1 \\ 4 & 1 & 1 \\ 4 & 5 & 1 \\ 4 & 8 & 1 \\ 5 & 1 & 1 \\ 5 & 2 & 1 \\ 5 & 8 & 1 \\ 5 & 9 & 1 \\ 6 & 2 & 1 \\ 6 & 3 & 1 \\ 6 & 5 & 1 \\ 6 & 9 & 1 \\ 7 & 8 & 1 \\ 8 & 5 & 1 \\ 8 & 9 & 1 \\ 9 & 5 & 1 \\ 9 & 8 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 5 & 1 \\ 2 & 1 & 1 \\ 2 & 5 & 1 \\ 3 & 2 & 1 \\ 4 & 1 & 1 \\ 4 & 5 & 1 \\ 5 & 1 & 1 \\ 5 & 2 & 1 \\ 5 & 9 & 1 \\ 6 & 2 & 1 \\ 6 & 5 & 1 \\ 6 & 9 & 1 \\ 8 & 5 & 1 \\ 8 & 9 & 1 \\ 9 & 5 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \\ 4 & 1 & 1 \\ 5 & 1 & 1 \\ 6 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 1 \\ 5 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{bmatrix}$
	OUT3CSR	$\begin{bmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{bmatrix}$	$\begin{bmatrix} 4 & 4 & 1 \\ 6 & 6 & 1 \\ 7 & 7 & 1 \end{bmatrix}$	$\begin{bmatrix} 4 & 4 & 1 \\ 4 & 8 & 1 \\ 6 & 6 & 1 \\ 6 & 3 & 1 \\ 7 & 7 & 1 \end{bmatrix}$	$\begin{bmatrix} 4 & 4 & 1 \\ 4 & 8 & 1 \\ 4 & 5 & 1 \\ 6 & 6 & 1 \\ 6 & 3 & 1 \\ 6 & 9 & 1 \\ 7 & 7 & 1 \end{bmatrix}$	$\begin{bmatrix} 4 & 4 & 1 \\ 4 & 8 & 1 \\ 4 & 5 & 1 \\ 4 & 2 & 1 \\ 6 & 6 & 1 \\ 6 & 3 & 1 \\ 6 & 9 & 1 \\ 7 & 7 & 1 \end{bmatrix}$	$\begin{bmatrix} 4 & 4 & 1 \\ 4 & 8 & 1 \\ 4 & 5 & 1 \\ 4 & 2 & 1 \\ 4 & 1 & 1 \\ 6 & 6 & 1 \\ 6 & 3 & 1 \\ 6 & 9 & 1 \\ 7 & 7 & 1 \end{bmatrix}$

되고, 병합 상황을 OUT3CSR에 업데이트하면 OUT3CSR의 요소 수는 9로 최초의 기본 공간단위 개수와 동일하게 된다. 최종적인 OUT3CSR에 포함되어 있는 정보는 다음과 같이 해석될 수 있다. 행ID가 4, 6, 7인 것은 이 세 개 공간단위 중심의 합역체가 생성되었다는 것을 의미하고, 각각의 행ID와 결부되어 있는 열ID는 개별 합역체의 구성 공간단위를 보여준다. 즉, 행ID 4와는 열ID 4, 8, 5, 2, 1이 연결되어 있는데, 이는 4번 중심 최종구역이 4, 8, 5, 2, 1번 기본 공간단위로 구성된다는 것을 보여주는 것이다.

### 3. 무작위합역 절차의 다양성

지금까지 설명한 기본 알고리즘이 무작위합역의 유일한 절차인 것이 아니라는 사실을 이해하는 것은 매우 중요하다. 무작위합역은 당연히 무작위성에 기반해야 하며, 앞에서 설명한 기본 알고리즘은 세 가지 세부 과정에서 이 무작위성의 원칙이 적용되었다. 그 세부 과정 중 하나는 시드 공간단위를 선정하는 과정이고, 또 다른 하나는 시드 공간단위나 라운드구역의 순서를 정하는 과정이고, 그리고 마지막은 연결한 기본 공간단위 중 병합을 위한 공간단

위를 선정하는 과정이다. 그런데 처음의 두 과정과 달리 마지막 과정에서 무작위성을 적용하는데 다양성이 존재할 수 있다. 특히, 라운드구역에 병합할 후보 공간단위가 두 개 이상 존재하는 경우 이 다양성은 실질적인 의미를 가질 수 있다. 무작위성에 다양성을 가미하는 것에는 다음의 두 가지 차원이 있을 수 있다.

첫 번째 차원은 병합의 과정에 무선정 옵션을 어떻게 적용하느냐에 의거한 차원이다. 이차원에 근거하면, 세 가지의 서로 다른 방식이 가능한데, 하나는 앞에서 설명한 기본 알고리즘과 동일하게 후보 공간단위 중 반드시 하나를 무작위로 선정하는 것이다(방식 A). 두 번째 옵션은 무선정을 마치 후보 공간단위 중 하나인 것으로 취급하는 것이다(방식 B). 즉, 특정 공간단위가 선정되면 그 공간단위를 병합하겠지만, 무선정이 선정되면 해당 라운드구역에 대해서는 병합을 실행하지 않는 것이다. 세 번째는 선정/무선정을 우선적으로 결정한 후, 선정이 결정된 경우에만 첫 번째 방식처럼 후보 공간단위 중 하나를 반드시 선정하는 방식이다(방식 C). 이를 설명하기 위해 그림 1의 (c)~(d)의 예를 들고자 한다. 그림 1의 (c)~(d)는 48 라운드구역에 5가 병합되는 과정을 보여주고 있는데, 48 라운드구역에 병합될 후보 공간단위는 1, 5, 9로 모두 3개이다. 이때 5번이 선정될 확률이 위에서 설명한 세 가지 방식에 따라 달라지게 된다. 우선 기본 알고리즘에 따르면 그 확률은 1/3이다. 그러나 무선정 옵션을 포함시키는 두 번째 방식에 의거하면 그 확률은 1/4로 변화한다. 선정/무선정을 우선 결정하고 난 후 후보 공간단위 중에서 선정하는 세 번째 방식에 따르면 그 확률은 1/6이 되는 것이다.

두 번째 차원은 후보 공간단위 중 하나가 선정되는 과정에서 무작위성을 어떻게 정의하느냐에 의거한 차원이다. 여기에는 두 가지 방식이 있을 수 있다. 첫 번째 방식은 위에서 설명한 기본 알고리즘과 같이 완전 무작위성을 적용하는 방식이다(방식 1). 즉, 선정될 확률은 후보 공간단위 사이에서 동일하다. 두 번째 방식은 후보 공간단위의 라운

드구역에 대한 연결성 강도에 따라 확률을 다르게 적용하는 것이다(방식 2). 이 두가지 방식의 차이를 도해하기 위해 다시 그림 1의 (c)~(d)의 예로 돌아가고자 한다. 48 라운드구역의 병합 후보는 1, 5, 9번이고 각각이 선정될 확률은 모두 1/3로 동일하다(첫 번째 차원의 A 방식을 적용한 것으로, 나머지 두 가지 방식을 적용하여도 확률이 동일하다는 점은 변하지 않는다). 그런데 5번은 1번이나 9번과 달리 48 라운드구역에 대한 연결도(contiguousness)가 강하다고 말할 수 있다. 즉, 5번은 48 라운드구역의 구성 공간단위인 4번과 8번 모두와 연결하지만, 1번은 4번과만, 9번은 8번과만 연결해 있다. 따라서 5번이 선정될 확률을 1번이나 9번이 선정될 확률의 두 배가 되도록 설정할 수 있다(5번은 2/3, 1번과 9번은 각각 1/3). 이러한 고려는 최종 구역의 형태가 보다 원형에 가까워지도록 하는 효과가 있을 것으로 기대된다(Flowerdew *et al.*, 2001). 비록 무작위합역이라 하더라도 보다 원형도(compactness)가 높은 최종 구역이 산출되도록 해주는 알고리즘을 선택하는 것이 연구 결과의 현실유관성을 향상시킨다는 점에서 보다 더 합리적이 선택될 수 있다.

표 2는 이러한 두 가지 차원을 결합하여 여섯 가지의 서로 다른 무작위합역 유형(A1, B1, C1, A2, B2, C2)을 정의하고 있다. 각각의 무작위합역 유형이 어떻게 다른 방식으로 작동하는지를 설명하기 위해 그림 1의 (c)~(d)에서 5번 공간단위가 선정될 확률을 예를 들고자 한다. 이미 앞에서 A1, B1, C1에 해당하는 확률이 1/3, 1/4, 1/6로 서로 다르다는 점을 설명하였다. 이제 방식 2가 적용된 A2, B2, C2의 해당 확률이 어떻게 되는지 살펴보고자 한다. 우선 A2의 경우, 후보 공간단위가 1, 5, 9의 세 개가 아니라 연결도로 인해, 1, 5, 5, 9처럼 5가 하나 더 있는 경우로 변화한다. 따라서 A2에서 5가 선정될 확률은 1/2이다. B2의 경우는 후보 공간단위가 무선정, 1, 5, 5, 9인 것처럼 취급되기 때문에, 5가 선정될 확률은 2/5가 된다. 마지막으로 C2의 경우는 선정/무선정 확률이 1/2, 3개의 공간단위에서 선정될 확률

표 2. 무작위합역 절차의 다양화: 두 차원과 여섯 가지 유형

		차원 1: 무선정 옵션의 적용 방식		
		방식 A: 후보 공간단위 중 하나를 반드시 선정	방식 B: 무선정 옵션을 후보 공간단위로 취급	방식 C: 선정/무선정 결정 후, 후보 공간단위 중 하나를 반드시 선정
차원 2: 후보 공간단위의 선정 확률 결정 방식	방식 1: 동일 확률 적용	A1	B1	C1
	방식 2: 연결도에 기반한 상이한 확률 적용	A2	B2	C2

이 1/2, 따라서 결과적으로 확률은 1/4이 된다. 공교롭게 B1과 동일한 확률이지만, 내용은 완전히 다르다. 따라서, 실질적인 의미에서는, 여섯 가지 무작위합역 유형이 동일한 상황에 대해 서로 다른 확률을 규정한다고 볼 수 있다. 다음 장에서는 여기에서 규정한 서로 다른 여섯 가지의 무작위합역 유형이 결론적으로 어떤 결과를 보여줄지를 시뮬레이션 분석을 통해 밝히고자 한다.

### III. 시뮬레이션 연구

#### 1. 시뮬레이션 디자인

시뮬레이션을 위해 두 종류의 기본 공간단위 체계를 설정했다. 둘 다 1,024개 공간단위로 이루어진 정다각 테셀레이션(regular tessellations)인데, 각각 정육각형과 정사각형으로 이루어져 있다(자세한 설명은 이상일 등, 2015; Boots and Tiefelsdorf, 2000; Lee *et al.*, 2019 참조). 정육각형 테셀레이션은 행정구역과 같은 현실 세계의 구역 체계를 가장 잘 반영하는 것으로 인정되고 있으며, 연결성 정의가 명확하게 이루어 진다는 장점이 있다. 이에 비해 정사각형 테셀레이션은 원격탐사 데이터와 같은 래스터 데이터 유형의 특성을 가장 잘 반영하는 것으로 인정되고 있다. 그러나 연결성 정의의 두 가지 방식(루크 방식과 퀴 방식)에 따라 GIS 오퍼레이션 결과나 공간통계분석의 결과가 달라지는 문제가 있다. 이상일 등의 연구에 따르면, 일변량 공간연관성통계량의 가능치 범위가 연결성 정의의 방식에 따라 상당히 달라진다. 사실상 정사각형 테셀레이션의 경우 퀴 방식을 적용하게 되면, 육각형 테셀레이션과 유사한 결과를 보여주는 것으로 드러났다(이상일 등,

2015, 표 2; 이상일 등, 2016). 그래서 정사각형 테셀레이션의 경우는 루크 방식을 통한 연결성 정의를 적용하여 분석에 투입하였다. 이상적인 평균 이웃의 수는 정육각형 테셀레이션의 경우 6개, 정사각형 테셀레이션의 경우 4개이지만, 이웃의 수를 상대적으로 적게 갖는 주변부 다각형의 존재로 인해 본 연구에서 사용된 테셀레이션의 경우 실질적인 평균 이웃의 수는 정육각형의 경우 5.752개, 정사각형의 경우 3.875개이다.

보다 상세한 시뮬레이션 디자인은 정육각형 테셀레이션을 중심으로 설명하고자 한다. 시뮬레이션은 그림 2에 나타나 있는 1,024개 정육각형 기본 공간단위를 64개의 최종구역으로 합역하는 것이다. 표 2에서 나타나 있는 6가지의 무작위합역 유형 각각에 대해 1,000개의 합역체를 시뮬레이션한다. 예를 들어, 표 2에서 A1 유형으로 정의된 무작위합역 알고리즘을 적용하여 1,024개의 육각형을 합역하여 64개의 최종구역을 생성하는 1,000개의 서로 다른 방법을 시뮬레이션한다. 결국 정육각형 테셀레이션에 대해 총 6,000개의 합역체가 생성된다. 동일한 시뮬레이션 절차를 정사각형 테셀레이션에 적용하면 또 다른 총 6,000개의 합역체가 생성된다.

서로 다른 여섯 가지의 무작위합역 유형이 어떻게 서로 다른지를 평가하기 위해 두 가지의 준거를 설정하였다. 하나는 연결도(contiguity)이고, 또 다른 하나는 원형도(compactness)이다. 우선, 연결도는 최종구역이 얼마나 많은 이웃과 연결되어 있는가의 정도를 의미한다. 64개의 최종구역으로 구성된 합역체 각각에 대해, 개별 64개 최종구역별로 연결한 최종구역의 개수를 구한다. 그리고 나서 연결이웃수의 평균과 분산을 구한다. 이렇게 하면 개별 합역체별로 연결도와 관련된 두 가지 통계량이 구해지는데, 각각을 '연접도-평균'과 '연접도-분산'이라고 부르하고자 한다.

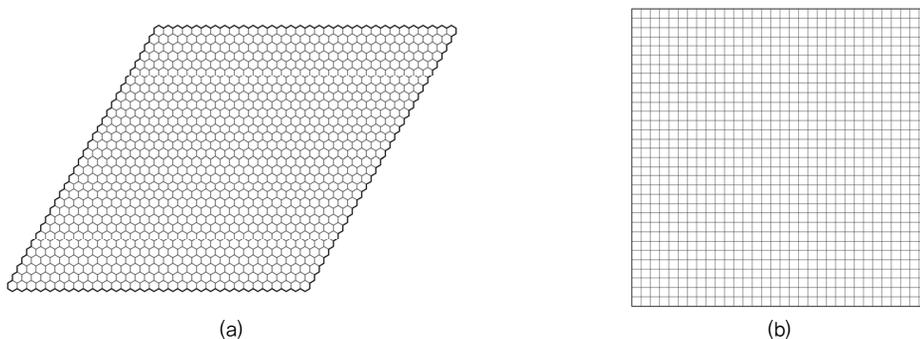


그림 2. 정다각 테셀레이션(n=1,024): (a) 정육각형, (b) 정사각형

전자는 서로 다른 무작위합역 알고리즘이 최종구역의 연결도에 영향을 끼치는 지의 여부를 평가할 수 있게 해주며, 후자는 서로 다른 무작위합역 알고리즘의 상대적 안정성 혹은 일관성을 연결도라는 측면에서 평가할 수 있게 해준다.

원형도 지표는 연결도 지표와는 또 다른 차원을 평가할 수 있게 해주며, 실질적인 측면에서는 보다 중요한 것이라 말할 수 있다. 원형도는 최종구역이 얼마나 원에 가까운 형상을 가지는가를 평가한다. 최종구역의 원형도는 최종구역이 얼마나 현실유관성이 있는 것이냐를 평가하는 중요한 지표가 된다. 무작위합역 알고리즘은 그 성격상 매우 불규칙적인 형상을 가진, 그러므로 실질적 유효성이 매우 낮은 최종구역을 생성할 가능성이 충분히 있기 때문에, 합역 알고리즘의 적절성을 평가하는데 원형도 가지는 의미는 적지 않다. 다시 말해, 특정한 무작위합역 알고리즘이 다른 무작위합역 알고리즘에 비해 실질성이 낮은 최종구역을 산출할 가능성이 충분히 낮다면, 그 알고리즘을 우선적으로 고려하는 것은 매우 합리적인 결정일 수 있다. 역형 개체의 형태를 측정하는 매우 다양한 지표가 존재하지만(Kim, 2012, 496, 표 1 참조), 여기서 사용할 형태지수(SI, shape index)는 콕스의 원형도(Cox's circularity)(Cox, 1927) 혹은 폴스비-포퍼 인덱스(Polsby and Popper, 1991)라고 불리는 지표이다.

$$SI = A / (P^2 / 4\pi) = 4\pi A / P^2 \quad (2)$$

여기서  $A$ 와  $P$ 는 해당 역형 개체(여기서는 최종구역)의 면적과 둘레이다. 이 인덱스는 해당 역형 개체의 면적과, 해당 역형 개체와 동일한 둘레를 갖는 원의 면적의 비로 정의된다. 이것은 둘레가 동일한 경우 원이 면적이 가장 큰 도형이라는 기하학적 공리를 이용한 것으로, 0~1의 값을 가지며, 1에 가까울수록 형태가 원에 가깝다는 것을 의미한다. 원형도는 다음과 같은 방식으로 구해진다. 64개의 최종구역으로 구성된 합역체 각각에 대해, 개별 64개 최종구역별로 식 (2)에 따른 형태지수 값을 구한다. 그리고 나서 형태지수 값의 평균과 분산을 구한다. 이렇게 하면 개별 합역체별로 원형도와 관련된 두 가지 통계량이 구해지는데, 각각을 '원형도-평균'과 '원형도-분산'이라고 부르려 한다. 전자는 서로 다른 무작위합역 알고리즘이 최종구역의 원형도에 영향을 끼치는 지의 여부를 평가할 수 있게 해주며, 후자는 서로 다른 무작위합역 알고리즘의 상대적 안정성 혹은 일관성을 원형도라는 측면에서 평가할 수 있게 해준다.

## 2. 분석 결과

그림 3은 정육각형 테셀레이션에 대한 연구 결과를 보여주고 있는데, 네 가지 준거인 연결도-평균, 연결도-분산, 원형도-평균, 원형도-분산 각각에 대한 박스플롯(boxplot)이 나타나 있다. 그림 3의 (a)와 (b)는 연결도 기준의 시뮬레이션 결과를 보여주고 있다. 연결도-평균(그림 3(a))의 경우 6개의 서로 다른 무작위합역 유형 사이에 큰 차이가 없다는 것을 알 수 있다. 연결도-평균의 중앙값이 모두 5.04~5.07에 위치해 있고, 6,000개 합역체 대부분의 연결도-평균값이 4.80~5.30 범위 안에 위치하고 있음을 알 수 있다. 첫 번째 차원에 따른 A, B, C 간의 차이나, 두 번째 차원에 따른 1, 2 간의 차이도, 특정한 일반화가 가능할 만큼의 유의미한 차이를 보여주고 있지 않다. 다음으로, 연결도-분산의 경우를 살펴보면(그림 3(b)), 연결도-평균과 마찬가지로 6개의 서로 다른 무작위합역 유형 사이에 큰 차이가 없다는 것을 알 수 있다. 즉, 연결도-분산의 중앙값이 모두 2.25~2.50에 위치해 있고, 6,000개 합역체 대부분의 연결도-분산값이 1.50~3.50 범위 안에 위치하고 있음을 알 수 있다. 그러나 좀 더 주의 깊게 살펴볼 필요가 있는 사항들도 확인된다. 1과 2 사이의 차이는 거의 없지만(A1과 A2, B1과 B2, C1과 C2는 매우 유사), B가 A나 C에 비해 연결도-분산이 상대적으로 낮다. 중앙값도 가장 낮을 뿐만 아니라 이례치로 확인되는 최대값도 다른 두 유형에 비해 낮은 경향이 있다( $t$ -검정의 결과, 통계적으로 유의함). 이는 B가 다른 두 방식에 비해 연결도라는 측면에서 상대적으로 안정적인 혹은 일관성 있는 결과를 산출해 준다는 점을 함축하고 있다.

그림 3의 (c)와 (d)는 원형도 기준의 시뮬레이션 결과를 보여주고 있다. 원형도-평균(그림 3(c))의 경우, A, B, C 사이에는 큰 차이가 없지만, 1, 2 사이에서 뚜렷한 차이가 있음을 보여주고 있다. 즉, 2가 1에 비해 원형도가 훨씬 더 높다. 2의 경우는 A, B, C 유형에 상관없이 원형도-평균값의 중앙값이 0.38~0.40에 위치하고 있지만, 1의 경우는 0.32~0.34에 위치하고 있다( $t$ -검정의 결과, A1과 A2, B1과 B2, C1과 C2의 차이는 통계적으로 유의함). 앞에서 후보 공간 단위의 라운드구역에 대한 연결도에 따라 확률을 다르게 적용하면 최종구역의 형태가 보다 원형에 가까워질 것으로 기대했는데(Flowerdew *et al.*, 2001), 그것을 충족하는 결과가 도출된 것이다. 이것은 무작위합역의 기본 공리는 준수하면서 연결성 강도를 부가적으로 고려하는 것만으

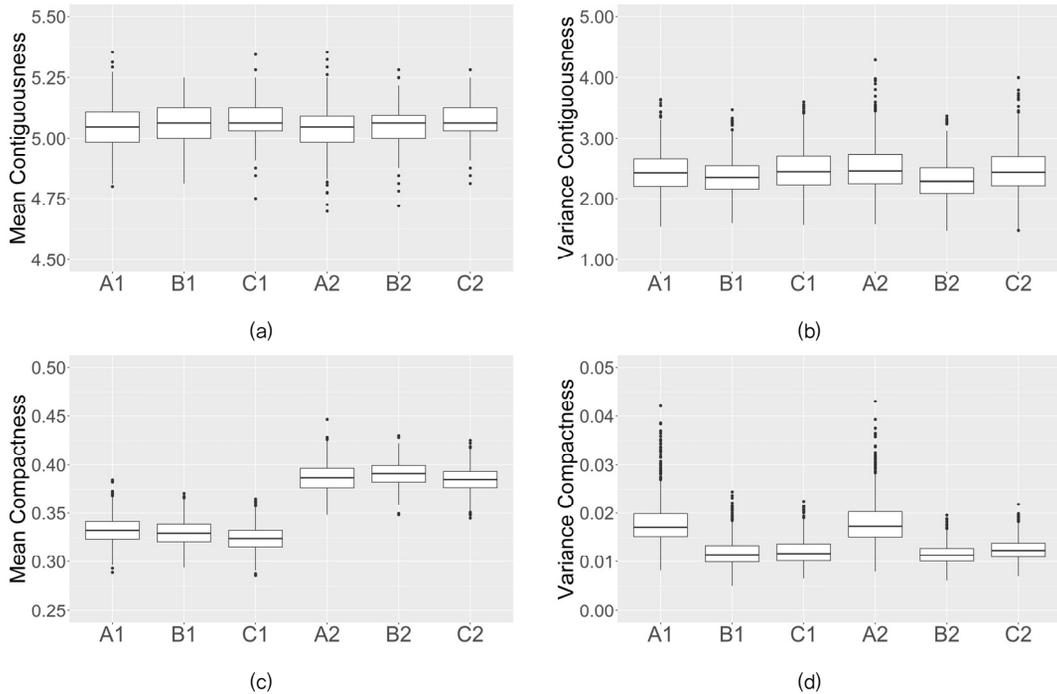


그림 3. 정육각형 테셀레이션에 대한 시뮬레이션 결과 분석: (a) 연결도-평균, (b) 연결도-분산, (c) 원형도-평균, (d) 원형도-분산

로도 최종구역에서의 원형도를 유의미하게 향상시킬 수 있다는 것을 보여준다는 측면에서 고무적인 결과가 아닐 수 없다. 다음으로 원형도-분산의 경우를 살펴보면(그림 3(d)), 1과 2 사이에는 유의미한 차이가 없음을 알 수 있다(A1과 A2, B1과 B2, C1과 C2는 매우 유사). 그러나 A는 B, C와 확실히 다른 특징을 보여주고 있음을 확인할 수 있다. 즉, A는 B, C에 비해 중앙값도 크고, 이레치의 개수도 많고, 값의 크기도 훨씬 크다. 이는 B, C가 A에 비해, 원형도라는 측면에서 상대적으로 안정적인 혹은 일관성 있는 결과를 산출해 준다는 점을 함축하고 있다. 특히 2만 본다면, B가 C에 비해 더욱더 고무적인 결과를 보여주고 있다( $t$ -검정의 결과, 통계적으로 유의함).

그림 4는 동일한 시뮬레이션 분석을 정사각형 테셀레이션에 적용한 결과를 보여주고 있다. 연결도-평균의 경우(그림 4(a)), 연결도-평균값의 중앙값이 정육각형의 경우보다 다소 낮아졌다는 점을 확인할 수 있다. 이는 1,024개 기본 공간단위 테셀레이션에서 정육각형에 비해 루크 방식이 적용된 정사각형이 연결도에서 더 적었던 것이 영향을 미친 것으로 보인다. 이는 최초 공간단위의 연결도가 최종구역의 연결도에 일정 정도 영향을 준다는 점을 보여

주는 것이다. 그러나, 여섯 개의 무작위합역 유형 사이에 유의미한 차이가 없다는 점은 동일하다. 연결도-분산(그림 4(b))의 경우, 분산값의 중앙값이 정육각형의 경우보다 다소 낮아졌다는 점을 확인할 수 있다. 전자가 2.5에 근접한 값을 보인 반면, 후자는 2.0에 근접한 값을 보여주고 있다. 이는 평균이 낮기 때문에 분산도 낮은 것으로 단순 해석할 수 있지만, 최초 공간단위의 연결도가 낮으면, 최종 구역의 연결도를 낮출 뿐만 아니라 변동폭도 낮춘다는 것을 함축하고 있다고도 해석할 수 있다. 그런데, 정육각형의 경우 B가 A, C에 비해 연결도-분산이 다소 낮게 나타났는데, 정사각형의 경우는 그러한 경향이 잘 나타나지 않는다. 대신에 C가 A, B에 비해 연결도-분산이 다소 높게 나타나는 경향은 보다 잘 드러난다( $t$ -검정의 결과, 통계적으로 유의함). 이는 A, B가 C에 비해, 원형도라는 측면에서, 상대적으로 안정적인 혹은 일관성 있는 결과를 산출해 준다는 점을 함축하고 있다.

그림 4의 (c)와 (d)는 원형도 기준의 시뮬레이션 결과를 보여주고 있다. 원형도-평균(그림 4(c))의 경우, 정육각형의 경우에 비해 정사각형의 원형도가 더 높다는 것을 알 수 있다. 즉, 여섯 가지 무작위합역 유형 모두에 대해 0.05 이

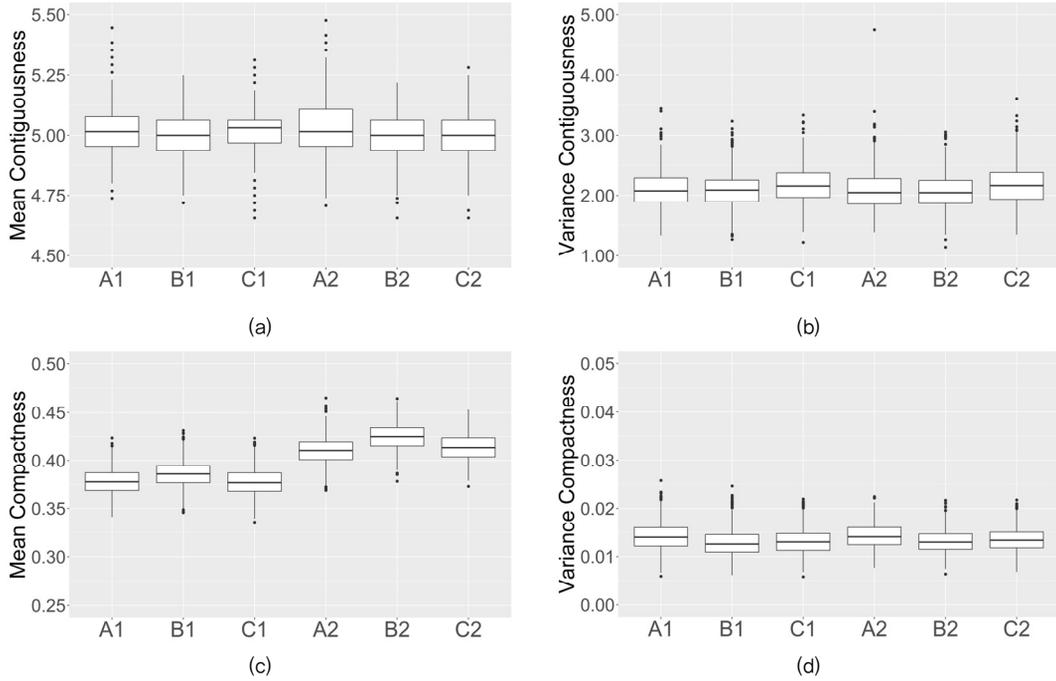


그림 4. 정사각형 테셀레이션에 대한 시뮬레이션 결과 분석: (a) 연결도-평균, (b) 연결도-분산, (c) 원형도-평균, (d) 원형도-분산

상의 차이를 보여주고 있다. 여섯 가지 무작위합역 유형 간의 차이는 정육각형의 경우와 동일하다. 즉, 2가1에 비해 원형도가 훨씬 더 높게 나타나고 있다( $t$ -검정의 결과, A1과 A2, B1과 B2, C1과 C2의 차이는 통계적으로 유의함). 그런데 여기에 부가적으로 B가A, C에 비해 원형도가 높게 나타나는 경향을 보이고 있다. 정육각형의 경우는 2에서 다소 그러한 경향이 나타나긴 했는데, 정사각형의 경우는 1, 2 모두에서 B가A, C에 비해 더 높은 원형도를 보여주고 있다( $t$ -검정의 결과, 통계적으로 유의함). 다음으로 원형도-분산의 경우를 살펴보면(그림 4(d)), 전체적으로 분산도가 정육각형에 비해 낮아진 것을 확인할 수 있다. 여섯 가지 무작위합역 유형 간의 차이는 정육각형의 경우와 유사하게 나타났다. 즉, B, C가A에 비해 낮은 변동성을 보여주었다( $t$ -검정의 결과, 통계적으로 유의함). 또한 1, 2 모두에서 B가C에 비해 좀 더 고무적인 결과를 보여주고 있다( $t$ -검정의 결과, B1와 C1, B2와 C2의 차이는 통계적으로 유의함). 이는 B가A, C에 비해, 원형도라는 측면에서, 상대적으로 안정적인 혹은 일관성 있는 결과를 산출해 준다는 점을 함축하고 있다.

이 모든 시뮬레이션 결과를 종합해 볼 때, B2 유형이 가

장 적절한 무작위합역 알고리즘인 것으로 평가된다. B2 유형은 무선정을 후보 공간단위 중 하나로 취급하여 무선정을 병합 가능한 공간단위와 동일한 선택 확률을 갖도록 함과 동시에 연결한 공간단위의 병합 확률을 연결도에 기반하여 다르게 부여하는 방식이다. 이 결과 B2 유형은 상대적으로 원형도가 높은 최종구역을 산출하고, 연결도와 원형도 두 가지 측면 모두에서 상대적으로 안정적인 혹은 일관성 있는 결과를 산출하는 것으로 드러났다. 요약하면 B2 유형은 통상적으로 생각하는 무작위합역 절차에 비해 무작위성과 현실유관성 모두를 향상시켰다.

#### IV. 결론

본 논문의 주된 연구목적은 무작위합역 절차의 다양성을 검토하고, 서로 다른 무작위합역 절차의 상대적 특성을 비교 분석함으로써 상대적으로 적절성이 높은 무작위합역 절차를 선정하여 제시하는 것이었다. 기본적인 무작위합역 절차가 정식화되었는데, 시드 공간단위를 중심으로 연결 공간단위를 무작위성에 기반해 병합해 나가는 일종

의 확장 과정을 통해 최종구역을 생성하는 방식이다. 이러한 기본적인 무작위합역 절차를 실행하는 효율적인 알고리즘으로 공간근접성행렬에 기반한 실행 알고리즘이 소개되었다. 이 실행 알고리즘은 매 라운드마다 투입 공간근접성행렬에서는 삭감, 산출 공간근접성행렬에서는 첨가가 이루어지는 일종의 마이너스-플러스 기법이다. 무작위합역 절차의 다양성을 제공해주는 두 가지 차원으로, 병합의 과정에 무선정 옵션을 어떻게 정의하느냐에 의거한 차원과 후보 공간단위 중 하나가 선정되는 과정에서 무작위성을 어떻게 정의하느냐에 의거한 차원이 설정되었다. 이를 바탕으로 A1, B1, C1, A2, B2, C2의 총 여섯 가지 서로 다른 무작위합역 유형이 도출되었다.

이 여섯 가지 무작위합역 유형의 서로 다른 특성을 파악하기 위해 시뮬레이션 실험을 실시하였다. 상이한 무작위합역 알고리즘이 최종 결과에 어떤 영향을 끼치는지를 평가하기 위해 연결도와 원형도의 두 가지 기준을 설정하였고, 각각에 대해 평균과 분산을 고려하였다. 시뮬레이션 실험의 주요 결과를 요약하면 다음과 같다. 첫째, 방식 B가 A, C에 비해 연결도-분산이 낮다. 이는 B가 다른 두 방식에 비해 연결도라는 측면에서 상대적으로 안정적인 혹은 일관성 있는 결과를 산출해 준다는 점을 함축하고 있다. 둘째, 방식 2가 방식 1에 비해 원형도-평균값이 훨씬 더 높다. 이는 2가 최종구역을 형태를 보다 원형에 가까워지도록 하는 효과가 있음을 보여주는 것이다. 셋째, 방식 A는 방식 B, C에 비해 원형도의 중앙값도 크고, 이레치의 개수도 많고 값의 크기도 훨씬 크다. 이는 B와 C(특히 B)가 원형도라는 측면에서 상대적으로 안정적인 혹은 일관성 있는 결과를 산출해 준다는 점을 함축하고 있다. 이러한 시뮬레이션 연구의 결과는 결국 B2 유형이 가장 적절한 무작위합역 절차를 보여주고 있다. B2 유형은 상대적으로 원형도가 높은 최종구역을 산출하고, 연결도와 원형도 두 가지 측면 모두에서 상대적으로 안정적인 혹은 일관성 있는 결과를 산출하는 것으로 드러났다.

본 연구는 무작위합역 절차가 고정불변의 것이 아니라 다양하게 정의될 수 있으며, 그 다양성 속에서도 무작위성과 현실유관성의 관점에서 상대적으로 더 적절한 선택이 가능할 수 있다는 점을 실험 연구를 통해 보여주었다는 점에서 그 학문적인 의의가 크다고 평가할 수 있다. 합역은 단순히 지도학적 일반화의 한 종류 혹은 GIS 오퍼레이션의 한 종류가 아니라, 지역구분, 재구획, MAUP과 같은 보다 광범위한 지리학과 지리정보과학의 연구 주제와 연결

되어 있다. 따라서 합역 절차에 대한 연구는 단순히 합역 절차의 다양성이나 실행 알고리즘의 향상으로 끝나는 것이 아니라 보다 심오한 이론의 발전과 관점의 전환을 이끌 수 있다. 앞으로 합역 관련 연구가 보다 더 활성화되기를 기대한다.

## 참고문헌

- 강영옥·조선희, 2012, “국가기초구역 설정 방법 비교 연구: 종로구, 아산시, 태안군을 사례로,” 한국지도학회지, 12(1), 113-127.
- 김갑영, 2011, “공간 상호작용 모델에 대한 공간단위 수정 가능성 문제(MAUP)의 영향,” 대한지리학회지, 46(2), 197-211.
- 김갑영, 2014, “경계 결합 가능성 기반 구역설정 최적화 모델,” 대한지리학회지, 49(3), 423-437.
- 김갑영·신정엽·이건학·조대현, 2009, “농촌지역 노년인구를 위한 방문 의료서비스 구역 설정 모델 및 알고리즘,” 대한지리학회지, 44(6), 813-831.
- 김갑영·이건학·신정엽, 2010, “공간분석 및 모델링을 이용한 지방행정구역 재설정에 관한 연구,” 한국지역지리학회지, 16(6), 673-688.
- 김명진, 2012, “새로운 인접성 식별 방법을 이용한 의회선거구 구획연구 알고리즘 개발,” 한국지도학회지, 12(3), 81-94.
- 김영훈, 2009, “구역설정문제에 있어 P-median 알고리즘 적용 연구,” 한국지도학회지, 9(2), 73-80.
- 이상일, 1999, “기능지역의 설정과 ‘공간단위 수정가능성의 문제(MAUP)’,” 한국지리환경교육학회지, 7(2), 757-783.
- 이상일·김갑영·제갈영, 2012, “지오컴퓨터이션 접근에 의한 주택시장지역의 설정: 우리나라 수도권에의 적용,” 한국도시지리학회지, 14(3), 59-75.
- 이상일·조대현·이민파, 2015, “일변량 공간연관성통계량에 대한 비교 연구 (I): 전역적 S 통계량을 중심으로,” 한국지리학회지, 4(2), 329-345.
- 이상일·조대현·이민파, 2016, “일변량 공간연관성통계량에 대한 비교 연구 (II): 국지적 S<sub>i</sub> 통계량을 중심으로,” 한국지리학회지, 5(3), 375-396.
- Adams, M.D., Kanaroglou, P.S., and Coulibaly, P.,

- 2016, Spatially constrained clustering of ecological units to facilitate the design of integrated water monitoring networks in the St. Lawrence Basin, *International Journal of Geographical Information Science*, 30(2), 390-404.
- Amrhein, C.G., 1995, Searching for the elusive aggregation effect: evidence from statistical simulations, *Environment and Planning A*, 27(1), 105-119.
- Anselin, L. and Smirnov, O., 1996, Efficient algorithms for constructing proper higher order spatial lag operators, *Journal of Regional Science*, 36(1), 67-89.
- Bailey, T.C. and Gatrell, A.C., 1995, *Interactive Spatial Data Analysis*, Harlow, Essex: Longman.
- Bivand, R.S., Pebesma, E., and Gómez-Rubio, V., 2013, *Applied Spatial Data Analysis with R*, 2nd Edition, New York: Springer.
- Bolstad, P., 2019, *GIS Fundamentals: A First Text on Geographic Information Systems*, 6th Edition, Ann Arbor: XanEdu.
- Boots, B. and Tiefelsdorf, M., 2000, Global and local spatial autocorrelation in bounded regular tessellations, *Journal of Geographical Systems*, 2(4), 319-348.
- Chang, K.-T., 2019, *Introduction to Geographic Information Systems*, 9th Edition, New York: McGraw-Hill Education.
- Cope, C.R., 1971, Regionalisation and the electoral districting problem, *Area*, 3(3), 190-195.
- Cox, E.P., 1927, A method of assigning numerical and percentage values to the degree of roundness of sand grains, *Journal of Paleontology*, 1(3), 179-183.
- Flowerdew, R., Geddes, A., and Green, M., 2001, Behaviour of regression models under random aggregation, in Tate, N. and Atkinson, P. A., eds., *Modelling Scale in Geographical Information Science*, Chichester: John Wiley & Sons, 89-104.
- Fotheringham, A.S. and Wong, D.W., 1991, The modifiable areal unit problem in multivariate statistical analysis, *Environment and Planning A*, 23(7), 1025-1044.
- Green, M. and Flowerdew, R., 1996, New evidence on the modifiable areal unit problem, in Longley, P. and Batty, eds., *Spatial Analysis: Modelling in a GIS Environment*, Cambridge: GeoInformation International, 41-54.
- Jaegal, Y., 2013, Delineating housing market areas in the Seoul metropolitan area using a geo-computational approach, *Journal of the Association of Korean Geographers*, 2(1), 7-20.
- Kim, M. J., 2012, Comprehensive review of compactness measures to political redistricting problems, *The Geographical Journal of Korea*, 46(4), 403-413.
- Kim, K., Dean, D.J., Kim, H., and Chun, Y., 2016, Spatial optimization for regionalization problems with spatial interaction: a heuristic approach, *International Journal of Geographical Information Science*, 30(3), 451-473.
- Koo, H., 2012, Improved hierarchical aggregation methods for functional regionalization in the Seoul metropolitan area, *Journal of the Korean Cartographic Association*, 12(2), 25-35.
- Lee, S.-I., Lee, M., Chun, Y., and Griffith, D.A., 2019, Uncertainty in the effects of the modifiable areal unit problem under different levels of spatial autocorrelation: a simulation study, *International Journal of Geographical Information Science*, 33(6), 1135-1154.
- Macmillan, W., 2001, Redistricting in a GIS environment: An optimisation algorithm using switching-points, *Journal of Geographical Systems*, 3(2), 167-180.
- Polsby, D.D. and Popper, R.D., 1991, The third criterion: compactness as a procedural safeguard against partisan gerrymandering, *Yale Law & Policy Review*, 9(2), 301-353.
- Stillwell, J., Daras, K., and Bell, M., 2018, Spatial aggregation methods for investigating the MAUP effects in migration analysis, *Applied Spatial Analysis and Policy*, 11(4), 693-711.
- Sun, M. and Wong, D.W., 2017, Spatial aggregation as a means to improve attribute reliability, *Computers, Environment and Urban Systems*, 65, 15-27.
- Tiefelsdorf, M., Griffith, D.A., and Boots, B., 1999, A variance-stabilizing coding scheme for spatial link

matrices, *Environment and Planning A*, 31(1), 165-180.

Wade, T. and Sommer, S. eds., 2006, *A to Z GIS*, Redlands: ESRI Press.

Wong, D., 1996, Aggregation effects in geo-referenced data, in Arlinghaus, S. L. ed., *Practical Handbook of Spatial Statistics*, Boca Raton: CRC Press, 83-106.

교신: 이몽현, 18448, 경기도 화성시 삼성전자로 1 삼성전자 메모리사업부(이메일: monghlee@gmail.com)

Correspondence: Monghyeon Lee, Memory Business, Samsung Electronics Co. Ltd., 1 Samsungjeonja-ro, Hwaseong-si, Gyeonggi-do, 18448, Republic of Korea (Email: monghlee@gmail.com)

**투 고 일: 2020년 12월 14일**

**심사완료일: 2020년 12월 20일**

**투고확정일: 2020년 12월 21일**

