

Journal of
the Korean Geographical Society
대한지리학회지

Volume 54, Number 5 (Series No. 194), October 2019

Articles

- Landform Classification Using Geomorphons on the Middle Yangsan Fault System, Southeastern Korea..... Dong-Eun Kim · Jeong-Sik Oh (493)
- Lineament Extraction and Its Comparison Using DEMs based on LiDAR, Digital Topographic Map, and Aerial Photo in the Central Segment of Yangsan Fault Jeong-Sik Oh · Dong Eun Kim (507)
- A Study on the Delineation and Application of Region on the Quasi Proportional Representation System - Simulation Based on the 20th General Election Lee Chung Sup · Cho Hanseok · Chi Sang-Hyun (527)
- Decomposing the Pearson's Correlation Coefficient Based on the Eigenvector Spatial Filtering Approach Sang-II Lee (545)
- Analysis of Influential Factors on the Educational Achievement by School Units using the Geographically Weighted Regression Model (GWR) Jiwoo Kim · Gunhak Lee (561)



대한지리학회지
Journal of the Korean Geographical Society

제54권 제5호 (통권 194호) 2019. 10

논문

- 지형패턴 분류방법(geomorphons)을 이용한 지형분류
- 중부 양산단층을 대상으로 - 김동은 · 오정식 (493)
- LiDAR, 수치지형도, 항공사진 기반 수치표고모델을 활용한 중부 양산단층의
선형구조 추출과 비교 오정식 · 김동은 (507)
- 준연동형 비례대표제의 권역 설정과 적용에 관한 연구
- 제20대 총선의 투표 결과에 기초한 시뮬레이션 이정섭 · 조한석 · 지상현 (527)
- 고유벡터공간필터링 접근법에 기반한 피어슨 상관계수의 요소분해 이상일 (545)
- 지리기중회귀 모델을 이용한 학교급별 학업성취도 영향 요인 분석 김지우 · 이진학 (561)



대한지리학회지

Journal of the Korean Geographical Society

Vol. 54, No. 5 October 2019

대한지리학회

지형패턴 분류방법(geomorphons)을 이용한 지형분류 - 중부 양산단층을 대상으로 -

김동은* · 오정식**

Landform Classification Using Geomorphons on the Middle Yangsan Fault System, Southeastern Korea

Dong-Eun Kim* · Jeong-Sik Oh**

요약 : 본 연구는 정량적인 지형분류 방법 중 형태적 특징에 따라 지형을 분류하는 지형패턴 분류방법(geomorphons)을 이용하여 양산단층 중부에 있는 활성단층인 벽계단층 일대를 대상으로 지형분류를 수행하였다. 입력 자료인 DEM과 분석을 수행할 때 설정할 수 있는 분석반경 및 평탄임계치 값을 조절하여 지형분류를 수행할 경우 분석결과에 어떠한 영향을 미치는지 살펴보았다. 수 km의 공간을 대상으로 지형패턴을 이용한 지형분류를 할 때, 5.0m 해상도의 DEM은 정밀한 지형분석에 어려움이 있었다. 분석반경은 10~30셀 사이에서 적절한 결과를 얻을 수 있었다. 여기서 분석반경이 너무 작으면 실제 지형을 반영하기 힘들며 분석반경이 너무 크면 분석결과에 차이는 없고 분석 시간이 오래 걸렸다. 평탄임계치는 기본값을 사용할 경우 실제 지형을 반영하는 데 어려움이 있었다. 지형패턴을 이용한 지형분류는 지형의 형태적인 특징에 기반하여 지형을 분류하기 때문에 현장답사와 시각적 지형분류 결과를 이용하여 검증하는 단계가 요구된다.

주요어 : 지형분류, 지형패턴 분류법(geomorphons), LiDAR, 제4기 단층, 양산단층

Abstract : The construction of high-resolution Digital Elevation Model (DEM) with cell size smaller than 1m and the development of GIS technology enabled quantitative topographic classification. This study carried out the landform classification using the geomorphons, which categorize the terrain according to its morphological characteristics. The study area is around the Byeokgye fault in the middle Yangsan Fault System. DEM, the analysis radius and the flatness threshold were changed to see how the result of classification would be different. The result indicated that it was difficult to perform precise topographic analysis when using a 5m resolution. In addition, the appropriate analysis radius were obtained between 10 and 30cells. The result also found it was difficult to reflect the actual landscape when the analysis radius was too small. If Outer Search Radius was set over 30cells, the result shows no difference regardless of Outer Search Radius. Flatness Threshold needs a optimized value rather than the default value (1°). In conclusion, the field surveys and visual classifications is essential because geomorphons has classified the landform based on the shape of the terrain.

Key Words : landform classification, geomorphons, LiDAR, Quaternary fault, Yangsan Fault

* 고려대학교 미래국토연구소 전임연구원(Full-time Researcher, Research Institute of Future Land, Korea University, roehd@dms@korea.ac.kr)

** 고려대학교 지리교육과 연구교수(Research Professor, Department of Geography Education, Korea University, ojs0201@korea.ac.kr)

1. 서론

2016년 경주지진과 2017년 포항지진은 한반도의 활성단층 존재 가능성을 환기했으며, 이를 연구하기 위해 조사연구기술의 선진화도 이루어지게 하였다(오정식, 2019). 한반도와 같은 판의 내부에 있는 지역의 경우 중규모의 지진이 일어나며, 이들은 주로 활성단층(active fault)을 따라 발생한다(Kim *et al.*, 2011; Choi *et al.*, 2017). 활성단층에 대한 정의는 국가마다 다양하나 국내에서는 제4기 층을 절단하는 단층을 4기 단층(Quaternary fault), 5만 년 이내에 한번 이상 단층 활동이 있거나 50만 년 내에 두 번 이상의 움직임이 있었던 단층을 활동성 단층(Capable fault) 등으로 구분하여 사용하기도 한다(Im *et al.*, 2004; Kim *et al.*, 2011). 판의 내부에 위치하고, 더불어 지표의 풍화, 침식이 활발한 우리나라의 경우 비교적 최근(제4기 이후) 형성된 지진의 흔적이 남아있기 힘들어 활성단층을 추적하기 어려운 환경이다.

선형구조 분석과 지형분류도 작성은 조사와 추적이 까다로운 활성단층 연구에 있어서 가장 먼저 수행되는 지표분석 작업이다. 선형구조 분석은 지진을 유발했던 단층의 흔적을 찾는 것으로 활성단층 가능성이 있는 추정 단층선을 찾고, 굴착조사나 물리탐사를 위한 기초자료를 제공하기 때문이다(오정식, 2019). 이 중 지형분류도는 지형의 형태, 지표구성물질의 공간적 분포, 물질의 이동 경로 등을 표현하기 위해 작성하는 지도로써(Griffiths, 2004; 탁한명, 2014b) 지역에 대한 기초정보를 제공하기 때문에 여러 조사사업에서 기초자료로써 이용된다. 원격탐사 기술의 발달에 따른 고해상도 수치고도모델(digital elevation model, 이하 DEM)은 지형분류도가 단순히 기초자료를 제공하는 기능뿐만 아니라 생태자원의 지속 가능한 보존, 이용, 개발과 관련된 지형분류도 제작을 가능하게 하였다(김태호, 2003; 박수진, 2004; 이민부, 2010). 활성단층 조사에서 지형분류도는 과거 지진을 유발한 단층 활동으로 인해 형성된 지표면의 흔적을 추적할 수 있는 정보를 제공하므로 매우 중요하다. 기존의 지형분류도 작업은 주로 수치지도, DEM,

위성영상, 항공사진 등을 이용하여 전문가가 주관적으로 분류하고 현장조사를 통해 검증하는 방식으로 제작되기 때문에 많은 시간과 노력이 소요되었다.

지형분류 방법의 하나인 지형패턴(geomorphons)을 이용한 지형 분류법은 2011년 폴란드 지리학자인 Stepinski와 Jasiewicz에 의해 제안되었다(Stepinski and Jasiewicz, 2011). 국내에서는 김동은 등(2012)과 탁한명(2014a, 2014b)에 의해 방법론적 적용과 한반도를 대상으로 지형분류를 할 때 최적 변수를 설정하는 것에 관한 연구가 이루어진 바 있다. 이 방법은 지형의 연속성을 잘 반영하는 장점이 있으며, 대용량의 자료 처리가 가능하므로 기존의 연구들이 진행하지 못했던 큰 공간 규모의 지형분류가 가능하다는 장점이 있다(탁한명, 2014a, 2014b). 다만, 기존 연구들은 모두 수십 km 이상 되는 공간을 대상으로 분석을 진행하였다.

연구지역인 벽계단층은 중부 양산단층대에 속하며, 최근 굴착조사를 통해 제4기 퇴적층을 절단하는 새로운 단층핵이 발견되었다. 송영식 등(2018)은 벽계단층의 북측 연장선에서 실시한 굴착조사 결과를 토대로 벽계단층은 양산단층의 가지단층이 아니라 주단층이며, 제4기 동안 최소 3회 이상 재활하였음을 보고 했다. 본 연구는 고해상도 DEM을 이용한 자동화된 지형분류 방식이 활성단층 운동의 영향을 받아 변형된 지형을 추적할 수 있는지를 살펴보기 위해 지형패턴 분석을 이용한 지형분석을 수행하였다. 분석을 위해 LiDAR(Light Detection And Ranging) 기반의 고해상도 DEM을 입력자료로 이용하였고, 입력변수를 바꿔가며 변수값의 변화에 따른 지형패턴 분석 결과의 차이를 살펴보았다. 또한, 분석결과와 검증을 위해 기 보고된 선형구조 및 제4기 지형분류 결과와 비교하여 지형패턴을 이용한 제4기 지형면 분류의 가능성에 대해 살펴보았다.

2. 패턴을 이용한 지형분류 원리와 변수(parameter) 설정

1) 패턴을 이용한 지형분류의 원리

패턴을 이용한 지형분류의 원리에 대한 설명은 Jasiewicz and Stepinski(2013)과 탁한명(2014b)의 논문에서 자세하게 설명되어 있다. 본 고에서는 간단한 원리에 관하여 설명하고자 하며, 국문 용어에 대해서는 탁한명(2014b)의 연구에서 사용한 것을 참조하였다. 패턴을 이용한 지형분류는 지표의 기복을 분석을 수행하는 지점을 기준으로 하여 지표의 기복을 ‘기준보다 높다(+), ‘낮다(-), ‘기준과 비슷하다(0)’의 3개로 나누는 LTP(local ternary patterns) 개념을 이용하여 분석을 수행한다(Liao, 2010). 이를 8개의 방향에서 수행할 경우 이론적으로 38(6,561)개의 패턴을 얻을 수 있게 된다. 그러나 이 패턴은 좌우의 대칭과 같은 중복된 패턴이 존재하기 때문에, 이를 제거하면 498개의 패턴이 나온다. 이 중에서 비슷한 패턴을 하나의 지형으로 묶어 총 10개의 대표지형(그림 1)으로 분류한다(Jasiewicz and Stepinski, 2013).

패턴을 이용한 지형분류 방법은 몇 가지 장점이 있다. 이 분류방법은 규모의 유연성(scale flexibility)을 갖기 때문에 스케일 종속성(scale dependency)으로 인해 발생할 수 있는 문제점을 개선하였다(Jasiewicz

and Stepinski, 2013). 기존 미분기하학 방법을 이용한 지형분류법들은 고정된 분석반경을 이용하기 때문에 분석반경의 크기를 어떻게 설정하느냐에 따라서 전혀 다른 지도가 만들어질 수 있다(Wood, 2002; Fisher *et al.*, 2004). 고정된 분석반경을 획일적으로 적용하는 기존 방법론들과 달리 시준선 원리(line of sight)를 따라 국소적인 지형의 기하를 반영하여 동적으로 지형패턴이 설정되기 때문에 변수설정의 변화에 따른 편차가 작고, 출현하는 지형이 안정적으로 나타나며, 지형분류 감도와 처리 속도의 트레이드오프를 고려한 최적 변수설정이 용이하다(Jasiewicz and Stepinski, 2013; 탁한명, 2014b; 그림 5). 또한, 오픈소스(Open source)인 GRASS GIS를 이용하여 분석을 수행하여 저작권 문제에서 자유롭다. 또한, 분석을 수행하는 데 있어 giga-cell (109 개) 이상의 대용량 자료를 처리하는데도 문제없이 분석이 가능하다(Jasiewicz and Stepinski, 2013).

2) 패턴을 이용한 지형분류법 변수 설정

(1) 분석지역

연구지역은 벽계단층 일대로 행정구역상 경상북도 경주시 강동면 단구리에 속한다(그림 2a). 이 지역은 양산단층의 주단층대가 통과하는 위치로 양산단층의 제4기 단층운동이 보고된 바 있다(Choi *et al.*, 2012). 벽계단층은 양산단층의 주단층과 평행한 북북동-남

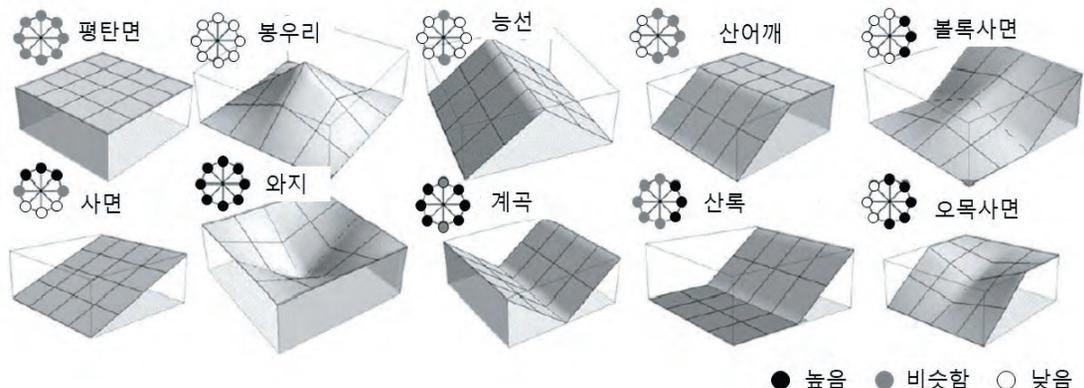


그림 1. 3D로 유형화한 대표지형 10가지
출처: Jasiewicz and Stepinski(2013)을 수정

남서 방향의 주향과 거의 수직에 가까운 경사를 보이며, 수평에 가까운 단층조선이 관찰되는 우수향 주향 이동단층이다. 이 단층을 경계로 동측은 세립질 화강암과 안산암이 발달해 있고, 서측은 제4기 퇴적층이 발달하여 서로 접촉하고 있다(Ryoo *et al.*, 1999).

벽계단층 노두의 제4기 퇴적층을 대상으로 실시한 OSL 연대측정 결과는 75 ± 3 ka를 보고되었으며(Kee *et al.*, 2009), 최근 인접 지역에서 굴착조사를 통해 발견된 새로운 단층(벽계단층 연장) 역시 제4기 이후 최소 3회 이상의 단층운동이 있었음이 밝혀졌다(송영석 등, 2018). 이러한 연구결과들은 연구지역에서 제4기 이후 단층운동이 있었음을 시사하며, 수백 m 이상 수 km에 달하는 단층의 연장을 보일 수도 있음을 지시한다(류충렬 등, 2018). 본 연구는 지형패턴을 이용한 분석을 수행하여 단층활동의 영향을 받은 제4기 지형을 찾기 위해 LiDAR를 이용하여 구축된 고해상도 DEM을 사용하였다. DEM의 경계는 벽계단층 지점을 포함하는 동-서 3.5 km, 남-북 7.7 km의 지역이며, 면적은 약 27km²이다.

(2) 분석환경

DEM을 이용하는 연구에서 DEM의 해상도는 분석

결과의 정확성을 결정짓는 중요한 요소 중 하나이다. 지금까지 지형패턴 분류법을 사용한 연구들은 30m 해상도의 DEM을 사용했다(김동은 등 2012, 탁한명, 2014b). 이는 분석하고자 하는 공간의 규모가 평균군, 영남 알프스, 새만금 간척지와 같이 컸기 때문이다. 하지만 본 연구에서는 앞서 언급한 연구지역보다 작은 면적을 대상으로 하며, DEM의 해상도 역시 앞선 연구에서 사용한 30m보다 정밀한 DEM을 이용하여 지형분류를 수행하였다.

분석에 사용한 DEM의 해상도는 0.5m, 1.0m, 5.0m로 총 3종류이다. ‘국가 활성단층 연구단’에서 2018년 촬영한 항공 LiDAR를 이용하여 구축한 0.5m 해상도 DEM를 이용하였다. 항공 LiDAR는 식생의 영향을 최소화하기 위해 겨울에 촬영된 자료를 이용하였다. 원시 자료 측량결과는 측량기준점과 비교하여 검증하였으며, DEM을 제작한 후 평면위치의 정확도(RMSE 0.2m 이내)와 수직위치 정확도(RMSE 0.1m 이내)를 검증하였다. 마지막으로 육안으로 검사하고 오류가 확인되면 수정하였다(국가활성단층연구단, 2019).

0.5m 해상도 DEM을 바탕으로 다시 1.0m와 5.0m 해상도의 DEM을 제작하였다. DEM의 해상도 차이

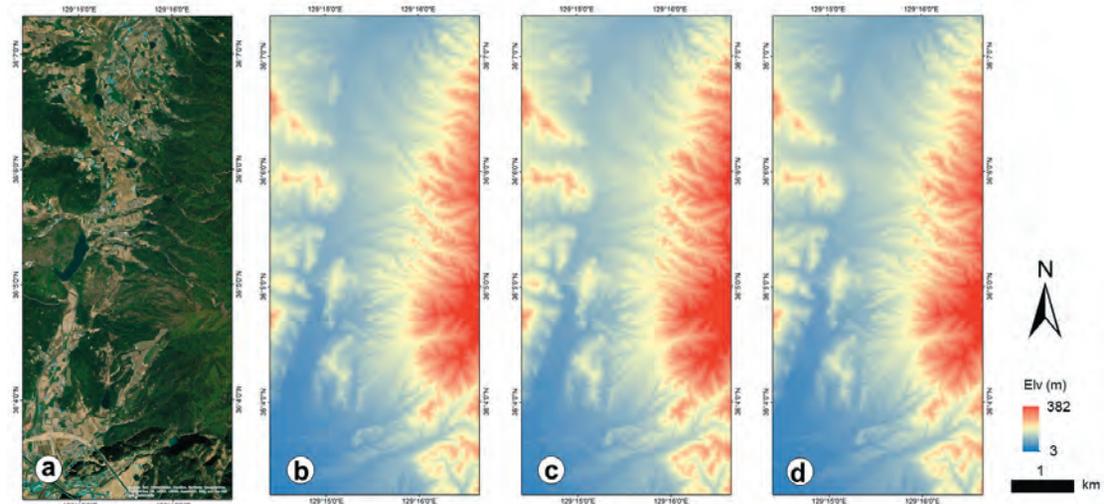


그림 2. 연구지역 위성영상과 분석에 사용된 DEMs. 각 그림은 위성영상(a)과 0.5m 해상도(b), 1.0m 해상도(c), 5.0m 해상도(d)의 DEM이다.

위성사진 출처: ESRI base map

가 분석결과에 미치는 영향을 살펴보기 위해 입력변수는 r.geomorphons 내에서 기본으로 제공하는 기본값을 사용하였다. 또한, 탁한명(2014b)의 연구에서 언급하였듯이 셀 값이 없는 지역에서 지형패턴 분석도중 오류가 날 가능성을 배제하기 위해 연구지역보다 더 넓은 범위를 포함하는 분석을 시행한 후 결과물을 연구지역의 범위에 맞게 편집하는 작업을 진행하였다.

래스터를 편집한 결과 5.0m 해상도의 DEM은 698×1,554개의 셀이, 1.0m 해상도의 DEM은 3,489×7,772, 0.5m 해상도의 DEM은 6,978×15,544개의 셀을 갖는다(그림 2). 분석에 사용된 소프트웨어는 Window용 GRASS 7.6.1이며, addon extension에서 r.geomorphons를 설치하여 분석을 수행하였다.

3) 변수의 종류 및 변수설정

패턴을 이용한 지형분류는 크게 4가지 변수를 설정할 수 있다. 변수의 종류와 각각의 특성에 대해서는 기존연구(Jasiewicz and Stepinski, 2013; 탁한명, 2014b)에 자세하게 나와 있으며, 다음은 각 변수에 관한 간략한 설명이다. 변수의 종류는 분석반경(outer search radius), 평탄임계치(flatness threshold), 배제거리(inner search radius), 평탄거리(flatness distance)가 있다. 분석반경은 지형패턴 분석의 범위를 정의하는 변수이며 기본값으로 3 cells로 되어있다. 평탄임계치는 분석하고자 하는 중심 셀을 기준으로 고도가 같다고 할 수 있는 임계각(threshold angle)을 정의하며 기본값은 1°로 되어있다. 이 임계치보다 낮을 경우 고도가 같다(0)고 할 수 있다. 배제거리는 분석할 때 지형의 최소 규모를 설정하는 변수로 규모가 너무 작은 지형을 제거하는 역할을 한다. 평탄거리는 배제거리와는 반대로 해상도가 너무 낮아 분석반경에 따른 고도차를 찾지 못할 때 나타나는 영향을 제거하기 위해 사용하는 변수이다. 분석반경과 평탄임계치는 분석을 위해 필요한 변수들이다(탁한명, 2014b). 그러나 배제거리와 평탄거리는 특정한 환경(관심의 대상이 되지 않는 지형을 필터링할 때 혹은 1×1km 이상의 저해상도 DEM을 사용할 때)

에서 사용되는 변수들이다. 본 연구지역은 배제거리와 평탄거리를 이용하는 특정 환경에 속하지 않으므로 분석반경과 DEM의 해상도의 차이에 초점을 맞추어 변수를 설정하였다.

3. 지형패턴을 이용한 지형분류

1) DEM 해상도 차이에 따른 지형패턴 변화

기본값으로 설정한 상태에서 DEM의 해상도만 변화를 준 결과는 그림 3과 같다. DEM의 해상도가 높을수록 저수지와 같이 평탄한 지형면도 지형분류가 되어있음을 관찰할 수 있다. 해상도에 따른 지형패턴들이 차지하는 비율의 변화는 그림 4와 같다. 해상도가 낮아질수록 즉, 셀의 크기가 커질수록 사면의 비율은 낮아지며 다른 지형들의 비율이 높아지는 것을 볼 수 있다. 특히, 블록사면의 비율이 다른 지형분류에 비해 크게 증가했다. 평탄면의 경우 해상도가 0.5m에서 1.0m로 증가할 때 비율이 증가하는 경향을 보이다가 1.0m에서 5.0m로 증가할 경우 다시 감소하는 경향을 보였다. 이처럼 셀 크기에 따른 사면의 감소 경향은 영남 알프스를 대상으로 지형패턴 분석을 수행한 기존연구(탁한명, 2014b) 결과와도 유사하다. 하지만 사면의 감소를 제외한 다른 지형패턴들의 변화는 영남 알프스의 분석결과와 약간의 차이를 보인다. 이러한 경향성 차이는 영남 알프스와 본 연구지역의 공간 규모 차이 및 지형 기복의 차이로 인해 발생한다고 해석할 수 있다.

2) 변수 변경에 따른 지형패턴의 결과

분석반경에 따른 지형패턴 결과를 살펴보기 위해 3 종류의 DEM을 가지고, 5, 10, 15, 20, 25, 30, 35, 40 셀 총 8종류의 분석반경을 가지는 지형패턴 분석을 수행하였으며 그중 1.0m와 5.0m 해상도의 DEM을 이용하여 10, 20, 40셀의 분류결과를 그림 5에 표현하였다. 그림 5에서 볼 수 있듯이 분석반경이 커질수

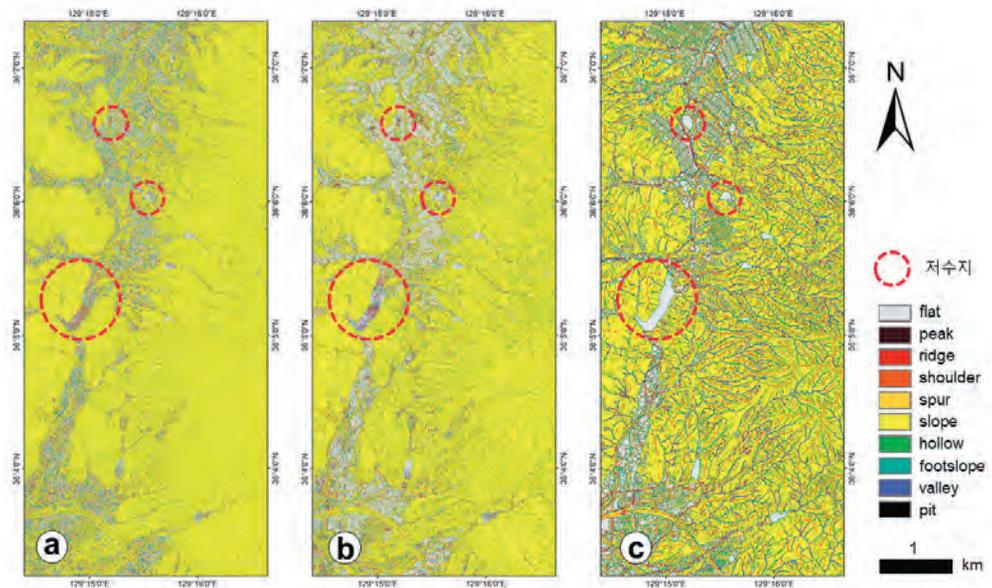


그림 3. DEM 해상도 차이에 따른 지형패턴 분석결과. 각 그림은 0.5m 해상도 DEM(a), 1.0m 해상도 DEM(b), 그리고 5.0m 해상도 DEM(c)을 이용하여 분석한 결과이며, 변수는 모두 기본값으로 설정함(분석반경: 3cell, 평탄임계치: 1°).

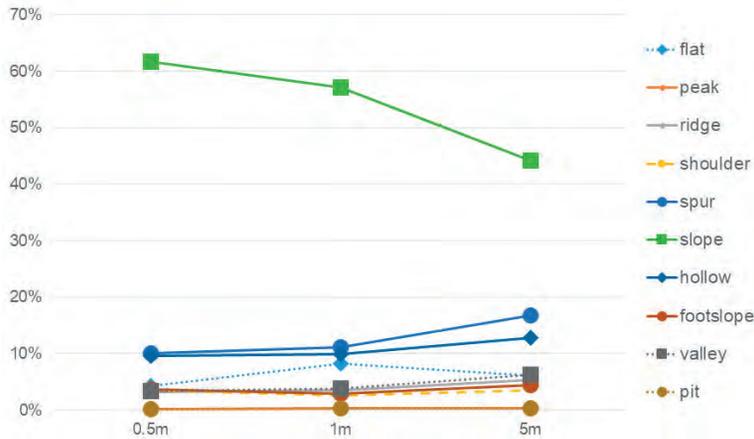


그림 4. DEM 해상도 차이에 따른 지형패턴 비율 변화.

록 계곡, 능선과 같은 선형의 지형들이 굵게 나타나는 것을 볼 수 있다. 이러한 현상은 5m의 DEM을 이용했을 때 더욱 두드러지게 나타났다. 10셀의 분석반경을 사용한 그림 5d와 40셀의 분석반경을 사용한 그림 5f를 비교했을 때, 저수지 평탄면으로 나타났던 지역이 계곡으로 분류되었다.

정량적인 지형패턴의 분포 변화를 살펴보기 위해

지형패턴들이 얼마나 변화하였는지 비율을 분석하였다. 분석결과 봉우리와 와지는 전체 지형패턴에서 차지하는 비율이 매우 낮으며, 변화 폭이 작았기 때문에 비율 변화를 살펴보는 분석에서 제외하였다. 분석반경이 커질수록 사면의 비율은 점차 줄어들었으며, 다른 지형들이 증가하는 경향을 보였다(그림 6). DEM 해상도에 따른 작은 차이는 있으나 전반적으로 5셀에

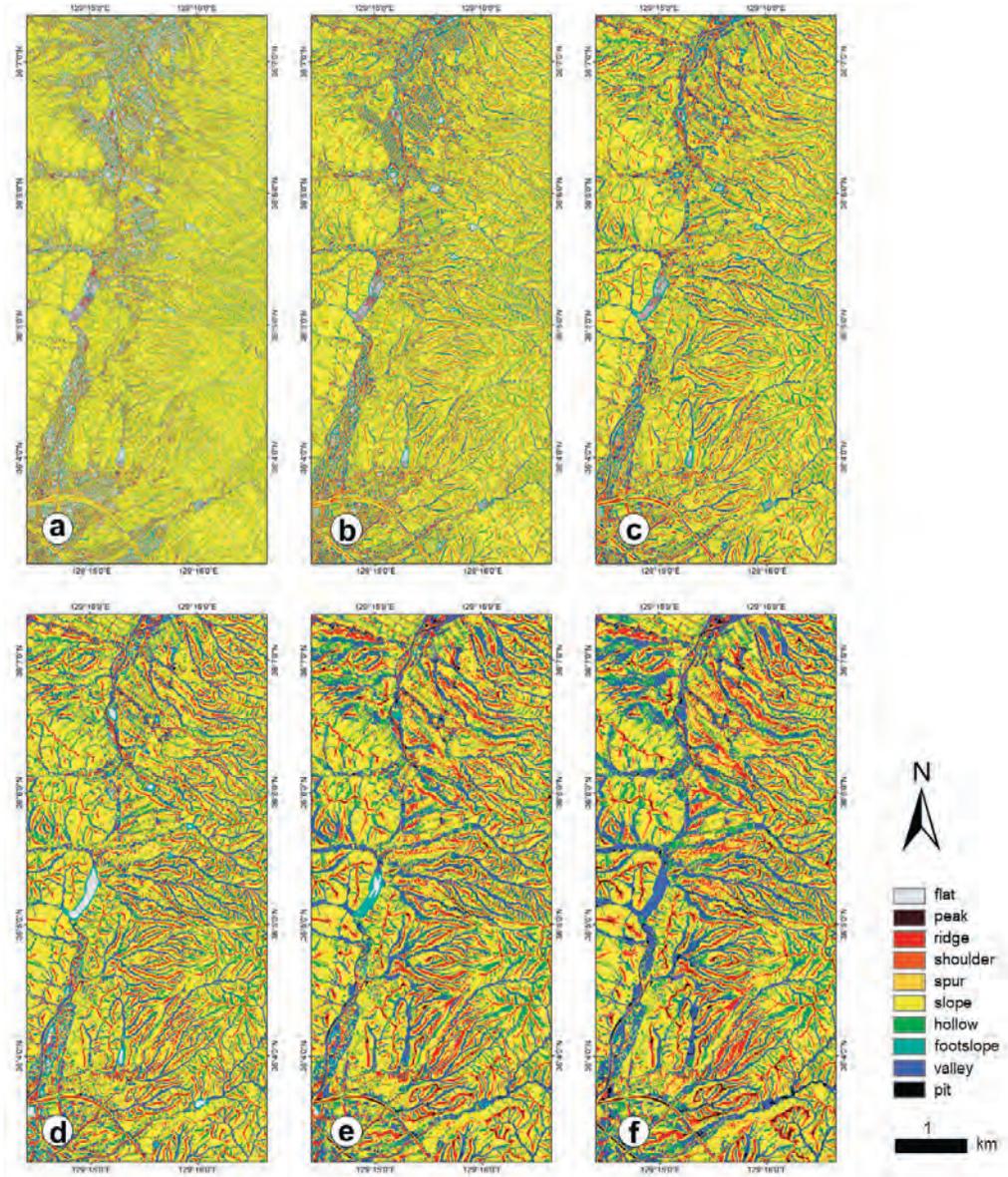


그림 5. 분석반경에 따른 지형패턴 변화. 1.0m 해상도 DEM을 이용하여 분석한 결과(a~c)와 5.0m 해상도 DEM을 이용하여 분석한 결과(d~f)를 보여주며, 평탄임계치는 1°로 고정되었다.

서 10셀로 분석반경이 변화하는 과정에 크게 변화하며, 30셀을 경계로 변화의 폭이 완만해지는 것을 확인하였다.

평탄임계치의 변화에 따른 결과를 살펴보기 위해 분석반경은 동일하게 설정하고, 평탄임계치는 기본값인 1°에서 5°로 바꿔 분석을 수행하였다. 값을 5°

로 설정한 이유는 한반도를 대상으로 지형패턴을 이용하여 분석을 수행한 결과 기본값 1°보다는 5°가 적절하다는 탁한명(2014a, 2014b)의 제언을 참조했다. 결과는 그림 7과 같으며, 평탄임계치를 조정할 결과 저수지와 같은 지형들은 모두 평탄면으로 분류가 되었다. 아울러 논과 밭으로 되어있는 면들도 모두 평탄

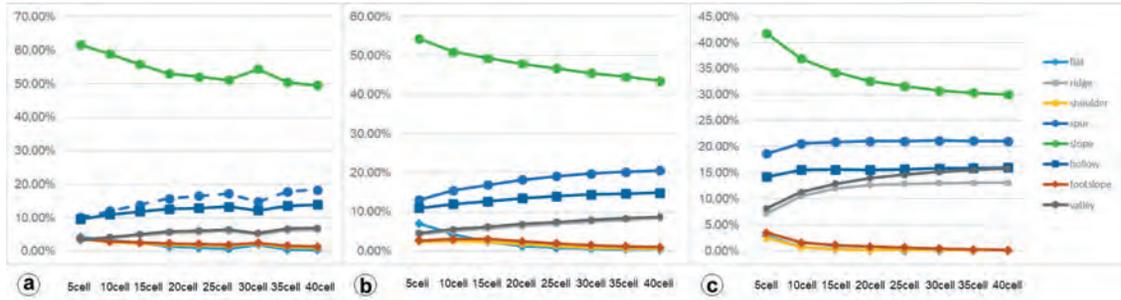


그림 6. 분석반경에 따른 지형패턴 비율 변화. 각 그림은 0.5m 해상도 DEM(a), 1.0m 해상도 DEM(b), 5.0m 해상도 DEM(c)을 이용한 결과이며, 평탄임계치는 1°로 고정되었다.

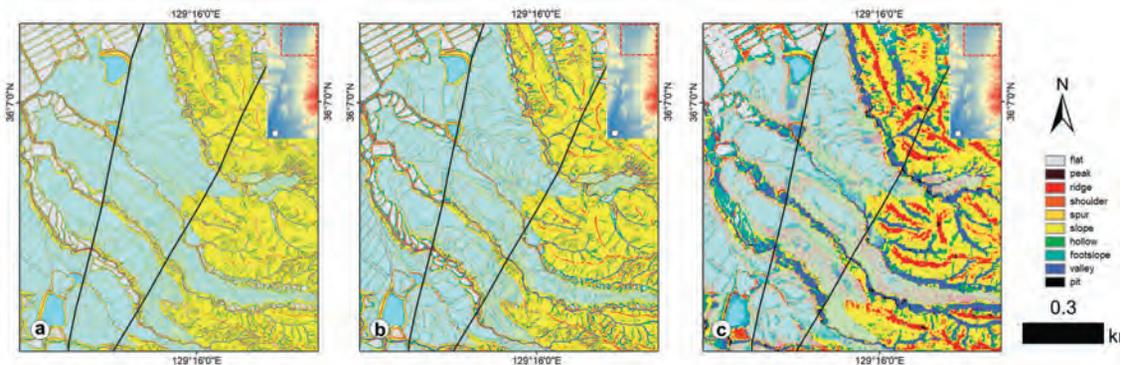


그림 7. 분석반경은 동일하고 평탄임계치를 5°로 수정하여 분석한 결과. 각각 0.5m 해상도 DEM(a), 1m 해상도 DEM(b), 5m 해상도 DEM(c).

면으로 분류 되었음을 확인하였다(그림 7).

3) 지형면 분류도와의 비교

고해상도 LiDAR를 이용하여 면분류한 결과와 지형패턴 분류를 통해 나온 결과를 비교했다. 앞서 분석 반경에 따른 임계값의 차이를 반영하여 20셀을 기준으로 분석한 결과를 바탕으로 현장조사와 시각적 관독을 바탕으로 분류한 지형분류도와 비교하였다(그림 8). 그림 8a는 LiDAR를 이용하여 선상지면을 분류한 것이다. 하늘색 실선은 1:25,000 수치지형도에서 추출한 하천이다. 검은색 실선은 LiDAR를 이용하여 선형구조를 추출한 것이다. 그림 8에서 볼 수 있듯이 DEM 해상도에 따라 시각적 관독 결과와 큰 차이를 보인다. 특히 계곡, 능선과 같은 지형은 같은 값을 설정했음에도 불구하고 DEM 해상도에 따라 차이가

났다. 이러한 특징은 5.0m 해상도를 이용하여 분석한 결과에서 더욱 두드러지게 나타났다. 0.5m 해상도 DEM을 이용한 그림 8b와 1.0m 해상도 DEM을 이용한 그림 8c는 계곡으로 표시된 부분이 실선으로 나타났다. 5.0m 해상도 DEM을 이용한 그림 8d의 경우 논으로 되어있는 부분이 계곡으로 분류되었다. 이는 선상지를 개척하는 하곡을 포함하여 계곡으로 분류한 것으로 볼 수 있으나, 본 연구지역과 같이 작은 공간 규모에서 지형분류가 필요할 경우 적절하지 않다. 또한, 선상지면으로 분류된 그림 8a의 하늘색 부분은 지형패턴 분류방법을 이용한 분석을 수행했을 때 분명하게 구분할 수 없었다. 이는 형태적인 특징에 따라 지형(landform)을 구성하고 있는 요소들로 분석을 수행했기 때문에 형태적 특징이 다른 지형요소들만 분석되었을 뿐 선상지라는 지형을 반영하지는 못하였기 때문이다.

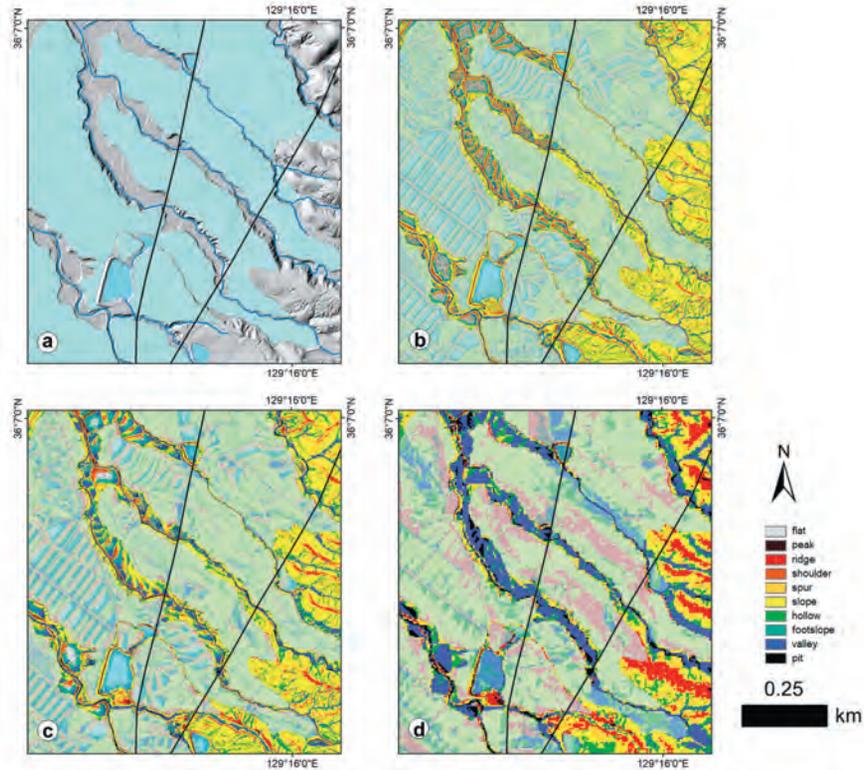


그림 8. LiDAR를 이용한 지형면 분류 결과와 분석반경을 20셀로 분석한 결과, 각 그림은 선상지 분류(a), 0.5m 해상도(b), 1.0m 해상도(c), 5.0m 해상도(d)를 이용하여 분석한 결과를 보여준다.

4. 논의

지형패턴 분류방법을 이용하여 분석한 결과는 DEM, 분석반경, 평탄임계치의 변화에 따른 변화만을 살펴보았으며, 연구지역에 최적화된 매개 변수에 관한 검증은 생략되었다. 따라서, 지형분류결과와 분류된 지형과 단층 관련성을 직접 논하기는 어렵다. 본 고에서는 이러한 변수들이 지형패턴 분류결과에 어떠한 영향을 미치고, 지형패턴 분류방법을 이용하여 지형면을 분류할 때 어떠한 점들이 고려되어야 하는 지에 대해서 살펴보았다.

1) DEM 해상도와 분석반경 변수 변경에 따른 지형패턴 분류결과

한반도를 대상으로 지형패턴 분류방법을 이용한 연구들(김동은 등, 2012; 탁한명, 2014a, b)은 모두 30m 해상도의 DEM을 이용해 분석을 수행하였다. 이들은 평창군이나 한반도 전체와 같이 넓은 지역을 분석하였기 때문에 30m보다 더 작은 DEM을 이용하여 분석을 수행하지 않았다. 그러나 본 연구와 같이 작은 규모의 지형분석을 수행할 경우 DEM의 해상도에 따라 지형분석의 결과가 큰 차이를 나타낼 수 있으므로 DEM의 해상도는 중요하다.

연구결과 비교적 좁은 공간(수 km 내의 범위)의 지형분석을 지형패턴 분류방법으로 분석할 경우 5.0m 해상도 DEM보다 정밀한 고해상도 DEM이 필요한 것을 확인할 수 있었다(그림 5, 7, 8). 5.0m 해상도

DEM을 이용할 경우 발생할 수 있는 가장 큰 문제점은 계곡지형을 명확하게 분류할 수 없다는 점이다. 수치지형도에서 추출한 계곡을 비교했을 때, 5.0m 해상도 DEM을 이용한 지형패턴 분류 결과는 논으로 이용되고 있는 지역도 계곡으로 분류하였다(그림 5, 7, 8). 보다 대규모 지역을 대상으로 분석할 경우에 이러한 분류는 분석에 큰 영향을 미치지 않을 수 있지만, 수 km의 규모에서는 황성단층과 관련된 지형의 변위나 단층지형을 찾기 위한 정밀한 분석을 방해한다. 따라서 수 km 규모에서 지형패턴을 이용한 지형분류를 수행할 때는 1.0m 이하의 고해상도 DEM을 사용하는 것이 적절하다.

분석반경 변수 변경에 따른 지형패턴 분류 결과의 변화를 종합하면 다음과 같다. 우선 분석반경 범위는 영남 알프스를 대상으로 분석을 수행한 탁한명(2014b)에서 제시한 범위와도 일치한다. 탁한명(2014b)에서 30m 해상도의 DEM을 이용하여 450m에서 750m 사이에서 임계값을 제안한 바 있다. 이를 셀로 바꾸면 15셀에서 25셀로 나오며 이는 온 임계값의 범위와 유사하게 나타난다. 30셀 이상의 값을 설정하여 분석을 수행할 경우 지형패턴의 변화는 거의 없지만 분석에 요구되는 시간이 급격하게 늘어나게 된다. 이는 30셀 이상에서 패턴의 변화가 나타나지 않는다고 한 기존의 연구결과와도 일치하였다(Jasiewicz and Stepinski, 2013).

평탄임계치 역시 지형패턴 분류방법을 이용할 때 중요한 변수이다. 기존연구들(김동은 등, 2012; Jasiewicz and Stepinski, 2013)은 평탄임계치 값을 기본값으로 설정하고 설정반경과 DEM 해상도 차이에 따른 분석결과 차이만 논의하였다. 그러나 탁한명(2014b)은 한반도를 대상으로 분석을 수행할 경우 평탄임계치를 5°로 설정하는 것이 적절하다고 제시한 바 있다. 본 연구는 이 값을 이용하여 분석을 수행하였으며, 그 결과 5°로 설정했을 경우 평탄한 저수지, 논, 그리고 밭과 같은 지형이 평탄면으로 분류되는 것을 확인 했다. 그러나 앞서 언급한 바와 같이, 본 연구는 매개 변수에 대한 정량적 검증을 생략했다. 따라서, 연구지역의 경우 평탄임계치를 기본값인 1°로 사

용하기보다 지역의 지형면을 적절하게 반영하기 위한 값으로 변경할 필요가 있다. 하지만 그 값으로 5°가 적절한지는 추가적인 매개 변수 검증연구가 필요하다.

2) 지형패턴 분류방법을 이용한 지형분류 논의

지형패턴을 이용한 지형분류는 지형을 구성하고 있는 요소들을 그룹화하여 대표적인 10개의 지형으로 분류하는 것이다. 국내지형을 대상으로 지형요소를 구분한 연구(김남신 등, 2014)를 기준으로 지형패턴을 나누면 크게 선형적으로 나타나는 계곡과 능선, 점으로 나타나는 정상부와 와지, 면으로 나타나는 나머지 지형패턴 요소로 나누어 볼 수 있다. 계곡이나 능선과 같이 선형으로 나타나는 지형은 지형패턴의 결과를 바탕으로 단층지형을 추적하는데 이용할 수 있을 것으로 판단된다. 특히 계곡의 경우 수치지형도를 이용한 계곡과 비교했을 때 일치하는 모습을 보였다. 그러나 지형패턴을 이용한 방법은 제4기 지형면을 분류할 때 분명한 한계가 있다. 이러한 분석은 지형의 형태적인 특징에 의해 지형요소만을 분류하였기 때문에 학문적으로 분류할 수 있는 지형들과 차이가 있을 수 있다(이민부, 2010). 예를 들어 제4기 지형면은 단구, 선상지와 같은 면으로 된 지형들이다. 이러한 지형들은 단순히 지형패턴만으로 표현할 수 있는 지형이 아니라 지형패턴을 구성하고 있는 요소들이 결합하여 나타나는 지형이다.

형태적인 특징에 의해 분류된 지형패턴을 학문적 의미의 지형으로 분류하기 위해서는 지형패턴들을 조합하여 학문적 의미의 지형으로 제시할 수 있는 기준이 필요하다. 여기서 GIS를 이용하여 지형을 정량적으로 분석한 후, 현장답사를 통해 지형요소를 8가지(배후산지, 피복형완사면, 선형완사면, 저기복구릉, 고립구릉, 개석곡저면, 하안단구, 범람원)로 상세하게 구분한 김남신 등(2003)의 연구는 좋은 사례이다. DEM을 이용하여 자동화된 방법으로 지형분류를 하는 접근법은 현 단계에서는 개별 지형요소만을 추출할 뿐, 인간이 개념화한 지형을 분류하는 데 한계가 있는 것은 사실이지만, 지형패턴으로 추출된 개별 지

형요소와 현장답사를 통해 확인한 지형 단위 간의 상호관계를 살펴볼 수 있는 자료가 축적된다면 장기적인 관점에서 정확한 지형분류에 이바지할 수 있을 것으로 기대한다.

또한, 지형면 분류가 이루어진 후 시각적으로 분류한 지형분류도와 반드시 교차검증이 필요하다. 그동안 저해상도 DEM을 이용하여 큰 규모의 공간에서 대략적인 지형분석 및 지형면 분류를 수행할 수 있었다. 하지만 활성단층 추적과 같이 수 km 규모의 비교적 좁은 공간에서 단층의 영향을 받은 지형을 찾기 위해서는 고해상도 DEM이 필요하다. 이를 바탕으로 수행한 선형구조 분석결과와 고해상도 DEM을 이용하여 육안으로 분류한 지형분류도를 반드시 비교하고 검증하는 작업이 필요하다.

5. 요약 및 결론

본 연구는 지형패턴 분류방법의 개념과 원리에 대해 간단히 설명하고, 분석에 사용되는 입력자료인 DEM의 크기와 분석반경 그리고 평탄임계치 같은 변수들의 설정값 변화에 따른 지형패턴의 변화를 분석하였다. 연구지역은 중부 양산단층에 속한 벽계단층 일대로 한반도와 같이 넓은 지역을 대상으로 분석을 수행한 기존연구들과 달리 작은 규모의 지역을 대상으로 하여 분석을 수행하였다. 또한, 이를 바탕으로 얻어진 지형패턴 분류결과를 시각관독을 통해 분류한 결과와 비교하였다.

입력자료인 DEM은 LiDAR를 이용하여 구축한 DEM을 이용하였으며, 분석하는 공간 규모에 맞게 0.5m, 1.0m, 5.0m의 해상도를 가지는 DEM을 분석에 이용하였다. 분석결과 0.5m, 1.0m의 DEM의 경우 계곡과 같은 지형을 표현하는데 적절한 것으로 나타났다. 그러나 5m 해상도를 이용하여 지형분류를 수행한 결과 계곡과 같은 지형을 표현할 때 다른 지형, 예를 들어 논으로 이용되고 있는 지역들도 계곡으로 표현을 하는 것을 관찰하였기 때문에 연구지역의 지형을 표현하는 데 한계가 있음을 알 수 있었다.

따라서 본 연구에서 수행한 지역에서는 0.5m 이상의 해상도를 가지는 DEM을 사용하면 정확한 지형분류가 힘들다는 결론을 도출하였다. 설정값 중 하나인 분석반경은 이 분석방법에서 중요한 변수이다. 연구지역에서 분석반경에 따른 지형패턴의 변화를 살펴보기 위해 3종류의 DEM을 대상으로 분석반경 값을 5, 10, 15, 20, 25, 30, 35, 40셀 총 8종류로 조절하여 분석을 수행하였다. 분석반경의 변화에 따라 지형패턴의 분포비율은 급격하게 변화하는 양상을 보이지는 않았으나, 대체로 10셀과 30셀을 경계로 지형패턴의 분포비율이 변화하는 것을 관찰할 수 있다. 이는 기존 연구결과에서 제시한 분석반경의 한계와 비슷한 결과다.

평탄임계치는 지형패턴 중 평탄면의 분포에 가장 큰 영향을 미치는 변수이다. 기존 연구자들은 기본값(1°)으로 설정하였으나, 본 연구는 한반도를 대상으로 분석을 수행한 연구결과를 반영하여 기본값과 더불어 한반도를 대상으로 수행하여 얻은 값 5°를 모두 이용하여 차이를 살펴보았다. 연구지역은 개간이 많이 이루어진 지역으로 저수지, 논, 그리고 밭이 많은 곳이다. 기본값을 이용하여 분석한 결과 지형패턴 분석이 이루어지는 것을 관찰하였고, 5°를 이용하였을 경우 이 지역들이 평탄하게 나오는 것을 관찰할 수 있었다. 따라서, 평탄임계치를 기본값으로 설정하지 않고 변수를 조정해야 함을 알 수 있다.

시각관독법에 따라 분류된 지형과 지형패턴을 이용하여 분류된 지형을 비교한 결과 계곡, 능선, 정상부 같은 지형은 대체로 일치하는 것을 확인하였으나, 지형패턴 분류법은 단순히 볼록사면, 사면, 오목사면과 같은 지형요소들로 분류하였기 때문에 시각관독으로 분류된 지형들(선상지, 범람원 등)을 직접 비교하는데 한계가 있다. 이를 비교하기 위해서는 지형패턴들을 구성요소로 조합하여 지형분류를 수행하는 작업이 추가로 필요하다. 또한, 시각관독법과 교차 검증하는 작업을 반드시 거쳐야 한다.

지형패턴 분류법을 이용하여 지형분석을 수행할 경우 분석의 정확도를 높이기 위해 입력자료인 DEM 해상도가 우선으로 중요하다. DEM 해상도는 분석 대상 지역의 규모에 따라 연구자가 적절하게 설정해야

한다. 본 고에서 연구지역의 공간 규모는 약 4×8km의 비교적 좁은 공간이며, 개발이 많이 이루어진 곳이기 때문에 5.0m 해상도의 DEM을 이용할 경우 정확한 지형분류가 힘들다. 따라서, 작은 규모의 지형분류를 수행할 경우 정밀한 DEM을 이용해야 정확한 지형분류를 할 수 있다. 변수설정의 경우에도 10~30셀 사이에서 분석반경을 설정해야 지형패턴의 분류를 정확하게 할 수 있다. 평탄입계치 역시 연구지역과 같이 저수지, 논, 그리고 밭이 많은 곳에서는 기본값인 1°를 사용하는 것이 아니라 적절한 값을 설정해야 정확한 지형패턴 분류를 할 수 있다. 최종적으로, 활성단층 조사가 있어 지형패턴 분류와 같은 자동분류 방식은 연구지역의 공간 규모에 따른 적절한 변수 값을 설정해야 하고, 현장을 확인하여 검증은 반드시 수행해야 하는 단점이 있으나 직관적이고 쉬운 변수설정, 빠른 처리 시간과 같은 장점이 있기 때문에 확인된 지형단위의 상호관계를 살펴볼 수 있을 정도로 자료가 축적된다면, 장기적인 관점에서 지형분류에 기준을 제시할 수 있을 것으로 기대한다. 적절한 변수설정과 관련하여 향후 통계적 검증과 같은 방법을 이용하여 후속 연구를 이어가고자 한다.

사사

본 연구는 행정안전부 한반도 단층구조선의 조사 및 평가기술 개발 사업의 연구지원(2017-MOIS31-006)에 의해 수행되었습니다.

참고문헌

국가활성단층연구단, 2019, 한반도 단층구조선의 조사 및 평가기술 개발 표준 절차 및 지침.
 김남신·이민부·신근하, 2003, "GIS를 이용한 거창, 가조 분지의 선사유적 입지 지형요소별 유형화," 한국지형학회지, 10(2), 267-281.
 김남신·조용찬·오승환·권혜진, 2014, "지형요소 지도화

를 위한 기호제작에 관한 연구," 한국지역지리학회지, 20(3), 245-255.
 김동은·성영배·손학기·최광희, 2012, "지형패턴(Geomorphons)을 이용한 새로운 지형분류 방법," 한국지형학회지, 19(4), 139-155.
 김태호, 2003, "제주도 해안지대의 지형분류," 한국지형학회지, 10(1), 33-47.
 류충렬·강지훈·강희철, 2018, "양산단층 북부 일대의 최후기 지구조환경에 대해," 암석학회지, 27(3), 173-184.
 박수진, 2004, "생태환경 특성 파악을 위한 지형분류기법의 개발," 대한지리학회지, 39(4), 495-513.
 송영석·하상민·이성준·서경환·강경걸·강희철·성영배·김영석·손문, 2018, "굴착 및 지표조사로 새롭게 발견된 중부 양산단층의 제4기 운동특성," 대한지질학회 학술대회, 129-129.
 오정식, 2019, "가음단층계의 선형구조 추출과 선형구조와 단층활동의 관련성," 한국지형학회지, 26(2), 69-84.
 이민부, 2010, "위성영상을 이용한 추가령열곡 DMZ 지역의 지형면 분석," 한국지형학회지, 17(1), 1-14.
 탁한명, 2014a, "GIS를 이용한 한반도 산지의 지형학적 분류에 관한 연구," 부산대학교 대학원 박사학위논문.
 탁한명, 2014b, "한반도의 지형패턴 분류를 위한 최적변수 설정에 관한 연구," 한국지형학회지, 21(1), 165-183.
 Choi, J. H., Kim, Y. S., and Klinger, Y., 2017, Recent progress in studies on the characteristics of surface rupture associated with large earthquakes, *Journal of the Geological Society of Korea*, 53(1), 129-157 (in Korean).
 Choi, S.-J., Jeon, J.-S., Song, K.-Y., Kim, H.-C., Kim, Y.-H., Choi, P.-Y., Chwae, U.C., Han, J.-G., Ryoo, C.-R., Sun, C.-G., Jeon, M.S., Kim, G.-Y., Kim, Y.-B., Lee, H.-J., Shin, J.S., Lee, Y.-S., and Kee, W.-S., 2012, *Active faults and seismic hazard map*. National Emergency Management Agency, Seoul.
 Fisher, P., Wood, J., and Cheng, T., 2004, Where is Helvellyn? Fuzziness of multi-scale landscape morphometry, *Transactions of the Institute of British Geographers*, 29(1), 106-128.

- Griffiths, J. S., 2004, Geomorphological mapping, in Goudie, A. S. (ed), *Encyclopedia of Geomorphology*, Routledge.
- Im, C. B., Noh, M., Shim, T. M., Lee, H., Choi, H. S., and Kim, H. J., 2004, Capable fault criteria applied to nuclear power plant sites, and their technical background, *Journal of the Geological Society of Korea*, 40(2), 279-284 (in Korean).
- Jasiewicz, J. and Stepinski, T. F., 2013, Geomorphons - a pattern recognition approach to classification and mapping of landforms, *Geomorphology*, 182, 147-156.
- Kee, W.-S., Kim, Y.-H., Lee, H.-J., Choi, D.-L., Kim, B.-C., Song, K.-Y., Koh, H.-J., Lee, S.R., Gwang, Y.Y., Hwang, S.-H., Park, K.-G., and Sung, N.-H., 2009, *South eastern fault variable research and DB Construction*, KIGAM, Daejeon.
- Kim, Y. S., Jin, K., Choi, W. H., & Kee, W. S., 2011, Understanding of active faults: A review for recent researches. *Journal of the geological society of Korea*, 47(6), 723-752.
- Liao, W. H., 2010, Region Description Using Extended Local Ternary Patterns, 2010 20th International Conference on Pattern Recognition.
- Ryoo, C.-R., Lee, B.-J., Cho, D.-L., Chwae, U.-C, Choi, S.-J. and Kim J.-Y., 1999, Quaternary fault of Dangu-ri, Gangdong-myeon, Gyeongju: Byeokgye fault. The Korean Society of Economic and Environmental Geology ·The Korean Society of Mineral and Energy Resources Engineers·Korean Society of Earth and Exploration Geophysicists, Spring Joint Conference on East-North Asia Ore Deposit of Cause of Formation, (abstract), 334.
- Stepinski, T. F., Jasiewicz, J., 2011, Geomorphons-a new approach to classification of landforms. *Proceedings of Geomorphometry, 2011*, 109-112.
- Wood, J., 2002, LandSerf: visualisation and analysis of terrain models, <http://www.landserf.org> (2002)
- 교신: 오정식, 02841, 서울시 성북구 안암로 145 고려대학교 지리교육과(이메일: ojs0201@korea.ac.kr, 전화: 02-3290-2360)
- Correspondence: Jeong-Sik Oh, Research Professor, Department of Geography Education, Korea University, Anam-ro 145, Seongbuk-gu, 02841, Korea (e-mail: ojs0201@korea.ac.kr, phone: 82-2-3290-2360)
- 최초투고일 2019. 10. 11
수정일 2019. 10. 26
최종접수일 2019. 10. 31

LiDAR, 수치지형도, 항공사진 기반 수치표고모델을 활용한 중부 양산단층의 선형구조 추출과 비교

오정식* · 김동은**

Lineament Extraction and Its Comparison Using DEMs based on LiDAR, Digital Topographic Map, and Aerial Photo in the Central Segment of Yangsan Fault

Jeong-Sik Oh* · Dong Eun Kim**

요약 : 중부 양산단층에서 보고된 벽계단층과 단구단층 일대 약 27km² 구역을 대상으로 LiDAR, 수치지형도, 항공사진 기반의 DEM을 구축하고, DEM 유형에 따른 선형구조 인지 특징과 선형구조 추출 결과를 비교했다. 구축된 DEM은 유형별로 다른 시기의 지표를 구현하며, 각각의 해상도는 L-DEM(LiDAR 기반 DEM)이 0.50m, T-DEM(수치지형도 기반 DEM)이 5.00m, 그리고 A-DEM(항공사진 기반 DEM)이 0.46m이다. PCI Geomatica 소프트웨어의 라인 모듈을 활용하여 선형구조를 자동추출한 결과 모든 유형의 DEM에서 산지를 중심으로 높은 선형구조 분포밀도를 보였지만, 분절 길이와 주향, 그리고 추출된 선형구조의 개수는 DEM 유형에 따라 현저한 차이를 나타냈다. 시각적 관독을 통해 추출된 선형구조는 DEM 유형에 따른 추출 개수의 차이가 크지 않았다. 특히, 단층와지, 단층구릉, 삼각말단면과 같이 주변 지형과의 기복 차이가 뚜렷하거나, 수평 변위를 동반하는 굴절하천으로 인지된 선형구조는 DEM 유형과 관계없이 추출 결과가 유사했다. 이러한 결과는 시각적 관독이 단층활동과 직접 관련성이 높은 선형구조만을 감독 분류하여 추출하기 때문이다. 선형구조 추출 결과와 야외조사를 바탕으로 단층활동과 연관성이 높은 21개의 선형구조가 분리 추출되었으며, 이들 중에서 3개는 단층애와 굴절하천과 같은 제4기 지형 변위가 인지된 활성도 높은 선형구조로 평가됐다. DEM 유형별 선형구조 추출 결과를 종합할 때, T-DEM은 선형구조의 거시적 특성을 파악하는데 유효하며, 1m 이하의 고해상도로 구축 가능한 L-DEM과 A-DEM은 제4기 지형의 변위를 수반한 선형구조를 추적할 때 유효함을 알 수 있었다. 특히, 활성도 높은 선형구조를 추적하고 선형구조 간 관계와 정확한 분절을 파악하는데 L-DEM의 활용성이 크며, 인위적 영향으로 지형의 변화가 심한 지역은 원지형에 가까운 지표를 구현하는 A-DEM이 효과적인 보완자료로 활용될 수 있다.

주요어 : 선형구조, 선형구조 추출, 수치표고모델, 양산단층, 활성단층

Abstract : LiDAR, digital topographic map, aerial photo-based DEMs were generated for the area of 27km² including Byuckgae and Dangu fault site reported from the central segment of Yangsan fault. We compared the characteristics of lineament detection and the results of lineament extraction based on type of DEMs. The generated DEMs are subjected to different time and have different resolutions of 0.50m for L-DEM

* 고려대학교 지리교육과 연구교수(Research Professor, Department of Geography Education, Korea University, ojs0201@korea.ac.kr)

** 고려대학교 미래국토연구소 전임연구원(Full-time Researcher, Research Institute of Future Land, Korea University, roe-hddms@korea.ac.kr)

(LiDAR-based DEM), and 5.00m for T-DEM (Digital topographic map-based DEM), 0.46m for A-DEM (Aerial photo-based DEM). Automatic extraction of lineaments using the 'Line Module' of the PCI Geomatica software resulted in high density of lineament distributions in the mountainous regions, showing similar results in the all types of DEM. However, segment length, strikes (directions), and the number of extracted lineaments differed significantly depending on the type of DEMs. The lineaments extracted by visual detecting showed similar numbers of extractions in all types of DEM. In particular, the lineaments recognized as fault-related landforms with distinct relief and horizontal displacement such as fault saddles, kernbutts, triangular facets, and deflected streams, had similar extraction results in all types of DEM. These results are because visual detection only classifies the lineament related to fault activation. We extracted 21 lineaments related to fault activation based on the results of lineament extraction and field survey. Three of them are identified as a highly active lineament, and displacements of Quaternary landforms, such as fault scarps or deflected streams are identified. We believe that T-DEM is effective for understanding the macroscopic characteristics of lineaments. L-DEM and A-DEM, which can be generated with high resolutions of less than 1m, are more effective when tracking highly active lineaments with displacement of Quaternary landforms. In particular, L-DEM is useful for tracing active lineament and identifying precise segmentation pattern and the relationship between the lineaments. In areas where the landforms have changed due to anthropogenic effects, A-DEM representing initial landforms can be practically used as a supplementary data.

Key Words : lineament, lineament extraction, digital elevation model, Yangsan fault, active fault

1. 서론

지난 2016년과 2017년 경주(M5.8; 기상청)와 포항(M5.4; 기상청)에서 잇달아 발생한 지진으로 한반도의 활성단층 존재와 지진 방재에 관한 관심이 크게 고조되었다. 그간 한반도는 지진 발생이 빈번한 판의 경계에서 떨어진 판의 내부에 위치하는 지리적 특성상 지진으로부터 비교적 안전하다는 인식이 있었다. 그러나 두 차례 지진을 계기로 한반도도 더는 지진으로부터 안전지대가 아니라는 사회적 인식이 커졌다. 경주와 포항 지진이 우리 사회에 더 큰 충격으로 다가왔던 이유는 지진의 규모가 컸기 때문이기도 하지만, 진앙과 인접한 곳에 대도시 인구 밀집 지역과 원자력 발전소를 비롯한 주요 국가기간시설이 자리하고 있어 하마터면 대규모 재난이 될 수도 있었기 때문이다.

지진은 지각에 작용하는 응력이 축적되었다가 해소되는 현상으로 판의 경계뿐만 아니라 판의 내부에서도 응력이 가해지는 지구조환경이면 어디서든 발생할 수 있다(Keller and Pinter, 2002; 김영석 등, 2011). 언제 어디서 발생할지 모르는 대규모 지진재해에 더욱 능동적으로 대처하기 위해서는 지진 유발

가능성이 있는 단층의 정확한 위치와 특성을 파악하는 것이 중요하다. 일반적으로 대규모 지진재해를 유발할 가능성이 큰 육상지진은 활성단층(active fault)에서 발생하며, 활성단층에 의해 변형된 지형과 지표지질을 분석하여 고지진의 위치, 시기, 규모 등을 추론하고 지진의 재발특성에 관한 정보를 얻을 수 있다(김영석 등, 2011; 최진혁 등, 2017). 활성단층에 관한 정의는 다양하지만, 일반적으로 현생의 지구조와 지진원 환경에서 단층운동을 겪었고, 가까운 미래에 지진을 유발할 가능성이 큰 단층을 의미한다(임창복 등, 2004; 김영석 등, 2011; 최진혁 등, 2017).

판의 경계나 지구조운동이 활발한 지역은 지표과열과 같은 지진의 흔적이 잘 남아있어 활성단층 연구에 많은 정보를 얻을 수 있다(최진혁 등, 2017). 하지만 한반도와 같이 규모(>M6.0)가 큰 지진의 발생 빈도가 낮고, 지표의 풍화·침식이 활발한 환경에서는 지표에 남겨진 활성단층의 흔적을 찾고 연구함에 많은 어려움이 따른다(오정식, 2019a). 과거 단층활동의 흔적을 추적하고, 활성단층 여부를 객관적으로 판단하기 위해서는 물리탐사나 굴착조사를 이용한 고지진학적 연구(paleoseismology)가 활용되며, 이러한 연구에서 가장 먼저 수행되는 것이 선형구조(lin-

ement) 분석이다(이광률 등, 2018; 오정식, 2019a). 여기서 선형구조(또는 선구조선)란, 단층, 단층 파쇄대, 암석의 경계와 같은 지구조적 특성과 하천, 바다와 같은 지표프로세스의 침식 및 퇴적 작용으로 인하여 지표상에 발달한 선형의 지형을 의미한다(O'Leary *et al.*, 1976; 국가활성단층연구단, 2019).

선형구조 분석은 주로 구조지형학과 구조지질학 분야에서 지구조운동이 지형발달에 미친 영향이나 국지적 혹은 광역적 규모의 지질구조선 발달 특성을 밝히는 목적으로 진행되었으며(오경섭·양재혁, 2007; 민석규·기근도, 2011; 박충선·이광률, 2018a; 이광률 등, 2018; 이광률 등, 2019; 오정식, 2019a), 선형구조의 자동추출 기법과 관련해서는 원격탐사와 지리정보체계 분야에서도 활발하게 다루고 있다(김상완 등, 1999; 이민부·김남신, 2002; 박서우 등, 2018). 분석 방법은 전통적으로 종이지형도와 항공사진을 활용한 아날로그 분석 기법을 사용했지만, 수치지형도와 고해상도 항공·위성영상이 보편화 되면서 GIS와 같은 소프트웨어를 이용한 디지털 분석이 아날로그 분석을 대체하고 있다. 가장 최근에는 지표를 수십 cm 수준의 초고해상도 수치표고모델(Digital Elevation Model, 이하 DEM)로 구현할 수 있는 LiDAR(Light Detection And Ranging)와 UAV(Unmanned Aerial Vehicle) 촬영 영상의 활용도 증가 추세이다.

선형구조의 분석방법은 크게 지형요소를 고려한 시각적 판독과 알고리즘을 이용한 자동추출로 구분된다(김상완 등, 1999; 이민부·김남신, 2002; 박서우 등, 2018). 이들 두 가지 방법 모두 DEM 또는 DEM으로부터 작성된 음영기복도를 이용하며, 각각의 장단점이 공존한다. 분석 결과의 객관성 측면에서는 연구자의 주관적 판단이 개입될 여지가 큰 시각적 판독보다 자동추출이 더 객관적 분석 결과를 제시한다. 그러나 자동추출은 인위적 지형(도로, 제방, 사면절개지 등)과 단층 관련성이 낮은 선형구조까지 추출되는데 단점이 있고, 시각적 판독은 단층활동과 직접 관련성이 높은 선형구조를 연구자가 감독 분류할 수 있다는 장점이 있다(이민부·김남신, 2002; 오정식, 2019a).

선형구조 분석에 있어서 DEM은 가장 보편적으로 활용되는 자료이다. 그러나 DEM의 유형과 특성이 선

형구조 추출 결과에 어떻게 반영되고 어떠한 영향을 미치는지에 관한 논의와 더불어 활성단층 조사에 있어서 신뢰도 높은 선형구조 추출을 위한 DEM의 효과적 활용방안에 관한 논의도 부족한 실정이다.

본 연구는 DEM 유형에 따른 선형구조 추출 결과를 비교하고, 유형별 DEM의 선형구조 분석 활용성에 관하여 논하고자 했다. 이를 위해 중부 양산단층에서 보고된 벽계단층과 단구단층 일대 약 27km² 구역을 대상으로 LiDAR, 수치지형도, 항공사진 기반의 DEM을 구축하고, DEM 유형에 따른 선형구조 인지 특성과 선형구조 추출 결과를 비교·검토하였다. 본 논의는 활성단층 조사·연구의 가장 첫 단계인 선형구조 분석에 있어서 신뢰도 있는 선형구조 추출을 위한 DEM의 유형별 활용성에 관하여 시사점을 갖는다.

2. 연구지역

연구지역은 행정구역상 경상북도 포항시 신광면과 경주시 강동면 일대로 경주 지진 진앙으로부터 북쪽으로 약 34km, 포항 지진 진앙으로부터 서쪽으로 약 8km 떨어진 지역이다(Figure 1-a). 연구지역 북부는 비학산(762m m.s.l)을 주봉으로 하는 해발 300m 내외의 산지로 둘러싸인 폐쇄형 분지(신광분지)를 이루고 있다. 분지 내부는 넓은 선상지와 범람원이 분포하며, 곡강천과 신광천이 각각 북쪽과 남쪽 수계를 이룬다. 이들 하천은 분지의 동쪽 가장자리에서 합류한 뒤 좁은 협곡을 통해 홍해 평야로 흘러나간다. 연구지역 남부는 도음산(383m m.s.l) 서사면의 계곡을 따라 좁고 긴 형태의 선상지가 분포하며, 선상지 말단은 형산강의 지류인 기계천에 의해 발달한 범람원과 점이적으로 만난다(Figure 1-b). 신광분지와 도음산 서사면에 발달한 선상지는 선정과 선양을 중심으로 20m 내외의 깊은 침식곡이 발달했으며, 하천 양안에는 2~3단의 선상지성 하안단구가 분포한다.

이 일대는 경상분지의 백악기 퇴적암류를 기반으로 이후의 백악기 화산암류와 고제3기 화강암이 관입하고 있으며, 포항분지에 널리 분포하는 신생대 마이

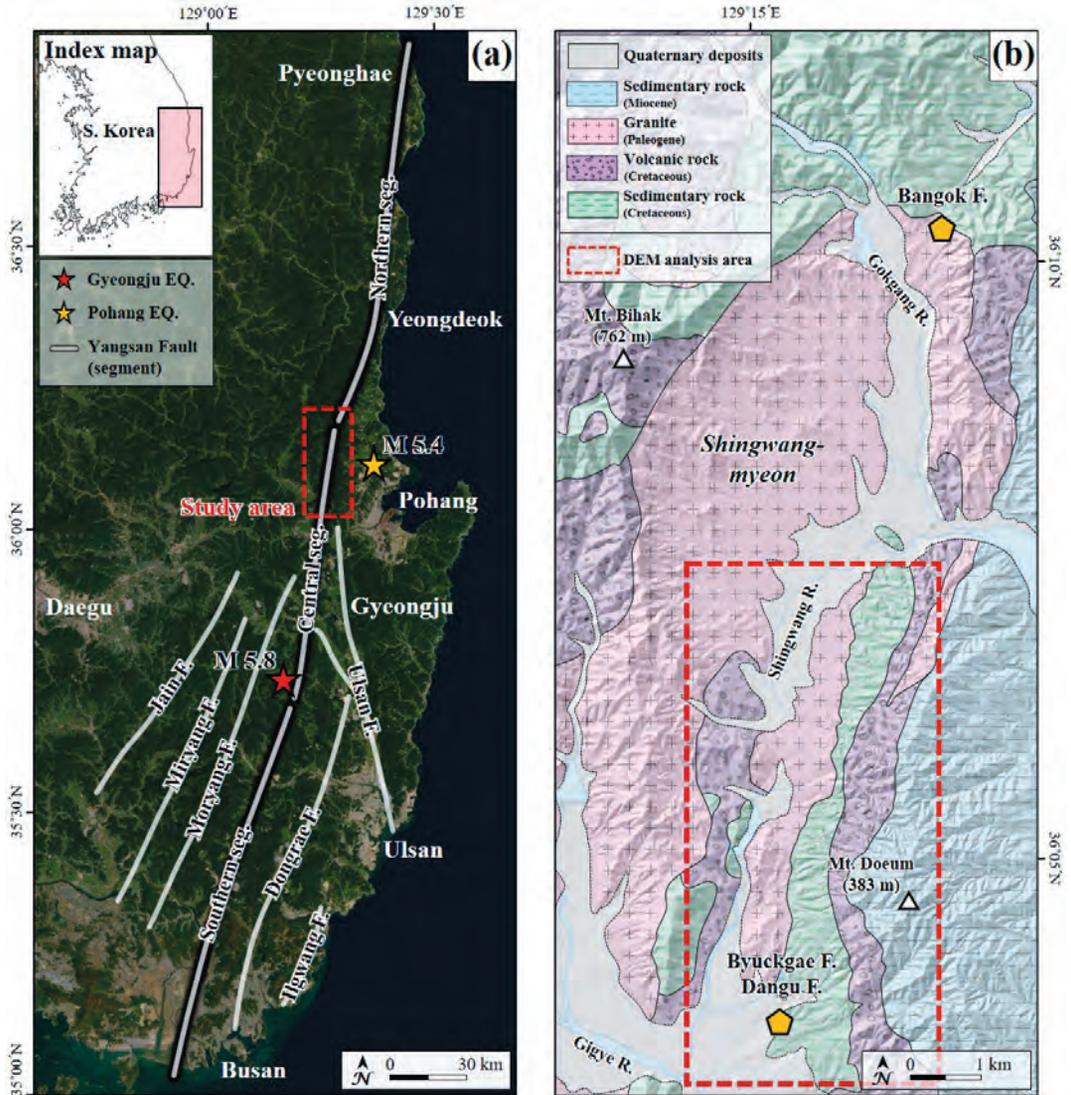


Figure 1. Topography and geological settings in the study area.

Source: 최진혁 등, 2017; KIGAM, <https://mgeo.kigam.re.kr>, Modified

오세 퇴적암이 도음산 일대에 분포한다(Figure 1-b). 낙동강 하구로부터 영덕군 영해면(또는 울진군 기성면)까지 한반도 남동부 지역을 북북동-남남서 방향으로 가로지르는 대규모 주향이동단층인 양산단층이 연구지역을 통과하며, 북북동과 북서 주향을 보이는 여러 조의 지질구조선이 알려져 있으나 주단층대의 정확한 위치는 명확히 밝혀진 바 없다(이진현 등, 2015; 최진혁 등, 2017; 류충렬 등, 2018). 지질구조

선과 더불어 이 지역의 산릉과 수계(선형곡)는 남북 혹은 북북동 방향의 발달 양상을 보이는데, 이는 제4기 이후 지형발달이 연구지역을 통과하는 양산단층과 여러 부수단층에 수반된 구조운동과 밀접하게 관련되어 있음을 간접적으로 지시한다.

양산단층은 주단층과 사교하거나 평행하게 발달한 수많은 부수단층으로 이루어진 단층대(fault zone)이며(최진혁 등, 2009; 류충렬 등, 2018; Figure 1-a),

지진학적 특성과 단층의 기하 및 특징을 기준으로 적게는 3개에서 많게는 9개의 분절로 구분된다(Lee and Jin, 1991; 장천중·장태우, 2009; 최진혁 등, 2017). 양산단층을 3개의 분절로 이해할 때, 연구지역 일대는 양산단층 중부 분절의 북쪽 끝에 해당한다.

양산단층과 관련하여 연구지역의 남쪽에는 벽계단층과 단구단층, 그리고 북쪽에는 반곡단층이 각각 보고되었다(Ryoo *et al.*, 1999; 소방방재청, 2012; 이진현 등, 2015; 류충렬 등, 2018). 기 보고된 단층의 기하와 운동특성에 따르면 벽계단층은 N12°E 주향의 역이동성 주향이동 운동감각을 보이며, 이로부터 약 50~60m 북쪽으로 떨어진 곳의 단구단층은 벽계단층과 유사한 N10~20°E 주향의 역이동성 우수향 주향이동 운동감각을 보인다. 아울러 신광분지의 북쪽 산록에서 보고된 반곡단층은 N16°W 내지 N36°W 주향의 역이동성 좌수향 주향이동단층으로 앞선 2개의 단층과 기하와 운동특성이 차이를 보인다.

운동특성과 변형사에 관한 그간의 연구결과를 통해 양산단층은 에오세 이후 주로 우수향 운동을 경험했으며, 좌수향 및 경사이동 운동도 함께 경험한 것으로 해석된다(채병근·장태우, 1994; 황병훈 등, 2004). 그리고 단층대를 따라 확인된 화강암류(고제3기 A-type 화강암)의 암상과 분포 특징을 통해 양산단층 중부 분절은 약 20~35km의 우수향 수평 변위를 지시하는 것으로 알려져 있다(황병훈 등, 2004; 최진혁 등, 2017). 연구지역의 단층 특성과 양산단층의 운동특성을 종합적으로 판단할 때, 벽계단층과 단구단층은 양산단층의 주단층대와 높은 관련성을 지니며, 반곡단층은 부수단층 혹은 또 다른 단층 분절과 관계된 것으로 볼 수 있다.

3. 연구방법

1) 수치표고모델 구축

연구지역의 DEM은 해상도 차이와 토지이용의 시계열 차이가 가장 뚜렷하게 구분되는 세 가지 유형을

구축·사용했다. 현재의 지표를 가장 정밀하게 표현할 수 있는 것은 LiDAR로부터 획득한 DEM(이하, L-DEM)이며, ‘국가 활성단층 연구단’에서 구축(2018년 촬영)한 항공 LiDAR 기반의 0.5m 해상도 L-DEM을 분석에 활용했다(국가활성단층연구단, 2019).

LiDAR 자료는 고해상도(수 cm ~ 수십 cm)의 DEM 구축이 가능하지만 많은 시간과 비용이 든다는 단점으로 접근이 제한적이다. 따라서 선형구조 분석에 전통적이면서도 가장 보편적으로 활용되는 자료는 수치지형도(Digital Topographic Map)를 기반으로 구축한 DEM(이하, T-DEM)이다. T-DEM은 LiDAR보다 지표를 구현하는 해상도가 낮지만, DEM 구축에 필요한 시간과 비용을 절약할 수 있으며, 우리 국토의 대부분 지역에서 DEM 구축에 필요한 자료 획득이 편리하다(Table 1). 본 연구는 국토지리정보원에서 제공하는 1:5,000 수치지형도(2017년 제작)를 기반으로 ArcGIS를 이용하여 T-DEM을 구축하고 분석에 활용했다.

항공사진은 특정 지역의 지표 변화를 시간의 흐름에 따라 추적할 수 있으며, 특히 우리나라와 같이 개발로 인해 단기간에 급격한 지형 변화가 발생한 곳에서는 과거에 촬영된 항공사진을 바탕으로 원지형에 가까운 지표의 공간정보를 획득할 수 있다는 장점을 지닌다(Table 1). 선형구조 분석에 있어서 항공사진은 실체경(stereoscope)을 이용한 육안 분석이 많이 활용되었다(이광률 등, 2018; 오정식, 2019a). 그러나 최근 항공사진 영상분석에 활용되는 소프트웨어 기술이 빠르게 발달하고 보급됨에 따라, 디지털화된 과거항공사진 혹은 드론을 이용한 항공사진측량(Photogrammetry Processing)이 새로운 지형분석 기법으로 주목받고 있다(최현·홍순현, 2007; 박충선·이광률, 2018b; 오정식, 2019b).

연구지역은 1970년대 이후 하천 및 수로정비, 농경지 개간, 도로개설, 그리고 마을개량 사업 등으로 지형이 급격하게 변화했다. 따라서 변화 이전의 지형이 촬영된 과거항공사진 중에서 이미지 품질(해상도, 기상, 촬영시각에 따른 음영 등을 고려)이 가장 우수한 1967년 촬영 영상을 과거항공사진 기반 DEM(이하, A-DEM) 구축에 사용했다(Table 2). A-DEM 구

Table 1. Advantages and disadvantages of using DEMs for extraction lineaments

| Type of DEM | Advantage | Disadvantage |
|---------------------------|--|---|
| L-DEM (DTM ^a) | <ul style="list-style-type: none"> · High resolution (>1 m) with high accuracy · Generation of a real topographic surface by removing man-made structures · Capable of collecting elevation data in a dense forest | <ul style="list-style-type: none"> · High operating costs · Requires long time for post-processing · Requires skilled data analysis techniques · Very large datasets |
| T-DEM (DTM) | <ul style="list-style-type: none"> · Low resolution (<5 m) · Easy to collect source data · Data processing (DEM generation) is quick and easy | <ul style="list-style-type: none"> · Can not detect fault-related landforms with small displacements (<5 m) · Surface distortion of low-relief landforms such as floodplain and alluvial fan |
| A-DEM (DSM ^b) | <ul style="list-style-type: none"> · High resolution (>1 m) · Easy to collect source data · Possible to reconstruct landforms prior to anthropogenic effects | <ul style="list-style-type: none"> · Irregular photograph timing · Image noise of the source data (printing or scanning image) affects quality of A-DEM · Requires complex accuracy verification for photogrammetry processing |

^aDigital Terrain Model

^bDigital Surface Model

Table 2. Details of aerial photo and ground control point

| Aerial photo ^a | | | | GCPs | | |
|---------------------------|--------|-----------------------|---------------------------|----------|-----------|-------------------|
| Code ^b | Scale | Resolution (cm/pixel) | Origin of coordinate | DD (°N) | DD (°E) | Altitude (m.s.l.) |
| 196700A0009 5-615 | 37,500 | 52.91667 | 36.05276°N 129.23879°E | 36.04869 | 129.24553 | 21.4 |
| | | | | 36.03980 | 129.25228 | 21.1 |
| | | | | 36.05843 | 129.26230 | 81.9 |
| 196700A0009 5-616 | 37,500 | 52.91667 | 36.08598°N 129.23955°E | 36.06285 | 129.24691 | 42.2 |
| | | | | 36.07029 | 129.26098 | 115.4 |
| | | | | 36.05981 | 129.25404 | 47.3 |
| | | | | 36.07592 | 128.24843 | 52.9 |
| 196700A0009 5-617 | 37,500 | 52.91667 | 36.11921°N 129.24030°E | 36.11770 | 129.25737 | 76.0 |
| | | | | 36.12678 | 129.26547 | 69.5 |
| | | | | 36.13481 | 129.24543 | 135.8 |
| | | | | 36.08379 | 129.24792 | 79.3 |
| | | | | 36.10651 | 129.25463 | 92.1 |
| | | | | 36.12857 | 129.27950 | 61.3 |

^aMeta data of aerial photo is provided by the National Geographic Information Institute (<http://map.ngii.go.kr>).

^bIdentification code of aerial photo used for DEM generation.

측은 Agisoft의 PhotoScan Pro(version 1.3)를 이용했으며, 항공사진 정사영상 및 DEM 제작과 관련한 일련의 과정은 소프트웨어 개발사인 Agisoft에서 제공하는 지침을 따랐다(<https://www.agisoft.com>). 항공사진을 활용한 DEM 구축에 있어서 지상기준점

(Ground Control Point; 이하, GCP)은 영상보정과 정사영상 제작을 포함한 정밀한 DEM 구축에 중요한 역할을 한다(유재진 등, 2016; 오정식, 2019b). 연구지역의 A-DEM 구축에 활용한 GCP는 1967년 항공사진과 가장 최근에 촬영된 항공사진의 비교를 통

해 지형의 변화가 없는 총 13개 지점으로 선정했으며, GCP 좌표와 해발고도는 정밀측량 및 오차 검증이 완료된 LiDAR 자료로부터 획득하였다(Table 2).

2) 선형구조 추출

선형구조 추출은 시각적 판독과 알고리즘을 활용한 자동추출 방법을 사용했다. 시각적 판독은 3가지 유형의 DEM으로부터 각기 제작된 음영기복도를 기본도로 하고 1:50,000 수치지질도와 위성영상을 중첩 분석했다. 음영기복도는 태양광 조사 각도에 따라 인지되는 선형구조가 다를 수 있다는 점을 고려하여 총 4개의 방위각(북동, 북서, 남동, 남서)에서 조사된 음영기복도를 모두 분석했다(김원균 등, 1996; 이민부·김남신, 2002). 또한, 축척에 따라 인지되는 선형구조의 길이와 주향도 다르므로 소축척에서 대축척으로 분석 규모를 달리하며 판독을 진행했다(오정식, 2019a).

시각적 판독을 통해 추출된 선형구조는 분석에 활용한 DEM과 같은 시기에 해당하는 위성·항공사진과 비교를 통해 인공의 선형구조(도로, 하천제방, 논두렁, 사면절개지 등)를 결과에서 제외하였으며, 제4기 지형면의 변위가 의심되고 두 가지 유형 이상의 단층지형(fault-related landforms)이 선형구조를 따라 인지된 것을 단층 관련성이 높은 선형구조로 재분류했다. 단층지형의 유형과 분류 기준은 국가활성단층연구단(2019)의 표준화 조사지침과 박충선·이광률(2018a)이 제시한 분류 유형을 적용했다. 아울러 시각적 판독이 지니는 연구자의 주관 개입 문제를 최소화하고자 선형구조 추출 결과는 국가활성단층연구단 내 지형·구조지질 전문가 3인 이상의 동료평가를 토대로 교차 검증을 진행했다.

시각적 판독과 더불어 객관적 수준으로 DEM 유형별 선형구조 인지 특성을 비교하고자, 선형구조 자동추출 기법을 활용했다. 본 연구에서 선형구조 자동추출의 목적은 주관 개입이 가능한 시각적 판독 결과의 DEM 유형별 차이가 객관화된 분석 기법에서도 동일 혹은 유사하게 나타나는지 비교·검토하는 데 있다. 선형구조 자동추출은 고안된 알고리즘을 바탕으

로 위성영상이나 DEM의 래스터 정보로부터 벡터를 추출하는 방법으로 ‘호프 변환(Hough Transform; Hough, 1962)’과 ‘캐니 윤곽선 검출(Canny Edge Detector; Canny, 1986)’ 방법이 많이 활용되고 있다(이민부·김남신, 2002; Šilhavý *et al.*, 2016). 본 연구는 호프 변환 원리를 이용하는 PCI geomatica 소프트웨어의 라인 모듈을 활용하여 음영기복도에서 선형구조를 추출했다. 라인 모듈은 래스터 이미지에서 선형구조를 추출하고 6개의 매개 변수를 조정하여 선형구조를 벡터 형태로 변환하는 알고리즘을 이용한다(Abdullah *et al.*, 2010; PCI Geomatics, 2011).

6개의 매개 변수는 분석대상이 되는 DEM의 해상도 혹은 면적과 기복량에 따라 최적값이 다르다. 본 연구는 DEM 유형에 따라 선형구조의 추출 결과가 어떠한 차이를 보이는지 비교·분석하는 데 목적이 있으므로 변수의 최적값을 찾는 과정은 생략하고, 소프트웨어에서 제공하는 기본값을 사용했다. 각 매개 변수의 간략한 의미와 분석과정에서의 역할은 다음과 같다.

① ‘Filter Radius’는 윤곽선 검출 필터(edge detection filter)의 반경을 지정하는 변수이다. 입력된 이미지에서 선으로 감지할 최소 수준을 픽셀 단위로 결정하며, 설정값이 작을수록 상세한 수준의 선을 감지할 수 있다. ② ‘Edge Gradient threshold’는 앞서 언급한 필터반경에서 결정된 선형 데이터를 바탕으로 선형구조 추출에 필요한 가장자리 픽셀의 최소 수준에 관한 임계치를 지정하는 변수이다. ③ ‘Curve Length Threshold’는 하나의 선형으로 간주할 곡선의 최소길이를 지정한다. ④ ‘Line Fitting Error Threshold’는 픽셀 곡선에 선(polyline)을 맞춤(fitting) 때 허용하는 최대 오류(픽셀)를 지정한다. 이 값을 낮게 설정하면 결과는 잘 일치하지만, 선형구조의 길이는 더 짧은 분절로 표현된다. ⑤ ‘Angular Difference Threshold’는 각 분절의 최대 각도를 지정한다. 해당 변수를 지정하지 않으면 각도가 다른 두 선은 각각의 벡터로 분할된다. 해당 매개 변수의 지정 범위는 0-90도이며, 분석에는 기본값 30도를 사용했다. ⑥ ‘Linking Distance Threshold’는 두 벡터의 최소 거리(끝점 사이의 거리)를 픽셀 단위로 지정할 수

있으며, 해당 변수의 설정값에 따라 주향이 유사한 선형구조를 하나의 벡터 혹은 분절된 두 개 이상의 벡터로 표현될 수 있다.

4. 결과

1) 수치표고모델의 유형별 특징

LiDAR, 수치지형도, 과거항공사진을 활용하여 구축된 DEM은 서로 다른 지표 구현 특징을 갖는다. DEM의 지표 구현 차이는 지표의 형상을 얼마나 실제와 가깝게 구현하는가와 어떠한 시기의 지표를 구현하는가로 구분할 수 있다. 우선, 선형구조 인지에 가장 큰 영향을 줄 것으로 생각된 DEM의 최대 해상도는 L-DEM이 0,50m이며, A-DEM은 L-DEM과

거의 같은 수준인 0,46m로 구축되었다. T-DEM은 분석에 활용한 1:5,000 축척의 수치지형도가 갖는 최대 해상도를 벗어날 수 없으므로 5,00m 해상도로 구축되었다(Figure 2).

원자료와 DEM 구축과정이 최종 결과물의 품질에 미치는 영향이 적은 L-DEM 및 T-DEM과 달리, A-DEM의 경우 원자료의 특징에 따라 구축된 A-DEM의 해상도와 왜곡(noise) 정도가 다소간 차이를 보였다. 이는 과거항공사진의 촬영 시기에 따라 필름 인화 품질과 디지털화(scanning) 과정의 해상도, 촬영 증첩률 등이 다르기 때문이다(박충선·이광률, 2018b). 특히, 인화 이미지에 나타난 구름과 그림자 등의 왜곡(noise)은 A-DEM의 품질에 가장 큰 영향을 미쳤다(Figure 3). 아울러 소프트웨어를 이용한 A-DEM 구축 과정(Dense Point Cloud와 Mesh 구축 단계 등)의 설정값(parameters)도 A-DEM의 해상도와 품질에 영향을 준다(박충선·이광률, 2018b).

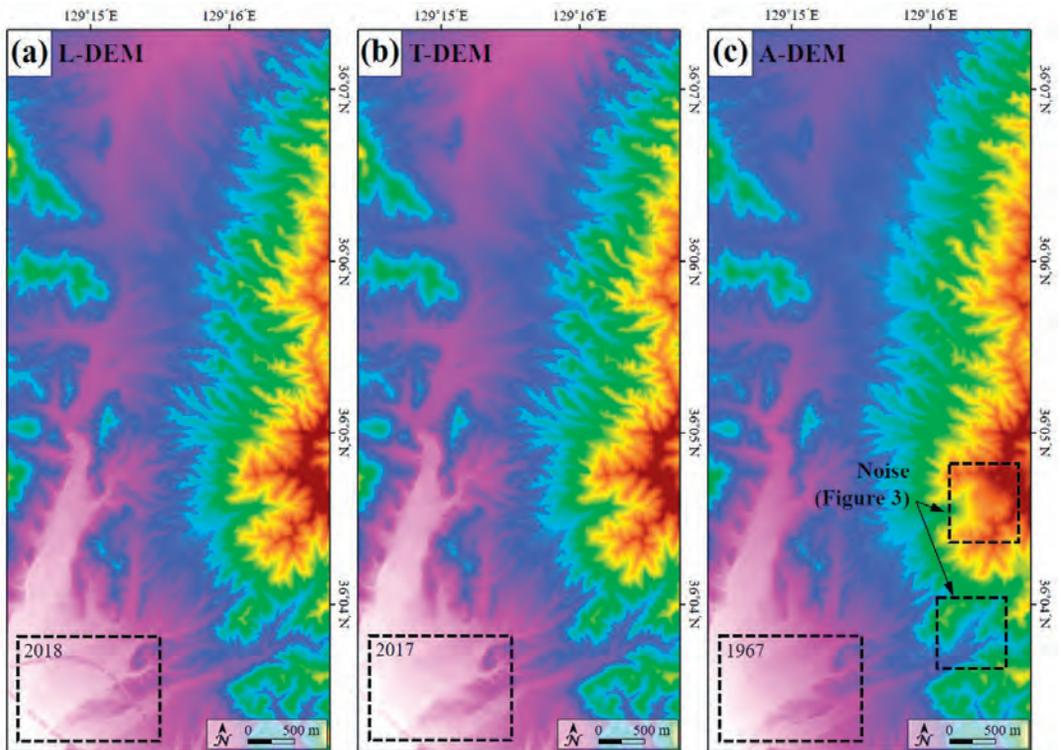


Figure 2. DEMs generated based on different source. Each figure shows the results of DEM generation based on LiDAR (a), digital topographic map (b), and aerial photo (c).

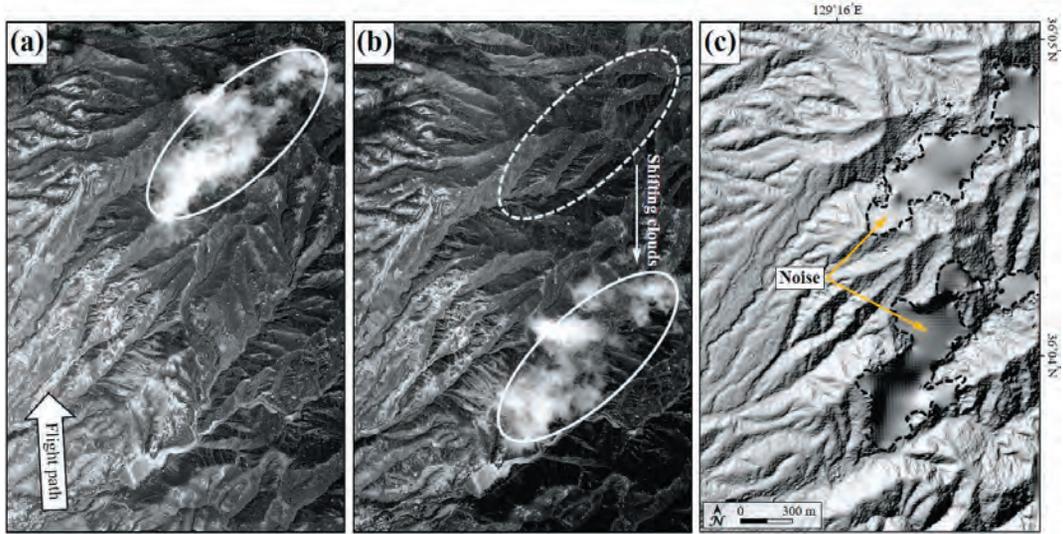


Figure 3. Image noise in aerial photo and A-DEM. Figure (a) and (b) show a cloud shifting and surface covering. Image noise occurred at two points in the A-DEM (c), which are generated by combining aerial photos taken at different times and positions.

PhotoScan Pro를 이용할 경우, 일반적으로 dense point cloud의 구축 품질을 높게 하고 심도 감식(depth filtering)을 공격적(aggressive)으로 설정하면 지표 형상을 더 정교하게 구현할 수 있지만, 이미지 왜곡이 심해진다. 또한, mesh 구축 단계에서 polygon count를 증가시킬 경우도 지표 형상을 더 정교하게 표현하지만, 이미지 왜곡도 함께 증가한다.¹⁾

구축된 DEM은 서로 다른 시기의 지표를 구현하고 있다. L-DEM은 가장 최근의 지표 형상을 가장 정밀하게 표현한다. T-DEM은 L-DEM과 유사한 시기의 자료를 바탕으로 구축되었지만, 낮은 해상도로 인하여 L-DEM이 구현한 지표의 기복이 반영되지 않은 지점(도로, 하도, 제방 등)이 다수 확인됐다(Figure 2-b). A-DEM의 경우 지형의 인위적 변형이 상대적으로 덜한 1967년의 지표를 구현했지만, 유사한 해상도의 L-DEM보다 평지(주로 농경지)에서의 왜곡이 크게 발생하여 실제 해상도에 못 미치는 수준의 지표 구현을 보였다(Figure 2-c). 각각의 DEM이 갖는 이러한 지표 구현 차이는 연구지역 남서부의 범람원과 선상지에 조성된 농경지(Figure 2의 파선으로 된 영역)에서 뚜렷하게 비교된다.

아울러, 구축된 A-DEM에서는 몇 가지 문제점이 확인되는데, 구름이나 그늘과 같이 항공사진 이미지에 나타난 왜곡이 구축된 A-DEM에 그대로 반영된다는 점이다(Figure 2-c, Figure 3-c). 항공사진 촬영 빈도(시기)가 제한적임에 따라 특정 지역을 대상으로 제공되는 과거항공사진이 한정되어 있음을 고려할 때, 이러한 문제점은 A-DEM 구축에 장애 요인으로 판단된다. 또한, 연구지역 북부에 분포하는 범람원과 선상지의 해발고도가 과대 추정되는 문제점이 확인되는데(Figure 2-c), 이는 A-DEM 구축에 활용한 항공사진의 표정(orientation)이나 GCP의 공간 분포가 고르지 못한 것에서 발생한 오류로 판단된다. 다만, 해당 지형의 상대고도 분포가 L-DEM과 비교할 때 유사한 형태를 보이며, 선형구조 분석에 있어서 지형의 절대고도는 분석을 저해하는 결정적 장애요소가 아닌 것으로 판단하여 A-DEM의 측량 정확도에 관한 논의는 제외하였다.

2) 선형구조 추출 결과

음영기복도를 이용한 시각적 관독 결과 모든 유형

의 DEM에서 남북 주향의 선형구조가 우세하게 나타났으며, 그 외 부수적으로 인지되는 선형구조는 L-DEM에서 북북동 내지 북서, T-DEM에서 북동 내지 북북서, A-DEM에서 북북동 내지 북북서 주향이 분포 빈도가 높았다(Figure 4; Table 3). 비교적 연속성이 좋은 길이 2km 이상의 선형구조는 모든 유형의 DEM에서 연구지역 중앙과 이로부터 남쪽과 북쪽에 분포하는 산록부 능선과 선상지를 가로질러 통과하며, 연구지역 동쪽의 도읍산 산지는 선형구조의 분포 빈도가 낮다(Figure 4).

추출된 선형구조의 수는 최소 21개(L-DEM)에서 최대 24개(T-DEM)로 DEM 유형별로 차이가 크지 않았지만, 길이와 분절 양상은 다소 차이를 보

였다(Table 2). 선형구조의 평균 길이는 L-DEM과 T-DEM에서 2km 내외로 다소 길게 인지되며, A-DEM은 이보다 약 31~36% 짧게 인지된다. 선형구조의 최대와 최소길이 또한 A-DEM이 짧게 인지되며, L-DEM 및 T-DEM과 비교할 때 최대길이는 18~20%, 최소길이는 54~73% 짧다. A-DEM은 1km 미만의 짧은 선형구조가 전체의 41% 비율을 차지해 L-DEM(0%)과 T-DEM(20%)보다 작은 분절의 선형구조가 우세한 것으로 확인된다. 추출된 선형구조를 중첩한 결과 총 9개 지점에서 주향이 일치($\pm 2.5^\circ$)하는 선형구조가 확인되었다(Figure 4-d). 이는 각기 다른 유형의 DEM에서 공통으로 인지되는 선형구조이며, 남북 내지 북북동 주향이 우세한 특징을 보인다.

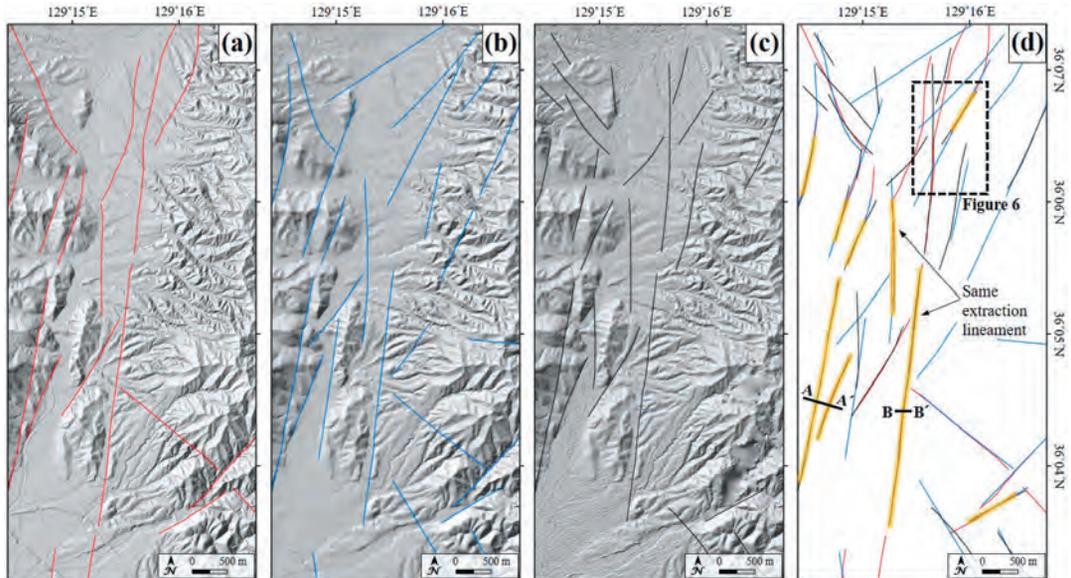


Figure 4. Results of lineament extraction based on visual detecting. Each figure shows a lineament extracted from L-DEM (a), T-DEM (b), and A-DEM (c). The result of overlaying lineaments extracted from different DEMs. A yellow solid line means lineament of the same strike (d).

Table 3. Details of lineament visual extracted from different DEMs.

| Type | Count | Length (m) | | | | | Strike | |
|-------|-------|------------|-------|-------|-------|-----|--------|---------|
| | | Total | Ave. | Max. | Min. | SD. | Major | Sub. |
| L-DEM | 22 | 45,080 | 2,049 | 4,254 | 1,111 | 868 | NS | NNE/NW |
| T-DEM | 25 | 47,624 | 1,905 | 4,318 | 881 | 906 | NS | NE/NNW |
| A-DEM | 24 | 31,863 | 1,327 | 3,494 | 414 | 799 | NS | NNE/NNW |

PCI geomatica의 라인 모듈을 활용한 선형구조의 자동추출 결과, 선형구조의 수와 길이, 그리고 분포와 분포밀도가 DEM 유형별로 차이를 보였다(Figure 5; Table 4). 특히, 자동추출된 선형구조의 수는 L-DEM에서 20,279개로 최대이었으며, A-DEM이 2,232개 그리고 이어서 T-DEM이 378개 순으로 매우 큰 차이를 보였다. 평균 길이는 L-DEM이 가장 짧은 37m이었으며, A-DEM이 112m, 그리고 T-DEM이 가장 긴 330m였다. 추출된 선형구조의 수와 길이는 반비례 관계를 보인다. 선형구조의 주향은 모든 유형의 DEM에서 북동 방향이 우세한 것으로 파악됐다.²⁾ 다만, 이들 결과는 연구방법에서 언급한 바와 같이 분석과정의 매개 변수 조정이 없었으며, 추출된 선형구조

에 대한 감도 분류(인공지물, 계곡, 능선 등 단층 관련성이 낮은 선형구조의 감도 분류)가 이루어지지 않았으므로 추출된 선형구조의 단층 관련성이 낮고, 시각적 판독 결과와 직접 비교가 어렵다.

자동추출 결과는 DEM의 해상도와 DEM 구축에 사용된 원자료의 제작 시기에 따라 선형구조 추출 결과가 다르게 나타남을 보여준다. 가장 해상도가 높은 L-DEM에서는 제방, 도로, 농로, 그리고 논두렁과 같은 인공지형이 선형구조로 추출됐다. 또한, 연구지역 동쪽의 도읍산 산지에 발달한 우곡(gully)은 높은 밀도의 선형구조로 표현됐으며, 이러한 양상은 A-DEM에서도 관찰된다(Figure 5-a). 선상지와 범람원과 같은 저기복 지형에서도 다수의 선형구조가

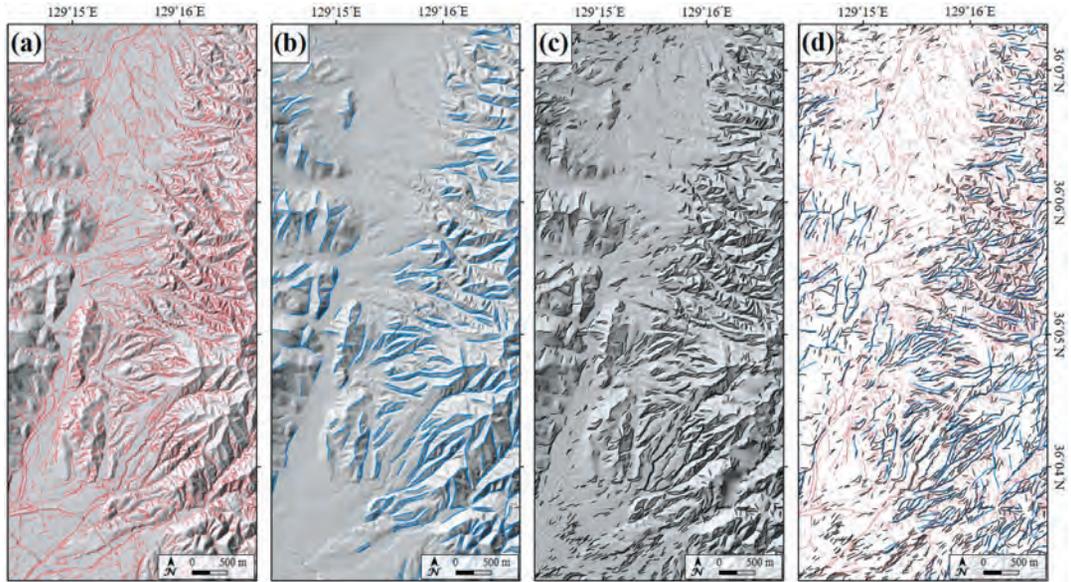


Figure 5. Results of lineament extraction using auto-extraction analysis. Each figure shows a lineament auto-extracted from L-DEM (a), T-DEM (b), and A-DEM (c). The result of overlaying lineaments auto-extracted from different DEMs (d).

Table 4. Details of lineament auto extracted from different DEMs.

| Type | Count | Length (m) | | | | | Strike | |
|-------|--------|------------|------|-------|------|-----|--------|---------|
| | | Total | Ave. | Max. | Min. | SD. | Major | Sub. |
| L-DEM | 20,279 | 759,222 | 37 | 641 | 15 | 31 | NE | NNW/ENE |
| T-DEM | 378 | 124,623 | 330 | 1,490 | 150 | 184 | NE | NNE/ENE |
| A-DEM | 2,232 | 250,433 | 112 | 689 | 56 | 63 | NE | NNE/WNW |

자동추출 되었지만, 이러한 지역의 선형구조는 대부분 하천과 농로와 같은 지형으로 확인되어 단층과의 직접 관련성은 낮다.

해상도가 가장 낮은 T-DEM은 나머지 두 유형보다 선형구조의 수(20~30% 수준)와 분포밀도가 현저히 낮게 나타났지만, 추출된 선형구조의 길이는 200% 이상 길다. 아울러 L-DEM과 달리 저기복 선상지와 범람원 지역에서 선형구조가 단 한 건도 자동추출되지 않았다(Figure 5-b). 이러한 결과는 자동추출에 적용된 매개 변수의 조정에 따라 달라질 수 있으므로 5m 해상도의 T-DEM을 활용한 저기복 지형의 선형구조 자동추출 가능 여부는 본 연구에서 판단할 수 없다.

A-DEM은 L-DEM과 유사한 해상도를 가졌으나 선형구조 자동추출 결과는 차이를 보였다. 특히, 자동추출된 선형구조의 분포밀도와 분포양상이 큰 차이를 보이는데, 이러한 차이점은 연구지역의 산지 능선과 계곡에서보다 선상지와 범람원 같은 저기복 지형에서 더 뚜렷하게 관찰된다(Figure 5-c). 도로와 제방 등 길이가 긴 인공지형에 의해 추출된 선형구조를 제외하면, A-DEM에서 자동추출된 선형구조의 평균 길이는 L-DEM과 유사한 수준(약 120%)이다. 아울러, T-DEM에서 단 한 건도 추출되지 않았던 선상지와 범람원 상의 선형구조는 A-DEM에서 다소나마 추출되었다. A-DEM의 선상지와 범람원에서 자동추출된 선형구조는 소규모 하곡(세류) 또는 선상지와 범람원이 만나는 곳에 형성된 단구애를 지시하지만(연구지역 북부), 일부 선형구조는 음영기복도에 반영된 DEM의 왜곡(연구지역 남서부)이 영향을 준 것으로 확인된다(Figure 5-c).

세 가지 유형의 DEM으로부터 자동추출된 선형구조를 모두 중첩한 결과 연구지역 동쪽 도음산 산지의 하곡을 중심으로 공통의 선형구조가 확인되며, 서쪽 산지 산록부를 따라 분포하는 삼각말단면과 선형곡(또는 단층와지)에도 곳에 따라 공통의 선형구조가 관찰됐다(Figure 5-d). 그러나 공통의 선형구조가 관찰되는 지점이라도 추출된 선형구조의 길이와 분절양상은 DEM 유형에 따라 다르다.

5. 논의

1) 선형구조 인지 특성

자동추출된 선형구조는 연구지역에 최적화된 매개 변수에 관한 검증은 생략했고 소프트웨어에서 제공하는 기본값을 사용하였으므로 추출된 선형구조의 단층 관련성을 직접 논하기는 어렵다. 다만, DEM 구축에 사용된 원자료의 시계열 차이와 구축된 DEM의 유형별 해상도 차이에 따른 선형구조 추출 결과를 시각적 관독보다 객관화된 수준에서 비교할 수 있다(이민부·김남신, 2002; Abdullah *et al.*, 2010; Šilhavý *et al.*, 2016).

자동추출된 선형구조는 DEM 유형과 관계없이 능선과 계곡을 중심으로 높은 분포밀도를 보인다. 다만, DEM의 해상도에 따라 분절의 길이가 다르고 추출되는 선형구조의 수도 차이가 있다. 이러한 차이는 0.5m 해상도의 L-DEM과 5.0m 해상도의 T-DEM에서 가장 현저하게 나타난다(Figure 5). A-DEM(0.48m)의 경우 L-DEM과 유사한 해상도지만, 음영기복도에서 발생한 왜곡(특히, 저기복 평탄지의 요철)으로 실제 지표면에 존재하지 않는 짧은 분절의 선형구조가 추출되는 오류가 있다. 이러한 추출 오류는 DEM 구축에 활용된 PhotoScan Pro 소프트웨어가 형상기준(feature based) 영상접합 방식인 SIFT(Scale-Invariant Feature Transform) 기법을 이용(박진환·이원희, 2016; 김경덕 등, 2017)하기 때문에 필름 인화로 제작된 과거항공사진 이미지에 존재하는 왜곡(얼룩, 구름, 그림자 등)이 구축된 DEM의 왜곡으로 반영되기 때문이다.

선형구조 자동추출 결과는 선상지와 범람원 같은 저기복 지형에서 가장 큰 차이를 보였다(Figure 5). L-DEM에서는 농로나 농수로 같은 미기복 인공지형(상대비고 1m 내외)도 짧은 분절로 자동추출 되었으며, 이보다 규모가 큰 제방과 도로는 수백 미터의 연속성을 갖는 선형구조로 추출되었다. 상대적으로 저해상도인 T-DEM의 경우 선상지와 범람원 상에서 단 한 건의 선형구조도 자동추출 되지 않았는데, 이

는 연구지역의 선상지와 범람원 상에 분포하는 미지형의 상대비고(<5m)가 T-DEM의 해상도에 미치지 못하기 때문에 해석된다. A-DEM 또한 선상지와 범람원 상에서 자동추출 빈도가 낮았는데, 이는 A-DEM의 음영기복도 이미지에 발현된 왜곡이 비고가 작은 지형으로 이루어진 선형구조의 감지를 저해하거나, 실제 지형과 다른 비정상 감지를 유도하기 때문으로 판단된다.

선형구조의 시각적 관독은 연구자의 주관이 개입될 수 있으나, 단층지형의 분포와 제4기 지형 변위 등 지형요소를 기반으로 동정하므로 단층운동과 관련된 선형구조를 감동 분류할 수 있다(이민부·김남신, 2002; 박충선·이광률, 2018a; 이광률 등, 2018; 이초희 등, 2019; 오정식, 2019a). 따라서, 본 연구의 시각적 관독 결과는 자동추출 결과보다 선형구조의 단층 관련성이 크다.

선형구조의 시각적 관독 결과 지형의 수직·수평 변위 혹은 기복량이 큰 형태적 특징으로 인하여 인지된 선형구조는 모든 유형의 DEM에서 유사한 추출 결과를 보였다. 특히, 단층와지(saddle)³⁾, 단층구릉(kernbut), 삼각말단면, 단층애와 같이 기복이 뚜렷하거나(Figure 6), 수 미터 이상의 수평 변위를 동반하는 굴절하천(deflected stream)을 바탕으로 인지되는 선형구조(Figure 7)는 DEM의 유형과 관계없이 유

사한 양상으로 추출되었으며, 이러한 선형구조는 5° 이내의 주향 차이를 보였다. 아울러 선형구조의 길이와 분절 양상도 유사하다(Figure 4-d).

시각적 관독 결과가 가장 현저한 차이를 보이는 곳은 연구지역 북부로 기복이 작은 선상지와 범람원이 분포하며, 농경지 개간과 하천 정비 등으로 지형의 인위적 변형도 많이 진행된 곳이다(Figure 8). 해당 지역의 시각적 관독 결과는 상대적으로 낮은 해상도의 T-DEM이 선형구조의 거시적인 특성을 파악할 때는 유효하나, 저기복의 단층지형 또는 단층에 의한 제4기 지형의 변위 지점을 수반한 선형구조를 추적하고자 할 때는 더 높은 해상도의 DEM이 필요함을 보여준다. 상대적으로 높은 해상도의 L-DEM과 A-DEM도 일부 지점에서 선형구조의 분절 양상이 다르게 나타났다. 이러한 차이는 L-DEM의 선상지에서 인지된 지형의 수직 변위(추정 단층애)가 A-DEM에서 명확하게 인지되지 않기 때문이며, A-DEM의 음영기복도에 발생한 이미지 왜곡이 선형구조의 시각적 인지를 방해한 것으로 판단된다(Figure 8-a, 8-c).

다만, L-DEM의 경우 지표의 인위적 변화가 진행된 현재 지표의 모습을 구현하기 때문에 원지형의 특성을 파악하기에는 부적합하다. 이런 경우 지표면의 인위적 변화가 비교적 덜한 A-DEM을 활용하는 것이 유용하지만, A-DEM은 항공사진의 해상도와 이

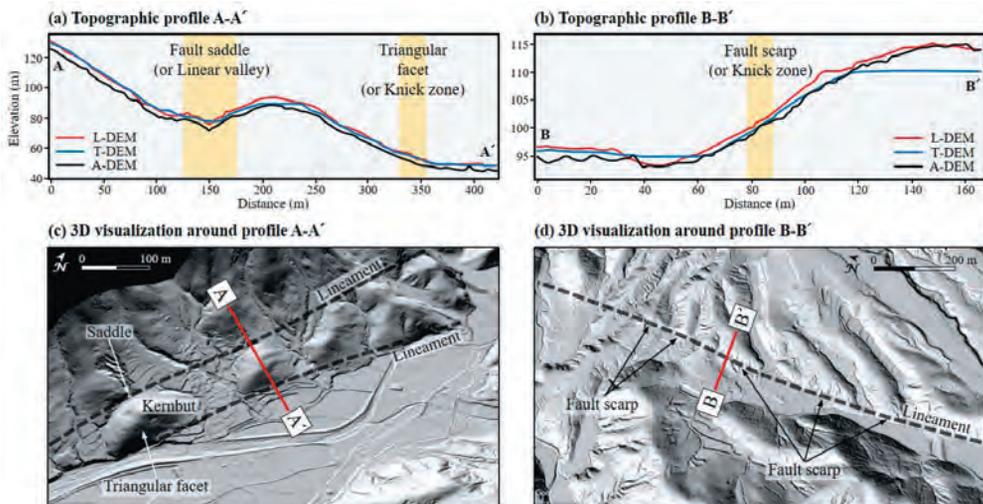


Figure 6. Topographic profile of equally detected lineaments in the fault saddle, triangular facet, and fault scarp.

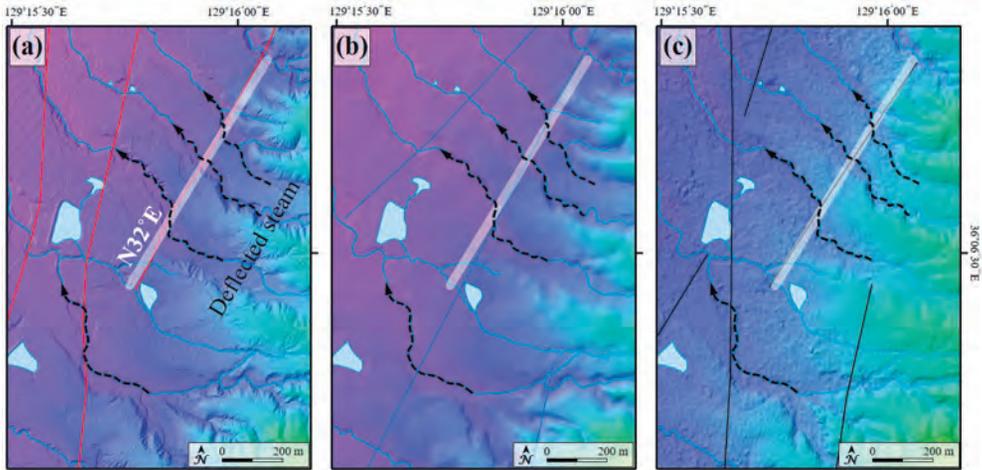


Figure 7. Equally detected lineament along the deflected streams in alluvial fan. White bold lines in each figure shows equally detected lineament. The Base map of each figure is from L-DEM (a), T-DEM (b), and A-DEM (c).

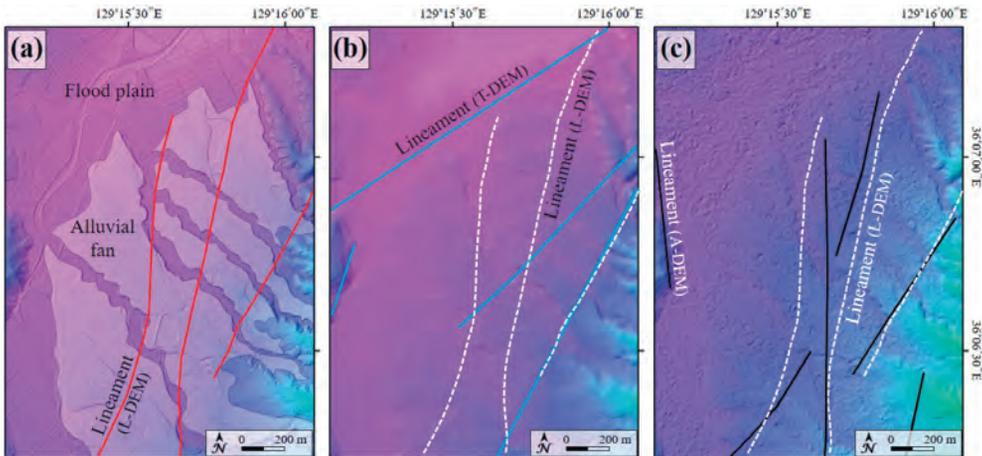


Figure 8. Lineament detected differently in low-relief landforms depending on the type of DEMs. White dashed lines are lineaments in L-DEM. Base map of each figure is L-DEM (a), T-DEM (b), and A-DEM (c).

미지 품질에 따라 왜곡 정도의 차이가 발생하므로 이미지 프로세싱을 포함한 DEM 구축과정 전반에 걸쳐 더 정교한 항공사진측량 기법의 활용이 동반되어야 한다.

2) 선형구조의 활성화

시각적 관독과 야외조사를 바탕으로 단층활동과 직접 연관성이 높은 선형구조를 분리 추출하고, 단층 관련 지형(삼각말달면, 단층와지, 단층구릉)의 분포

와 제4기 지형(단층애, 굴절하천)의 범위 여부를 기준으로 선형구조 활성도를 높고 낮음으로 구분하였다 (Figure 9). 이와 관련하여 ‘국가 활성단층 연구단’은 선형구조의 활성도를 1등급(활성도 높음)과 2등급(활성도 낮음)으로 구분하고 있는데, 여기서 활성도가 높은 선형구조란 단층 관련 지형이 선형구조를 따라 최소 2개 지점 이상 분포함과 동시에 제4기 지형 범위가 추정되거나 확인된 것을 말한다(국가활성단층연구단, 2019; 오정식, 2019a).

연구지역에서 21개의 선형구조가 최종 추출되었

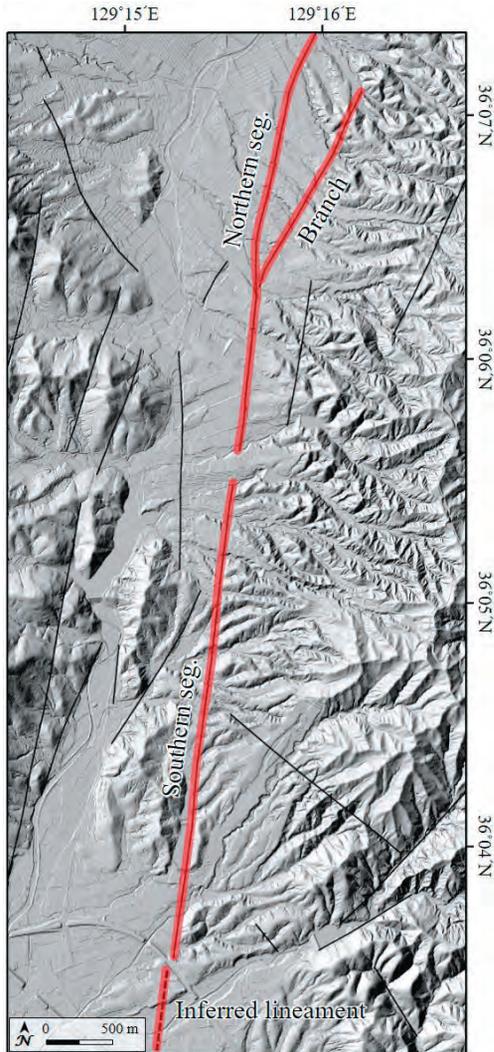


Figure 9. Lineament extracted by overlapping DEMs. Red solid lines are estimated high activity lineaments based on fault-related landforms and Quaternary landforms displacement.

으며, 이들은 총연장 29,080m, 평균 1,384m, 최대 3,614m, 최소 254m로 확인된다. 선형구조는 남북 주향과 북북동 주향이 우세하며, 남동부와 북서부 지역에서는 북서 또는 북북서 주향의 선형구조도 확인된다. 연구지역에서 지배적인 양상을 보이는 남북 또는 북북동 주향의 선형구조는 양산단층 중부 분절의 주향과도 잘 일치한다(최진혁 등, 2017; 이초희 등, 2019). 연구지역 중앙을 남북으로 통과하는 3개의 선

형구조는 활성도가 높은 선형구조로 해석되며, 그 외는 제4기 이전에 형성된 기반암 단층이나 암석의 경계와 관련된 것으로 활성도가 낮은 선형구조이다 (Figure 9).

활성도가 높은 3개의 선형구조 중에서 남북 주향의 2개는 남부 분절(3,614m)과 북부 분절(3,249m)로 구분된다. 남부 분절은 모든 유형의 DEM에서 2~5m의 수직 변위를 보이는 단층애가 선형구조를 따라 연속 분포하며(Figure 6-b, 6-d), 북부 분절은 선상지상에 발달한 지형의 수직 변위(2m 내외 추정 단층애)가 고해상도의 L-DEM과 A-DEM에서 확인된다. 이들 2개 분절은 하나의 선형구조로 볼 수 있지만, 분절 사이(gap) 약 300m 구간에서 단층 관련 지형이나 제4기 지형의 변위가 확인되지 않아 각각의 분절로 구분하였다. 남부와 북부 분절에서 관찰되는 지형의 수직 변위는 선형구조를 기준으로 동쪽 지괴가 일관되게 2m 이상 높은 특징을 보이는데, 이는 벽계 단층(N11°E/80°SE; 소방방재청, 2012)과 단구 단층(N10-20°E/75-79°SE; 이진현 등, 2015)에서 보고된 역이동성(역단층성) 주향이동단층의 운동감각과도 잘 일치하는 지형학적 특징으로 해석된다.

북부 분절의 동편에는 북북동 주향의 가지선형구조(branched lineament; 1,676m)가 확인된다. 가지선형구조는 선상지 고위면을 침식하여 발달한 굴절하천의 선형 분포를 통해 인지되며, T-DEM보다 L-DEM과 A-DEM에서 더욱 뚜렷하게 관찰된다 (Figure 7, 9). 하천의 굴절은 총 4개 지점에서 확인되며, 모두 우수향의 수평 변위를 지시한다. 굴절하천의 수평 변위량은 남쪽이 약 80m로 최대이고, 북쪽으로 갈수록 변위량은 점차 줄어들어 약 60m 수준에 이른다. 가지선형구조에서 확인된 굴절하천의 지형 변위 양상은 인접한 벽계 단층과 단구 단층의 우수주향이동 운동감각과도 잘 일치하며(이진현 등, 2015; 류충렬 등, 2018), 연구지역에서 북쪽으로 약 10km 떨어진 유계-보경사 지역(북부 양산단층)에 분포하는 굴절하천의 변위 양상과도 같다(이초희 등, 2019).

다만, 가지선형구조에서 확인된 굴절하천의 변위량(약 60~80m)은 인접한 유계-보경사 지역의 굴절하천 변위량(0.5~2.68km; 이초희 등, 2019)과 큰 차

이를 보인다. 이러한 차이는 연구지역 굴절하천이 선상지를 통과하는 것으로 비교적 가까운 시기(제4기)에 발생한 단층의 수평 변위를 지시하지만, 유계-보경사 지역은 오랜 침식으로 형성된 산릉 사이를 통과하는 굴절하천으로 장기간의 누적 변위를 지시하기 때문에 볼 수 있다.

6. 결론

중부 양산단층에서 보고된 벽계단층과 단구단층 북부 지역을 대상으로 LiDAR, 수치지형도, 항공사진을 이용하여 DEM을 구축하고, DEM 유형별 선형구조 추출 결과를 비교·분석하였다. 구축된 DEM의 최대 해상도는 L-DEM이 0.50m, T-DEM이 5.00m, 그리고 A-DEM이 0.46m이며, 각기 다른 시기의 지표를 구현했다. L-DEM은 가장 최근의 지표 형상을 고도로 정밀하게 표현했으며, T-DEM은 L-DEM과 유사한 시기의 자료를 바탕으로 구축되었지만 낮은 해상도로 L-DEM 수준의 지표 형상을 구현하지 못했다. A-DEM은 지표의 인위적 변화가 상대적으로 덜한 과거의 지표를 구현할 수 있었지만, DEM 구축과정에 이미지 왜곡을 발생시키는 각종 변수가 많아 이론적 해상도보다 낮은 수준의 지표 구현력을 보였다.

자동추출된 선형구조는 연구지역에 최적화될 때개 변수에 관한 검증이 생략됐으므로 추출된 선형구조와 단층의 관련성은 논할 수 없었다. 다만, 사용된 원자료의 구축 시기와 DEM 해상도 차이에 따른 자동추출 결과 차이는 시각적 관독보다 객관적으로 비교할 수 있다. 자동추출된 선형구조는 DEM 유형과 관계없이 능선과 계곡을 중심으로 높은 분포밀도를 보였으며, DEM 해상도에 따라 분절의 길이와 추출된 선형구조의 개수는 현저한 차이가 있다. DEM 유형별 자동추출 결과 차이는 선상지와 범람원 같은 저기복 지형에서 더 크게 발생했다.

시각적 관독을 통해 추출된 선형구조는 DEM 유형에 따른 추출 개수의 차이가 크게 없었다. 특히, 단층와지, 단층구릉, 삼각말단면과 같이 주변 지형과의 기

복 차이가 뚜렷하거나, 수평 변위를 동반하는 굴절하천으로 인지된 선형구조는 DEM 유형과 관계없이 추출 결과가 유사하다. 그러나 선상지와 범람원과 같이 기복이 작은 지형에서는 선형구조의 길이와 분포 측면에서 다소간의 차이가 있었으며 특히, 선형구조의 주향과 분절 양상은 곳에 따라 큰 차이를 보였다.

선형구조 추출 결과와 야외조사 결과를 종합적으로 판단하여 단층활동과 직접 연관성이 높은 21개의 선형구조가 분리 추출되었다. 이들 중 3개의 선형구조는 제4기 지형의 변위가 인지되는 활성도 높은 선형구조로 평가된다. 연구지역에서 확인된 활성도 높은 선형구조는 지형학적 특성을 고려할 때, 모두 역이동 성분을 가진 우수주향이동단층 운동을 경험한 것으로 추정되며, 기존에 보고된 벽계단층과 단구단층의 운동감각과도 잘 일치한다.

연구결과 T-DEM은 선형구조의 거시적인 특성을 파악할 때 유효하지만, 소규모 단층 지형이나 제4기 지형의 변위를 수반한 선형구조를 추적할 때는 L-DEM이나 A-DEM과 같이 1m 이하의 고해상도 DEM이 필요함을 알 수 있었다. 특히, 활성도 높은 선형구조를 추적하고 선형구조 간 관계와 정확한 분절을 파악하는데 L-DEM의 활용성이 크며, 개발 등으로 지표의 인위적 변화가 심한 지역의 경우 원지형에 비교적 가까운 모습을 구현할 수 있는 A-DEM이 보완자료로 활용될 수 있다.

시각적 관독 결과의 객관성을 높이기 위해서는 자동추출 기법을 활용한 분석 결과의 교차 검증이 수반되어야 하며, 이를 위해 국내 지형발달 환경과 국지적인 지형 특성에 맞춘 자동추출 매개 변수 설정에 관한 논의가 이어져야 한다.

사사

본 연구는 행정안전부 한반도 단층구조선의 조사 및 평가기술 개발 사업의 연구 지원(2017-MOIS31-006)에 의해 수행되었습니다.

주

- 1) 선형구조 추출을 위해 특정 유형의 DEM을 구축하고 분석에 활용할 경우, DEM 구축과정에 발생하는 각종 오차(혹은 오류)가 선형구조 추출 결과에도 영향을 미치므로 DEM 구축 결과의 정확도와 신뢰도 검증이 필요하다. 본 연구는 자료의 전처리와 DEM 구축과정에서 발생 가능한 오차도 DEM의 유형 차이로 전제했다. 특히, 항공사진 자료의 선정과 항공사진측량 과정에서 설정값(parameter)을 달리함에 따라 발생하는 DEM의 품질과 해상도 변화도 DEM 유형의 차이로 해석하였다.
- 2) 선형구조의 자동추출 결과는 DEM의 공간해상도에 영향을 받으며, 동시에 자동추출 방법이나 매개변수의 조정에 따라 그 결과가 달라질 수 있다(박서우 등, 2018; Meixner et al., 2018). 따라서 본 연구의 DEM 유형별 자동추출 결과 차이는 항상성을 지니지 않는다. 그럼에도 본 연구에 활용한 DEM의 공간해상도를 같은 수준으로 조정하지 않은 이유는 다음과 같다. 우선 활성단층 추적에 있어서 선형구조 분석의 주요한 목적은 제4기 지형의 수직·수평 변위를 수반한 선형구조를 찾는 것인데, DEM의 해상도가 높을수록 지형 변위의 인지 가능성이 크기 때문에 구축 가능한 최대 해상도의 DEM을 사용하는 것이 일반적이다. 아울러 본 연구에서 선형구조 자동추출의 목적은 연구자의 주관 개입 여지가 있는 시각적 관독의 DEM 유형별 결과 차이가 객관성 있는 자동추출에서도 나타나는지 단순 비교하고자 함에 있었다.
- 3) 'Kernbut'에 대응되는 개념으로 'Kerncol'이 사용되지만, 형태적 측면에서 단층으로 인하여 단절(변위)된 능선이나 구릉지를 통과하는 단층와지를 'saddle', 그리고 인접한 지형과 단절된 선형의 구릉을 'linear ridge'로 부르기도 한다. 다만, 이들 용어는 연구자에 따라 선택적으로 사용되고 있으므로 용어와 관련된 깊이 있는 논의와 개념 정의가 필요하다. 본 연구에서는 국가활성단층연구단(2019)의 조사 표준화 지침에 따라 kernbut을 단층구릉, saddle을 단층와지로 번역하여 사용하였다.

참고문헌

국가활성단층연구단, 2019, 한반도 단층구조선의 조사 및 평가기술 개발 표준 절차 및 지침.
 김경덕·이동국·유영걸·이현직, 2017, "무인항공사진측량 절차를 적용한 항공사진영상의 처리 및 소규모 변화 지역의 수치지도 수정/갱신 방안," 한국

지형공간정보학회지, 25(4), 15-24.
 김상완·원중선·김원균·민경덕, 1999, "자동 선구조 추출 알고리즘을 이용한 경북 의성지역의 선구조 분석," 자원환경지질, 32(1), 19-31.
 김원균·원중선·김상완, 1996, "원격탐사자료와 수치표고모형을 이용한 지질학적 선구조 분석기술: 경상분지 의성지역을 중심으로," 대한원격탐사학회지, 12(2), 139-154.
 김영석·진광민·최원하·기원서, 2011, "활성단층의 이해: 최근의 연구에 대한 고찰," 지질학회지, 47(6), 723-752.
 류충렬·강지훈·강희철, 2018, "양산단층 북부 일대의 최후기 지구조환경에 대해," 암석학회지, 27(3), 173-184.
 민석규·기근도, 2011, "웅천천 유역에 나타나는 지형선구조의 구조지형학적 특성," 한국지형학회지, 18(4), 261-270.
 박서우·김건일·신진호·홍상훈, 2018, "수치표고모형 공간해상도에 따른 선구조 자동 추출 연구," 대한원격탐사학회지, 34(3), 439-450.
 박진환·이원희, 2016, "저사양 무인항공기를 이용한 소규모 경사지역의 정사영상 및 수치표고모델 제작," 한국측량학회지, 34(3), 283-290.
 박충선·이광률, 2018a, "양산 단층곡 경주 지역의 단층 지형 분석," 한국지형학회지, 25(1), 19-30.
 박충선·이광률, 2018b, "항공사진을 이용한 DEM 제작에 관한 연구," 한국지형학회지, 25(3), 105-120.
 소방방재청, 2012, 활성단층지도 및 지진위험 지도 제작.
 오경섭·양재혁, 2007, "구조지형학 관점에서 본 한반도 중부의 지형선구조 체계," 한국지형학회지, 14(2), 1-19.
 오정식, 2019a, "가음단층계의 선형구조 추출과 선형구조와 단층활동의 관련성," 한국지형학회지, 26(2), 69-84.
 오정식, 2019b, "무인항공기를 활용한 고기복지형의 고해상도 수치표고모델 구축과 지형분석," 한국사진지리학회지, 29(2), 115-127.
 유재진·박현수·양윤정·장동호, 2016, "해안지역의 시계열 지형변화 분석을 위한 UAS 활용 가능성 평가: 안면도 바람아래해변을 대상으로," 한국지형학회지, 23(4), 113-126.
 이광률·박충선·신재열, 2018, "울산단층대 주변의 단

- 층 지형 및 선구조 분포,” 한국지형학회지, 25(3), 89-103.
- 이광률·박충선·신재열, 2019, “동래 단층 중부 지역 응촌-용상 일대의 단층 지형과 지형 발달,” 한국지형학회지, 26(1), 79-91.
- 이민부·김남신, 2002, “Hough변환과 음영기복을 이용한 추가령구조곡의 선형구조 분석,” 지질학회지, 38(4), 457-469.
- 이진현·Sowreh Rezaei·홍예지·최진혁·최정현·최원학·이근우·김영석, 2015, “경북 경주시 단구리 일대에서의 트렌치 조사를 통한 북·부 양산단층의 제4기 단층운동 특성분석,” 지질학회지, 51(5), 471-485.
- 이초희·성영배·오정식·김동은, 2019, “양산단층대 북부 유계-보경사지역의 조구조지형학적 분석,” 한국지형학회지, 26(1), 93-106.
- 임창복·노명현·심택모·이현우·최호선·김효정, 2004, “원자력발전소 부지에 적용하는 활동성 단층의 기준 및 그 배경,” 지질학회지, 40(2), 279-284.
- 장천중·장태우, 2009, “단층슬립의 기하분석에 의한 양산단층의 거동 특성,” 지질공학, 19(3), 277-285.
- 채병곤·장태우, 1994, “청하-영덕지역 양산단층의 운동사 및 관련단열 발달상태,” 지질학회지, 30(4), 379-394.
- 최진혁·김영석·Yann Klinger, 2017, “대규모 지진에 수반된 지표과열특성에 관한 최근 연구동향,” 지질학회지, 53(1), 129-157.
- 최진혁·양석준·김영석, 2009, “울산시 상천리 일대에 발달하는 남부 양산단층의 단층대 분류와 지질구조적 특성,” 지질학회지, 45(1), 9-28.
- 최현·홍순헌, 2007, “항공사진을 이용한 수치표고모형 생성과 활용방안,” 한국콘텐츠학회논문지, 7(3), 168-175.
- 황병훈·이준동·양경희, 2004, “양산단층 주변에 분포하는 화강암질암의 암석학적 연구: 양산단층의 수평 변위량,” 지질학회지, 40(2), 161-178.
- Abdullah, A., Akhir, J. M., and Abdullah, I., 2010, Automatic mapping of lineaments using shaded relief images derived from digital elevation model (DEMs) in the Maran - Sungi Lembing area, Malaysia, *Electronic Journal of Geotechnical Engineering*, 15(6), 949-958.
- Canny, J., 1986, *A computational approach to Edge Detection*, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI-8(6), 679-698.
- Hough, P. V., 1962, Method and means for recognizing complex patterns, United States Patent US 3,069,654, December 18.
- Keller, E. A. and Pinter, N., 2002, *Active tectonics: Earthquakes, Uplift, and Landscape* (2nd ed.), Prentice Hall Inc.
- Lee, K. and Jin, Y. G., 1991, Segmentation of the Yangsan fault system: Geophysical studies om major faults in the Kyeongsang basin, *Journal of the Geological Society of Korea*, 27(4), 434-449.
- Meixner, J., Grimmer, J. C., Becker, A., Schill, E., and Kohl, T., 2018, Comparison of different digital elevation models and satellite imagery for lineament analysis: Implications for identification and spatial arrangement of fault zones in crystalline basement rocks of the southern Black Forest, *Journal of Structural Geology*, 108, 256-268.
- O’Leary, D. W., Friedman, J. D., and Pohn, H. A., 1976, Lineament, linear, lineation: Some proposed new standards for old terms, *Geological Society of America Bulletin*, 87, 1463-1469.
- PCI Geomatics, 2011, *Geomatics help*, PCI Geomatics Enterprises Inc, Richmond Hill, Ontario, Canada.
- Ryoo, C. R., Lee, B. J., Cho, D. L., Chwae, U. C., Choi, S. J., and Kim, J. Y., 1999, Quaternary fault of Dangu-ri, Gangdong-myeon, Gyeongju: Byeokgae fault, *Abstract, Spring Joint Conference on East-North Asia Ore Deposit of Cause of Formation*, 334.
- Šilhavý, J., Minár, J., Mentlík, P., and Sládek, J., 2016, A new artefacts resistant method for automatic lineament extraction using Multi-Hillshade Hierarchic Clustering (MHHC), *Computers & Geosciences*, 92, 9-20.
- 국토지리정보원, 국토정보플랫폼, <http://map.ngii.go.kr>
기상청, 지진화산감시센터, http://www.weather.go.kr/weather/earthquake_volcano/report.jsp
지질자원연구원, 지질정보서비스시스템, <https://mgeo.kigam.re.kr>
Agisoft, PhotoScan Pro1.3 Tutorial, <https://www>

agisoft.com

02841, Korea (e-mail: roehddms@korea.ac.kr, phone: 82-2-3290-2360)

교신: 김동은, 02841, 서울시 성북구 안암로 145 고려대학교 미래국토연구소(이메일: roehddms@korea.ac.kr, 전화: 02-3290-2360)

Correspondence: Dong Eun Kim, Research Institute of Future Land, Korea University, Anam-ro 145, Seongbuk-gu,

최초투고일 2019. 10. 11

수정일 2019. 10. 29

최종접수일 2019. 10. 31

준연동형 비례대표제의 권역 설정과 적용에 관한 연구 - 제20대 총선의 투표 결과에 기초한 시뮬레이션

이정섭* · 조한석** · 지상현***

A Study on the Delineation and Application of Region on the Quasi Proportional Representation System - Simulation Based on the 20th General Election

Lee Chung Sup* · Cho Hanseok** · Chi Sang-Hyun***

요약 : 현재 운영되고 있는 소선거구제 다수대표제와 병립형의 전국단위 비례대표제로 인해 발생되고 누적된 문제점을 해결하기 위해, 다양한 선거제도 개혁안이 논의되고 있다. 이중 연동형 비례대표와 권역별 비례대표로 이루어진 소위 '준연동형 권역별 비례대표제'를 근간으로 하는 「공직선거법」 개정안이 국회에서 추진되고 있다. 이 논문은 제20대 총선의 투표 결과를 해당 개정안에 투입한 시뮬레이션을 통해, 각 정당 및 각 권역별 지역구-비례대표 의석수를 추정하였다. 개정안은 선거결과에 큰 변화를 가져오며, 실제 선거와 가상의 선거결과를 토대로 살펴본 결과는 다음과 같다. 첫째, 정당별 득표율이 의석수와 더 잘 연계되며, 둘째, 사표가 다수 감소하고, 셋째, 지역주의에 기반한 의석확보에 균열이 생겨났다. 이는 개정안 도입의 정치적 목표가 일정 부분 달성되었다는 점을 의미한다. 하지만 선거 이해당사자인 국회의원들의 임의적인 권역 설정 및 적용 때문에 매우 복잡한 의석배분과 당선자 결정을 비롯하여 다음과 같은 문제들이 나타날 수 있음을 확인하였다. 첫째, 권역별 의석 배분의 공정성이 훼손되고, 둘째, 권역별 의석배분으로 인해 새로운 개념의 사표가 다수 등장하며, 셋째, 지역주의 완화라는 성과에도 불구하고 비례대표제도의 중요성은 오히려 감소될 수도 있다. 이러한 문제점은 향후 현재의 개혁안에 대한 불만과 개정요구로 이어질 수 있다는 점에서 선거제도 및 선거의 공간적 단위에 대한 논의는 지속되어야 할 것이다.

주요어 : 선거제 개혁, 준연동형 비례대표제, 권역, 비례성, 사표, 시뮬레이션

Abstract : Various electoral reforms have been discussed to address the problems initiated and accumulated by the majoritarian system of electoral district and the national proportional representative system. The amendment to the Public Official Election Act, which is based on the so-called “quasi-proportional representation system”, which consists of a district voting linked proportional representative and a regional proportional representative, is being discussed in the National Assembly. This paper estimates the number of seats of regional proportional representatives of each political party based on the simulation of voting results of the 20th general election. The new system may induce major changes in the election results. Based on the examinations on the actual and simulated election results, some pivotal notions of the new election system can be summarized as follows. First, the earned votes by party is better associated with the number of seats.

* 경상대학교 지리교육과 부교수(Associate Professor, Department of Geography Education, Gyeongsang National University), yisup@gnu.ac.kr

** 동국대학교대학원 정치학과 박사과정(Ph. D. Candidate, Political Science Major, Dongguk University), hansok@gmail.com

*** 경희대학교 지리학과 부교수(Associate Professor, Department of Geography, Kyunghee University), hyungeo@khu.ac.kr

Second, the number of wasted votes is significantly reduced. Third, there is a positive change in the constellation of seats in some regions. These changes mean that some of the political goals of the amendment would be achieved. However, besides the problems of extremely complicated processes of confirming winners and consequent seat allocation and, arbitrariness of regional configuration, several problems are revealed. First, the fairness of the allocation of seats by regions is undermined. Second, the new category of wasted votes appears due to the distribution of seats by regions. As such problems may lead to dissatisfaction and demand for revision of the quasi-proportional representation system in the future, discussions on electoral systems and spatial units of elections should be continued.

Key Words : Reform of the Electoral System, Quasi Mixed-member Proportional Representation, Regional Party-list, Proportionality, Wasted Vote, Simulation

1. 서론

현행 국회의원 선거제도는 혼합형, 즉 소선거구 지역대표와 전국 명부의 비례대표가 결합된 구조이다. 비록 몇 차례 관련 법률의 개정이 있었지만, 1988년 제13대 총선부터 가장 최근의 2016년 제20대 총선까지 기본 골격은 그대로 유지되어 왔다.

그런데 지난 8차례 총선 동안, 다수 의석을 차지하는 지역대표의 단순다수제 때문에 큰 규모의 사표 발생, 각 정당의 득표율과 의석수 간 괴리라는 낮은 비례성, 그리고 특정 지역에 기반한 정당체제로 인한 영호남 중심의 지역주의 등 여러 부작용들이 발생했다. 이와 같은 현행 국회의원 선거제도로 인해 파생된 여러 문제들을 극복하고 대안을 모색하는 과정에서 ‘연동형 비례대표제’가 주목을 받았고, 제20대 국회에서 이것을 적용하기 위한 여러 「공직선거법」 일부개정법률안들이 발의되었다. 현재의 선거제도가 가지고 있는 문제점과 이를 보완하고 해결하기 위해 도입되는 연동형 비례대표제가 추구하는 가치에 대해서는 우호적인 의견이 많다. 독일의 선거제도로 대표되는 연동형 비례대표제는 단순다수제가 가지는 비(非)비례성을 완화하고 사표를 줄여 국민의 의사가 선거의 결과인 의석으로 보다 정확하게 이어진다는 장점이 있다. 현재 국회에서 논의되고 있는 연동형 혹은 준연동형 비례대표제는 이러한 독일의 제도를 상당부분 받아들인 것으로 보인다.

문제는 이 과정에서 어떤 이유에서인지 발의된 의

안 중 상당수가 ‘권역’이라는 낯선 공간 단위를 제시하고 있다는 것이다.¹⁾ 특히 2019년 4월 심상정 의원이 대표발의한 의안(의안번호 2019985, 이하 심상정 의원안)은 국회 ‘정치개혁특별위원회’에서 신속처리안건(일명 패스트트랙)으로 지정되어 8월 29일 해당 상임위원회를 통과한 상태인데, 비례대표 의석배분을 기존 병립형에서 지역구 당선자수와 50% 연동시키고, 비례대표후보자 명부를 6개 권역으로 작성하고 아울러 권역별 석패율 제도 도입을 추진하고 있다. 이는 공간적으로 보면 현행 ‘지역구-소선거구’와 ‘비례대표-전국’이라는 다계층(multi-tier) 선거구 체계를 ‘지역구-소선거구’와 ‘비례대표-6개 권역’으로 변동시키는 것이라고 할 수 있다(전용주 역, 2017; Farell, 2011).

게다가 심상정 의원안은 서울, 부산·울산·경남, 대구·경북, 인천·경기, 대전·세종·강원·충북·충남 그리고 광주·전북·전남·제주 등 6개 권역으로 구획하는 기준과 근거를 제시하지 않고 있다. 극단적 사례이지만, 직선거리로 약 300km 넘게 떨어진 강원도 고성과 충청남도 서천이 같은 권역으로 묶이는 것이 적절한 것일까? 서울을 제외하고는 2~5개 광역자치단체를 묶은 옥상옥의 공간 범위, 즉 권역이 과연 지방자치·분권이라는 헌법적 가치를 실현하기 위한 적절한 공간단위가 될 것인가? 절차적인 면에서도 지역구 선거구는 인구·행정구역·지리적 여건·교통·생활문화권 등(「공직선거법」 제25조 제①항) 획정 기준이 있고, 아울러 ‘국회의원선거구획정위원회’(동법 제24조)라는 획정의 주체가 존재하여 공청화와 여론 수

럼 등을 통해 선거구획정이 이루어진다. 하지만 비례대표제의 권역 획정에 대해서는 이러한 제도적 절차가 개입되지 못하고 있기에, 이해당사자들인 국회의원들의 일방적인 하향식(top-down) 의사결정이 될 수 있다.

결국 합리적 근거와 기준을 마련하지 못한 채 확정되는 권역은 또 다른 문제, 특히 공간·지역적 마찰과 불균형 문제라는 예상치 못한 부작용을 야기하여 연동형 비례대표제가 달성하고자 했던 원래의 목적을 훼손할 수도 있을 것이다. 이러한 배경에서 이 연구는 소위 '준연동형 권역별 비례대표제'로 일컬어지는 심상정 의원 대표 발의의 「공직선거법」 일부개정법률안이 제시하고 있는 권역이라는 공간의 적절성을 검토하고, 각 권역에 대해서 투표 전 비례대표 의석수 할당 및 투표 후의 비례대표 의석배분 과정에서 나타날 수 있는 제반 문제를 진단하고자 한다. 연구의 방법은 지난 2016년 제20대 총선의 투표 결과를 준연동형 권역별 비례대표제에 투입하여 그 결과를 시뮬레이션 하는 것이다. 이를 통해 심상정 의원안이 제시한 권역 설정이 내재한 문제점과 함께 이를 보완할 수 있는 논의를 이끌어내고자 한다.

2. 준연동형 권역별 비례대표제의 내용

1) 준연동형 권역별 비례대표제의 등장 배경과 주요 내용

국회 의안정보시스템(<http://likms.assembly.go.kr>)에서 확인되는 의안번호 2019985, 「공직선거법」 일부개정법률안의 제안이유를 살펴보면, 소선거구 단순다수제로 선출하는 257석 지역구국회의원과 병립형으로 선출하는 43석 비례대표를 골격으로 하는 지금의 국회의원 선거제도가 대량의 사표 발생, 정당 득표율과 의석점유율 사이 큰 폭의 불일치 그리고 지역별로 지배적인 정당이 그 지역의 의석 대부분을 독점하는 현상 지속, 즉 지역주의 등의 원인이라고 지적하고 있다. 이러한 지적은 그동안 정치권 뿐만 아니라

학계와 시민 사회 등에서 꾸준히 제기해 왔던 내용이며, 우리 정치의 발전을 위해서 반드시 극복해야만 하는 문제, 동시에 선거제도 개혁의 목표라고 할 수 있다(강원택, 2009: 중앙선거관리위원회, 2015).

따라서 현행 국회의원 선거제도가 내재한 문제점의 극복과 정치 개혁이라는 맥락에서 심상정 의원은 비례대표에 대한 의석수 확대, 의석배분 방식 변동, 권역 구획 등을 제시하고 있다. 구체적으로는 지금까지 전체 국회의원 의석수(299석 내지 300석) 중 지역구 의석수를 확정하고 남은 의석수만을 비례대표에 할당하는 것에서 탈피하여, 아예 「공직선거법」에 3:1, 즉 지역구 의석수 225석과 비례대표 의석수 75석을 명시하였다(안 제21조 제①항). 그리고 비례대표 선출 방식을 각 당의 지역구 당선인 수와 연동시키는데, 다만 비례대표의석 중 50%만 연동시키고 나머지는 현재와 같은 병립식을 유지한다(안 제189조 제②항 및 제③항). 이것이 '준연동형 비례대표제'로 일컬어지는 이유이다. 아울러, 각 당은 비례대표 후보자 명부를 전국 6개 권역으로 나누어 작성토록 하고(안 제49조 제②항), 근소한 차이로 낙선한 지역구 후보자를 비례대표의원으로 구제할 수 있는 권역별 석패율제 도입(안 제47조의2)을 제시하고 있기에 '권역별 비례대표제'라고도 칭할 수 있다.

2) 의석수 확대의 방지: 준연동형

사표를 줄이고, 비례성을 확대시키는 대표적인 선거제도는 심상정 의원안이 제시한 준연동형이 아니라, 독일과 뉴질랜드가 채택하고 있는 완전 연동형 비례대표제이다. 실제로 제20대 국회에서는 <표 1>과 같이 연동형 비례대표제 도입을 추진한 의안들이 다수 제출되었다. 그런데, 해당 의안들이 추구하는 사표 감소와 비례성 확대를 위해서는 첫째 비례대표 의석수를 늘리면서 전체 국회의원 정수를 확대, 둘째 전체 국회의원 정수를 확대하지 않는 대신에 비례대표 의석수를 늘리고 그 만큼 지역구 의석수를 감축, 셋째 앞의 두 경우 포함하여 반드시 투표 결과에 따른 초과 및 보정의석 발생이라는 세 가지 조건이 전제된다.

주지하다시피, 첫 번째 조건은 유권자들의 국회의

표 1. 제20대 국회에서 발의된 연동형 비례대표제 의안들이 제시한 국회의원 정수

| 의안 | 국회의원 정수에 관한 제시 내용 | 총 의석수 | 지역구 의석수 | 비례대표 의석수 |
|------------------------|---|-------|---------|----------|
| 소병훈 의원안 (2016. 7. 27) | 국회의 의원정수는 지역구국회의원과 비례대표국회의원을 합하여 300명으로 하되, 지역구국회의원과 비례대표국회의원의 비율은 2:1로 한다. | 300 | 200 | 100 |
| 김상희 의원안 (2017. 2. 14) | …인구 15만명당 1인을 기준으로 산출한 수로 하되, 지역구국회의원과 비례대표국회의원의 비율은 3:1로 하고… | 345 | 259 | 86 |
| 박주민 의원안 (2017. 2. 15) | 국회의 의원정수는 인구 14만명당 1명을 기준으로 산출한 수(소수점 이하는 버린다)로 하되, 지역구국회의원과 비례대표국회의원의 비율은 2:1로 한다. | 370 | 247 | 123 |
| 박주현 의원안 (2016. 10. 24) | 국회의 의원정수는 지역구국회의원과 비례대표국회의원을 합하여 316명으로 하고, 지역구국회의원정수는 253석, 비례대표국회의원정수는 63석으로 한다. | 316 | 253 | 63 |
| 심상정 의원안 (2017. 12. 12) | 국회의 의원정수는 지역구국회의원 240인과 비례대표국회의원 120인을 합하여 360인으로 하되… | 360 | 240 | 120 |

출처: 이정섭 · 조한석 · 지상현, 2018, “권역별 연동형 비례대표제에서 의석할당과 국회의원 의석수에 대한 공간적 접근,” 대한지리학회지, 53(1), p. 96.

원에 대한 불신 및 그들의 특권적 지위에 대한 비판적 정서에 크게 반하는 것이었고, 두 번째 조건은 지역구 국회의원의 극심한 반발과 선거구획정의 어려움이 예상되면서, 결국 지역구 의석수 감소를 최소화하는 방향(253석→225석)에서 총 의석수를 현행과 같이 300석으로 유지하고 있다. 그리고 세 번째 조건인 초과 및 보정 의석의 발생은 비례대표 의석배분을 연동 50%와 병립 50%로 하고, 부가적으로 ‘조정 의석’이라는 방식을 통해 방지하고 있다.

결국 사표 방지, 비례성 확대라는 가치 지향과 현행 국회의원 정수 300명 유지라는 현실 사이에서 타협한 것이 완전연동형을 대신한 준연동형 비례대표제라고 할 수 있다.

3) 지역주의 극복을 위한 공간 단위 도입: 권역별

심상정 의원안이 6개 권역 구획을 도입하고자 함은, “…비례대표 명부를 권역별로 작성하고, 정당별 열세 지역에서 근소한 차이로 낙선한 지역구 후보자를 비례대표의원으로 선출하는 석패율제를 도입함으로써 국회의 의석배분에 있어 국민의 의사의 왜곡

을 최소화함과 동시에 지역주의를 개선하며 다양한 정책과 이념에 기반한 정당의 의회 진출을 촉진하려는 것…”이라는 제안 이유를 통해 충분히 확인된다.

즉 해당 의안의 제안이유를 거꾸로 읽어 나가면, 지역주의 개선을 위해서는 석패율제가 필요하고, 석패율제를 운용하기 위해서 바로 전국이 아닌 권역이라는 공간 구획이 필요하다는 것이다. 다시 말해, 준연동형 권역별 비례대표라는 용어에서 ‘준연동형’이 사표의 최소화, 비례성의 확대라는 목적에 초점을 두었다면, ‘권역별’이라는 용어는 한국 정치의 고질적인 문제점인 지역주의를 완화하기 위한 공간적 기획인 것이다.

3. 시뮬레이션 자료와 과정

1) 시뮬레이션 투입 자료: 제20대 총선 투표 결과의 조정

준연동형 권역별 비례대표제를 근간으로 「공직선거법」이 개정된다면, 과연 사표가 줄어들고 정당득표

울과 의석점유율 간의 비례성이 확대되며 아울러 지역주의가 개선될 것인지를 살펴보기 위해서 이제 시뮬레이션을 진행하고자 한다.

물론 시뮬레이션은 과거 결과로 미래를 가늠하는 것이며, 두 시점 사이에서 발생한 여러 변화를 포착할 수도 없어 상당한 한계와 괴리 등이 있음은 주지의 사실이다. 그렇지만 이 연구는 제21대 총선의 정당별 획득 의석수를 예측하는 것이 아닌, 비례대표제를 준연동 및 권역별로 적용하는 과정에 초점을 맞추고 있기에 가장 최근 총선 결과인 2016년 제20대 총선 결과를 시뮬레이션 투입 자료로 채택하였다.

우선 준연동형이기에 지역구 국회의원 당선인수를 설정하여야 하는데, 제20대 총선의 지역구 의석수는 253석이었지만, 심상정 의원안은 225석으로 감축하도록 되어 있어서 이를 조정하여 시뮬레이션에 투입하였다. 이 때 조정치는 253:225를 단순 비례식으로

산출할 수도 있지만, 이 연구에서는 225석 지역구 의석을 2015년 10월 주민등록인구를 기준으로 17개 광역자치단체에 Hamilton 방식으로 의석할당(apportionment)하고, 각 광역자치단체별 할당된 의석수를 놓고 제20대 총선의 정당 및 무소속 당선자의 실제 득표수를 추가 확인하여 조정치를 구했다.

이와 같은 방식을 통해 4개 정당 및 무소속 지역구 당선인수 추정치를 <표 3>과 같이 17개 광역자치단체별로, 그리고 <표 4>와 같이 6개 권역별로 구했다.

다음으로는 제20대 총선의 정당별 비례대표 득표수를 <표 5>와 같이 정리하였다. 지난 총선에서 비례대표 후보자 명부를 작성한 정당은 총 21개이며, 이에 대한 전국 유권자의 투표수 합계는 2천 4백만 표 이상이다. 하지만 현행 「공직선거법」과 심상정 의원안 모두 비례대표국회의원선거에서 유효투표총수의 3% 이상 또는 지역구국회의원총선거에서 5석 이상

표 2. 제20대 총선의 정당별 지역구 당선인수

| 시도명 | 지역구 국회의원 의석수 | 정당별 지역구 국회의원 당선인수 | | | | |
|-------|--------------|-------------------|--------|------|-----|-----|
| | | 새누리당 | 더불어민주당 | 국민의당 | 정의당 | 무소속 |
| 서울 | 49 | 12 | 35 | 2 | 0 | 0 |
| 부산 | 18 | 12 | 5 | 0 | 0 | 1 |
| 대구 | 12 | 8 | 1 | 0 | 0 | 3 |
| 인천 | 13 | 4 | 7 | 0 | 0 | 2 |
| 광주 | 8 | 0 | 0 | 8 | 0 | 0 |
| 대전 | 7 | 3 | 4 | 0 | 0 | 0 |
| 울산 | 6 | 3 | 0 | 0 | 0 | 3 |
| 세종 | 1 | 0 | 0 | 0 | 0 | 1 |
| 경기 | 60 | 19 | 40 | 0 | 1 | 0 |
| 강원 | 8 | 6 | 1 | 0 | 0 | 1 |
| 충북 | 8 | 5 | 3 | 0 | 0 | 0 |
| 충남 | 11 | 6 | 5 | 0 | 0 | 0 |
| 전북 | 10 | 1 | 2 | 7 | 0 | 0 |
| 전남 | 10 | 1 | 1 | 8 | 0 | 0 |
| 경북 | 13 | 13 | 0 | 0 | 0 | 0 |
| 경남 | 16 | 12 | 3 | 0 | 1 | 0 |
| 제주 | 3 | 0 | 3 | 0 | 0 | 0 |
| 전국 합계 | 253 | 105 | 110 | 25 | 2 | 11 |

자료: 중앙선거관리위원회 선거통계시스템(<http://info.nec.go.kr>)

주: 제20대 총선(2016년 4월) 당시 정당별 기호 순으로 정리

표 3. 시도별-정당별 지역구 당선인수 조정치(가정)

| 시도 | 정당 | 새누리당 | 더불어민주당 | 국민의당 | 정의당 | 무소속 | 시도별 의석수 |
|----------|----|------|--------|------|-----|-----|---------|
| 서울 | | 10 | 30 | 2 | 0 | 0 | 42 |
| 부산 | | 10 | 4 | 0 | 0 | 1 | 15 |
| 대구 | | 7 | 1 | 0 | 0 | 3 | 11 |
| 인천 | | 4 | 7 | 0 | 0 | 2 | 13 |
| 광주 | | 0 | 0 | 6 | 0 | 0 | 6 |
| 대전 | | 3 | 3 | 0 | 0 | 0 | 6 |
| 울산 | | 2 | 0 | 0 | 0 | 3 | 5 |
| 세종 | | 0 | 0 | 0 | 0 | 1 | 1 |
| 경기 | | 18 | 38 | 0 | 1 | 0 | 57 |
| 강원 | | 5 | 1 | 0 | 0 | 1 | 7 |
| 충북 | | 4 | 3 | 0 | 0 | 0 | 7 |
| 충남 | | 5 | 4 | 0 | 0 | 0 | 9 |
| 전북 | | 1 | 2 | 5 | 0 | 0 | 8 |
| 전남 | | 1 | 1 | 6 | 0 | 0 | 8 |
| 경북 | | 12 | 0 | 0 | 0 | 0 | 12 |
| 경남 | | 11 | 3 | 0 | 1 | 0 | 15 |
| 제주 | | 0 | 3 | 0 | 0 | 0 | 3 |
| 정당별 당선인수 | | 93 | 100 | 19 | 2 | 11 | 225 |

표 4. 권역별-정당별 지역구 당선인수 조정치(가정)

| 권역 | 정당 | 새누리당 | 더불어민주당 | 국민의당 | 정의당 | 무소속 | 권역별 의석수 |
|---------------------|----|------|--------|------|-----|-----|---------|
| 1권역(서울) | | 10 | 30 | 2 | 0 | 0 | 42 |
| 2권역(부산·울산·경남) | | 23 | 7 | 0 | 1 | 4 | 35 |
| 3권역(대구·경북) | | 19 | 1 | 0 | 0 | 3 | 23 |
| 4권역(인천·경기) | | 22 | 45 | 0 | 1 | 2 | 70 |
| 5권역(광주·전북·전남·제주) | | 2 | 6 | 17 | 0 | 0 | 25 |
| 6권역(대전·세종·충북·충남·강원) | | 17 | 11 | 0 | 0 | 2 | 30 |
| 정당별 당선인수(추정) | | 93 | 100 | 19 | 2 | 11 | 225 |

표 5. 제20대 총선의 정당별 비례대표 득표수

| 시도명 | 시도별 선거인수 | 투표수 | 정당별 비례대표 득표수 | | | |
|-----|-----------|-----------|--------------|-----------|-----------|---------|
| | | | 새누리당 | 더불어민주당 | 국민의당 | 정의당 |
| 서울 | 8,423,654 | 5,034,474 | 1,522,417 | 1,280,881 | 1,424,383 | 420,292 |
| 부산 | 2,952,961 | 1,636,061 | 654,214 | 422,916 | 322,693 | 95,622 |
| 대구 | 2,031,478 | 1,113,056 | 571,775 | 175,726 | 187,765 | 65,502 |
| 인천 | 2,379,666 | 1,323,196 | 430,683 | 327,690 | 346,300 | 96,538 |
| 광주 | 1,158,598 | 713,886 | 20,124 | 200,628 | 374,308 | 51,390 |

| | | | | | | |
|-------|------------|------------|-----------|-----------|-----------|-----------|
| 대전 | 1,214,402 | 712,143 | 215,645 | 196,382 | 189,062 | 52,781 |
| 울산 | 937,421 | 554,631 | 197,388 | 122,468 | 113,350 | 46,907 |
| 세종 | 167,798 | 106,587 | 29,755 | 29,591 | 27,619 | 9,201 |
| 경기 | 10,034,919 | 5,768,325 | 1,821,246 | 1,513,849 | 1,521,240 | 439,071 |
| 강원 | 1,277,858 | 736,891 | 307,007 | 169,288 | 136,559 | 40,457 |
| 충북 | 1,287,549 | 737,249 | 274,497 | 196,055 | 152,411 | 40,147 |
| 충남 | 1,683,854 | 934,429 | 331,976 | 243,245 | 202,457 | 50,426 |
| 전북 | 1,520,032 | 956,033 | 69,868 | 298,537 | 395,984 | 75,361 |
| 전남 | 1,567,192 | 997,524 | 54,369 | 289,799 | 458,772 | 55,952 |
| 경북 | 2,242,016 | 1,270,113 | 703,928 | 156,227 | 179,477 | 63,081 |
| 경남 | 2,719,668 | 1,549,255 | 658,361 | 364,354 | 260,999 | 97,646 |
| 제주 | 501,332 | 286,893 | 97,019 | 82,108 | 62,193 | 19,517 |
| 전국 합계 | 42,100,398 | 24,430,746 | 7,960,272 | 6,069,744 | 6,355,572 | 1,719,891 |

자료: 중앙선거관리위원회 선거통계시스템(<http://info.nec.go.kr>)

주: 제20대 총선(2016년 4월) 당시 정당별 기호 순으로 정리

의 의석을 차지한 정당에게만 비례대표 의석을 배분하는 ‘봉쇄조항’이 있다. 이를 적용하면 제20대 총선 비례대표 의석할당 피제수는 당시 새누리당, 더불어민주당, 국민의당, 정의당에 투표한 22,105,479표이며, 전체 유효 투표수 중 약 10%에 해당하는 2백 3십만 이상의 사표가 발생했다.

2) 시뮬레이션 과정

현재 선거제도는 비례대표를 지역구 득표와 별개로 병립식으로 선출하기 때문에 의석 배분과 당선인 결정이 비교적 단순하다. 하지만, 심상정 의원안은 무려 7단계에 걸쳐 준연동형, 권역별 그리고 석패율제를 적용하고 나서야 각 당의 비례대표 당선인이 결정된다.

제1단계는 ‘연동배분의석수’를 구하는 것으로, 다음 식을 따른다(안 제189조 제2항 제1호).

$$\text{연동배분의석수} = [(\text{국회의원정수} - \text{의석할당정당이 추천하지 않은 지역구국회의원당선인수}) \times \text{해당 정당의 비례대표국회의원선거 득표비율} - \text{해당 정당의 지역구국회의원당선인수}] \div 2$$

그림 1. 연동배분의석수 계산식

계산식 중 ‘의석할당정당이 추천하지 않은 지역구 국회의원당선인수’는 무소속 또는 앞서 살펴본 봉쇄조항으로 비례대표 의석을 할당받지 못하는 정당의 지역구 당선인을 의미하는 것이며, ‘해당 정당’은 봉쇄조항을 넘어서는 의석할당정당을 의미한다. 그리고 ‘해당 정당의 비례대표국회의원선거 득표비율’은 각 의석할당정당의 전국득표수를 모든 의석할당정당의 전국 득표수의 합계로 나누어 산출하도록 되어 있기에 제20대 총선 기준으로 새누리당, 더불어민주당, 국민의당, 정의당에 투표한 22,105,479표가 피제수가 되고, 각 당의 득표가 제수, 즉 분자가 된다.

제2단계는 정당별 지역구국회의원당선자수와 무관한 비례의석 즉, ‘잔여배분의석수’를 산정하는 단계로서, 아래와 같은 식으로 구한다(안 제189조 제2항 제2호).

$$\text{잔여배분의석수} = (\text{비례대표국회의원의석정수} - \text{각 연동배분의석수의 합계}) \times \text{비례대표국회의원선거 득표비율}$$

그림 2. 잔여배분의석수 계산식

제3단계는 ‘조정석수’에 대한 계산인데(안 제189조 제2항 제3호), 이것은 제1단계 연동의석수 합계

가 비례대표국회의원 정수를 초과할 경우, 각 정당의 연동배분의석을 비례적으로 축소하여 초과의석을 방지하기 위한 조치이다(음선필, 2019).

$$\text{조정석수} = \text{비례대표국회의원의석정수} \times \text{연동배분의석수} \div \text{각 연동배분의석수의 합계}$$

그림 3. 조정석수 계산식

이처럼 제1단계 연동형과 제2단계 병립형이라는 준연동형 그리고 제3단계를 통한 초과의석 방지를 거쳐, 75개 비례대표 의석 중 각 당이 몇 석씩을 배분받게 되는지가 우선 결정된다.

다음으로 제4단계와 제5단계는 ‘권역별 연동배분의석수’ 및 ‘권역별 잔여배분의석수’를 구하는 단계로(안 제189조 제③항 제1호 및 제2호), 앞선 단계를 통해 각 당이 배분받은 비례대표의석을 6개 권역 중 어디에 다시 배분할 것인가를 결정짓는 과정이다.

$$\text{권역별 연동배분의석수} = [(\text{정당이 추천한 지역구국회의원 당선인수} + \text{제②항에 따라 정당에 배분한 총 비례대표국회의원 의석수}) \times \text{해당 권역별 비례대표국회의원선거 득표비율} - \text{해당 권역의 지역구국회의원당선인수}] \div 2$$

그림 4. 권역별 연동배분의석수 계산식

$$\text{권역별 잔여배분의석수} = (\text{제②항에 따라 정당에 배분한 총 비례대표국회의원의석수} - \text{각 권역별 연동배분의석수의 합계}) \times \text{권역별 비례대표국회의원선거 득표비율}$$

그림 5. 권역별 잔여배분의석수 계산식

그리고 제6단계는 앞의 제4단계에서 초과의석이 발생할 때, 이를 방지하기 위한 조치로 ‘권역별 조정석수’를 계산하는 것이다(안 제189조 제③항 제3호). 이는 제1단계 ‘연동배분의석수’에 대해서 제3단계 ‘조정석수’를 통해 초과의석을 방지하는 것과 동일한 장치로 볼 수 있다.

$$\text{권역별 조정석수} = \text{제②항에 따라 정당에 배분한 총 비례대표국회의원의석수} \times \text{권역별 연동배분의석수} \div \text{각 권역별 연동배분의석수의 합계}$$

그림 6. 권역별 조정석수 계산식

한편, 제4~5단계 계산식의 ‘해당 권역별 비례대표국회의원선거 득표비율’은 각 권역별 정당의 득표수를 해당 정당의 전국득표수로 나눈 것이다. 결국 i) 각 정당들에게 비례대표 의석수를 배분하는 제1~3단계에서는 해당 정당의 전국 득표율, 이후 ii) 각 정당이 배분받은 비례대표 의석수를 다시 권역별로 나누는 제4~6단계에서는 개별 정당이 전국적으로 얻은 전체 표에 대한 각 권역의 기여도/비율이라는 서로 다른 기준을 적용하도록 하고 있다.

이러한 이유로 전국의 비례대표의석수는 75개로 고정되어 있지만, 각 권역이 몇 개의 비례대표의석을 가지는지는 투표가 끝난 후에야 결정된다. 즉 각 권역의 비례대표의석 정수(定數)가 없고, 지역구 당선인수를 비롯한 전체 투표결과에 따라 변동되는 구조이다.

마지막 제7단계는 석패율제에 관한 것인데, 이 연구의 시뮬레이션에서는 이를 생략하고자 한다. 이것은 지역구에서 낙선한 후보자만이 대상이며, 석패율 적용대상 후보자를 누구로 선택할 것인지도 각 당의 몫이기에 현 시점에서는 시뮬레이션에 투입할 적절한 근거 자료(parameter)도 없고, 아울러 제1~6단계를 통해 얻어진 정당별·권역별 비례대표의석수 값에 변동을 끼치지 않아도 되기 때문이다.

4. 시뮬레이션 결과

1) 연동배분 및 잔여배분의석수

먼저 연동배분의석수를 계산하는데 국회의원정수 300, 의석할당정당이 추천하지 않은 지역구국회의원 당선인수 11 그리고 봉쇄조항을 넘은 4개 정당별 비

$$[(300 - 11) \times \begin{array}{|c|c|} \hline & \text{비례대표국회의원선거 득표비율} \\ \hline \text{새누리당} & 0.36010 \\ \hline \text{더불어민주당} & 0.27458 \\ \hline \text{국민의당} & 0.28751 \\ \hline \text{정의당} & 0.07780 \\ \hline \end{array} - \begin{array}{|c|c|} \hline & \text{지역구국회의원 당선자수} \\ \hline \text{새누리당} & 93 \\ \hline \text{더불어민주당} & 100 \\ \hline \text{국민의당} & 19 \\ \hline \text{정의당} & 2 \\ \hline \end{array}] \div 2 = \begin{array}{|c|c|} \hline & \text{계산값} \\ \hline \text{새누리당} & 5.535 \\ \hline \text{더불어민주당} & -10.323 \\ \hline \text{국민의당} & 32.045 \\ \hline \text{정의당} & 10.243 \\ \hline \end{array}$$

그림 7. 연동배분의석수 계산

$$(75 - 48) \times \begin{array}{|c|c|} \hline & \text{비례대표국회의원선거 득표비율} \\ \hline \text{새누리당} & 0.36010 \\ \hline \text{더불어민주당} & 0.27458 \\ \hline \text{국민의당} & 0.28751 \\ \hline \text{정의당} & 0.07780 \\ \hline \end{array} = \begin{array}{|c|c|} \hline & \text{계산값} \\ \hline \text{새누리당} & 9.723 \\ \hline \text{더불어민주당} & 7.414 \\ \hline \text{국민의당} & 7.763 \\ \hline \text{정의당} & 2.101 \\ \hline \end{array}$$

그림 8. 잔여배분의석수 계산

표 6. 정당별 지역구 및 비례대표 당선인수(추정)

| | 새누리당 | 더불어민주당 | 국민의당 | 정의당 | 무소속 |
|----------------|------|--------|------|-----|-----|
| 지역구 당선인수 | 93 | 100 | 19 | 2 | 11 |
| 비례대표 당선인수(A+B) | 16 | 7 | 40 | 12 | - |
| 연동배분(A) | 6 | 0 | 32 | 10 | - |
| 잔여배분(B) | 10 | 7 | 8 | 2 | - |
| 총 당선인수 | 109 | 107 | 59 | 14 | 11 |

례대표국회의원선거 득표비율 및 지역구국회의원 당선자수를 투입하여 계산값을 구하였다.

계산된 값에 대해서 1보다 작은 경우는 0으로, 그리고 소수점 첫째자리에서 반올림하여 산정하도록 해당 의안에 명시되어 있기에, 새누리당 6석, 더불어민주당 0석, 국민의당 32석, 정의당 10석 등 총 48석의 연동배분의석이 결정되었다.

다음으로 잔여배분의석의 계산은 비례대표국회의원의석정수는 75, 각 연동배분의석수의 합계는 48이며, 여기에 4개 정당의 득표비율을 투입하였다. 잔여배분의석수 계산에서 발생하는 소수점 이하에 대해서 전 단계와 다르게 Hamilton 방식으로 처리하도록 명시하고 있다. 이것은 현행 「공직선거법」의 비례대표의석을 배분하는 방식과 동일한 것으로, 이에 따라 4개 정당은 각각 10석, 7석, 8석, 2석을 배분받게 된다.

이번 시뮬레이션에서는 총 75석의 비례대표의원정수를 초과하는 의석수가 발생하지 않았기 때문에

조정의석을 계산할 필요 없으며, 4개 정당은 <표 6>과 같은 값의 지역구 및 비례대표 당선인수를 확보할 것으로 계산되었다.

2) 권역별 연동배분 및 권역별 잔여배분의석수

새누리당, 더불어민주당, 국민의당, 정의당 등 4개 정당들이 배분받은 각각 16석, 7석, 40석, 12석의 비례대표 의석을 6개 권역으로 배치하는 것은 권역별 연동 및 잔여배분의석수 산정식에 의해 결정된다. 먼저 권역별 연동배분의석수의 산정은 <그림 9>와 같이, 각 정당의 전체 지역구 당선인수와 비례대표당선인수(=연동배분의석수+잔여배분의석수), 그리고 각 정당의 6개 권역별 비례대표 득표비율과 권역별 지역구 당선인수에 따라 결정된다.

심상정 의원안은 권역별 연동배분의석수 계산으로 산출된 값(그림 10의 A.)을 소수점 첫째자리에서 반

올림하여 산정하고, 그 값이 1보다 작은 경우에는 0으로 처리(그림 10의 B.)하도록 명시하고 있다. 그런데 4개 정당 중 더불어민주당은 각 권역별 연동배분

의석수의 합이 14로, 앞 단계에서 배분받은 비례대표 당선인수 7보다 크다. 즉 초과의회가 발생하였기에 권역별 조정의석수를 다시 산출하였는데, 이때 소

$$\left[\begin{array}{|c|c|} \hline & \text{지역구 당선인수} \\ \hline \text{새} & 93 \\ \hline \text{민} & 100 \\ \hline \text{국} & 19 \\ \hline \text{정} & 2 \\ \hline \end{array} + \begin{array}{|c|c|} \hline & \text{비례대표 당선인수} \\ \hline \text{새} & 16 \\ \hline \text{민} & 7 \\ \hline \text{국} & 40 \\ \hline \text{정} & 12 \\ \hline \end{array} \right] \times \begin{array}{|c|c|c|c|c|c|} \hline & \text{1 권역} & \text{2 권역} & \text{3 권역} & \text{4 권역} & \text{5 권역} & \text{6 권역} \\ \hline \text{새} & 0.19125 & 0.18969 & 0.16026 & 0.28290 & 0.03032 & 0.14558 \\ \hline \text{민} & 0.21103 & 0.14988 & 0.05469 & 0.30340 & 0.14351 & 0.13750 \\ \hline \text{국} & 0.22412 & 0.10967 & 0.05778 & 0.29384 & 0.20317 & 0.11142 \\ \hline \text{정} & 0.24437 & 0.13965 & 0.07476 & 0.31142 & 0.11758 & 0.11222 \\ \hline \end{array} - \begin{array}{|c|c|c|c|c|c|} \hline & \text{1 권역} & \text{2 권역} & \text{3 권역} & \text{4 권역} & \text{5 권역} & \text{6 권역} \\ \hline \text{새} & 10 & 23 & 19 & 22 & 2 & 17 \\ \hline \text{민} & 30 & 7 & 1 & 45 & 6 & 11 \\ \hline \text{국} & 2 & 0 & 0 & 0 & 17 & 0 \\ \hline \text{정} & 0 & 1 & 0 & 1 & 0 & 0 \\ \hline \end{array} \right] \div 2$$

그림 9. 권역별 연동배분의석수 계산

A. 권역별 연동배분의석수 계산값

| | 1 권역 | 2 권역 | 3 권역 | 4 권역 | 5 권역 | 6 권역 |
|---|-------|-------|-------|-------|-------|-------|
| 새 | 5.42 | -1.16 | -0.77 | 4.42 | 0.65 | -0.57 |
| 민 | -3.71 | 4.52 | 2.43 | -6.27 | 4.68 | 1.86 |
| 국 | 5.61 | 3.24 | 1.70 | 8.67 | -2.51 | 3.29 |
| 정 | 1.71 | 0.48 | 0.52 | 1.68 | 0.82 | 0.79 |

B. 소수점 이하의 값 처리

| | 1 권역 | 2 권역 | 3 권역 | 4 권역 | 5 권역 | 6 권역 | 합계 |
|---|------|------|------|------|------|------|----|
| 새 | 5 | 0 | 0 | 4 | 1 | 0 | 10 |
| 민 | 0 | 5 | 2 | 0 | 5 | 2 | 14 |
| 국 | 6 | 3 | 2 | 9 | 0 | 3 | 23 |
| 정 | 2 | 0 | 1 | 2 | 1 | 1 | 7 |

C. 더불어민주당 초과의회 조정

| | 1 권역 | 2 권역 | 3 권역 | 4 권역 | 5 권역 | 6 권역 | 합계 |
|---|------|------|------|------|------|------|----|
| 새 | 5 | 0 | 0 | 4 | 1 | 0 | 10 |
| 민 | 0 | 2.5 | 1 | 0 | 2.5 | 1 | 7 |
| 국 | 6 | 3 | 2 | 9 | 0 | 3 | 23 |
| 정 | 2 | 0 | 1 | 2 | 1 | 1 | 7 |

그림 10. 각 정당별-권역별 연동배분의석수 배분

$$\left(\begin{array}{|c|c|} \hline & \text{비례대표 당선인수} \\ \hline \text{새} & 16 \\ \hline \text{민} & 7 \\ \hline \text{국} & 40 \\ \hline \text{정} & 12 \\ \hline \end{array} - \begin{array}{|c|c|} \hline & \text{각 정당별 연동배분 의석수의 합계} \\ \hline \text{새} & 10 \\ \hline \text{민} & 7 \\ \hline \text{국} & 23 \\ \hline \text{정} & 7 \\ \hline \end{array} \right) \times \begin{array}{|c|c|c|c|c|c|} \hline & \text{1 권역} & \text{2 권역} & \text{3 권역} & \text{4 권역} & \text{5 권역} & \text{6 권역} \\ \hline \text{새} & 0.19125 & 0.18969 & 0.16026 & 0.28290 & 0.03032 & 0.14558 \\ \hline \text{민} & 0.21103 & 0.14988 & 0.05469 & 0.30340 & 0.14351 & 0.13750 \\ \hline \text{국} & 0.22412 & 0.10967 & 0.05778 & 0.29384 & 0.20317 & 0.11142 \\ \hline \text{정} & 0.24437 & 0.13965 & 0.07476 & 0.31142 & 0.11758 & 0.11222 \\ \hline \end{array} = \begin{array}{|c|c|c|c|c|c|} \hline & \text{1 권역} & \text{2 권역} & \text{3 권역} & \text{4 권역} & \text{5 권역} & \text{6 권역} \\ \hline \text{새} & 1.148 & 1.138 & 0.962 & 1.697 & 0.182 & 0.873 \\ \hline \text{민} & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ \hline \text{국} & 3.810 & 1.864 & 0.982 & 4.995 & 3.454 & 1.894 \\ \hline \text{정} & 1.222 & 0.698 & 0.374 & 1.557 & 0.588 & 0.561 \\ \hline \end{array}$$

그림 11. 각 정당별-권역별 잔여배분의석수 계산

표 7. 정당별-권역별 비례대표 당선인수(추정)

| | 각 정당의 권역별 연동배분의석수 (A) | | | | 각 정당의 권역별 잔여배분의석수 (B) | | | | 각 정당의 권역별 비례대표 당선인수 (A+B) | | | | 권역별 비례대표 의석수 |
|------|-----------------------|-----|---|---|-----------------------|---|---|---|---------------------------|-----|----|---|--------------|
| | 새 | 민 | 국 | 정 | 새 | 민 | 국 | 정 | 새 | 민 | 국 | 정 | |
| 제1권역 | 5 | 0 | 6 | 2 | 1 | 0 | 4 | 1 | 6 | 0 | 10 | 3 | 19 |
| 제2권역 | 0 | 2.5 | 3 | 0 | 1 | 0 | 2 | 1 | 1 | 2.5 | 5 | 1 | 9.5 |
| 제3권역 | 0 | 1 | 2 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 3 | 1 | 6 |
| 제4권역 | 4 | 0 | 9 | 2 | 2 | 0 | 5 | 1 | 6 | 0 | 14 | 3 | 23 |
| 제5권역 | 1 | 2.5 | 0 | 1 | 0 | 0 | 3 | 1 | 1 | 2.5 | 3 | 2 | 8.5 |
| 제6권역 | 0 | 1 | 3 | 1 | 1 | 0 | 2 | 1 | 1 | 1 | 5 | 2 | 9 |

수점 이하에 대한 처리는 Hamilton 방식을 적용하여 배분된다(그림 10의 C.)²⁾

이상의 과정을 통해 구해진 4개 정당별 '각 당 비례대표 당선인수-각 당 권역별 연동배분의석수 합'의 값은 새누리당 6, 더불어민주당 0, 국민의당 17, 정의당 5이다. 이들 의석에 대해서 추가적으로 권역별 잔여배분의석수를 산정하여야 한다.

권역별 잔여배분의석수를 계산할 때 발생하는 소수점 이하 값은 각 정당이 각 권역에 대해서 Hamilton 방식으로 처리한다. 즉 정수(整數)의 의석을 먼저 배정하고 남은 의석은 소수점 이하 수가 큰 순으로 각 권역에 1석씩 배분하여, 정수의 합이 각 당 비례대표 당선인수에 도달할 때까지 반복하였다. 이처럼 각 당의 '권역별 연동배분의석수' 산정 → 더불어민주당의 '권역별 조정의석수' 산정 → 각 당의 '권역별 잔여의석수' 산정을 통해 얻어진 결과는 <표 7>과 같다.

5. 준연동형 권역별 비례대표제 시뮬레이션 평가

1) 비례성(득표율과 의석수의 연계): 정당 간 비례성은 개선되지만, 권역 간 비례성에 대한 의문

지금까지 지역구 225석에 대한 가정, 그리고 비례대표 75석의 정당별-권역별 의석배분 산정 단계를 거쳐 얻어진 결과가 <표 8>이다. 지난 제20대 총선의 결과는 새누리당 122석, 더불어민주당 123석, 국민의당 38석, 정의당 6석, 무소속 11석이었으나, 준연동형 권역별 비례대표제 시뮬레이션의 결과는 각각 109석, 107석, 59석, 14석, 11석이었다. 이러한 차이는 실제 정당별 득표율이 의석수에 더 잘 반영됨으로 인해 정당 간 비례성이 개선되었다고 해석될 수 있다. 즉 새누리당, 더불어민주당은 각각 13석, 16석의 감

표 8. 정당별-권역별 지역구와 비례대표 당선인수(추정)

| 권역/시도 | 정당 | 지역구 당선인수 (A) | | | | | 비례대표 당선인수 (B) | | | | | 총 당선인수 (A+B) | | | | | 권역별 의석수 |
|----------|----------|--------------|--------|--------|--------|--------|---------------|-----|----|----|---|--------------|-----|----|----|----|---------|
| | | 새 | 민 | 국 | 정 | 무 | 새 | 민 | 국 | 정 | 무 | 새 | 민 | 국 | 정 | 무 | |
| 1권역 | 서울 | 10 | 30 | 2 | 0 | 0 | 6 | 0 | 10 | 3 | 0 | 16 | 30 | 12 | 3 | 0 | 61 |
| 2권역 | 부산 | 10 | 4 | 0 | 0 | 1 | | | | | | | | | | | |
| | 울산 경남 | 2 11 | 0 3 | 0 0 | 0 1 | 3 0 | 1 | 2.5 | 5 | 1 | 0 | 24 | 9.5 | 5 | 2 | 4 | 44.5 |
| 3권역 | 대구 | 7 | 1 | 0 | 0 | 3 | | | | | | | | | | | |
| | 경북 | 12 | 0 | 0 | 0 | 0 | 4 | 1 | 3 | 1 | 0 | 20 | 2 | 3 | 1 | 3 | 29 |
| 4권역 | 인천 | 4 | 7 | 0 | 0 | 2 | | | | | | | | | | | |
| | 경기 | 18 | 38 | 0 | 1 | 0 | 6 | 0 | 14 | 3 | 0 | 28 | 45 | 14 | 4 | 2 | 93 |
| 5권역 | 광주 | 0 | 0 | 6 | 0 | 0 | | | | | | | | | | | |
| | 전북 | 1 | 2 | 5 | 0 | 0 | | | | | | | | | | | |
| | 전남 | 1 | 1 | 6 | 0 | 0 | 1 | 2.5 | 3 | 2 | 0 | 3 | 8.5 | 20 | 2 | 0 | 33.5 |
| | 제주 | 0 | 3 | 0 | 0 | 0 | | | | | | | | | | | |
| 6권역 | 대전 | 3 | 3 | 0 | 0 | 0 | | | | | | | | | | | |
| | 세종 | 0 | 0 | 0 | 0 | 1 | | | | | | | | | | | |
| | 충북 | 4 | 3 | 0 | 0 | 0 | 1 | 1 | 5 | 2 | 0 | 18 | 12 | 5 | 2 | 2 | 39 |
| | 충남 강원 | 5 5 | 4 1 | 0 0 | 0 0 | 0 1 | | | | | | | | | | | |
| 정당별 당선인수 | | 93 | 100 | 19 | 2 | 11 | 16 | 7 | 40 | 12 | 0 | 109 | 107 | 59 | 14 | 11 | 300 |

소되었지만 국민의당은 21석, 정의당은 8석 증가하였기에 소선거구제의 단순다수대표제로 인한 거대 양당의 과다 대표 문제와 소수정당에 대한 다수의 득표가 사표가 되는 문제는 완화되었다.

보다 구체적으로 ‘갤러거 최소제공지수’를 적용한 비비례성 수준³⁾에 대해 제20대 총선 결과와 시물레이션 간을 비교하면, 당초 8.9의 지수가 3.0까지 개선되는 것으로 나타났다.

그렇지만, 권역 사이의 비례성 문제는 여전히 의문으로 남아있다. 제20대 총선의 인구를 기준(2015년 10월 31일)으로 권역별로 지역구, 비례대표, 총 의석 등의 1개 의석당 평균 인구를 계산하였다. 우선 제1권역 서울 및 제2권역 인천·경기, 즉 수도권 총 의석수가 154석으로 전국 총의석수의 51.3%, 비례대표 의석수도 75석 중 42석으로 56%를 차지하는 것으로 나타난다. 실제 전국에서 수도권 인구가 차지하는 비중이 약 49.3% 수준임을 감안하면, 수도권의 과다대표와 이외 지역의 과소대표가 나타나고 있다. 이는 사표를 최소화하기 위해 고안된 절차로 인해 정당별 득표율이 크게 차이가 나지 않는 지역에 보다 많은 의석을 배분하는 결과로 나타났기 때문이다.

특히 비례대표 1개 의석당 평균 인구를 따져보면, 제1권역 서울은 약 53만 명, 제3권역인 대구·경북은 약 87만 명으로 가장 차이가 크게 나타났고, 두 권역 사이의 인구편차는 약 -23%와 +26% 이다. 물론 시물레이션을 위해 지역구 당선인수를 사전에 시도별로 의석할당하여 가정하였고, 준연동형이기에 이것이 다시 각 권역별 비례대표 의석배분에 영향을 주고 있다.

2) 사표: 전국 단위에서는 사표가 줄어들지만, 권역에서 새로운 개념의 사표 발생

앞서 시물레이션을 진행하는 과정에서 봉쇄조항으로 인해 비례대표 투표의 약 10%, 즉 비례대표의석을 배분받지 못하는 정당에 대한 투표는 사표 처리된다고 언급한 적이 있다. 그러나 해당 봉쇄조항은 현행 「공직선거법」과 동일하고, 연동형 비례대표제를 실시하고 있는 독일, 뉴질랜드 등과 비교해도 높은 수준은 아니다.

이보다는 준연동형 비례대표제를 권역이라는 공간으로 획정하고 적용시키면서 발생하는 새로운 개념

표 9. 의석할당 정당의 의석률과 득표율 비교

| 구분 | | 새누리당 | 더불어민주당 | 국민의당 | 정의당 | 비비례성 |
|------------|-----|------|--------|------|-----|------|
| 제20대 총선 결과 | 의석률 | 40.7 | 41.0 | 12.7 | 2.0 | 8.9 |
| | 득표율 | 38.9 | 33.9 | 22.5 | 4.8 | |
| 준연동형 시물레이션 | 의석률 | 36.3 | 35.7 | 19.7 | 4.7 | 3.0 |
| | 득표율 | 38.9 | 33.9 | 22.5 | 4.8 | |

표 10. 권역별 지역구, 비례대표 및 총의석의 1개 의석당 평균인구

| | 인구수 (2015. 10. 기준) | 지역구 의석수 | 비례대표 의석수 | 총 의석수 | 지역구 1개 의석 평균인구 | 비례대표 1개 의석 평균인구 | 총의석 1개 평균인구 |
|-----|-----------------------|------------|-------------|----------|-------------------|--------------------|----------------|
| 1권역 | 10,045,027 | 42 | 19 | 61 | 239,167.3 | 528,685.6 | 164,672.6 |
| 2권역 | 8,050,228 | 35 | 9.5 | 44.5 | 230,006.5 | 847,392.4 | 180,904.0 |
| 3권역 | 5,190,725 | 23 | 6 | 29 | 225,683.7 | 865,120.8 | 178,990.5 |
| 4권역 | 15,414,110 | 70 | 23 | 93 | 220,201.6 | 670,178.7 | 165,743.1 |
| 5권역 | 5,871,177 | 25 | 8.5 | 33.5 | 234,847.1 | 690,726.7 | 175,259.0 |
| 6권역 | 6,929,719 | 30 | 9 | 39 | 230,990.6 | 769,968.8 | 177,685.1 |

의 사표 문제를 지적하고자 한다.

지난 20대 총선에서 서울과 인천·경기에서 더불어민주당은 각각 약 128만 표와 184만 표의 정당명부 비례대표 투표를 얻었다. 그런데 시뮬레이션 결과는 더불어민주당의 해당 권역 비례대표 당선인수가 모두 0명이다. 이는 기술적으로, 해당 지역에서 더불어민주당이 지역구 선거에서 비교적 적은 표차로 다수의 의석을 확보하였다는 점을 의미하는 것이지만, 수도권 지역의 더불어민주당 비례대표 득표 약 312만 표의 가치를 어떻게 해석해야 할까? 즉 사표인지, 사표가 아닌지를 쉽사리 답할 수 없다.

심상정 의원안에 근거한다면, 공간 단위별로 사표가 될 수도 있고, 아닐 수도 있다. 우선 권역의 공간에서는 사표로 판단된다. 적어도 제1권역 서울과 제4권역 인천·경기에서 얻은 더불어민주당 비례대표의 득표가 의석수로 전혀 연결되지 못했기 때문이다. 그런데 전국 공간에서 접근하면 사표가 아닐 수도 있는데, 수도권의 더불어민주당 비례대표 득표가 다른 권역의 자당 비례대표 당선에는 기여하기 때문이다. 이처럼 비록 해당 권역의 비례대표 의석에는 기여가 없지만 전국적으로는 기여가 존재하기에 사표가 아니라는 해석도 어느 정도 가능해 보인다.

그렇지만 개별 유권자의 시각으로 접근한다면, 다시금 사표라고 판단된다. 왜냐하면, 유권자는 자신이 거주하는 권역의 각 정당명부 비례대표 후보자에 대해서 투표했기 때문이다. 즉 서울 혹은 인천·경기의 더불어민주당 비례대표 후보자 명부를 보고 유권자들은 선택하고 투표했는데, 그 명부의 후보자 당선에는 전혀 기여하지 못하고, 뜻밖에도 영남, 호남, 충청 등의 명부에 등재된 후보자, 게다가 기본 정보조차 제공받지 못한 다른 권역 후보자의 당선에 기여할 것이기 때문이다.

마치 대중가요의 가사처럼 ‘사표인 듯 사표 아닌, 사표 같은’ 비례대표 투표와 그에 따른 의석배분의 결과는 유권자의 선택권을 심각하게 침해하는 것일 수도 있다. 어느 유권자는 특정 정당에 대한 지지가 우선이고 비례대표 명부에 있는 후보자에 대한 지지는 부차적인 것일 수도 있지만, 어느 유권자는 그 반대일 수도 있다. 하지만 심상정 의원안은 전자의 경우만을

전제하고 있다. 따라서 내가 싫어하는 타 권역 비례대표 후보자가 같은 정당이라는 이유만으로 나의 정확한 정치적 의사표현을 통하지 않고, 내 표의 기여를 받아 당선될 수도 있다. 특정 지역에서 특정 정당과 그 후보자들에 대한 지지가 그 당 모든 후보자에 대한 지지라고 전제한다는 점은 권역별 비례대표제의 정신과는 거리가 있다.

준연동형 권역별 비례대표제가 상당 부분 참고한 것으로 보이는 독일의 선거제도에서는 연방을 구성하는 각 주라는 공간 단위에서 연동형으로 비례대표 의석을 배분하며 각 주의 투표결과는 해당 주의 의석 배분에만 기여한다. 작센의 비례대표 투표가 바이에른 의석배분에 절대 기여하지 않는데, 우리는 서울이 제주에 영향을 주어도 괜찮다고 할 수 있을까? 물론 연방국가 체제에서의 각 주와 우리의 권역을 같은 정치적 단위로 비교할 수는 없지만, 정치적 의사표현이 권역을 넘나드는 방식의 수학적 계산에 의존된다는 점은 큰 문제로 남는다.

그리고 분명한 사실은 제21대 총선에서 유권자가 받게 될 ‘선거공보물’에는 다른 권역의 정당별 비례대표자 명부가 아닌 해당 권역의 것만 들어 있다는 것이다.

3) 지역주의: 지역주의 완화의 효과, 그러나 석패율제의 함정

심상정 의원안을 시뮬레이션했을 때, 지역주의 문제가 개선되었는지에 답하기 위해 다음의 과정을 거쳤다. 그리고 6개 권역별로 의석수에 대한 각 정당의 당선인수, 즉 점유율로 표준화하여 함께 비교하였다.

〈표 11〉 및 〈표 12〉처럼, 준연동형 권역별 비례대표제는 일정 정도 지역주의를 완화시키는 것으로 나타났다. 구체적으로는 4개 정당 모두 제20대 총선에서 취약했던 권역에서 다수 의석을 배분받을 수 있고, 특히 거대 양당인 새누리당과 더불어민주당은 각자의 강세 지역에서는 당선인수가 줄어들지만, 상대적으로 약세 지역들인 영남과 호남에서 1~2석씩 확보하는 것으로 나타났다. 그리고 제20대 총선 때 새누리당은 제2·3·6권역, 더불어민주당은 제1·4권역,

표 11. 제20대 총선과 준연동형 비례대표제의 권역별-정당별 당선인수 비교

| 권역 | 새누리당 | | 더불어민주당 | | 국민의당 | | 정의당 | | 무소속 | |
|-----|------|-----|--------|-----|------|-----|-----|-----|-----|-----|
| | 20대 | 준연동 | 20대 | 준연동 | 20대 | 준연동 | 20대 | 준연동 | 20대 | 준연동 |
| 1권역 | 12 | 16 | 35 | 30 | 2 | 12 | 0 | 3 | 0 | 0 |
| 2권역 | 27 | 24 | 8 | 9,5 | 0 | 5 | 1 | 2 | 4 | 4 |
| 3권역 | 21 | 20 | 1 | 2 | 0 | 3 | 0 | 1 | 3 | 3 |
| 4권역 | 23 | 28 | 47 | 45 | 0 | 14 | 1 | 4 | 2 | 2 |
| 5권역 | 2 | 3 | 6 | 8,5 | 23 | 20 | 0 | 2 | 0 | 0 |
| 6권역 | 20 | 18 | 13 | 12 | 0 | 5 | 0 | 2 | 2 | 2 |
| 합계 | 105 | 110 | 110 | 107 | 25 | 59 | 2 | 14 | 11 | 11 |

주: 제20대 총선은 지역구 253석, 준연동형 비례대표제는 300석을 기준으로 비교하였음.

표 12. 권역별 의석에 대한 정당별 점유율 비교

(단위: %)

| | 제20대 총선 (A) | | | | | 준연동 시뮬레이션 (B) | | | | | B - A | | | | |
|-----|-------------|------|------|-----|------|---------------|------|------|-----|------|-------|-------|-------|-----|------|
| | 새 | 민 | 국 | 정 | 무 | 새 | 민 | 국 | 정 | 무 | 새 | 민 | 국 | 정 | 무 |
| 1권역 | 24.5 | 71.4 | 4.1 | 0.0 | 0.0 | 26.2 | 49.2 | 19.7 | 4.9 | 0.0 | 1.7 | -22.2 | 15.6 | 4.9 | 0.0 |
| 2권역 | 67.5 | 20.0 | 0.0 | 2.5 | 10.0 | 53.9 | 21.3 | 11.2 | 4.5 | 9.0 | -13.6 | 1.3 | 11.2 | 2.0 | -1.0 |
| 3권역 | 84.0 | 4.0 | 0.0 | 0.0 | 12.0 | 69.0 | 6.9 | 10.3 | 3.4 | 10.3 | -15.0 | 2.9 | 10.3 | 3.4 | -1.7 |
| 4권역 | 31.5 | 64.4 | 0.0 | 1.4 | 2.7 | 30.1 | 48.4 | 15.1 | 4.3 | 2.2 | -1.4 | -16.0 | 15.1 | 2.9 | -0.6 |
| 5권역 | 6.5 | 19.4 | 74.2 | 0.0 | 0.0 | 9.0 | 25.4 | 59.7 | 6.0 | 0.0 | 2.5 | 6.0 | -14.5 | 6.0 | 0.0 |
| 6권역 | 57.1 | 37.1 | 0.0 | 0.0 | 5.7 | 46.2 | 30.8 | 12.8 | 5.1 | 5.1 | -11.0 | -6.4 | 12.8 | 5.1 | -0.6 |

국민의당은 제5권역에서 과반 이상의 의석 점유율을 차지하였지만, 시뮬레이션 결과로는 제2·3권역의 새누리당, 제5권역의 국민의당만 과반을 유지하였고 점유율도 축소되었다.

그런데 심상정 의원안이 담고 있는 석패율제를 고려하면, 또 다른 문제가 발생할 수도 있다. 해당 의안에서는 각 정당은 각 권역에 최대 2명까지 지역구 후보이면서 비례대표 명부에도 등재하여 당선시킬 수 있는 석패율제 도입을 명시하고 있다. 이 경우 '순수' 비례대표 당선인수는 각 권역 및 각 정당별로 최대 2명씩 줄어들 수도 있다. 다만 해당 권역에서 30% 이상의 지역구 의석을 차지하는 정당에 대해서는 석패율제 적용을 배제하도록 명시하고 있어서, 이보다는 적은 규모로 감소할 것이다.

시뮬레이션에 석패율제를 적용하면, 순수 비례대표 당선인수는 전국적으로 75명이 아닌 46명이 될 수 있

고 이는 제20대 총선의 비례대표 정수 47명보다 적다. 권역별로는 제3권역은 2명, 제5·6권역은 3~4명만의 순수 비례대표 당선인만 존재하고, 반면 수도권인 제1권역과 제4권역은 각각 13명과 19명이다. 그리고 정당별로는 거대 양당인 새누리당은 제2·3·5·6권역, 그리고 더불어민주당은 거의 모든 권역에서 0~1명의 순수 비례대표 당선인만 존재하기에 해당 권역의 비례대표 명부작성이 무의미한 것일 수도 있다. 독일의 소위 인물화된 비례대표제(Personalised proportional system)는 '비례대표제'에 방점이 찍혀 있지만, 심상정 의원안은 석패율제로 인해 지역구 선거의 성격이 강화되어 소선거구+대선거구의 혼합적인 양상이 나타날 가능성이 있다. 즉 몇몇 권역에서는 투표 결과에 따라서 중복 지역구 선거로 변질될 수도 있으며, 이 과정에서 정당명부에 의한 비례대표제는 그 의미가 상당 부분 감소될 가능성이 있다고 할 것이다.

표 13. 권역별-정당별 순수 비례대표 당선인수(가정)

| | 각 정당의 권역별 비례대표 당선인수 | | | | 각 정당의 권역별 순수 비례대표 당선인수 | | | |
|------|---------------------|--------|------|-----|------------------------|--------|------|-----|
| | 새누리당 | 더불어민주당 | 국민의당 | 정의당 | 새누리당 | 더불어민주당 | 국민의당 | 정의당 |
| 제1권역 | 6 | 0 | 10 | 3 | 4 | 0 | 8 | 1 |
| 제2권역 | 1 | 2.5 | 5 | 1 | 1 | 0.5 | 3 | 0 |
| 제3권역 | 1 | 1 | 3 | 1 | 1 | 0 | 1 | 0 |
| 제4권역 | 6 | 0 | 14 | 3 | 6 | 0 | 12 | 1 |
| 제5권역 | 1 | 2.5 | 3 | 2 | 0 | 0.5 | 3 | 0 |
| 제6권역 | 1 | 1 | 5 | 2 | 1 | 0 | 3 | 0 |
| 합계 | 16 | 7 | 40 | 12 | 13 | 1 | 30 | 2 |

6. 결론

현재 국회에서 정치개혁특별위원회에서 신속처리 안건으로 지정하여 논의하고 있는 심상정 의원안은 현행 국회의원 선거제도에 비해서, 첫째 정당 득표율과 의석수 사이의 비례성을 확대하고, 둘째 사표를 축소하며, 셋째 지역주의 문제를 개선함을 목표로 하고 있다. 이러한 문제점의 상당 부분은 현행 선거제도가 가지는 단점에서 기인하며, 선거제도 개혁과 정치 발전을 위해 개선되고 해결되어야 하는 것임은 분명하다.

본 논문에서는 해당 의안이 제안한 준연동형 권역별 비례대표제라는 새로운 선거제도에 지난 총선의 투표 결과를 투입하여 시뮬레이션하여 현행 선거제도의 결과와 비교하였다. 이를 바탕으로 살펴보면, 정당 득표율과 의석수 간 비례성은 상당히 확대되었고 지역주의 문제에 있어서도 어느 정도 개선의 효과가 나타날 것으로 분석된다.

그러나 성과와 더불어 해결해야 할 문제점도 발생하는 것으로 나타났다. 첫째 정당 간 비례성은 개선되었지만, 비례대표 의석배분에 있어 권역 간 비례성/인구편차 격차 및 수도권에서 과다 대표의 문제가 확인되고, 둘째 권역별 비례대표 의석 배분과 당선자 확정 과정에서 권역을 초월하는 의석배분과 득표 기여 때문에 유권자의 선택권이 침해될 가능성, 마지막으로 몇몇 권역에서 엇보이는 ‘소선거구 + 대선거구’라는 중복 지역구제로의 변질 가능성 등이 시뮬레이

션 과정에서 확인할 수 있다.

이와 같은 문제의 직접적이고 기술적인 이유는 6개 권역에 투표 전 각각 몇 석의 비례대표 의석을 할당할 것인지를 생략했기 때문이다. 전국적으로는 75석의 비례대표 정수를 「공직선거법」에 명시하는 것으로 개정하지만, 각 권역의 비례대표 당선인 수는 투표가 끝난 후 6~7단계의 계산을 거쳐야만 확정될 수 있다. 유권자의 시각으로는 매우 복잡하고 이해하기 어려운 선거제도이며, 동시에 나의 투표가 어느 권역, 어느 비례대표 후보자의 당선에 기여할 것인지를 확인할 수도 없다는 문제가 있다.

이에 덧붙여, 본 연구자들은 본질적인 문제로서 과연 ‘권역’ 설정이 적절한지, 나아가 선거제도의 개혁을 위해 현재의 권역 설정이 반드시 필요한 것인지에 대한 문제의식을 제기한다. 선행 연구에서 본 연구자들은 ‘권역별 연동형 비례대표제’를 도입하고자 하는 논의들이 제시한 권역 설정에 따른 제반 문제들을 이미 지적한 바 있다. 연동형이 아닌 준연동형으로 개정한다고 하더라도 권역의 적절성과 필요성에 대한 의문은 해결되지 않았다.

물론 사표를 줄이고 정당 간 득표와 의석 점유의 비례성을 확대하기 위해서는 연동형 또는 준연동형 비례대표제 도입이 필요하다. 또 지역주의를 개선하기 위해서는 권역이라는 공간적 장치의 필요성은 인정된다. 게다가 초과 및 보정의석을 포함한 국회의원 정수 확대에 대한 국민의 비판적 정서, 각 정당의 이해득실 대립, 현재 253석 지역구를 대폭 감소시킬 때

나타날 엄청난 갈등 등이 '이상적' 개혁의 제약 요인으로 작용하고 있는 현실도 충분히 감안되어야 한다.

그렇지만 '현실적', '실현가능성'이라는 논리에 근거하여 준연동형과 권역을 연결하는 선거제도로의 정치적 합의는 오히려 선거제도 개선을 통해 이루고자 했던 목표 달성을 어렵게 할 수도 있다. 즉, 한 번의 선거는 가능하지만 이로 인해 발생하는 복잡함, 모호함, 이해관계의 충돌이 오히려 준연동형 비례대표제에 대한 개혁을 요구하는 목소리로 이어질 수도 있을 것이다.

향후 추가적인 시뮬레이션 등 분석과 연구가 필요함을 전제로, 뉴질랜드처럼 전국을 비례대표 의석배분의 공간적 단위로 설정한다면, 이 연구가 지적한 여러 문제들 중 지역주의 개선을 제외한 대부분을 해결할 수 있을 것으로 보인다. 지역주의 개선까지 목표한다면 독일의 주처럼, 우리의 광역자치단체, 즉 시도로 확정하는 것도 고려해 볼 필요가 있다. 단 300석 안팎의 국회의원 정수를 유지하기 위해서는 현행보다는 대폭적인, 대략 140~150석으로 지역구를 축소하고 이에 상응하는 비례대표 의석 확대 그리고 초과·보정의석 도입이 전제된다.

물론 이러한 대안 역시 수많은 토론과 정치적 논란과 정치세력 간 타협을 필요로 하는 것이지만, 현재의 복잡한 선거제도 개혁안이 유권자의 정치적 의사판단을 명확하게 반영하고 있다고 말하기가 분명 쉽지 않을 것이다. 이 때문에 정치적 실현가능성에도 불구하고 현재의 개정안에 대한 논의를 시작하는 것은 결코 대의제 민주주의의 발전을 가로막거나 미루는 행위로 폄하될 수는 없을 것이며, 동시에 선거제도 개혁에 있어 공간적 관점을 간과해서는 안 되는 이유일 것이다.

주

1) 권역이라는 선거구 공간은 우리에게만 낯선 것이 아니라, 전 세계적으로도 이례적인 것이라고 할 수 있다. 연동형 비례대표제를 채택하고 있는 국가 중, 독일은 연방을 구성하

는 각주(bundesland), 뉴질랜드는 전국을 비례대표 명부 작성 및 연동배분 공간으로 설정하고 있다(김영일·김종갑, 2014; 김종갑·이정진, 2017; 김형철, 2016).

2) 더불어민주당의 '권역별 조정의석수'를 산출하면, 제2권역과 제5권역의 값이 2.5로 동일하다. 심상정 의원안에는 '소수점 이하 수가 큰 순으로 각 권역에 1석씩 배분되되, 그 수가 같은 때에는 해당 권역 사이의 추첨에 의한다.'고 명시하였다. 따라서 제2권역 3석, 제5권역 2석으로 배분, 또는 그 반대로 배분될 수도 있다. 이 연구에서는 추첨을 생략하고 소수점 값을 포함하여 일단 2.5석씩 배분하였다.

3) 최소제곱지수(LSq: least squares index)

$$Lsq = \sqrt{\frac{1}{2} \sum (\text{득표율} - \text{의석률})^2}$$

참고문헌

- 강원택, 2009, "지역주의 극복과 정치 통합을 위한 선거제도 개혁의 방안," 입법과 정책, 1(1), 31-54.
- 김영일·김종갑, 2014, 2013년 독일 연방하원 선거제도의 개정내용 및 특징, 국회입법조사처(NARS) 현안 보고서 제225호.
- 김종갑·이정진, 2017, 뉴질랜드 비례대표제의 주요 특징 및 시사점, 이슈와 논점 제1254호, 국회입법조사처.
- 김형철, 2016, "뉴질랜드 선거제도 개혁 과정과 성공요인: 한국에 주는 시사점," 시민과 세계, 28, 61-91.
- 음선필, 2019, "이른바 준연동형 비례대표제에 관한 헌법적 검토: 국회 정개특위 선거제도 개편안을 중심으로," 홍익법학, 20(2), 29-62.
- 이정섭·조한석·지상현, 2018, "권역별 연동형 비례대표제에서 의석할당과 국회의원 의석수에 대한 공간적 접근," 대한지리학회지, 53(1), 91-109.
- 전용주(역), 2017, 선거제도의 이해, 한울 아카데미, 서울(Farrell, D. M., 2011, *Electoral Systems: A Comparative Introduction*, Palgrave macmillan.)
- 중앙선거관리위원회, 2015, 정치관계법 개정의견.
- 국회 의안정보시스템, <http://likms.assembly.go.kr>
- 중앙선거관리위원회 선거통계시스템, <http://info.nec.go.kr>
- 통계청 국가통계포털, <http://kosis.kr>

교신: 지상현, 02447, 서울특별시 동대문구 경희대로 26,
경희대학교 지리학과(이메일: hyungeo@khu.ac.kr)

Correspondence: Chi Sang-Hyun, Department of Geogra-
phy, Kyunghee University, 26, Kyungheedaero, Dongdae-
mun-gu, 02447, Seoul, Republic of Korea(e-mail: hyungeo
@khu.ac.kr)

최초투고일 2019. 10. 4

수정일 2019. 10. 29

최종접수일 2019. 10. 30

고유벡터공간필터링 접근법에 기반한 피어슨 상관계수의 요소분해

이상일*

Decomposing the Pearson's Correlation Coefficient Based on the Eigenvector Spatial Filtering Approach

Sang-Il Lee*

요약 : 본 논문의 주된 연구 목적은 고유벡터공간필터링 접근법에 기반한 피어슨 상관계수 요소분해 기법을 정련화하고, 이 기법이 공간적 자기상관이 전제된 상태에서의 이변량 상관관계 분석에 어떠한 공헌을 할 수 있을지를 실험 연구를 통해 검토하는 것이다. 피어슨 상관계수 요소분해 기법은 공간적 패턴 요소분해, 하위 상관계수의 산출, 결정계수의 요소분해의 과정을 거쳐 이루어지며, 최종적으로 피어슨 상관계수를 네 가지 상관관계 요소, 즉 '잔차-잔차 상관관계 요소(EE)', '공통-공통 상관관계 요소(CC)', '특수-잔차 상관관계 요소(UE)', '잔차-특수 상관관계 요소(EU)'로 분해한다. 피어슨 상관계수 요소분해 기법의 유용성을 검토하기 위해 동일한 피어슨 상관계수 값을 갖지만 서로 다른 수준의 이변량 공간적 자기상관을 보이는 가상의 8개 패턴쌍에 적용하였다. 실험 연구를 통해 밝혀진 주요 내용을 정리하면 다음과 같다. 첫째, 공간적 패턴 요소분해 결과, 공통 패턴 요소와 특수 패턴 요소의 존재/부존재의 양상이 매우 다양하게 나타난다. 둘째, 공간적 패턴 요소의 존재/부존재의 양상과 개별 변수의 이변량 공간적 자기상관의 정도에 따라, 하위 상관계수의 상대적 크기, 그리고 공통 결정계수와 특수 결정계수의 상대적 크기가 다양한 방식으로 나타난다. 셋째, 전체적인 이변량 공간적 자기상관의 수준뿐만 아니라 이변량 공간적 자기상관의 조합 양상에 따라 상관관계 요소분해의 결과는 달라진다. 본 연구는 공간데이터분석의 연구 관행에 새로운 제안을 하고자 하는데, 이변량 혹은 다변량 공간통계분석의 경우, 피어슨 상관계수와 그것의 유의성 검증 결과뿐만 아니라 피어슨 상관계수 요소분해 결과도 함께 병기한다면 공간적 자기상관이 상관계수의 팽창/위축에 어떠한 방식과 강도로 작동하는지에 대한 새로운 통찰력을 제공할 수 있을 것으로 기대된다.

주요어 : 피어슨 상관계수, 이변량 공간적 자기상관, 고유벡터공간필터링, 상관계수 요소분해

Abstract : The main objective of this paper is to propose a refined version of the Pearson's correlation coefficient decomposition technique and to examine how much contribution the technique can make to our understanding about what happens to the bivariate correlation when spatial autocorrelation is present. The technique employs sequential steps which are the spatial pattern decomposition, the sub-correlation coefficients calculation, and the determination coefficients decomposition, and it finally divides the Pearson's correlation coefficient into four correlation components, *the residual-residual correlation*, *the common-common correlation*, *the unique-residual correlation*, and *the residual-unique correlation* components. The applicability and practicality of the technique is assessed on a hypothetical data set composed of 8 pairs which are identi-

이 논문은 2016학년도 서울대학교 '인문·사회계열 학문전공교수 해외연수 지원 사업'의 후원을 받아 수행된 연구 결과의 일부분임. 이 논문의 일부 내용은 2017년 4월 미국 매사추세츠주 보스턴시(Boston, Massachusetts)에서 개최된 미국지리학회 연례학술대회(Annual Meeting of the American Association of Geographers)에서 발표되었음.

* 서울대학교 지리교육과 교수(Professor, Department of Geography Education, Seoul National University, si_lee@snu.ac.kr)

cal in terms of the Pearson's correlation coefficient but are different in terms of the level of bivariate spatial autocorrelation. Main findings from the experimental study are as follows. First, individual variables involved in the 8 pairs are diverse in terms of presence/absence of particular spatial pattern components. Second, the relative size and proportion of the sub-correlation coefficients and the sub-determination coefficients turn out to be substantially influenced by the presence/absence of particular spatial pattern components and the relative strength of spatial autocorrelation of two variables. Third, the correlation decomposition results are significantly subject to the relative strength of univariate spatial autocorrelation of each variable as well as the overall level of bivariate spatial autocorrelation. In conclusion, this paper proposes a new research practice that encourages researchers conducting a bivariate or multivariate spatial statistical analysis to report some of the results from the Pearson's correlation coefficient decomposition analysis in addition to Pearson's correlation coefficients and their associated p -values, which may lead to a new insight into our understanding about how spatial autocorrelation is involved in the process of inflating/deflating correlation coefficients.

Key Words : Pearson's correlation coefficient, bivariate spatial autocorrelation, eigenvector spatial filtering, correlation coefficients decomposition

1. 서론

피어슨 상관계수(Pearson's correlation coefficient)는 이변량 연관성을 측정하고 그것의 통계적 유의성을 검토하기 위한 지배적인 통계량의 지위를 누려 왔다. 그러나 다른 대부분의 표준 통계 기법과 마찬가지로 이 통계 기법 역시 공간 데이터의 본질적 특성을 반영하지 못하는 취약성을 보유하고 있다(Haining, 1991). 연구 데이터에 이변량 공간적 의존성(spatial dependence) 혹은 공간적 자기상관(spatial autocorrelation)이 존재하는 경우, 피어슨 상관계수는 다음 두 가지의 중대한 문제점을 갖게 된다. 첫째, 두 공간적 패턴 간의 연관성을 측정하는 도구로서 피어슨 상관계수가 갖는 가치가 심대하게 훼손된다. 즉, 피어슨 상관계수는 두 변수 간의 공변동(co-variation)의 정도는 측정할 수 있지만 두 공간적 패턴 간의 공형성(共型性, co-patterning)의 정도는 측정하지 못한다(Lee, 2001; 이상일 등, 2018). 둘째, 피어슨 상관계수가 상정하고 있는 '독립관측가정(independent observations assumption)'이 심대하게 훼손된다(Bivand, 1980; Griffith, 1980; Haining, 1980; Richardson and Hémon, 1981). 독립관측가정의 훼손은 유효표본크기(effective sample size)의 삭감 혹은 자유도(degree of freedom)의 하락으로

이어지고, 결국 표준 가설검정에 기반한 통계학적 결론의 오류(제1종 오류) 가능성이 증대되는 결과가 초래된다(Lee, 2017; 이상일 등, 2018).

이러한 피어슨 상관계수의 문제점을 해결하기 위한 노력은 세 갈래로 나뉘어 진행되어 온 것으로 보이는데(이상일 등, 2018), 각각은 수정 t -검정(Clifford and Richardson, 1985; Clifford *et al.*, 1989; Dutilleul, 1993; Griffith and Paelinck, 2011), 공간필터 상관계수(Getis and Griffith, 2002; Griffith 2003; 2010; 2017; Griffith and Chun, 2014), 그리고 이변량 공간적 자기상관 통계량(Wartenberg, 1985; Griffith, 1993; Reich *et al.*, 1994; Lee, 2001; 2004, 2009; 2017; 이상일, 2007; 2008)이다. 이상일 등의 연구(2018)는 다소 독립적으로 발전해 온 세 기법이 상당한 정도의 상호 일관성 있는 결과를 보여준다는 점을 실험 연구를 통해 밝힌 바 있다. 즉, 이변량 공간적 자기상관이 높은 패턴쌍일수록 낮은 공간필터 상관계수, 작은 유효표본크기(자유도), 높은 유의확률을 보여준다는 것이다. 특히 고유벡터공간필터링(eigenvector spatial filtering, ESF) 접근에 기반한 공간필터 상관계수가 이변량 공간적 자기상관 통계량으로서의 L^* 와 가장 높은 상호 조응성을 보여주었다.

이러한 결과는 ESF 방법론이 피어슨 상관계수를 공간화하는데 있어 상당한 잠재력을 보유하고 있음

을 시사하는 것이다. ESF 기법은 다니엘 그리피스(Daniel A. Griffith) 교수의 선구적인 연구(1996; 2000; 2003)에 기반하고 있는데, ‘조정된 공간근접성행렬(adjusted spatial proximity matrix)’로부터 도출된 공간적 고유벡터(spatial eigenvectors)를 창의적으로 활용함으로써 다양한 공간통계학의 난제를 해결하고자 하는 총체적인 접근법이다. 특히 공간적 회귀분석에서의 기여도는 매우 높는데, 표준 선형 회귀분석(OLS 회귀분석)의 구조와 추정 방식은 그대로 따르는 대신 공간적 고유벡터를 일종의 통제변수(control variables)로 투입함으로써 잔차에서의 공간적 자기상관의 문제를 해결하고자 한다(이상일 등, 2017). 이러한 ESF 접근법에 기반한 공간필터링 기법을 피어슨 상관계수에 적용한 것이 바로 ‘공간필터 상관계수’인데(이상일 등, 2018), 이를 좀 더 심화시킨 것이 바로 ‘피어슨 상관계수 요소분해(Pearson’s correlation coefficient decomposition)’ 기법이다(Griffith, 2010; Griffith and Paelinck, 2011; Chun and Griffith, 2013).

본 연구는 ESF에 기반한 피어슨 상관계수 요소분해 기법을 정련화하고, 이 기법이 공간적 자기상관이 전제된 상태에서의 이변량 상관관계 분석에 어떠한 공헌을 할 수 있을 지를 실험 연구를 통해 검토해 보고자 한다. 이를 통해 ‘진정한’ 혹은 ‘순수한’ 상관계수가 공간적 자기상관의 존재로 인해 어떠한 방식과 어느 정도의 크기로 팽창 혹은 위축되어 피어슨 상관계수에 이르게 되는지에 대한 체계적인 이해가 가능하게 될 것으로 기대된다.

2. 연구 방법 및 절차

1) 피어슨 상관계수 요소분해 기법의 원리

ESF 기법은 표준 선형회귀분석의 기본 구조와 모수 추정(parameter estimation) 방식은 고수하면서도 조정된 공간근접성행렬로부터 추출된 공간적 고유벡터를 일종의 ‘인공(synthetic)’ 변수 혹은 ‘대

리(proxy)’ 변수로 투입함으로써 잔차에서의 공간적 자기상관의 문제를 해결하고자 한다(Griffith, 2003; 2010; 2017; Tiefelsdorf and Griffith, 2007; Griffith and Chun, 2014; Thayne, 2017; Griffith et al., 2019). 공간적 고유벡터는 “실제로 존재하는 변수가 아니라 만들어진 변수라는 점에서 ‘인공적’이며, 특정한 공간 패턴을 보유한 누락된 독립변수를 대신 한다는 의미에서 ‘대리적’이며, 공간적 효과를 배제한 순수한 독립변수의 설명력을 평가할 수 있게 해준다는 의미에서 ‘통제적’이다”(이상일 등, 2017). 피어슨 상관계수 요소분해 기법은 크게 세 단계로 나누어진다. 첫 번째 단계는 이변량 상황을 상정한 상태에서의 개별 변수의 공간적 패턴을 분해하는 ‘공간적 패턴 요소분해(spatial pattern decomposition)’ 단계이고, 두 번째 단계는 하위 상관계수와 하위 결정계수를 산출하는 단계이고, 세 번째 단계는 최종적으로 피어슨 상관계수를 4개의 구성요소로 분해하는 단계이다.

(1) 공간적 패턴 요소분해

ESF 기법에 기반한 공간적 패턴 요소분해는 상관관계 분석에 투입된 두 변수 각각에 대한 공간적 필터링(spatial filtering)의 적용으로부터 시작된다. 이것은 두 변수 각각을 종속변수에 두고 각각의 종속변수에 대해 선정된 고유벡터를 독립변수로 하는 ESF 모형을 구성함으로써 이루어진다(이상일 등, 2018)

$$\mathbf{x} = \mu_x \mathbf{1} + \mathbf{E}_x \boldsymbol{\beta}_x + \boldsymbol{\eta}_x \quad (1)$$

$$\mathbf{y} = \mu_y \mathbf{1} + \mathbf{E}_y \boldsymbol{\beta}_y + \boldsymbol{\eta}_y \quad (2)$$

여기서 $\mathbf{1}$ 은 요소가 모두 1인 열벡터(column vector), \mathbf{E}_x 와 \mathbf{E}_y 는 변수 x 와 변수 y 각각에 대해 선정된 고유벡터의 매트릭스, $\boldsymbol{\beta}_x$ 와 $\boldsymbol{\beta}_y$ 는 고유벡터와 결부된 회귀계수 벡터, $\boldsymbol{\eta}_x$ 와 $\boldsymbol{\eta}_y$ 는 공간적으로 독립적인 잔차 벡터이다. 두 잔차 벡터 간의 피어슨 상관계수를 공간적 효과가 배제된 두 변수간의 진정한 상관계수인 것으로 규정하며(Griffith, 2010), 이를 ESF에 기반한 ‘공간필터 상관계수(spatially filtered correlation coefficient)’로 개념화할 수 있다(이상일 등, 2018).

그런데 \mathbf{E}_x 와 \mathbf{E}_y 를 비교해보면, 어떤 고유벡터는 두

변수 모두에 대해 선정되어 있을 수 있고, 또 어떤 고유벡터는 특정한 변수 하나에 대해서만 선정되어 있을 수 있다. 이 때 전자를 ‘공통(common) 고유벡터’ 후자를 ‘특수(unique) 고유벡터’라 부를 수 있고, 이를 이용하면 식(1)과 식(2)는 다음과 같이 변형된다(Griffith, 2010).

$$\mathbf{x} = \mu_x \mathbf{1} + \mathbf{E}_C \boldsymbol{\beta}_{C_x} + \mathbf{E}_{U_x} \boldsymbol{\beta}_{U_x} + \boldsymbol{\eta}_x \quad (3)$$

$$\mathbf{y} = \mu_y \mathbf{1} + \mathbf{E}_C \boldsymbol{\beta}_{C_y} + \mathbf{E}_{U_y} \boldsymbol{\beta}_{U_y} + \boldsymbol{\eta}_y \quad (4)$$

여기서 \mathbf{E}_C 는 공통 고유벡터의 매트릭스, $\boldsymbol{\beta}_{C_x}$ 와 $\boldsymbol{\beta}_{C_y}$ 는 공통 고유벡터를 변수 x 와 변수 y 각각에 적용했을 때의 회귀계수 벡터, \mathbf{E}_{U_x} 와 \mathbf{E}_{U_y} 는 변수 x 와 변수 y 각각에 대한 특수 고유벡터 매트릭스, $\boldsymbol{\beta}_{U_x}$ 와 $\boldsymbol{\beta}_{U_y}$ 는 특수 고유벡터를 변수 x 와 변수 y 각각에 적용했을 때의 회귀계수 벡터, $\boldsymbol{\eta}_x$ 와 $\boldsymbol{\eta}_y$ 공간적으로 독립적인 잔차벡터이다.

그런데 식(3)과 식(4)는 결국 개별 변수가 네 개의 벡터의 합으로 규정된다는 것을 나타내므로, 다음과 같이 보다 간단한 방식으로 정식화할 수 있다.

$$x \equiv M_x + C_x + U_x + E_x \quad (5)$$

$$y \equiv M_y + C_y + U_y + E_y \quad (6)$$

이러한 정식화는 이변량 상황을 상정한 상태에서 해당 개별 변수에 대한 ‘공간적 패턴 요소분해(spatial pattern decomposition)’를 가능하게 해준다. 즉, 상관관계 분석에 투입된 개별 변수의 공간적 패턴은 ‘평균 패턴 요소(M)’, ‘공통 패턴 요소(C)’, ‘특수 패턴 요소(U)’, ‘잔차 패턴 요소(E)’로 분해될 수 있다. 이 네 패턴 요소들은 식 (3)과 식(4)에 주어져 있는 것처럼, 공통 고유벡터와 특수 고유벡터 모두를 독립변수로 하고 개별 변수를 종속변수로 하여 OLS 회귀분석을 실행한 후, 그 결과를 바탕으로 재구성하면 얻을 수 있다. 즉, 평균 패턴 요소(M)는 회귀식의 절편값 자체이며, 잔차 패턴 요소(E)는 회귀분석 결과 도출되는 잔차와 동일하다. 공통 패턴 요소(C)는 공통 고유벡터 각각에 해당 회귀계수를 곱하고 그것을 모든 공통 고유벡터에 대해 합산함으로써 구해지고, 특수 패

턴 요소(U)도 개별 특수 고유벡터와 해당 회귀계수를 곱하고 그것을 모든 고유벡터에 대해 합산함으로써 구해진다. 보다 기술적으로 설명하면, 공통 패턴 요소(C)는 공통 고유벡터만을 가지고 각 변수에 대해 회귀분석을 실행했을 때의 예측값에서 평균값을 뺀 것이고, 특수 패턴 요소(U)는 특수 고유벡터만을 가지고 각 변수에 대해 회귀분석을 실행했을 때의 예측값에서 평균값을 뺀 것이다. 이러한 방식이 가능한 이유는 모든 고유벡터들이 서로 직교하여 상관성이 존재하지 않기 때문이다(Griffith, 2003; 이상일 등, 2017).

(2) 하위 상관계수의 산출과 결정계수의 요소분해

그리피스 교수와 그의 동료가 제안한 정식화에 따르면, 식(5)와 식(6)에 나타나 있는 패턴 요소들 간의 상관관계를 고려할 때, 서로 다른 다섯 개의 하위 피어슨 상관계수가 산출된다(Griffith, 2010; Griffith and Paelinck, 2011; Chun and Griffith, 2013). 다섯 개의 하위 상관계수는 두 잔차 요소 간 상관계수(E_x 와 E_y 간의 피어슨 상관계수, r_{ee}), 공통의 공간적 자기상관으로부터 야기된 상관계수(C_x 와 C_y 간의 피어슨 상관계수, r_{cc}), x 변수의 특수한 공간적 자기상관과 y 변수의 잔차 간의 상관계수(U_x 와 E_y 간의 피어슨 상관계수, r_{ue}), x 변수의 잔차와 y 변수의 특수한 공간적 자기상관 간의 상관계수(E_x 와 U_y 간의 피어슨 상관계수, r_{eu}), 그리고 특수한 공간적 자기상관 간의 상관계수(U_x 와 U_y 간의 피어슨 상관계수, r_{uu})이다.

그런데 이러한 개념화는 다소간의 혼동을 야기할 우려가 있다고 보아 본 논문은 일종의 개념적 정련화를 시도하고자 한다. 우선 그리피스 교수의 5개 하위 상관계수 외에 4개의 부가적인 하위 상관관계가 더 가능하다는 점을 강조하고자 한다. 즉, C_x 와 U_y 간의 상관계수(r_{cu}), U_x 와 C_y 간의 상관계수(r_{uc}), C_x 와 E_y 간의 상관계수(r_{ce}), E_x 와 C_y 간의 상관계수(r_{ec})가 그것들이다. 그런데 공간적 고유벡터의 속성과 OLS의 원리상, 이 네 개의 부가적 상관계수와 그리피스 교수의 다섯 번째 상관계수의 값은 항상 0이다. 그러므로 하위 상관계수는 모두 네 가지인 것으로 정식화하는 것이 보다 합당하며, 각각의 하위 상관계수를 ‘잔차-잔

차 상관계수(r_{ee}), ‘공통-공통 상관계수(r_{cc})’, ‘특수-잔차 상관계수(r_{ue})’, ‘잔차-특수 상관계수(r_{eu})’로 명명하고자 한다. 이 때 첫 번째의 ‘잔차-잔차 상관계수’가 바로 공간필터 상관계수를 의미함을 인식하는 것이 중요하다.

한편, 식(3)과 식(4)에 제시되어 있는 회귀 모형으로부터 두 개의 공간적 패턴 요소가 함께 혹은 개별적으로 원변수의 변량을 얼마나 설명하는지에 대한 정보, 즉 전체 결정계수 및 그것을 구성하고 있는 하위 결정계수 값을 산출할 수 있다. 우선 식 (3)으로부터, 공통 패턴 요소(C)와 특수 패턴 요소(U) 모두를 투입한 회귀분석을 실행했을 때의 결정계수 값을 구할 수 있고 이를 ‘전체 결정계수(R^2)’라 부르하고자 한다. 이 결정계수는 다시 공통 패턴 요소(C)(보다 정확하게는, 공통 고유벡터)만 투입한 회귀분석을 실행했을 때의 결정계수인 ‘공통 결정계수(R_{cc}^2)’와 특수 패턴 요소(U)(보다 정확하게는 특수 고유벡터)만 투입한 회귀분석을 실행했을 때의 결정계수인 ‘특수 결정계수(R_{uu}^2)’로 분해된다. 이 때 공통 결정계수와 특수 결정계수의 합이 정확히 전체 결정계수와 일치한다는 점을 인식할 필요가 있다. 이것이 가능한 것은, 앞에서 언급한 것처럼, 모든 공간적 고유벡터는 서로서로 직교하여 상관성이 없기 때문이다. 동일한 방식을 식(4)에 적용하면 또 다른 변수에 대한 전체 결정계수(R_y^2), 공통 결정계수(R_{yc}^2), 특수 결정계수(R_{yu}^2)를 얻을 수 있다.

이러한 결정계수의 요소분해가 필수적인 것은, 최종적인 피어슨 상관계수 요소분해를 위해서는 앞에서 산출된 네 개의 하위 상관계수에 대해 그것들 각각이 관여하고 있는 변량 비중에 의거해 일종의 정규화를 적용해야 하기 때문이다. 즉, 네 개의 하위 상관계수(r_{ee} , r_{cc} , r_{ue} , r_{eu})의 합이 전체 피어슨 상관계수 값인 것도 아니며, 하위 상관계수의 크기가 상대적인 중요성을 대변하고 있지도 않다. 예를 들어, 잔차-잔차 상관계수(r_{ee})에는 잔차 변동만 개입되어 있고, 공통-공통 상관계수(r_{cc})에는 공통 요소에 의해 설명된 변동만 개입되어 있다. 따라서 잔차 변동과 공통 요소에 의해 설명된 변동의 상대적 크기를 일종의 가중치로 사용하여 각각의 하위 상관계수 값을 정규화할 필요

가 있는 것이다. 이렇게 하여 정규화된 하위 상관계수의 합은 정확히 전체 상관계수와 같아지고, 각 요소의 상대적인 중요성도 논의될 수 있는 토대가 제공될 수 있다.

(3) 피어슨 상관계수 요소분해

하위 상관계수에 해당 결정계수로부터 산출된 가중값을 곱하여 일종의 하위 상관계수 정규화를 수행하면 최종적인 피어슨 상관계수 요소분해 기법이 정식화 된다. 식(7)은 개념적으로 가능한 9개의 하위 상관계수 모두를 동원해 표현한 것인데, 그리피스 교수는 9개의 상관계수 중 순서적으로 앞서 있는 5개의 하위 상관계수를 통해 자신의 기법을 제시하였다 (Griffith, 2010; Griffith and Paelinck, 2011; Chun and Griffith, 2013). 본 연구는 개념적으로 늘 0일수 밖에 없는 상관계수를 모두 배제하는 것이 훨씬 더 간명한 정식화에 도달할 수 있다는 점에서 식(8)에 나타나 있는 것처럼 4개의 하위 상관계수만으로 표현된 수식을 사용하고자 한다.

$$\begin{aligned}
 r &= r_{ee}\sqrt{(1-R_x^2)(1-R_y^2)} + r_{cc}\sqrt{R_{xc}^2R_{yc}^2} + r_{ue}\sqrt{R_{xu}^2(1-R_y^2)} \\
 &+ r_{eu}\sqrt{(1-R_x^2)R_{yu}^2} + r_{uu}\sqrt{R_{xu}^2R_{yu}^2} + r_{cu}\sqrt{R_{xc}^2R_{yu}^2} \\
 &+ r_{uc}\sqrt{R_{xu}^2R_{yc}^2} + r_{cc}\sqrt{R_{xc}^2(1-R_y^2)} + r_{ec}\sqrt{(1-R_x^2)R_{yc}^2} \quad (7) \\
 r &= r_{ee}\sqrt{(1-R_x^2)(1-R_y^2)} + r_{cc}\sqrt{R_{xc}^2R_{yc}^2} + r_{ue}\sqrt{R_{xu}^2(1-R_y^2)} \\
 &+ r_{eu}\sqrt{(1-R_x^2)R_{yu}^2} \quad (8)
 \end{aligned}$$

여기서 각 상관계수의 가중값은 해당 요소와 관련되어 있는 변동 비중의 곱, 즉, 해당 결정계수의 곱에 의해 산출된다. 예를 들어, 잔차-잔차 상관계수(r_{ee})는 두 개의 잔차 변동과 결부되어 있고, 각 잔차의 변동 비중은 각각 $(1-R_x^2)$ 와 $(1-R_y^2)$ 로 주어지기 때문에, 두 변동 비중을 곱하고 거기에 제곱근을 적용하면 잔차 간 상관계수의 가중값을 얻을 수 있다.

본 연구는 최종적인 정련 과정으로서, 식(8)을 보다 간명하게 표현한 식(9)를 최종적인 피어슨 상관계수 요소분해 기법으로 제안하고자 한다. 이에 의거하면 피어슨 상관계수는 네 개의 하위 ‘상관관계 요소’

로 나뉘어 진다.

$$r = EE + CC + UE + EU \quad (9)$$

즉, 피어슨 상관계수는 ‘잔차-잔차 상관관계 요소(EE)’, 두 변수의 공통적인 공간적 자기상관과 관련되어 있는 상관관계 요소, 즉 ‘공통-공통 상관관계 요소(CC)’, 그리고 개별 변수의 특수한 공간적 자기상관과 관련되어 있는 두 개의 상관관계 요소, 즉 x 변수의 공간적 자기상관과 관련되어 있는 ‘특수-잔차 상관관계 요소(UE)’와 y 변수의 공간적 자기상관과 관련되어 있는 ‘잔차-특수 상관관계 요소(EU)’로 나뉘어 진다. 결국 앞에서 정의된 네 개의 하위 상관계수가 정규화를 통해 네 개의 상관관계 요소로 바뀐 것이다.

이 네 요소의 합은 정확히 피어슨 상관계수와 같기 때문에, 각 요소를 전체 피어슨 상관계수에 대한 비중 값(각 요소를 피어슨 상관계수로 나누어 구한 %)으로 환산하면 각 요소가 어느 방향의, 어느 정도 크기의 영향력을 행사했는지에 대한 정량적 정보를 획득할 수 있다. 우선, 잔차-잔차 상관관계 요소(EE)의 상대적 비중은 공간적 자기상관의 작동으로 인해 순수한 상관계수가 얼마나 팽창(inflation) 혹은 수축(deflation)되었는지에 대한 정략적 지표를 제공할 것이다. 또한 공간적 자기상관과 관련되어 있는 세 요소의 상대적 비중은 각각의 요소가 전체적인 팽창 혹은 수축 과정에 어느 방향으로, 어느 정도의 크기로 가담했는지에 대한 정보를 제공할 것이다. 전체적으로 상관관계 팽창이 발생한 경우의 예를 들자면, 세 요소가 서로 다른 정도로 모두 팽창에 가담할 수도 있고, 어떤 요소는 팽창에, 다른 요소는 수축에 가담할 수도 있다. 이 피어슨 상관계수 요소분해 기법은 단순히 피어슨 상관계수의 전체적인 팽창 혹은 위축의 정도만 보여주는 것이 아니라(전체 피어슨 상관계수와 공간 필터 상관계수의 차 혹은 비를 통해 파악 가능), 그 팽창과 위축이 어떠한 방식으로 발생하는지에 대한 세부적인 정보를 제공한다는 점에서 이변량 공간적 자기상관 연구에 새로운 통찰력을 제시해 줄 것으로 기대된다.

2) 연구 디자인 및 분석 절차

분석을 위한 실험 디자인은 기본적으로 이전 연구(이상일 등, 2018)와 동일하네, 정리하면 다음과 같다. 37개의 육각형으로 구성되어 있는 가상의 연구 지역에 평균과 분산은 동일하지만 공간적 자기상관의 정도는 매우 상이한 두 패턴을 만들어 낸다(두 패턴의 모런 통계값은 0.681과 -0.186)이다(이상일 등, 2018, 그림 1 참조). 두 패턴 간의 피어슨 상관계수는 0.422로 유의미한 양의 상관관계를 보여준다. 이 데이터에 무작위 재배열을 적용하여 서로 다른 이변량 공간적 자기상관을 나타내는, 혹은 서로 다른 공형성 수준을 보여주는 8개의 패턴쌍을 도출한다. 이 패턴쌍의 이변량 공간적 자기상관 통계값(L^* 로 측정)은 0.000, 0.050, 0.100, 0.150, 0.200, 0.250, 0.302, 0.371인데, PAIR1~PAIR4는 이변량 공간적 자기상관을 보여주지 않지만, PAIR5~PAIR8은 통계적으로 유의미한 이변량 공간적 자기상관을 보여준다. 패턴쌍을 구성하고 있는 16개 변수는 다양한 수준의 이변량 공간적 자기상관을 보여준다. 모런 통계량을 기준으로 할 때, PAIR1~PAIR4를 구성하고 있는 어느 변수도 통계적으로 유의미한 공간적 자기상관을 보여주지 않는다. PAIR5는 y 변수만 통계적으로 유의미한 공간적 자기상관을 보여주지만(각각 0.019와 0.253), PAIR6는 두 변수 모두 상당한 정도의 공간적 자기상관을 보여주고 있다(각각 0.143과 0.240). PAIR7는 x 변수는 극단적으로 높은 공간적 자기상관을 보여주지만(0.461), y 변수는 통계적으로 유의미한 공간적 자기상관을 보여주지 않는다(0.079). PAIR8은 두 변수 모두 극단적으로 높은 공간적 자기상관을 보여준다(각각 0.407과 0.359)(이상일 등, 2018, 표 1 참조).

분석 절차는 다음과 같다. 첫째, 식(5)와 식(6)에 나타나 있는 정식화를 적용해 변수들에 대한 공간적 패턴 요소분해를 실행한다. 사용된 공간근접성행렬(spatial proximity matrix)은 연결성(contiguity)-기반의 이항 행렬로 연결한 공간단위 쌍에 대해서는 1의 값을, 그렇지 않은 공간단위 쌍에 대해서는 0의 값을 갖는다. 고유벡터의 추출을 위해 모런 통계량이 사용되며, 고유벡터의 선정은 최대 공간적 자기

상관의 최소한 25% 수준을 보여주는 고유벡터를 선정 후, 단계적 회귀분석을 실행하여 최종적인 고유벡터를 추출하는 방식에 의거해 이루어진다(Griffith, 2010; Chun and Griffith, 2013; Chun *et al.*, 2016; Griffith, 2017). 둘째, 산출된 공간적 패턴 요소를 이용하여 하위 상관계수를 산출하고, 결정계수의 요소분해를 실행한다. 셋째, 산출된 하위 상관계수와 결정계수를 이용하여 식(9)에 의거한 피어슨 상관계수 요소분해를 실행한다. 각 단계별로 결과에 대한 세세한 해석을 제공함으로써 본 연구가 제안하고 있는 분석 기법의 의의를 부각시키고자 한다.

3. 연구 결과

1) 공간적 패턴 요소분해 결과

표 1은 공간적 자기상관의 정도가 서로 다른 8개의 가상의 패턴쌍에 대해 식(5)와 식(6)을 적용한 결과를 보여주고 있다. 즉, 각 패턴쌍을 구성하고 있는 두 변수를 4개의 서로 다른 구성요소로 분해한 결과를 보여주고 있는 것이다. 평균 패턴 요소(M)는 각 변수의 전체 평균 값을 개별 육각형에 할당한 것으로 패턴 내

표 1. 8개의 패턴쌍과 공간적 패턴 요소분해 결과

| 패턴쌍 | 원변수 | 공간적 패턴 요소분해 | | | | |
|--------------------------|-----|-------------|----------|----------|----------|--|
| | | 평균 패턴 요소 | 공통 패턴 요소 | 특수 패턴 요소 | 잔차 패턴 요소 | |
| PAIR1 ($L^*=0,000$) | X | | | - | - | |
| | Y | | | - | - | |
| PAIR2 ($L^*=0,050$) | X | | | - | - | |
| | Y | | | - | - | |
| PAIR3 ($L^*=0,100$) | X | | | | | |
| | Y | | | | | |

| | | | | | | |
|--------------------------|---|--|--|--|---|--|
| PAIR4 ($L^*=0.150$) | X | | | | - | |
| | Y | | | | - | |
| PAIR5 ($L^*=0.200$) | X | | | | | |
| | Y | | | | | |
| PAIR6 ($L^*=0.250$) | X | | | | | |
| | Y | | | | - | |
| PAIR7 ($L^*=0.302$) | X | | | | | |
| | Y | | | | - | |
| PAIR8 ($L^*=0.371$) | X | | | | | |
| | Y | | | | - | |

(음영의 차이는 각 패턴 내에서의 상대적인 값의 높낮이만을 나타냄.)

에서 공간적 변동은 존재하지 않는다. 또한 표 1에 나타나 있는 모든 패턴쌍의 모든 변수가 동일한 평균값을 갖고 있기 때문에 평균 패턴 요소(M)는 16개 변수 모두에 대해 동일하다. 공통 패턴 요소(C)는 각 패턴쌍을 구성하고 있는 두 변수가 공통적으로 보유하고 있는 공간적 자기상관 패턴과 관련되어 있다. 실질적으로는 공통 고유벡터만을 가지고 각 변수에 대해 회귀분석을 실행했을 때의 예측값에서 평균값을 뺀 것이다. 특수 패턴 요소(U)는 각 패턴쌍을 구성하고 있는 두 변수 각각이 개별적으로 보유하고 있는 공간적 자기상관 패턴과 관련되어 있다. 실질적으로는 특수 고유벡터만을 가지고 각 변수에 대해 회귀분석을 실행했을 때의 예측값에서 평균값을 뺀 것이다. 마지막으로 잔차 패턴 요소(E)는 개별 원변수에서 공간적 자기상관과 관련된 패턴을 제거한, 즉 공간적으로 독립적인 변수이다. 실질적으로는 공통 고유벡터와 특수 고유벡터 모두를 독립변수로 회귀분석했을 때의 잔차와 동일하다.

표 1에 나타나 있는 개별 패턴쌍의 패턴 요소분해 결과를 좀 더 세밀하게 살펴볼 필요가 있다. 우선 PAIR1과 PAIR2는 나머지 패턴쌍과는 다른 양상을 보여주고 있는데, 네 개의 모든 패턴에서 공통 패턴 요소와 특수 패턴 요소가 존재하지 않는다. 이것은 두 변수에는 공간적으로 구조화된 부분이 존재하지 않으며 따라서 원변수와 잔차 패턴 요소가 본질적으로 동일하다는 점을 보여주고 있다. 그러므로 이 변수 쌍에서 피어슨 상관계수와 그것에 대한 유의성 검정 결과는 비공간적인 변수에 대한 것과 동일한 방식으로 해석되어도 무방하다. PAIR3~PAIR8는 PAIR1과 PAIR2와는 달리 모든 변수들이 공통 패턴 요소와 특수 패턴 요소 중 최소한 하나에 대해서는 하위 패턴 요소를 보유하고 있다. PAIR3~PAIR8을 구성하고 있는 12개 변수 모두는 공통 패턴 요소를 가지고 있고, 그 중 7개 변수는 특수 패턴 요소도 함께 가지고 있다. 예를 들어 PAIR3을 구성하고 있는 두 변수는 거의 동일한 패턴을 보여주고 있는 공통 패턴 요소와 매우 상이한 패턴을 보여주고 있는 특수 패턴 요소를 각각 보유하고 있다. 이에 비해 PAIR4를 구성하고 있는 두 변수는 공통 패턴 요소만을 보유하고 있으며,

PAIR6~PAIR8의 경우는 한 변수는 두 패턴 요소를 모두 다 가지고 있지만 다른 변수는 공통 패턴 요소만을 가지고 있다. 여기서 중요한 것은 특정한 패턴쌍이 공통 패턴 요소만을 가지면 공통 패턴 요소와 특수 패턴 요소 모두를 가지면 간에 두 잔차 요소간의 피어슨 상관계수를 의미하는 공간필터 상관계수의 값은 각 변수에 내재된 공간적 자기상관으로 인해 그 값이 평균 혹은 축소되어 최종적인 원변수간 피어슨 상관계수에 이르게 된다는 점이다. 이것은 뒤에서 보다 자세히 다루도록 한다.

2) 하위 상관계수의 산출과 결정계수의 요소분해 결과

표 2와 표 3은 식(8)에 나타나 있는 수식을 계산할 수 있게 해주는 하위 상관계수와 결정계수들에 관한 것이다. 표 2는 네 개의 서로 다른 하위 상관계수를 보여주고 있는데, 앞에서 ‘잔차-잔차 상관계수(r_{ee})’, ‘공통-공통 상관계수(r_{cc})’, ‘특수-잔차 상관계수(r_{ue})’, ‘잔차-특수 상관계수(r_{eu})’로 명명한 바 있다. 표 3은 각 상관계수의 가중값을 계산하기 위해 필요한 하위 결정계수(R^2)값을 나타내고 있다. 각 변수별로 3개씩, 모두 6개의 하위 결정계수 값이 계산되는데, 앞에서 ‘전체 결정계수(R_{cc}^2 와 R_{ee}^2)’, ‘공통 결정계수(R_{cc}^2 와 R_{ee}^2)’, ‘특수 결정계수(R_{uu}^2 와 R_{uu}^2)’라 명명한 바 있다.

표 2가 보여주는 사항 중 가장 중요한 것은, 이변량 공간적 자기상관의 정도가 높아질수록 진정한 상관관계를 의미하는 잔차-잔차 상관계수의 값이 체계적으로 낮아진다는 점이다(PAIR6과 PAIR7의 순서가 바뀐 것을 제외). 이는 공간필터 상관계수에 대한 이전 연구에서 이미 밝혀진 사항이다(이상일 등, 2018, 표 2 참조). 표 2를 좀 더 자세히 살펴보면, 공간적 패턴 요소의 존재/부존재의 양상과 개별 변수의 공간적 자기상관의 정도에 따라 하위 상관계수 값이 매우 다양하게 나타난다는 점을 알 수 있다. PAIR1과 PAIR2는 표 1에 나타나 있는 것처럼 두 변수 모두 공통 패턴 요소와 특수 패턴 요소를 보유하고 있지 않은데, 그 두 패턴 요소와 결부되어 있는 세 개의 하위 상관계수 값이 모두 0으로 나타나 있는 것을 확인할 수 있다.

표 2. 하위 상관계수의 산출

| 패턴쌍 | 피어슨 상관계수 (r) | 하위 상관계수 | | | |
|-----------------------|------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | | 잔차-잔차 상관계수 (r_{ee}) | 공통-공통 상관계수 (r_{cc}) | 특수-잔차 상관계수 (r_{ue}) | 잔차-특수 상관계수 (r_{eu}) |
| PAIR1 ($L^*=0.000$) | 0.4219 | 0.4219 | 0 | 0 | 0 |
| PAIR2 ($L^*=0.050$) | | 0.4219 | 0 | 0 | 0 |
| PAIR3 ($L^*=0.100$) | | 0.3567 | 0.9990 | -0.0769 | 0.0439 |
| PAIR4 ($L^*=0.150$) | | 0.3028 | 1.0000 | 0 | 0 |
| PAIR5 ($L^*=0.200$) | | 0.1898 | 1.0000 | 0.1507 | 0.3127 |
| PAIR6 ($L^*=0.250$) | | 0.1197 | 0.9851 | 0.0718 | 0 |
| PAIR7 ($L^*=0.302$) | | 0.1277 | 1.0000 | 0.3100 | 0 |
| PAIR8 ($L^*=0.371$) | | 0.0110 | 1.0000 | 0.0416 | 0 |

표 3. 결정계수의 요소분해

| 패턴쌍 | 변수 X | | | 변수 Y | | |
|-----------------------|---------------------|------------------------|------------------------|---------------------|------------------------|------------------------|
| | 전체 결정계수 (R_c^2) | 공통 결정계수 (R_{cc}^2) | 특수 결정계수 (R_{uu}^2) | 전체 결정계수 (R_c^2) | 공통 결정계수 (R_{cc}^2) | 특수 결정계수 (R_{uu}^2) |
| PAIR1 ($L^*=0.000$) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| PAIR2 ($L^*=0.050$) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| PAIR3 ($L^*=0.100$) | 0.3075 (3) | 0.2210 (2) | 0.0864 (1) | 0.2565 (3) | 0.1352 (2) | 0.1213 (1) |
| PAIR4 ($L^*=0.150$) | 0.1349 (1) | 0.1349 (1) | 0 (0) | 0.2240 (1) | 0.2240 (1) | 0 (0) |
| PAIR5 ($L^*=0.200$) | 0.2442 (2) | 0.1287 (1) | 0.1155 (1) | 0.4363 (3) | 0.0702 (1) | 0.3661 (2) |
| PAIR6 ($L^*=0.250$) | 0.3379 (3) | 0.2649 (2) | 0.0730 (1) | 0.4338 (2) | 0.4338 (2) | 0 (0) |
| PAIR7 ($L^*=0.302$) | 0.6354 (5) | 0.1714 (1) | 0.4639 (4) | 0.1393 (1) | 0.1393 (1) | 0 (0) |
| PAIR8 ($L^*=0.371$) | 0.4937 (3) | 0.3760 (1) | 0.1177 (2) | 0.4365 (1) | 0.4365 (1) | 0 (0) |

(결정계수값 아래의 괄호 안의 숫자는 선정된 고유벡터의 개수임)

두 변수 모두가 특수 패턴 요소를 보유하고 있지 않은 PAIR4의 경우는 특수 패턴 요소와 결부되어 있는 두 하위 상관계수 값이 모두 0으로 나타나 있고, 두 변수 모두 공통 패턴 요소와 특수 패턴 요소 모두를 가지고 있는 PAIR3과 PAIR5의 경우는 모든 하위 상관계수가 0이 아닌 값이 나타나 있음을 확인할 수 있다. 한 변수는 공통 패턴 요소와 특수 패턴 요소 모두를 가지고 있고 다른 한 변수는 공통 패턴 요소만을 가지고 있는 PAIR6~PAIR8의 경우는 누락된 특수 패턴 요소는 결부되어 있는 하위 상관계수 값이 0으로 나타나 있음을 확인할 수 있다.

또한 표 2는 하위 상관계수 값의 상대적 크기가 패턴쌍을 구성하고 있는 개별 변수의 일변량 공간적 자기상관과 특정한 방식으로 관련되어 있음을 시사하고 있다. 예를 들어, PAIR5의 경우 특수-잔차 상관계수에 비해 잔차-특수 상관계수의 값이 배 이상 크다. 이것은 PAIR5를 구성하고 있는 x 변수에는 공간적 자기상관이 존재하지 않지만 y 변수에는 통계적으로 유의미한 공간적 자기상관이 존재하기 때문일 것으로 보인다(이상일 등, 2018, 표 1 참조). 더 나아가 공통-공통 상관계수(r_{cc})는 공통 요소가 존재하는 한 거의 1에 가깝다는 점 또한 주목할 필요가 있다. 이는 공통 고유벡터로 각 변수를 회귀분석 했을 때 산출되는 예측값이 두 변수 사이에서 매우 유사하게 나타난다는 것을 함축하고 있다. 더 나아가 공간필터 상관계수의 팽창/위축 과정에 관여하는 세 개의 하위 상관계수 중 가장 큰 영향력을 행사할 가능성이 높다는 점을 함축하는 것이기도 하다. 그러나 실질적인 하위 상관계수의 영향력은 하위 상관계수의 절대값이 아니라 표 3에 나타나 있는 하위 결정계수를 고려한 이후에야 정확히 말할 수 있다는 점을 명확히 인식할 필요가 있다.

표 3에서 일차적으로 중요한 사항은 공통 결정계수와 특수 결정계수의 합은 정확히 전체 결정계수와 같다는 것이며, 이러한 점에서 결정계수 요소분해라는 명칭을 사용하고자 한다. 다음으로 중요한 사항은, 공간적 패턴 요소의 존재/부존재의 양상이 결정계수 값이 0인지 아닌지를 기본적으로 결정한다는 점이다. 즉, 각 패턴쌍에서 누락된 패턴 요소의 종류와 개수에

따라 그와 결부된 결정계수 값이 0으로 나타나는 것이다. 예를 들어 두 변수 모두에서 특수 패턴 요소를 가지지 않는 PAIR4의 경우는 특수 패턴 요소와 관련된 두 결정계수 값이 0으로 나타나 있다. 그런데 보다 더 중요한 사항은 각 변수별로 공통 결정계수와 특수 결정계수의 상대적 크기가 해당 변수 내에서의 패턴 요소 별 상대적 중요성을 나타내는 지표로 해석할 수 있다는 점이다. 예를 들어, PAIR5의 x 변수의 경우는 공통 패턴 요소(0.1287)가 특수 패턴 요소(0.1155)보다 상대적 중요성에서 약간 더 크지만, y 변수의 경우는 특수 패턴 요소(0.3661)가 공통 패턴 요소(0.0702)보다 상대적 중요성에서 훨씬 더 크다. 이러한 차이는 각 패턴쌍을 구성하고 있는 변수들의 상대적인 공간적 자기상관의 크기와 관련되어 있는 것으로 보인다. 앞서서도 언급한 것처럼 PAIR5의 경우 x 변수에는 공간적 자기상관이 존재하지 않지만 y 변수에는 통계적으로 유의미한 공간적 자기상관이 존재한다. 이러한 점은 PAIR7에서도 동일하게 드러나는데, x 변수의 특수 결정계수의 값이 매우 크게 나타나고 y 변수의 경우 0인 것은, 전자는 통계적으로 유의미한 공간적 자기상관을 보이지만 후자는 그렇지 않기 때문인 것으로 보인다.

3) 피어슨 상관계수 요소분해 결과

표 4는 식(9)에 나타나 있는 정식화에 의거해 표 2와 표 3에 나타나 있는 값을 이용해 최종적인 피어슨 상관계수 요소분해를 실행한 결과이다. 예를 들어 PAIR8의 요소 분해는 식(8)과 식(9)에 의거한 다음의 계산을 통해 이루어진 것이다.

$$0.4219 = 0.0110 \times \sqrt{(1-0.4937)(1-0.4365)} + 1 \times \sqrt{0.3760 \times 0.4365 + 0.0416 \times \sqrt{0.1177 \times (1-0.4365)}} + 0 \times \sqrt{(1-0.4937) \times 0} = 0.0059 + 0.4051 + 0.0109 + 0 \quad (10)$$

표 4는 각 상관관계 요소가 상관계수의 팽창과 위축 어느 쪽에, 어느 정도의 강도로 가담하는지에 대한 간명한 요약치를 제공한다. 양의 상관관계의 경우, 상관

관계 요소가 양수면 팽창에, 음수면 위축에 가담한 것이고, 강도는 절대값에 비례한다. 음의 상관관계인 경우는 정반대이다. 이러한 방식의 해석은 그리피스 교수가 제시하고 있는 기존의 방식(Chun and Griffith, 2013, 117-118), 즉 식(10)과 같은 변환식에서 각 요소의 팽창/위축 가담 여부와 강도를 하위 상관계수를 기준으로 설명하는 방식에 비해 좀 더 간명하다는 장점이 있다.

표 4에 나타나 있는 상관관계 요소들 중 가장 중요한 것은 잔차-잔차 상관관계 요소(EE)이다. 우선 이 값은 상관계수가 아님을 다시 한번 명확히 할 필요가 있다. 순수한 상관계수를 의미하는 공간필터 상관계수는 표 2에 나타나 있는 잔차-잔차 상관계수이다. 표 4의 잔차-잔차 상관관계 요소(EE)는 잔차-잔차 상관계수를 표 3에 나타나 있는 결정계수 값을 이용하여 정규화한 것이다. 식(10)에서 볼 수 있는 것처럼 PAIR8에 대한 잔차-잔차 상관계수는 0.0110이지만 변동 비중을 고려한 가중치를 곱함으로써 0.0059

라는 잔차-잔차 상관관계 요소(EE) 값이 산출된 것이다. 따라서 잔차-잔차 상관계수 값은 잔차-잔차 상관관계 요소(EE) 값보다 항상 크거나 같다. 그러나 두 값의 차이에도 불구하고, 이변량 공간적 자기상관의 정도가 높아질수록 잔차-잔차 상관계수 값이 체계적으로 작아지는 경향성은 잔차-잔차 상관관계 요소(EE) 값에도 그대로 나타난다는 점은 매우 중요한 사실이다(표 2에 나타나 있는 PAIR6와 PAIR7의 순위 역전도 사라졌음). 잔차-잔차 상관관계 요소(EE)의 절대값뿐만 아니라 그것의 비중값도 매우 중요한 통찰력을 제공해 주는 것으로 판단된다. 이 비중값은 개별 상관관계 요소값을 피어선 상관계수로 나누어 백분율(%)을 구한 것인데, 요소 별 상대적 중요성을 판단하는 가장 중요한 지표로 활용될 수 있을 것으로 보인다. 잔차-잔차 상관관계 요소(EE)의 경우 절대값뿐만 아니라 비중값도 이변량 공간적 자기상관의 정도가 높아질수록 낮아지는 경향성을 뚜렷이 보여주고 있다.

표 4. 피어슨 상관계수 요소분해 결과

| 패턴쌍 | 피어슨 상관계수 (r) | 상관관계 요소 | | | |
|--------------------------|------------------|-------------------|------------------|-------------------|------------------|
| | | 잔차-잔차 (EE) | 공통-공통 (CC) | 특수-잔차 (UE) | 잔차-특수 (EU) |
| PAIR1 ($L^*=0.000$) | 0.4219 | 0.4219 (100.0) | 0 | 0 | 0 |
| PAIR2 ($L^*=0.050$) | | 0.4219 (100.0) | 0 | 0 | 0 |
| PAIR3 ($L^*=0.100$) | | 0.2560 (60.7) | 0.1727 (40.9) | -0.0195 (-4.6) | 0.0127 (3.0) |
| PAIR4 ($L^*=0.150$) | | 0.2481 (58.8) | 0.1738 (41.2) | 0 | 0 |
| PAIR5 ($L^*=0.200$) | | 0.1239 (29.4) | 0.0950 (22.5) | 0.0384 (9.1) | 0.1645 (39.0) |
| PAIR6 ($L^*=0.250$) | | 0.0733 (17.4) | 0.3340 (79.2) | 0.0146 (3.5) | 0 |
| PAIR7 ($L^*=0.302$) | | 0.0715 (16.9) | 0.1545 (36.6) | 0.1958 (46.4) | 0 |
| PAIR8 ($L^*=0.371$) | | 0.0059 (1.4) | 0.4051 (96.0) | 0.0109 (2.6) | 0 |

(하위 상관계수값 아래의 괄호 안의 값은 전체 피어슨 상관계수에서 개별 하위 상관관계 요소가 차지하는 백분율(%)임.)

표 4에 나타나 있는 것을 이용해 각 패턴쌍을 좀 더 자세히 분석하면 다음과 같다. PAIR1과 PAIR2는 피어슨 상관계수와 잔차-잔차 상관관계 요소(EE)가 동일하다. 즉, 이 두 패턴쌍에는 공간적 자기상관이 개입되지 않았으므로 피어슨 상관계수가 두 패턴의 진정한 상관성을 대변한다고 말할 수 있다. PAIR3의 경우는 매우 흥미로운 결과에 해당하는데, 순수한 상관관계를 대변하고 있는 잔차-잔차 상관관계 요소(EE)는 피어슨 상관계수가 표현하는 것의 60.7% 정도 수준인데, 이것에 공통-공통 상관관계 요소(CC)에 의한 41.2%의 팽창, y 변수의 특수한 공간적 자기상관과 결부되어 있는 잔차-특수 상관관계 요소(EU)에 의한 3.0%의 팽창, x 변수의 특수한 공간적 자기상관과 결부되어 있는 특수-잔차 상관관계 요소(UE)에 의한 4.6%의 위축이 결합되어 최종적인 피어슨 상관계수에 이르게 되었다는 것을 보여주고 있는 것이다. 특히 위축의 요소가 포함되어 있는 유일한 패턴쌍이라는 점에서 독특한 예라고 할 수 있다. PAIR4의 경우는 PAIR3의 경우에 비해 훨씬 단순한 구조를 보여주고 있다. 순수한 상관관계의 정도는 피어슨 상관계수가 표현하는 것의 58.8% 수준이고 나머지 41.2%는 공통의 공간적 자기상관에 의해 팽창된 것이다. PAIR3과 PAIR4의 예는 공통점과 차이점을 동시에 보여주고 있는데, 공통점은 두 경우 모두 피어슨 상관계수가 보여주는 상관성의 강도 중 대략 40% 정도는 공간적 자기상관에 의해 부풀려졌다는 점이고, 차이점은 그러한 팽창을 야기한 원인의 구성이나 상대적 비중이 서로 다르다는 점이다.

PAIR5의 경우도 상당히 흥미로운 사례이다. 우선 PAIR1~PAIR4의 경우 이변량 공간적 자기상관이 통계적 유의성을 보이지 않지만 PAIR5~PAIR8은 통계적 유의성을 보인다는 점을 상기할 필요가 있다. 내부적으로 보면 x 변수는 공간적 자기상관이 없지만 y 변수는 통계적으로 유의한 양의 공간적 자기상관을 보여주고 있다. 이러한 특성이 상관관계 요소분해 결과에 그대로 드러나 있다. 우선 순수한 상관관계의 정도는 피어슨 상관계수가 표현하는 것의 29.4% 정도 수준에 불과하고 나머지 70.6%는 공간적 자기상관에 의해 팽창된 것이다. 팽창의 구성을 자세히 살펴보면,

공통-공통 상관관계 요소(CC)가 22.5%, 특수-잔차 상관관계 요소(UE)가 9.1%, 잔차-특수 상관관계 요소(EU)가 39.0%를 차지하고 있다. y 변수의 특수한 공간적 자기상관과 결부되어 있는 마지막 요소의 비중이 가장 큰 것에서도 알 수 있듯이 패턴쌍에서 한 변수에게만 공간적 자기상관이 존재하는 경우 그 변수가 상관관계 팽창 과정에서 주도적인 역할을 한다는 점을 명확히 보여주고 있다.

PAIR6와 PAIR7은 유사한 수준의 이변량 공간적 자기상관을 보여주고 있고, 순수한 상관관계가 피어슨 상관계수에서 차지하는 비중이 각각 17.4%와 16.9%로 역시 유사하지만, 내부 구조에서는 매우 큰 차이를 보여주고 있다. PAIR6의 경우는 공통의 공간적 자기상관에 의한 팽창 비중이 79.2%로 x 변수의 특수한 공간적 자기상관에 기인한 팽창 비중인 3.5%를 압도하지만, PAIR7의 경우는 두 요소의 팽창 비중이 각각 36.6%와 46.4%로 엇비슷하다. 이는 각 패턴쌍을 구성하고 있는 두 변수의 공간적 자기상관 정도가 다르기 때문에 발생한 것인데, PAIR6의 경우는 두 변수 모두가 상당한 정도의 양의 공간적 자기상관을 보이고 있는 반면, PAIR7의 경우는 x 변수는 극단적으로 높은 공간적 자기상관을 보여주고 있지만, y 변수는 통계적으로 유의미한 공간적 자기상관을 보여주고 있지 않다. 이것은 유사한 수준의 이변량 공간적 자기상관을 보이는 패턴쌍의 경우라도 패턴쌍을 구성하고 있는 변수들의 일변량 공간적 자기상관의 정도에 따라 상관관계 요소분해의 결과가 달라진다는 점을 명확히 보여주고 있는 것이다.

마지막으로 PAIR8의 예는 극단적으로 높은 이변량 공간적 자기상관을 보이는 패턴쌍의 경우 피어슨 상관계수가 얼마나 무의미한 통계량인가를 극적으로 보여주고 있다. PAIR8의 경우 만일 공간적 자기상관을 제거한다면 두 변수 간의 진정한 상관관계의 정도는 피어슨 상관계수가 제시하는 수준의 1.4%에 불과하다. 더 나아가 공통의 공간적 자기상관에 기인한 팽창 비중이 96.0%로 극단적으로 높다. 이는 PAIR6의 결과와 일치하는 것으로 두 변수 모두 높은 수준의 양의 공간적 자기상관을 보여주는 경우, 공통의 공간적 자기상관이 기여하는 비중이 개별 변수의 공간적 자

기상관이 기여하는 비중에 비해 월등히 높게 나타난다는 점을 분명히 보여주고 있는 것이다.

4. 요약 및 결론

본 연구는 기본적으로 ESF 접근법에 기반한 피어슨 상관계수 요소분해 기법을 정련화하고자 했다. 우선 공간적 패턴 요소분해를 통해 개별 변수의 공간적 패턴을 ‘평균 패턴 요소(M)’, ‘공통 패턴 요소(C)’, ‘특수 패턴 요소(U)’, ‘잔차 패턴 요소(E)’로 분해한다. 이 패턴 요소들의 변수 간 교차를 통해 네 개의 하위 상관계수, 즉 ‘잔차-잔차 상관계수(r_{ee})’, ‘공통-공통 상관계수(r_{cc})’, ‘특수-잔차 상관계수(r_{ue})’, ‘잔차-특수 상관계수(r_{eu})’를 산출한다. 각 변수별로 공통 패턴 요소와 특수 패턴 요소의 상대적 중요성을 회귀분석의 맥락에서 평가하기 위해, ‘전체 결정계수’와 그것의 하위 결정계수인 ‘공통 결정계수’와 ‘특수 결정계수’를 산출한다. 이 하위 결정계수값을 이용해 하위 상관계수의 정규화를 수행하면 최종적인 피어슨 상관계수 요소분해 기법이 정식화되는데, 피어슨 상관계수는 네 가지 상관관계 요소, 즉 ‘잔차-잔차 상관관계 요소(EE)’, ‘공통-공통 상관관계 요소(CC)’, ‘특수-잔차 상관관계 요소(UE)’, ‘잔차-특수 상관관계 요소(EU)’로 분해된다.

이렇게 정련화된 피어슨 상관계수 요소분해 기법을 동일한 피어슨 상관계수 값을 갖지만 서로 다른 수준의 이변량 공간적 자기상관을 보이는 가상의 8개 패턴쌍에 적용함으로써 해당 기법의 유용성을 검토하였다. 실험 연구를 통해 밝혀진 주요 내용을 정리하면 다음과 같다. 첫째, 8개 패턴쌍의 16개의 변수에 대해 공간 패턴 요소분해를 적용한 결과 공통 패턴 요소와 특수 패턴 요소의 존재/부존재의 양상이 매우 다양하게 나타났다. 공간적 자기상관이 매우 낮은 변수의 경우는 공간적 자기상관과 관련된 두 패턴이 모두 나타나지 않은 경향이 있고, 공간적 자기상관이 어느 정도 수준 이상이 되면 구성 변수의 특성에 따라 존재/비존재의 양상이 다양하게 나타난다. 둘째, 공

간적 패턴 요소의 존재/부존재의 양상과 개별 변수의 일변량 공간적 자기상관의 정도에 따라, 하위 상관계수 값의 상대적 크기, 그리고 공통 결정계수와 특수 결정계수의 상대적 크기가 다양한 방식으로 나타난다. 셋째, 전체적인 이변량 공간적 자기상관의 수준뿐만 아니라 일변량 공간적 자기상관의 조합 양상에 따라 상관관계 요소 분해의 결과는 달라진다. 패턴쌍에서 한 변수에게만 공간적 자기상관이 존재하는 경우 그 변수가 상관관계 평가 과정에서 주도적인 역할을 하며, 두 변수 모두 높은 수준의 공간적 자기상관을 보유한 경우, 공통의 공간적 자기상관이 기여하는 비중이 개별 변수의 공간적 자기상관이 기여하는 비중에 비해 월등히 높게 나타난다.

본 연구는 공간데이터분석의 연구 관행에 새로운 제안을 하고자 한다. 이변량 혹은 다변량 공간통계분석의 경우, 변수들간의 상관성에 대한 기술적, 탐색적 분석 결과를 제시할 때 단순히 피어슨 상관계수와 그것의 유의성 검정 결과만을 제시할 것이 아니라 표 4에 나타나 있는 결과를 병기하고 간결한 서술을 첨가할 것을 제안하고자 한다. 최소한 잔차-잔차 상관관계 요소의 값 및 비중을 제시해 주면, 독자들은 공간적 자기상관이 통제되었을 때 어느 정도 수준의 상관성을 보이는지, 그리고 공간적 자기상관으로 인해 순수한 상관관계가 어느 정도 부풀려졌는지를 금방 확인할 수 있게 된다. 더 나아가 표 4에 나타나 있는 전체 정보가 제공된다면 상관관계 분석에 투입된 개별 변수의 공간적 자기상관의 상대적 강도에 대한 정보와 상대적 영향력에 대한 보다 심도 깊은 통찰력도 함께 얻을 수 있을 것이다.

본 연구의 중요한 한계점 중의 하나는 기본적으로 가상의 패턴쌍에 대해 방법론을 적용했다는 것이다. 따라서 연구의 다음 단계는 완전한 시뮬레이션 프레임워크 하에서 이변량 공간적 자기상관 통제량과 잔차-잔차 상관관계 요소의 관련성을 정식화하는 것이다. 공간적 랜덤화(spatial randomization) 과정을 통해 동일한 피어슨 상관계수를 보이지만 이변량 공간적 자기상관에서는 서로 다른 수준을 보이는 수많은 패턴쌍을 도출하고, 그 패턴쌍에 본 연구가 제안하고 있는 피어슨 상관관계 요소분해 기법을 적용하면 정

식화를 위한 기본적인 데이터를 생산할 수 있게 될 것이다. 그 데이터를 바탕으로 특정한 L^* 값을 보이는 패턴상의 경우 어느 정도 규모의 상관계수에서의 팽창/위축이 발생하는지에 대한 회귀식을 획득할 수 있을 것이다. 그 회귀식은 공간적 자기상관과 이변량 상관관계의 관련성에 대한 논의(Lee, 2017)에서 새로운 통찰력을 제공해 줄 수 있을 것으로 기대된다.

참고문헌

- 이상일, 2007, “거주지 분화에 대한 공간통계학적 접근 (I): 공간 분리성 측도의 개발,” *대한지리학회지*, 42(4), 616-631.
- 이상일, 2008, “거주지 분화에 대한 공간통계학적 접근 (II): 국지적 공간 분리성 측도를 이용한 탐색적 공간데이터 분석,” *대한지리학회지*, 43(1), 134-153.
- 이상일·조대현·이민과, 2017, “공간적 자기상관 통계량의 고유벡터 간 비교 연구: 공간근접성행렬의 효과와 공간적 회귀분석에의 함의를 중심으로,” *대한지리학회지*, 52(5), 645-660.
- 이상일·조대현·이민과, 2018, “피어슨 상관계수의 공간화: 세 관련 기법 간의 비교 실험 연구,” *대한지리학회지*, 53(5), 761-776.
- Bivand, R., 1980, A Monte Carlo study of correlation coefficient estimation with spatially correlate observations, *Quaestiones Geographical*, 6, 5-10.
- Chun, Y. and Griffith, D. A., 2013, *Spatial Statistics & Geostatistics*, SAGE, Los Angeles.
- Chun, Y., Griffith, D. A., Lee, M., and Sinha, P., 2016, Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters, *Journal of Geographical Systems*, 18(1), 67-85.
- Clifford, P. and Richardson, S., 1985, Testing the association between two spatial processes. *Statistics and Decisions*, 2 (Supplementary issue), 155-160.
- Clifford, P., Richardson, S., and Hémon, D., 1989. Assessing the significance of the correlation between two spatial processes, *Biometrics*, 45(1), 123-134.
- Dutilleul, P., 1993, Modifying the t test for assessing the correlation between two spatial processes, *Biometrics*, 49(1), 305-314.
- Getis, A. and Griffith, D. A., 2002, Comparative spatial filtering in regression analysis, *Geographical Analysis*, 34(2), 130-140.
- Griffith, D. A., 1980, Towards a theory of spatial statistics, *Geographical Analysis*, 12(4), 325-339.
- Griffith, D. A., 1993, Which spatial statistics techniques should be converted to GIS functions? in Fischer, M. and Nijkamp, P. *New Directions in Spatial Econometrics*, Springer, Berlin, 101-114.
- Griffith, D. A., 1996, Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data, *The Canadian Geographer*, 40(4), 351-367.
- Griffith, D. A., 2000, A linear regression solution to the spatial autocorrelation problem, *Journal of Geographical Systems*, 2(2), 141-156.
- Griffith, D. A., 2003, *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*, Springer, Berlin.
- Griffith, D. A., 2010, Spatial filtering, in Fischer, M. M. and Getis, A. (eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, Springer, Berlin, 301-318.
- Griffith, D. A., 2017, Spatial filtering, in Shekhar, S., Xiona, H., and Zhou, X. (eds.), *Encyclopedia of GIS*, Volume 3, 2nd edition, Springer, Berlin, 2018-2031.
- Griffith, D. A. and Chun, Y., 2014, Spatial autocorrelation and spatial filtering, in Fischer, M. M. and Nijkamp, P. (eds.), *Handbook of Regional Science*, Springer, Berlin, 1477-1507.
- Griffith, D. A., Chun, Y., and Li, B., 2019, *Spatial Regression Analysis Using Eigenvector Spatial Filtering*, Cambridge, MA: Academic Press.
- Griffith, D. A. and Paelinck, J. H. P., 2011, *Non-standard Spatial Statistics and Spatial Econometrics*, Springer, Berlin.
- Haining, R. P., 1980, Spatial autocorrelation problems, in Herbert, D. T. and Johnston, R. J. (eds.), *Geography and the Urban Environment*, Wiley, New York,

- 1-44.
- Haining, R. P., 1991, Bivariate correlation with spatial data, *Geographical Analysis*, 23(3), 210-227.
- Lee, S.-I., 2001, Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I , *Journal of Geographical Systems*, 3(4), 369-385.
- Lee, S.-I., 2004, A generalized significance testing method for global measures of spatial association: an extension of the Mantel test, *Environment and Planning A*, 36(9), 1687-1703.
- Lee, S.-I., 2009, A generalized randomization approach to local measures of spatial association, *Geographical Analysis*, 41(2), 221-248.
- Lee, S.-I., 2017, Correlation and spatial autocorrelation, in Shekhar, S., Xiona, H., and Zhou, X. (eds.), *Encyclopedia of GIS*, Volume 1, 2nd edition, Springer, New York, 360-368.
- Reich, R. M., Czaplewski, R. L., and Bechtold, W. A., 1994, Spatial cross-correlation of undisturbed, natural shortleaf pine stands in northern Georgia, *Environmental and Ecological Statistics*, 1(3), 201-217.
- Richardson, S. and Hémon, D., 1981, On the variance of the sample correlation between two independent lattice processes, *Journal of Applied Probability*, 18(4), 943-948.
- Thayn, J. B., 2017, Eigenvector spatial filtering and spatial autocorrelation, in Shekhar, S., Xiona, H., and Zhou, X. (eds.), *Encyclopedia of GIS*, Volume 1, 2nd edition, Springer, New York, 511-522.
- Tiefelsdorf, M. and Griffith, D. A., 2007, Semiparametric filtering of spatial autocorrelation: The eigenvector approach, *Environment and Planning A*, 39(5), 1193-1221.
- Wartenberg, D., 1985, Multivariate spatial correlation: a method for exploratory geographical analysis. *Geographical Analysis*, 17(4), 263-283.
- 교신: 이상일, 08826, 서울특별시 관악구 관악로 1, 서울대학교 사범대학 지리교육과 (이메일: si_lee@snu.ac.kr, 전화: 02-880-9028)
- Correspondence: Sang-Il Lee, Department of Geography Education, College of Education, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea (e-mail: si_lee@snu.ac.kr, phone: +82-2-880-9028)
- 최초투고일 2019. 10. 14
수정일 2019. 10. 28
최종접수일 2019. 10. 29

지리가중회귀 모델을 이용한 학교급별 학업성취도 영향 요인 분석

김지우* · 이진학**

Analysis of Influential Factors on the Educational Achievement by School Units using the Geographically Weighted Regression Model (GWR)

Jiwoo Kim* · Gunhak Lee**

요약 : 학업성취도는 학습자의 학습 역량 뿐 아니라 국가의 교육 역량을 종합적으로 가늠할 수 있는 중요 지표이다. 이러한 학업성취도는 개인적 특성 뿐 아니라 학교, 지역 등 다양한 요인이 종합적으로 영향을 미친다. 그럼에도 불구하고 기존의 연구들은 대체로 개인적 역량이나 가정환경과 같은 일부 요인에 집중하고 있으며, 학교나 학교가 위치한 지역적 특성이 미치는 효과에 대해서는 크게 주목하지 않고 있다. 이를 고려한 연구라 하더라도 미시적 변이를 관찰하기에 제한적인 공간 단위에 기반하고 있어 지역에 따라 이질적인 학업성취도 영향 요인에 대한 미시적 특성을 밝혀주지 못하고 있다. 이러한 맥락에서 본 연구는 지리가중회귀 모델을 사용하여 학업성취도에 영향을 줄 수 있는 교육 환경, 지역적 특성 요인의 공간적 이질성을 학교급 단위의 미시적 관점에서 살펴보고자 한다. 분석 결과, 일반적인 회귀 모델보다 공간적 이질성을 고려한 국지적 지리가중회귀 모델이 지역 및 학교 등급별 학업성취도를 보다 잘 설명하고 있으며, 고등학교를 대상으로 한 모델이 중학교 모델에 비해 높은 설명력을 보여주고 있다. 또한 영향 요인과 관련하여 학업성취도는 지역에 따라 상이한 독립변수에 의해 설명될 수 있으며, 학교급 단위의 미시적 접근이 보다 유의미함을 알 수 있었다. 본 연구는 국가 수준의 학업성취도에 영향을 미치는 국지적인 변이 요인들을 보다 미시적으로 확인하고, 개별 학교 단위가 미치는 영향에 대한 의미 있는 통찰을 제공할 수 있다는데 의의가 있다.

주요어 : 학업성취도, 지리가중회귀 모델, 공간적 의존성, 공간적 이질성, 영향 요인

Abstract : The educational achievement is an important index for the comprehensive evaluation of the national level of educational performance as well as the personal capability of learners. The educational achievement can be broadly influenced by various factors such as personal characteristics, institution, and regional characteristics. However, existing studies have mostly focused on the limited factors as personal learning ability and family environments of students. They have paid little attention to the effect of regional characteristics that schools are located or school properties. Even studies considering those factors are often limited because their analyses are typically based on the macro-scale units which are not suited to small variations. Therefore, they cannot reveal the spatial variation of influential factors on the educational achievement which is heterogeneous over regions. In this context, this study seeks to explore the spatial heterogeneity of influential factors on the educational achievement, such as educational environment and regional characteristics based on the school units at the micro-level using the geographically weighted regression model (GWR). As a result of empirical analysis, the local GWR considering the spatial heterogeneity explains the educational achieve-

* 서울대학교 지리학과 박사과정(Doctor student, Department of Geography, Seoul National University, kwangtane@snu.ac.kr)

** 교신저자, 서울대학교 지리학과 부교수 및 국토문제연구소 겸무연구원(Associate Professor, Department of Geography and Institute for Korean Regional Studies, Seoul National University, gunhlee@snu.ac.kr)

ment better than the general OLS model, and the high school focused model shows higher performance than the middle school focused model. Also, regarding influential factors, different sets of significant independent variables are identified over space, and therefore a micro approach based on the school units could be useful. This study allows us to identify local variable factors influencing the educational achievement at the micro-level and provides a meaningful insight for the influence of individual school units.

Key Words : educational achievement, Geographically Weighted Regression(GWR), spatial dependency, spatial heterogeneity, influential factors

1. 연구 배경

학교는 한 국가의 교육 체계에 있어 핵심 기능을 수행하는 주체로 다양한 역할을 담당하고 있다. 학교의 역할에 대한 견해는 학자마다 조금씩 상이한데, Barrow(1979)는 학교의 기능을 학생 보호, 사회적 역할 선별, 제도 및 사상 전달, 지식 전달의 수행 주체로 판단하였고, 양미경(2004)은 학교의 주요 역할을 선별 기능, 교육적 기능으로 파악하였으며, 김정원(2007)은 교과 수업을 통한 지식 전달과 교육복지를 포함한 돌봄 기능을 학교의 주요 기능으로 판단하고 있다. 이를 종합해보면 학교는 사회 구성 기능으로서 다양한 역할을 수행하고 있지만, 공통적으로 지적하고 있는 부분은 지식 전달 기능으로 학교 존재 목적의 가장 큰 이유임을 알 수 있다(Young, 2010).

학교의 교육 기능이 중요한 이유는 학교 교육을 통한 학생의 성장이 최종적으로 사회 발전의 원동력으로 작용하기 때문이다(Feinberg, 2012). 즉, 학교 교육의 경쟁력은 결국 국가 경쟁력 강화를 위한 초석으로 이어지기 때문에 지속적으로 관심을 기울일 대상이며, 제대로 기능을 수행하고 있는지 꾸준히 관리 및 감독을 해야 할 대상이기도 하다. 이러한 점에서 교육 당국은 학교 교육의 내실을 쌓기 위해 학교 교육의 질을 점검하고 개선하고자 하는 노력을 책임지고 수행하여야 한다. 교육 시스템을 점검하고 개선하기 위한 방안으로는 여러 가지가 있겠지만 가장 직관적이고 확실한 방법은 시험을 통해 학생들의 학업 수준을 정량적으로 측정하고 이를 통해 기존의 교육 방법이 학생 역량 신장에 올바르게 작동하고 있는지 평가하는 것이다. 이러한 취지에서 교육 당국은 학교 교육의 질

을 체계적으로 점검하고 개별 학교 및 학생의 기초학력을 점검하기 위해 국가수준 학업성취도 평가를 도입하였다. 국가수준 학업성취도 평가는 1998년 기본 계획이 수립된 이후 1999년 첫 시행을 거쳐 현재까지 시행되고 있는 전국 단위 평가로, 평가 결과를 활용하여 학교 현장에서 기초학력에 도달하지 못한 학생들의 지도에 도움을 줄 수 있으며 적절한 학력 향상 지원 방안을 위한 기초자료로 활용할 수 있다. 특히, 국가수준 학업성취도 평가는 학교 교육의 말미에 이루어지는 대학수학능력시험과 달리 중학교 3학년과 고등학교 2학년에 실시하기 때문에 학업 진행에 대한 중간 점검 평가 기능이 강하며, 이에 대한 결과를 통해 미비점을 확인하고 보완을 통한 향후 개선이 가능하여 그 중요도가 상당하다.

국가수준 학업성취도 평가의 높은 중요성처럼 이와 관련된 연구 또한 많이 진행되고 있으며, 특히 학업성취도에 영향을 주는 요인에 대한 연구가 활발하게 이루어졌다. 이는 국가적 측면에서는 학업성취도 향상을 통해 국가 경쟁력을 키워야 하기 때문이며, 개인적 측면에서는 사회적 지위 향상을 위한 수단으로 학업성취도 향상이 중요하기 때문이다(임다희 등, 2012). 하지만 기존의 연구들은 학업성취도에 영향을 주는 요인을 관별함에 있어 지역적 환경 특성이나 차이를 제대로 규명하지 못하고 있으며, 이를 고려한 연구라 하더라도 시군구처럼 미시적 변이를 관찰하기에 제한적인 공간 단위를 기반으로 학업성취도에 대한 요인 분석을 진행한 경우가 대부분이다. 따라서 이러한 기존 연구들은 지역에 따라 이질적인 학업성취도 영향 요인에 대한 미시적 특성을 제대로 밝혀줄 수 없는 한계를 가지고 있다.

이러한 맥락에서 본 연구는 학업성취도에 영향을 줄 수 있는 교육 환경, 지역적 특성 등 다양한 요인들의 공간적 이질성(spatial heterogeneity)을 학교급 단위의 미시적 관점에서 살펴보고자 한다. 이를 위해 본 연구에서는 지리가중회귀(Geographically Weighted Regression: GWR) 모델을 사용하여 지역 및 학교 등급별 변이 요인을 도출할 것이다. 지리가중회귀 모델은 지역마다 변수들 간의 관계가 다를 때 지역별로 고유의 회귀식을 도출하여 지역 특유의 성격을 제시할 수 있다는 장점이 있다(Fotheringham *et al.*, 2002). 보다 구체적으로 서울에 위치한 382개의 중학교와 226개의 고등학교를 대상으로 학업성취도의 변이 요인을 살펴볼 것이며, 학교급간의 학업성취도 차이를 보다 체계적으로 살펴보기 위해 중학교와 고등학교를 분리하여 분석하고자 한다. 중학교의 경우 대부분의 학생이 집 근처에 있는 학교에 배정을 받으며, 공통 교육과정에 기반한 동일 교육과정으로 수업이 이루어지기 때문에 모든 중학교를 분석 대상에 포함하였다. 반면 고등학교의 경우 특수목적 고등학교를 제외한 일반 고등학교 및 자율형 고등학교만 분석 대상으로 선정하였다. 특수목적 고등학교는 서울 전역에서 학생을 모집하는데, 본래의 특수목적에 따른 설립 취지와 다르게 학군에 상관없이 학업 성적이 좋은 학생들이 선별적으로 들어가는 학교로 기능함에 따라 지역의 일반적인 학력이 아니라 서울 전역에서 선발된 최상위권 학생의 학력을 반영한다고 볼 수 있어 분석의 왜곡을 가져올 수 있다(류주현, 2006; 전은정 등, 2015).

2. 선행연구

학업성취도에 영향을 미치는 요인에 대한 연구는 개인적 측면에서부터 제도적 환경에 이르기까지 매우 다양한 접근들이 있어 왔다. 학생 개인 수준의 특성에 따른 학업성취도 차이를 보여주는 연구들로 정충대(2015)는 학급 동료의 수준에 따라 소속 학생의 학업성취도가 달라질 수 있고 특히 하위권 동료의 영

향력이 더 강하기 때문에 저학력 학생들의 기초학력 신장이 필요함을 언급하고 있다. 김혜숙(2011)은 성별에 따라 우리나라 초중고 학생들의 사회과 영역에서의 학업성취도가 다르다고 지적하며, 교과별로 다소 차이는 있지만 초중고 모두 여학생의 성적이 상대적으로 우수하며, 특히 기초학력미달 비율은 초중고 모두 남학생 비율이 높다고 지적하였다. 한편, 학교나 환경적 특성에 따른 학업성취도 차이에 관한 연구들도 있는데, 임다희 등(2012)은 가정환경, 학교환경, 학습동기가 학업성취도에 중요한 요인으로 작용한다고 밝혔으며, 김현철 등(2012)은 특히, 가정 배경이 학업성취도에 영향을 주는 주요 요소라고 주장하였다. 이와 유사한 관점에서, 이정우(2013)는 다문화가정 학생의 학업성취도가 일반 학생에 비해 낮음을 보여주고 있으며, 오정란(2007), 박찬선 등(2007)은 부모의 직업이 학업성취도에 영향을 미치며, 특히 전문직 중심의 고소득층이 많은 지역일수록 학업성취도가 높다고 주장하였다. 김감영 등(2014)은 초등학교 저학년의 순이동률이 학업성취도와 정적인 상관성을 보인다고 분석하였다. 구체적으로 순이동률과 학업성취도는 인과성을 보이는 것은 아니지만, 인구가동을 지역 간 교육격차가 발생하는 사회적 과정에 대한 결과로 파악하였으며, 이는 강태중(2007)의 논조와 유사하다. 한편, 학교 교사 특성과 관련하여 박수정 등(2011), 이광현 등(2012)은 학교장의 공모 방식이, 조현국(2013)은 기간제 교사의 비율이 학업성취도에 영향을 미침을 보여주고 있다. 보다 종합적인 관점에서 정혜원 등(2018)은 앞선 연구들에서 언급한 학생, 부모, 학급 및 학교와 관련된 대부분의 변수들이 학업성취도에 영향을 준다고 분석하며 학생마다 위치한 상황 및 환경이 다르기 때문에 학업성취도를 향상하기 위한 맞춤형 방안이 필요하다고 언급하고 있다.

이러한 학업성취도에 영향을 미치는 다양한 요인에 대한 연구뿐 아니라 학업성취도의 공간적 분포 특성과 지역적 격차에 관한 연구도 진행되었다. 성기선(2004)은 학군이 학업성취도에 유의미한 영향을 주는 것은 아니지만 학업성취도에 대한 지역적 차이는 분명히 나타나고 있음을 보여주었고, 추경모(2012) 역시 유사한 결론을 내리고 있다. 손준중(2004)은 강남

지역의 교육적 의미에 대한 연구를 통해 교육과 관련한 공간적 불평등이 학업성취도의 분포 양상을 결정하며, 이로 인해 지역마다 학업성취도가 상이하다고 주장하였다. 특히, 학업성취도와 큰 연관성이 있는 대학진학률이 대졸 이상 인구 비율, 전문직 인구 비율, 학원 비율 등과 높은 상관성을 보이는 것으로 나타났다. 류주현(2006)은 학업성취도, 학력 수준, 주거 비용이 높은 상관 관계를 보이고 있으며, 세 조건을 모두 충족하는 지역이 공간적으로 군집되어 나타나고 있음을 보여주었다. 이러한 공간적 연계성을 윤희호 등(2008)은 거주지 효과로 설명하고 있으며 이들 지역은 한국의 명문대 진학률에 유의미한 변수로 작용하는 것을 보여주었다. 주경식 등(2012)은 동일한 행정구역으로 묶인 도시입에도 불구하고 구시가지와 신시가지간의 교육 격차가 존재하며, 특히 사회·경제적 특성의 영향을 많이 받는 과목의 경우 그 편차가 큰 것으로 파악하였다. 이와 유사하게 최은영(2004) 역시 서울 내 자치구간에도 소득 수준과 같은 경제적 차이와 학업성취도에는 유의미한 상관성이 있음을 보여주었다.

이러한 기존 연구들은 우리나라 학생들의 학업성취도와 관련한 매우 다양한 요인들을 개인적 차원에서 서부터 제도 및 환경적 특성까지 포괄하고 있음을 알 수 있다. 그럼에도 불구하고 이러한 변수들을 보다 종합적으로 접근하는 노력은 상대적으로 적은 편이며 특히, 대부분 전역적 모델링에 그치고 있어 학업성취도와 변이 요인들간의 관계에 있어 지역적인 이질성에 대한 공간적 고찰은 전무한 실정이다. 학업성취도에 대한 다양한 영향 요인들이 지역적으로 어떤 상이한 패턴들을 보여주고 있는지, 지역적으로 유의미하게 다른 영향 요인들은 무엇인지 살펴보는 것은 향후 학업성취도 평가에 있어 국지적, 미시적 접근을 통한 새로운 직관과 함의를 제시해 줄 것으로 기대한다.

3. 연구 방법 및 데이터

1) 연구 방법

최소제곱법(Ordinary Least Square: OLS) 기반의 전통적인 선형 회귀분석은 종속변수에 영향을 주는 독립변수들간의 관계를 선형 회귀식을 통해 파악하는 것으로 전역적인 현상을 설명할 수 있는 변수별 부분 회귀계수가 일괄적으로 도출된다(수식 1). 하지만 특정 현상을 지역적 특성이 이질적인 하위 지역 단위로 설명하고자 한다면 기존의 OLS 기법은 하나의 대푯값만 제공하기 때문에 제한이 있다.

$$y = B_0 + \sum_{k=1}^n B_k x_k + \varepsilon \quad (\text{수식 1})$$

이에 본 연구에서는 OLS 기법의 제한점을 보완하기 위해 GWR 분석을 병행하고자 한다. 수식 2는 전형적인 GWR 모델로 y_i 는 i 지역의 종속변수값에 해당한다. (u_i, v_i) 는 i 지역의 공간적 좌표를, $B_0(u_i, v_i)$ 는 i 지역의 절편계수, $B_k(u_i, v_i)$ 는 i 지역의 k 번째 회귀계수, x_{ik} 는 i 지역에 적용되는 k 번째 변수를 의미한다. 만약 종속변수가 유의미한 공간 의존성(spatial dependence)을 보인다면, 전역적 모델(global model)을 통해 지역성을 무시하고 설명하기 보다 국지적 모델(local model)을 통해 작은 지역 단위로 그 변화를 탐지하는 것이 현상 이해에 도움이 될 것이다(Fotheringham *et al.*, 2002; Oyana and Margai, 2016). 따라서 GWR은 동일한 독립변수를 사용한다 하더라도 각 하위 지역별로 개별적인 모델을 도출함에 따라 변수의 설명력에 해당하는 부분 회귀계수값이 달라지며, 이를 통해 기존의 전역적 모델이 파악하기 어려운 공간의 이질성을 정량적으로 파악할 수 있게 한다.

$$y_i = B_0(u_i, v_i) + \sum_{k=1}^n B_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (\text{수식 2})$$

한편, GWR은 공간적 이질성뿐 아니라 공간 단위에 존재하는 공간적 의존성을 암묵적으로 고려하기 때문에 인접한 공간 단위와의 거리 관계에 따

라 상이한 가중치를 부여한다. 거리가 짧을수록 높은 가중치를 부여하는 역거리 가중(inverse distance weighting)을 통해 연산되며 이때 공간적 커널을 활용하게 된다. 커널의 유형은 연구 상황에 따라 고정 커널(fixed kernels)과 가변 커널(adaptive kernels)을 사용할 수 있는데, 점 데이터에 대한 분포가 일정하다면 고정 커널을 적용하며 데이터의 분포가 고르지 않은 형태라면 그 정도에 따라 커널의 폭을 각기 적용하여 분석을 진행할 수 있다. 본 연구에서는 학업성취도 데이터 분포가 비교적 불균등하다고 판단하여 가변 커널 방식을 적용하고자 한다. 연구에 사용된 모델의 구현과 검증은 R 3.5.2와 ArcGIS 9.3을 사용하였다.

2) 데이터

최근 국가기관 및 지방자치단체가 보유하고 있는 정보를 국민에게 공개하고자 하는 움직임이 확산되고 있으며(윤광석 등, 2010), 이러한 움직임은 공공기관에 대한 관리 감독을 강화하고 국민의 알 권리를 실현할 수 있기 때문에 여러 긍정적인 효과를 창출한다(김승태, 2006). 이는 교육 분야에도 동일하게 적용되고 있으며 교육 관련 기관의 정보공개에 관한 특례법을 근거로 교육정보 공시 제도를 실시하고 있다(표 1). 전국의 모든 초·중등학교는 교육정보 공시 제도를 통해 학교 전반의 정보를 공개하고 있으며 이를 통해 학교 교육의 경쟁력 강화를 추구하고 있다. 교육정보 공시 제도는 각 단위 학교의 각종 정보를 한국교육학술정보원이 수집하여 이를 관리 및 검증하는 구조로 운영된다. 특히 각종 정보에 대한 입력은 각 학교의 업무 담당 교사가 직접 실시하고, 업무 담당 부서의 부장 교사가 직접 확인하며, 해당 시·도 교육청이 이에 대해 상시 모니터링 및 교육정보 공시 관련 연수를 진행하고 있기 때문에 정보의 정확도가 높다고 할 수 있다(한국교육개발원, 2008). 이렇게 수집된 교육정보는 학교알리미(www.schoolinfo.go.kr)를 통해 열람 및 구득할 수 있으며 6개의 대분류, 26개의 중분류, 52개의 세부 항목 형태로 정보를 제공하고 있다(표 2).

표 1. 교육정보 공시 제도

| | |
|-------|--|
| 공시 기관 | 초·중등교육법에 따라 설립된 각급 학교, 그밖에 다른 법률에 따라 설치된 각급 학교 |
| 공시 대상 | 전국 초·중·고등학교 약 12,000개교 |
| 공시 횟수 | 공시항목별로 매년 1회 이상 |
| 공시 시기 | 수시 공시 및 정시 공시(4월, 5월, 9월, 11월) |

출처: 교육관련기관의 정보공개에 관한 특례법 시행령

표 2. 교육정보 공시 항목

| 대분류 | 중분류 |
|---------|--|
| 학생 현황 | 학생 재학 현황 진·출입 및 학업중단 학생 수 입학생 현황 졸업생 진로 및 장학금 수혜 현황 학생 체력 증진 |
| 교원 현황 | 직위별 교원 현황 자격종별 교원 현황 |
| 교육 활동 | 각종 규정 교육계획 및 편성·운영 학교운영위원회 동아리활동 및 방과후학교 등 |
| 교육 여건 | 학교시설 및 개방 현황 학교급식에 관한 사항 학교폭력대책 및 학생·학부모 상담실적 학교보건 및 환경위생 현황 학교도서관 직원 현황 학교 평가 교복 구매 유형 및 단가 자율 공시 |
| 예·결산 현황 | 학교회계 예·결산서 사립학교 예·결산서 학교발전기금 |
| 학업 성취도 | 평가기준 및 계획 학년별 교과별 성적사항 |

출처: 학교알리미 홈페이지

본 연구는 학교알리미를 통해 공시되고 있는 다양한 교육 정보를 학업성취도와 이를 설명하는 변수 구축에 활용하였다. 종속변수에 해당하는 학업성취도의 경우 이를 평가할 수 있는 여러 형태가 있을 수 있

지만, 본 연구에서는 각 학교 단위의 국가수준 학업성취도 평가 결과를 활용하였다. 국가수준 학업성취도 평가는 우수학력, 보통학력, 기초학력, 기초학력 미달 총 4개 구간으로 평가 등급을 제시하며 이를 학생에게 성적표를 통해 통지한다. 반면 대외적으로 공개되는 교육정보 공시 자료의 경우 각 학교별 등급 구간의 비율을 제시하는데 기초학력과 기초학력 미달 비율은 그대로 공시하지만 우수학력 비율과 보통학력 비율의 경우 이를 합쳐 보통학력 이상 비율이라는 새로운 계급 구간으로 재구성한 뒤 결과를 공시한다. 이에 본 연구는 국어, 수학, 영어의 보통학력 이상 비율의 합을 종속변수로 사용하여 각 학교별 학업성취도 수준을 판단하였다. 한편, 학업성취도 평가는 중학교 3학년과 고등학교 2학년을 대상으로 실시하기 때문에 해당 학년의 점수를 각 학교의 대푯값으로 선정하였다.

독립변수의 경우 학교알리미를 통해 구득할 수 있는 여러 자료 중 본 연구에 적용할 수 있는 변수들을 수합하여 다음과 같이 선정하였다(표 3). 특히 선행 연구에서 학업성취도를 설명하는데 사용하였던 변수

들을 종합적으로 반영하여 분석의 설명력을 높이고자 하였다. 대부분의 변수들은 학교 등급에 관계없이 공통적으로 사용하였지만, 중학교의 경우 외국어·과학 고등학교 진학률과 고등학교의 경우 대학교 진학을 변수 및 자립형 고등학교/비자립형 고등학교 더미 변수를 각기 적용하여 학교 등급별 특성을 반영할 수 있도록 모델을 구축하였다. 중학교의 경우 7개, 고등학교의 경우 11개의 독립변수가 선정되었고 이 중 여학생 구성 비율, 학생 1인당 장학금, 순전입 변수의 경우 중학교와 고등학교에서 공통적으로 적용되는 독립변수로 채택하였다. 본 연구의 분석 시점은 2016년을 대상으로 진행하였다. 이는 국가수준 학업성취도 평가가 2009년부터 2016년까지는 전수 평가로 이루어졌으나, 2017년부터 표집 평가로 바뀌었기 때문에 표본의 수를 충분히 확보할 수 있는 연도에 기반하여 분석하기 위함이다.

4. 학업성취도 요인 분석

1) 모델 적합도 비교

OLS 모델과 GWR 모델의 적합도는 표 4와 같다. 중학교의 경우 전통적인 OLS 모델보다 공간적 자기상관성을 고려한 국지적 모델인 GWR의 성능이 보다 우수한 것으로 나타났으며, 고등학교 역시 GWR이 OLS에 비해 다소 우수하지만 그 차이는 크지 않다. 이는 중학교의 경우 학생이 거주지로부터 가까운 학교에 배정되기 때문에 유사한 학업성취도의 공간의존적 특성이 잘 나타날 수 있지만, 고등학교의 경우 학생이 자율적으로 학교를 지원할 수 있기 때문에 중학교에 비해 통학거리가 상대적으로 길어져 특정 학교들의 위치적 특성이 GWR 모델에 약하게 반영된 결과로 판단된다. 한편, 중학교와 고등학교의 학업성취도를 설명함에 있어 대부분 공통적인 독립변수를 적용했음에도 불구하고 결정계수(R^2) 및 조정된 결정계수($Adjusted R^2$) 모두 고등학교를 대상으로 한 모델의 설명력이 더 높은 것으로 나타났다. 이는 모델간

표 3. 독립변수

| 구분 | 변수 | 비고 |
|-------------|----------------------|------|
| 학생 관련 변수 | 학급당 학생 수 | 공통 |
| | 전교생 수 | 공통 |
| | 여학생 구성 비율 | 공통 |
| | 순전입(전입 학생 수-전출 학생 수) | 공통 |
| | 학생 1인당 교육비 | 공통 |
| | 학생 1인당 복지 및 교육격차 해소비 | 공통 |
| | 학생 1인당 발전 기금 | 공통 |
| | 학생 1인당 장학금 | 공통 |
| | 학생 1인당 도서 수 | 공통 |
| | 학생 1인당 방과후학교 강좌 수강 수 | 공통 |
| 교사 관련 변수 | 외국어·과학 고등학교 진학률 | 중학교 |
| | 대학교 진학률 | 고등학교 |
| 교사 관련 변수 | 교사 1인당 수업 시수 | 공통 |
| | 기간제 교사 및 강사 구성 비율 | 공통 |
| 환경 관련 변수 | 급식비 | 공통 |
| | 국·공립 학교/사립 학교 | 공통 |
| | 자립형 고등학교/비자립형 고등학교 | 고등학교 |

표 4. OLS와 GWR 모델의 적합도 비교

| | 중학교 | | 고등학교 | |
|-------------------------|-----------|-----------|-----------|----------|
| | OLS | GWR | OLS | GWR |
| R ² | 0.3830 | 0.4960 | 0.6373 | 0.6663 |
| Adjusted R ² | 0.3715 | 0.4513 | 0.6186 | 0.6258 |
| AIC | 918.0730 | 877.6949 | 417.0544 | 417.4618 |
| 표준화된 잔차의 분포 형태 | clustered | clustered | clustered | random |
| 표준화된 잔차의 모란 I 지수 | 0.2400*** | 0.1672*** | 0.0753• | 0.0709 |

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, •: $p < 0.1$ 에서 유의미

적합도 및 설명력을 비교하는데 자주 사용되는 또 다른 통계량인 AIC(Akaike Information Criteria)를 통해서도 확인할 수 있으며, OLS와 GWR 모두 고등학교에 대한 AIC가 중학교에 비해 낮게 나타나고 있다.

2) OLS 기반 학업성취도 요인 분석

본 연구에서 최종적으로 채택한 학교 등급별 독립 변수는 표 5와 같다. 공통적으로 선정된 변수들을 살펴보면, 여학생 구성 비율이 높을 경우 학업성취도가 높게 나타나고 있는데, 이는 여학생이 남학생보다 성적이 우수하다는 김혜숙(2011)의 결과와 일치한다. 학생 1인당 장학금의 경우 학업성취도와 음의 관계를 보이고 있으며, 특히 중학교의 경우 여러 설명 변수 중 절댓값이 가장 큰 값을 보이고 있다. 이는 중·고등학교에서 지급하는 장학금은 성적이 우수하여 받는 장학금도 있지만 집안 형편을 고려한 복지 차원에서 지급하는 경우가 더 많기 때문에 이러한 결과가 도출된 것으로 보인다. 순전입의 경우 중학교와 고등학교가 각각 양의 값과 음의 값의 회귀계수가 산출된 것을 확인할 수 있다. 먼저 중학교의 경우 학교를 배정 받는 것에 있어 선택 제도가 아닌 거주지 위치를 기반한 강제 배정이기 때문에 원하는 중학교로 진학하기 위해 전학을 가는 사례가 많이 존재한다(김형미 등, 2011; 이화정 등, 2013; 김갑영 등, 2014). 실제로 표 6을 살펴보면, 대규모 택지 개발로 인해 2015년에 개교한 강서구의 ○○중학교와 2014년에 개교한 송파구의 △△중학교가 순전입 1위와 10위를 각각 차지

하였으나 나머지 2위에서 9위 학교는 모두 선호 학군인 강남구 및 양천구에 위치한 것을 확인할 수 있다. 반면, 고등학교의 경우 학생 스스로 진학하고자 하는 학교를 선택하는 제도이기 때문에 입학 이후에 자발적인 전학은 많지 않을 것으로 판단된다. 실제로 고등학교 순전입의 경우 중학교와 달리 특별한 공간적 경향성을 보이고 있지 않다.

중학교에만 적용된 학급당 학생 수의 경우, 대체로 선호도가 높은 학교에 학생들이 많이 몰리는 경향이 있지만, 이를 수용하기 위한 교실과 같은 물리적 공간 확충이 즉각적으로 이루어지기 어렵기 때문에 학급당 학생 수가 많아진 것으로 보인다. 외국어·과학 고등학교 진학률은 학업성취도를 설명하는데 가장 주요한 변수로 작용하고 있으며 이는 외국어 고등학교와 과학 고등학교에 학업성취도가 우수한 학생들이 진학하고 있음을 의미한다.

고등학교의 경우 전교생수, 기간제 교사 및 강사 구성 비율, 국·공립 학교/사립 학교, 자립형 고등학교/비자립형 고등학교를 설명 변수로 채택하였다. 먼저 사립 및 자립형 고등학교가 국·공립 및 비자립형 고등학교보다 학업성취도가 높은 것으로 나타났다. 이는 사립 및 자립형 고등학교가 성적이 우수한 학생들이 입학하여 우수한 성적을 거두는 선발 효과가 성적이 저조한 학생들이 입학하여 우수한 성적을 보이는 학력 증진 효과보다 높은 것에 기인할 수 있다(Coleman *et al.*, 1982; 김경년, 2012; 황지원, 2013). 또한 전교생수의 경우, 학교선택제로 인해 사립 및 자율고와 같이 학업성취도가 우수한 학생이 선

표 5. OLS 모델 분석 결과

| 변수 명 | 회귀계수(B) | |
|----------------------|-------------|-------------|
| | 중학교 | 고등학교 |
| 절편 | -0.13531 | -0.41818*** |
| 여학생 구성 비율 | 0.20084*** | 0.18125*** |
| 학생 1인당 장학금 | -0.50001*** | -0.22271* |
| 순전입(전입 학생 수-전출 학생 수) | 0.14910*** | -0.19760*** |
| 학급당 학생 수 | 0.41417*** | |
| 학생 1인당 교육비 | 0.08997 · | |
| 학생 1인당 방과후학교 강좌 수강 수 | 0.04406 | |
| 외국어 · 과학 고등학교 진학률 | 0.45992*** | |
| 전교생 수 | | 0.31754*** |
| 기간제 교사 및 강사 구성 비율 | | 0.10761* |
| 학생 1인당 복지 및 교육격차 해소비 | | 0.15151* |
| 급식비 | | 0.14241*** |
| 국 · 공립 학교 / 사립 학교 | | 0.64149*** |
| 자립형 고등학교 / 비자립형 고등학교 | | 0.53840*** |
| 학생 1인당 도서 수 | | 0.08115*** |
| 대학교 진학률 | | -0.07347*** |

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, · : $p < 0.1$ 에서 유의미

표 6. 2016년 중 · 고등학교 순전입 상위 10개 학교의 소속 자치구

| 순위 | 중학교 | | 고등학교 | |
|----|-----|-----|------|-----|
| | 자치구 | 순전입 | 자치구 | 순전입 |
| 1 | 강서 | 113 | 구로 | 30 |
| 2 | 강남 | 57 | 서초 | 26 |
| 3 | 강남 | 53 | 성북 | 25 |
| 4 | 강남 | 38 | 구로 | 25 |
| 5 | 강남 | 37 | 광진 | 24 |
| 6 | 강남 | 33 | 양천 | 24 |
| 7 | 강남 | 32 | 송파 | 22 |
| 8 | 강남 | 30 | 광진 | 22 |
| 9 | 양천 | 30 | 중랑 | 22 |
| 10 | 송파 | 29 | 은평 | 22 |

호하는 학교에는 학생들이 집중되었고 반대로 비선호학교는 정원 미달 현상이 발생함에 따라 학생 수가 많은 학교가 비교적 높은 학업성취도를 보였을 것이라고 추정한다(배정현, 2012; 전창완, 2016). 기간제

교사 및 강사 구성 비율 또한 학업성취도에 유의미하게 작용하였다. 사립 고등학교의 경우 국 · 공립 고등학교와 달리 신규 교사 채용에 대한 사회적 압력이 적기 때문에 학령인구 감소 대비를 위한 교원 인력 공급의 탄력성을 확보하고자 기간제 교원 채용을 선호하는 경향이 있는데, 서울에 소재한 고등학교 중 59%가 사립 고등학교이기 때문에 학업성취도에 대해 기간제 교사 및 강사 구성 비율이 유의미하게 작용한 것으로 보인다(남기곤, 2018). 또한 사립 고등학교 기간제 교원의 경우 정교사 전환 및 계약 연장을 위해 수업에 상대적으로 많은 노동력을 투입하기 때문에 질 높은 수업이 학업성취도에 긍정적으로 작용했을 것이다(조현국, 2013). 한편 학생 1인당 방과후학교 강좌 수강수의 경우 학업성취도에 유의미한 영향을 미치지 않은 것으로 나타났다. 이는 한국교육과정평가원(2017)의 결과와는 대비되는 모습인데, 본 연구는 서울을 대상으로 하며 한국교육과정평가원(2017)은 전국단위를 다루고 있기 때문에 상반된 결과가 나온 것으로 보인다.

3) 공간적 이질성을 고려한 학업성취도 요인 분석

학업성취도에 영향을 미치는 변이 요인들은 상이한 지역 특성에 따라 영향력의 크기나 관계성에 차이가 존재할 수 있다. 이러한 변이 요인들의 공간적 이질성을 고려하기 위해 지리가중회귀 모델을 적용하였다. GWR은 국지적 관측치에 초점을 맞추어 설명변수의 지역적 이질성을 파악할 수 있는 모델로 부분회귀계수의 크기와 방향을 통해 지역에 따라 서로 다른 독립변수의 학업성취도에 대한 영향력을 정량적으로 확인할 수 있다. 실제로 OLS 분석 결과를 나타내는 표 5의 경우, 하나의 회귀식을 도출하기 때문에 지역별 영향력의 차이를 파악할 수 없다. 반면, GWR 모델의 연산 결과를 보여주는 표 7의 경우 지역별로 변수별 부분 회귀계수를 도출하기 때문에 지역에 따라 설명변수들의 영향력이 상이함을 파악할 수 있다.

구체적으로 중학교 OLS 분석에서 도출된 회귀계수 중 가장 큰 절댓값을 보였던 학생 1인당 장학금의 경우, GWR 분석에 있어 부분 회귀계수의 표준편차 또한 가장 커 지역에 따른 영향력의 차이를 살필 필요성을 보여주고 있다. 반면 고등학교 GWR 분석에 대한 부분 회귀계수의 표준편차를 살펴보면, 전체적으로 그 값이 작아 지역에 따른 영향력의 크기가 크지 않음을 확인할 수 있다.

한편, 실제 분석은 각 학교 단위로 이루어졌으나 이에 대한 해석을 보다 용이하게 하기 위해 각 학교가 소속된 자치구를 기준으로 각 학교별로 상이하게 도출된 부분 회귀계수의 평균값을 산출하여 독립변수의 영향력이 유사한 지역들을 유형화 하였다(표 8, 9). 중학교의 경우 크게 3개 유형으로 나눌 수 있는데, (가) 유형에 속한 자치구는 학급당 학생 수가 학업성취도에 가장 큰 영향을 주는 지역으로 대형 학교일수록 학업성취도에 긍정적인 영향을 주는 지역에 해

표 7. 부분 회귀계수 분포

| | | 평균 | 표준편차 | 최댓값 | 최솟값 | |
|--------------------|----------|----------------------|----------|----------|----------|----------|
| 부분회귀계수(B) | 중학교 | 절편 | -0.39344 | 0.64595 | 0.16261 | -2.40023 |
| | | 학급당 학생 수 | 0.40660 | 0.06388 | 0.50950 | 0.23572 |
| | | 여학생 구성 비율 | 0.20860 | 0.01329 | 0.23907 | 0.18397 |
| | | 순전입(전입 학생 수-전출 학생 수) | 0.15217 | 0.12273 | 0.30809 | -0.06298 |
| | | 학생 1인당 교육비 | 0.08669 | 0.07615 | 0.20217 | -0.01901 |
| | | 학생 1인당 장학금 | -0.95544 | 1.09128 | -0.11178 | -4.46376 |
| | | 학생 1인당 방과후학교 강좌 수강 수 | 0.05006 | 0.03862 | 0.11303 | -0.04188 |
| | 고등학교 | 외국어·과학 고등학교 진학률 | 0.43436 | 0.09004 | 0.55427 | 0.28332 |
| | | 절편 | -0.46783 | 0.07662 | -0.33347 | -0.63566 |
| | | 전교생 수 | 0.32067 | 0.05302 | 0.39515 | 0.24238 |
| | | 여학생 구성 비율 | 0.18719 | 0.01967 | 0.21585 | 0.15730 |
| | | 순전입(전입 학생 수-전출 학생 수) | -0.19850 | 0.01376 | -0.17637 | -0.22299 |
| | | 기간제 교사 및 강사 구성 비율 | 0.10219 | 0.00977 | 0.12130 | 0.08442 |
| | | 학생 1인당 복지 및 교육격차 해소비 | 0.13980 | 0.01621 | 0.16288 | 0.10042 |
| 학생 1인당 장학금 | -0.19743 | 0.04724 | -0.09284 | -0.27706 | | |
| 학생 1인당 도서 수 | 0.05973 | 0.01961 | 0.10322 | 0.03224 | | |
| 급식비 | 0.12820 | 0.02177 | 0.17312 | 0.09829 | | |
| 국·공립 학교/사립 학교 | 0.65709 | 0.07656 | 0.79322 | 0.52465 | | |
| 자립형 고등학교/비자립형 고등학교 | 0.55249 | 0.01093 | 0.57542 | 0.53318 | | |
| 대학교 진학률 | -0.08125 | 0.00853 | -0.06796 | -0.09588 | | |

표 8. 자치구별 학업성취도 요인 유형(중학교)

| 유형 | 자치구 | 부분 회귀계수(B) | | | | | | | 비고 |
|-----|-----|------------|-----|-----|-----|-----|-----|-----|---|
| | | 1순위 | 2순위 | 3순위 | 4순위 | 5순위 | 6순위 | 7순위 | |
| (가) | 강북 | ㉠ | ㉡ | ㉢ | ㉣ | ㉤ | ㉥ | ㉦ | ㉠: 학급당 학생 수 ㉡: 여학생 구성 비율 ㉢: 순전입 ㉣: 학생 1인당 교육비 ㉤: 학생 1인당 장학금 ㉥: 학생 1인당 방과후학교 강좌 수강수 ㉦: 외국어·과학 고등학교 진학률 |
| | 노원 | ㉠ | ㉡ | ㉢ | ㉣ | ㉤ | ㉥ | ㉦ | |
| | 도봉 | ㉠ | ㉡ | ㉢ | ㉣ | ㉤ | ㉥ | ㉦ | |
| | 동대문 | ㉠ | ㉡ | ㉢ | ㉣ | ㉤ | ㉥ | ㉦ | |
| | 성동 | ㉠ | ㉡ | ㉢ | ㉣ | ㉤ | ㉥ | ㉦ | |
| | 성북 | ㉠ | ㉡ | ㉢ | ㉣ | ㉤ | ㉥ | ㉦ | |
| | 중 | ㉠ | ㉡ | ㉢ | ㉣ | ㉤ | ㉥ | ㉦ | |
| (나) | 관악 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 구로 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 금천 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 동작 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 미포 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 서대문 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 서초 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 양천 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 영등포 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 용산 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 은평 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 종로 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 강남 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| 강동 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | | |
| (다) | 강서 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 광진 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |
| | 송파 | ㉧ | ㉨ | ㉩ | ㉪ | ㉫ | ㉬ | ㉭ | |

표 9. 자치구별 학업성취도 요인 유형(고등학교)

| 유형 | 자치구 | 부분 회귀계수(B) | | | | | | | | | | 비고 | |
|-----|-----|------------|-----|-----|-----|-----|-----|-----|-----|-----|------|----|------|
| | | 1순위 | 2순위 | 3순위 | 4순위 | 5순위 | 6순위 | 7순위 | 8순위 | 9순위 | 10순위 | | 11순위 |
| (A) | 강동 | l | k | a | b | e | h | d | g | i | f | c | |
| | 송파 | l | k | a | b | e | h | d | g | i | f | c | |
| (B) | 강남 | k | j | a | b | e | h | d | g | i | f | c | |
| | 강북 | k | j | a | b | h | e | d | g | i | f | c | |
| | 강서 | k | j | a | h | b | e | d | g | i | f | c | |
| | 관악 | k | j | a | b | e | h | d | g | i | f | c | |
| | 광진 | k | j | a | b | e | h | d | g | i | f | c | |
| | 구로 | k | j | a | h | b | e | d | g | i | f | c | |
| | 금천 | k | j | a | h | b | e | d | g | i | f | c | |
| | 노원 | k | j | a | b | h | e | d | g | i | f | c | |
| | 도봉 | k | j | a | b | h | e | d | g | i | f | c | |
| | 동대문 | k | j | a | b | e | h | d | g | i | f | c | |
| | 동작 | k | j | a | b | e | h | d | g | i | f | c | |
| | 마포 | k | j | a | b | h | e | d | g | i | f | c | |
| | 서대문 | k | j | a | b | h | e | d | g | i | f | c | |
| | 서초 | k | j | a | b | e | h | d | g | i | f | c | |
| | 성동 | k | j | a | b | e | h | d | g | i | f | c | |
| | 성북 | k | j | a | h | b | e | d | g | i | f | c | |
| | 양천 | k | j | a | b | h | e | d | g | i | f | c | |
| | 영등포 | k | j | a | b | e | h | d | g | i | f | c | |
| 용산 | k | j | a | b | e | h | d | g | i | f | c | | |
| 은평 | k | j | a | b | h | e | d | g | i | f | c | | |
| 종로 | k | j | a | b | e | h | d | g | i | f | c | | |
| 중 | k | j | a | b | e | h | d | g | i | f | c | | |
| 중랑 | k | j | a | b | e | h | d | g | i | f | c | | |

- a : 전교생수
- b : 여학생 구성 비율
- c : 순전입
- d : 기간제 교사 및 강사 구성 비율
- e : 학생 1인당 복지 및 교육역차 해소비
- f : 학생 1인당 장학금
- g : 학생 1인당 도서수
- h : 급식비
- i : 대학교 진학률
- j : 자립형 고등학교/비자립형 고등학교
- k : 국·공립 학교/사립 학교

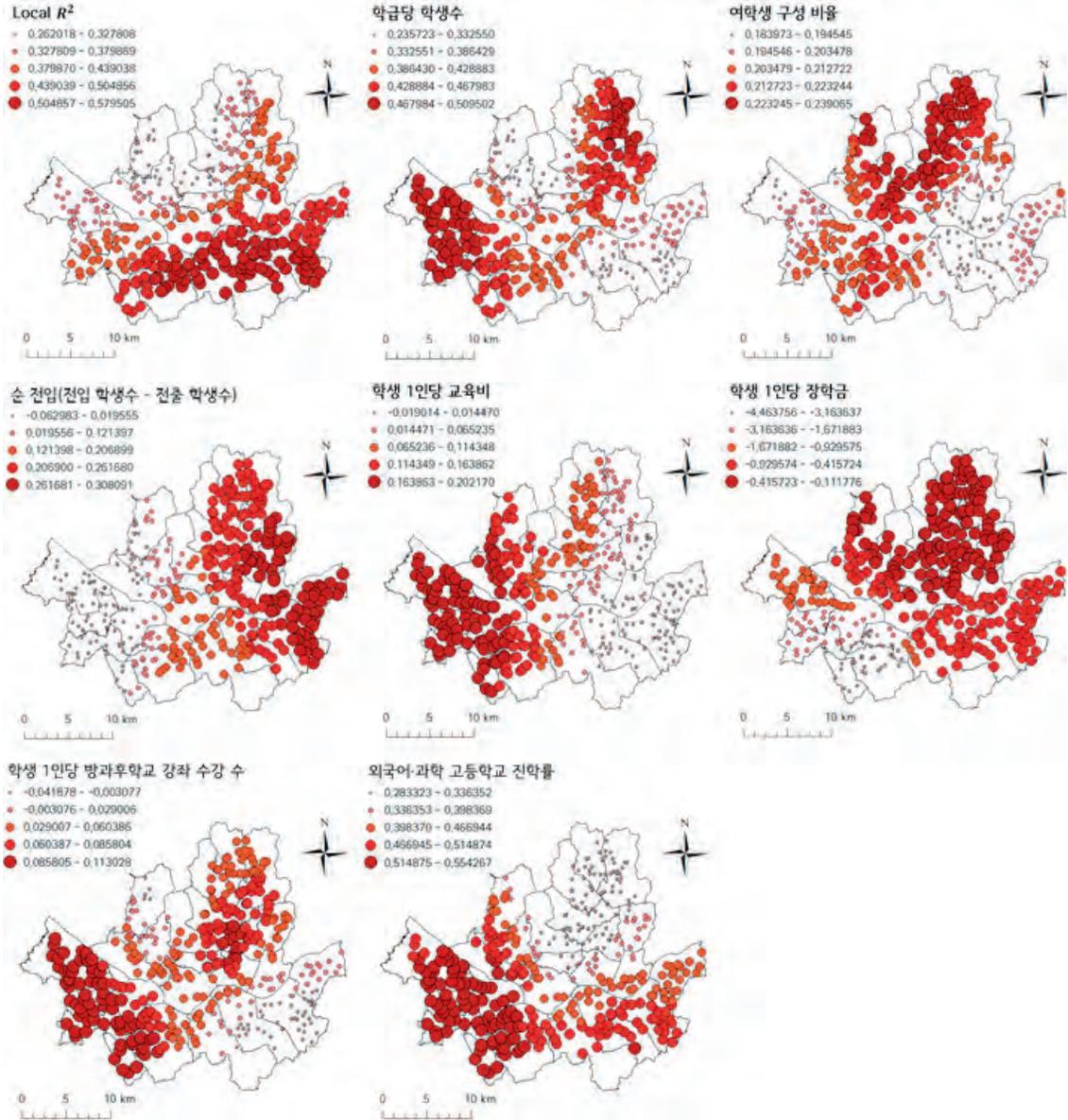


그림 1. 부분 회귀계수의 공간적 분포(중학교)

당한다. 또한, OLS 분석에서 학업성취도 설명에 있어 가장 주요하게 작용하였던 외국어·과학 고등학교 진학률이 두 번째로 크게 작용한다는 점이 특징이다. (나), (다) 유형은 (가) 유형과는 반대로 학업성취도에 가장 큰 영향을 주는 변수로 외국어·과학 고등학교 진학률이, 두 번째로 큰 영향을 주는 변수로는 학급당 학생 수가 선정되었다. 이는 OLS 분석 결과와 일치하

는데 (나) 유형은 여학생 구성 비율이, (다) 유형은 순 전입이 세 번째로 주요한 변수로 작용하고 있다. 중학교에 대한 GWR 분석 결과를 정리하면, 기존 OLS 분석에서 큰 영향력을 보였던 설명변수가 여전히 유의미하게 작용하고 있지만, 지역에 따라 각 변수별 중요도가 다소 다름을 확인할 수 있다. 이는 학업성취도에 대한 설명에 있어 국지적 또는 하위 지역단위별 해

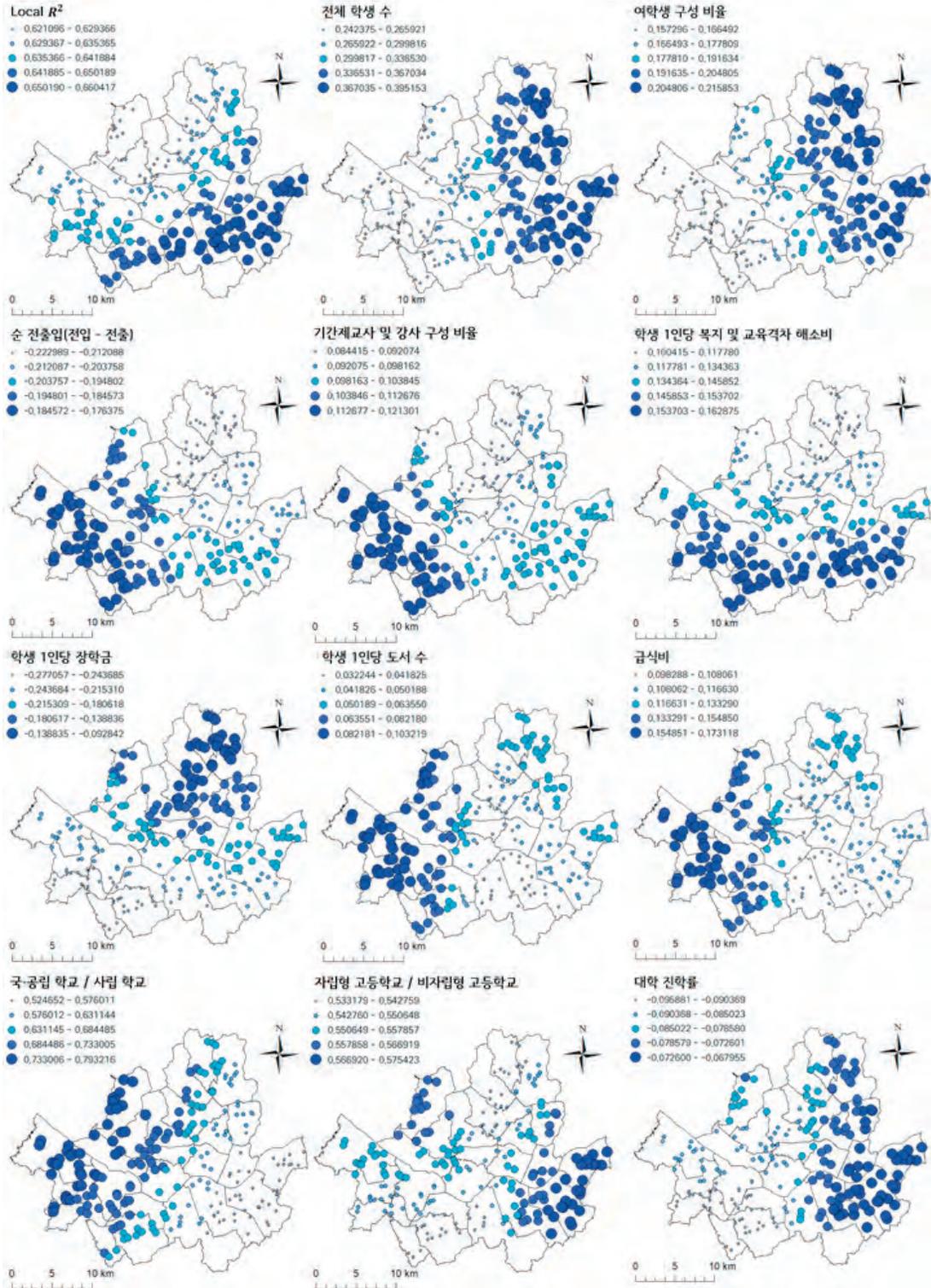


그림 2. 부분 회귀계수의 공간적 분포(고등학교)

석이 충분히 의미 있으며, 관련 정책을 시행함에 있어 지역 및 개별 학교에 따라 우선 순위 및 중요도를 다르게 설정해야 함을 시사한다.

반면 고등학교의 경우 두 개의 유형으로만 구분되었으며 이는 비교적 유의미한 독립변수군이 다양한 중학교와 대비된다. 실제로 자립형 고등학교/비자립형 고등학교 변수가 학업성취도에 대한 가장 큰 변수로 작용하는 (A) 유형은 2개의 자치구만 존재하였으며, 국·공립 학교/사립 학교 변수가 학업성취도에 대한 가장 큰 설명변수로 작용하는 (B) 유형은 23개로 대부분을 차지하고 있다. 두 번째로 큰 영향을 주는 변수의 경우, (A) 유형은 국·공립 학교/사립 학교, (B) 유형은 자립형 고등학교/비자립형 고등학교로 나타났다. 또한 (A), (B) 유형 모두 세 번째로 큰 영향을 주는 설명변수는 전교생수로 나타났다. 즉, 고등학교 학업성취도에 대한 지역별 이질성에 따른 설명변수 유형의 구분이 크게 두드러지지 않고 있다. 이는 앞서 진행한 OLS 기반 모델링 결과와 일치하는 것으로 고등학교의 경우 중학교와 달리 지역적 해석이 크게 유의미하지 않음을 의미한다. 한편, 그림 1과 그림 2는 각각 중학교와 고등학교의 GWR 결과에 대한 주요 통계량을 보여주고 있다. 대체로 같은 행정구역이더라도 학업성취도에 대한 개별 설명 요인의 크기가 상이함을 확인할 수 있다. 또한 거리가 인접한 학교간에는 일정 수준 이상의 공간적 자기상관성을 보이고 있음을 알 수 있다. 이는 미시적 공간 데이터를 활용한 지리중회귀적 접근이 학업성취도를 설명함에 있어 매우 적절한 방법임을 다시 한번 증명하는 부분이다.

5. 결론

학업성취도는 학생들의 교육 수준과 성취, 나아가 지역 및 국가의 교육 환경과 질을 전반적으로 평가할 수 있는 중요한 척도이다. 이러한 중요성과 관련하여 오래 전부터 많은 연구들이 수행되어 왔다. 기존의 연구들은 학업성취도에 영향을 미치는 개인, 사회, 지역적 요인에 대한 분석이 주를 이루고 있지만 대부분

지역적 특성에 상관없이 전역적으로 작동하는 요인에 초점을 맞추고 있다. 이에 본 연구는 지리가중모형을 이용하여 학업성취도에 대해 국지적 공간에서 이질적으로 작용하는 여러 영향 요인들을 체계적으로 고찰하였다. 특히, 학교급 단위의 교육정보 공시 자료를 활용함으로써 학업성취도에 대한 보다 미시적인 접근을 시도하였으며, 개인적 요인을 비롯하여 교육 환경적 요인에 이르는 다양한 측면을 고려한 분석을 수행하였다. 중학교의 경우 여학생 구성비율, 학생 1인당 장학금, 순전입, 학급당 학생 수, 학생 1인당 교육비, 학생 1인당 방과후학교 강좌 수강수, 외국어·과학 고등학교 진학률 변수를 고려하였고, 고등학교의 경우 여학생 구성 비율, 학생 1인당 장학금, 순전입, 전교생 수, 기간제 교사 및 강사 구성 비율, 학생 1인당 복지 및 교육격차 해소비, 급식비, 국·공립 학교/사립 학교, 자립형 고등학교/비자립형 고등학교, 학생 1인당 도서수, 대학교 진학률 변수를 설명변수로 고려하였다.

일반적인 OLS 분석과 설명변수의 공간적 이질성을 고려한 GWR 분석 결과, 중학교의 경우 외국어·과학 고등학교 진학률, 학급당 학생 수, 여학생 구성 비율이 학업성취도에 유의미한 설명 변수로 작용하였다. 고등학교의 경우 국·공립 학교/사립 학교, 자립형 고등학교/비자립형 고등학교 변수가 주요한 설명 변수로 작용하였다. 즉, 중학교의 경우 전체적으로 학생과 관련된 변수가 학업성취도에 크게 작용한 반면, 고등학교의 경우 학교 특성 및 교육 환경과 관련된 변수가 크게 작용하였다. 모델의 성능 측면에서는 대체로 일반 OLS 모델보다 공간적 의존성을 고려한 국지적 모델인 GWR 모델이 우수한 것으로 나타났다. 구체적으로, 중학교의 경우 고등학교에 비해 GWR이 OLS 보다 높은 설명력을 보여주고 있으며, 고등학교는 그 차이가 크지 않아 지역별 학업성취도 변이 요인이 크지 않음을 알 수 있다. 한편, OLS, GWR 모두 중학교 모델보다 고등학교 모델의 설명력이 더 높은 것으로 나타나 중학교에 비하여 고등학교가 학교 관련 변수 외에 외부 요인이 상대적으로 적음을 알 수 있다.

본 연구는 학업성취도의 분포가 지역적 이질성을 보임에도 불구하고 이를 전역적으로 다루어 학업성

취도에 대한 요인 분석이 제한적이었던 기존 연구에 비해, 국지적인 변이 요인들을 보다 미시적으로 확인하고 해석할 수 있었다는 점에서 의의가 있다. 또한, 개별 학교 단위의 고유한 특성을 반영하여 학교별 학업성취도 향상에 영향을 미치는 요인을 확인했다는 점에서 의의가 있다. 이는 지역별 학업성취도에 대한 교육 환경, 지역 특성 등에 기반한 접근을 넘어 개별 학교 단위가 어떠한 영향을 미치고 있으며, 어떤 측면이 개선되어야 할지에 대한 의미 있는 통찰을 제공할 수 있다. 한편, 본 연구는 대체로 학교알리미에 수록된 데이터에 의존하고 있어 학교 외적인 요소이면서 학업성취도에 영향을 줄 수 있는 소득, 사교육 관련 요인 등과 같은 학교 외부 변수를 고려하지 못하고 있는 한계가 있다. 따라서 향후에는 개인, 학교, 지역 특성을 아우르는 포괄적 모델을 구축하여 보다 다차원적 접근이 필요할 것으로 보인다.

참고문헌

- 강태중, 2007, “학업성취도의 지역 차이 분석: 인구 이동의 영향을 고려한 시도,” 한국청소년연구, 18(2), 315-344.
- 김감영·용환성, 2014, “도시 내 학업성취도 격차와 학령인구 이동과의 관련성: 대구광역시를 사례로,” 한국지도학회지, 14(2), 53-72.
- 김경년, 2012, “학교설립 유형이 학업성취도에 미치는 영향: 사회실험을 활용한 인과관계 분석,” 교육사회학연구, 22(2), 1-29.
- 김승태, 2006, 행정정보 공개제도의 집행요인 및 정책효과 분석, 한국학술정보.
- 김정원, 2007, “교육복지와 학교의 역할: 교육복지투자우선지역 지원사업을 중심으로,” 교육사회학연구, 17(4), 35-61.
- 김현철·장수정, 2012, “엔트로피 지수 분해를 통한 중학생 학업성취도 격차 요인의 분석,” 교육실천연구, 11(1), 1-23.
- 김형미·주경식, 2011, “경기도 남부지역 학생이동의 공간적 특성에 관한 연구,” 한국지리지리학회지, 17(6), 679-697.
- 김혜숙, 2011, “우리나라 초중고 학생들의 사회과 학업성취도의 성별 차이: 2009년 국가수준 학업성취도 평가 결과를 중심으로,” 한국지리환경교육학회지, 19(1), 67-87.
- 남기곤, 2018, “기간제 교사 규모 팽창의 원인 분석과 정책 대안,” 동향과 전망, 103, 190-228.
- 류주현, 2006, “거주지별 교육환경의 차별화,” 한국지리환경교육학회지, 14(1), 69-78.
- 박수정·황은희, 2011, “교장공모제 학교의 학업성취도와 특성 분석,” 한국교원교육연구, 28(2), 313-340.
- 박찬선·주경식, 2007, “학업성취도의 공간적 분포 패턴 연구 -안양시를 사례로-,” 한국지리환경교육학회지, 15(1), 65-76.
- 배정현, 2012, 학교선택제 도입과 학력격차 변동 추세: 수원시 경우, 서울대학교 대학원 석사학위논문.
- 성기선, 2004, “강남8학군, 정말 교육 효과 있나?,” 교육비평, 15, 188-210.
- 손준중, 2004, “교육공간으로서 강남(江南) 읽기,” 교육사회학연구, 14(3), 107-131.
- 양미경, 2004, “학교의 기능에 대한 사회적 요구의 성격: 다중성과 모호성,” 교육학연구, 42(2), 133-162.
- 오정란, 2007, “한국의 학교선택, 어떻게 이해할 것인가,” 교육사회학연구, 17(2), 89-113.
- 윤광석·김현성, 2010, 정보공개제도의 진단과 개선방안 연구, 한국행정연구원.
- 윤형호·강민정, 2008, “서울시 자치구별 상위학교 진학률에 대한 거주지 효과 분석,” 서울도시연구, 9(4), 87-106.
- 이광현·김민조, 2012, “일반고등학교에서 교장공모제가 학업성취도에 미치는 영향 분석,” 한국교원교육연구, 29(2), 195-221.
- 이정우, 2013, “국가수준 학업성취도 평가 결과를 통해 본 다문화가정 학생의 사회과 학업성취도 특성,” 시민교육연구, 45(2), 257-291.
- 이화정·이상일·조대현, 2013, “거주지 이동을 통한 학교 선택의 공간성에 관한 연구,” 대한지리학회지, 48(6), 897-913.
- 임다희·권기현, 2012, “서울시 고등학생의 학업성취도 결정요인: 절단형(censored) 자료에 대한 tobit model의 적용,” 정책분석평가학회보, 22(2), 43-63.
- 전은정·임현정·성태제, 2015, “고교 유형별 진학에 대한

가정배경 및 개인특성의 영향,” *교육학연구*, 53(1), 1-27.

전창완, 2016, “부산지역 일반계 고등학교의 학업 성취도에 대한 사립학교 효과 분석,” *교사교육연구*, 55(3), 313-326.

정충대, 2015, “동료효과와 학업성취도,” *교육종합연구*, 13(4), 1-22.

정혜원·원정은·박소영, 2018, “청소년의 학업성취도와 핵심역량 유형 분류 및 영향요인 검증,” *한국청소년연구*, 29(2), 185-215.

조현국, 2013, “학교 유형별 기간제 교사가 학업성취도에 미치는 영향 분석,” *경제학연구*, 61(2), 5-29.

주경식·정효선, 2012, “고등학생 학업성취도의 공간적 분포에 관한 연구: 성남시를 사례로,” *교원교육*, 28(3), 211-228.

최은영, 2004, “학력자본 재생산의 차별화와 빗장도시의 형성,” *대한지리학회지*, 39(3), 374-390.

추경모, 2012, “교육의 지역적 격차에 관한 연구: 중학교 교육현황을 사례로,” *한국지리학회지*, 1(1), 33-52.

한국교육개발원, 2008, *교육정보공시제 발전 체제 구축 방안*.

한국교육과정평가원, 2017, 2016년 국가수준 학업성취도 평가 결과: 고등학교 학업성취도 결과.

황지원, 2013, *일반계 고등학교의 학력격차 변화 추이 분석*, 서울대학교 대학원 박사학위논문.

Barrow, R., 1979, *Radical Education: A Critique of Free-schooling and Deschooling*, *British Journal of Edu-*

cational Studies, 27(3), 259-284.

Coleman, J. S., Hoffer, T. and Kilgore, S. B., 1982, *High School Achievement: Public, Catholic, and Private Schools Compared*, New York: Basic Books.

Feinberg, W., 2012, The Idea of a Public Education, *Review of Research in Education*, 36(1), 1-22.

Fotheringham, A. S., Brunsdon, C. and Charlton, M., 2002, *Geographically weighted regression: the analysis of spatially varying relationships*, John Wiley&Sons, Ltd., Chichester, England.

Oyana, T. J. and Margai, F. M., 2016, *Spatial analysis: statistics, visualization, and computational methods*, Boca Raton: CRC Press, Taylor & Francis Group.

Young, M. F. D., 2010, The future of education in a knowledge society: the radical case for a subject-based curriculum, *Pacific-Asian education*, 22, 21-32.

교신: 이건학, 08826, 서울특별시 관악구 관악로1, 서울대학교 사회과학대학 지리학과(이메일: gunhlee@snu.ac.kr, 전화:02-880-4019)

Correspondence: Gunhak Lee, Department of Geography, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Korea(e-mail: gunhlee@snu.ac.kr, phone: +82-2-880-4019)

최초투고일 2019. 9. 18
수정일 2019. 10. 23
최종접수일 2019. 10. 24

사단법인 대한지리학회 34대 임원(2019.1.1 ~ 2020.12.31)



| | | | |
|--------------|---------------|--------------|----------------|
| 회 장 | 이용우(국토연구원) | | |
| 부 회 장 | 최재현(건국대) | 강병수(부산남고) | 이영아(대구대) |
| | 박수진(서울대) | 황철수(경희대) | |
| 감 사 | 이정훈(경기연구원) | 정희선(상명대) | |
| 지리연구소장 | 김용창(서울대) | | |
| 지리교육특별위원장 | 정성훈(강원대) | | |
| 지리올림피아드추진위원장 | 이영아(대구대) | | |
| 지리학발전위원장 | 남기범(서울시립대) | 박정재(서울대) | |
| 총무이사 | 진중현(공주대) | 최진무(경희대) | 이정섭(경상대) |
| 학술이사 | 이승철(동국대) | 전병운(경북대) | 김화환(전남대) |
| 국제이사 | 정현주(서울대) | 박선엽(부산대) | 이종호(경상대) |
| 홍보이사 | 김영훈(한국교원대) | 박지훈(공주대) | 신혜란(서울대) |
| 편집이사 | 이자원(성신여대) | 오충원(남서울대) | 임은진(공주대) |
| 이 사 | 강영옥(이화여대) | 구동회(부산대) | 구자용(상명대) |
| | 권정화(한국교원대) | 기근도(경상대) | 김선기(주부른길) |
| | 김선배(산업연구원) | 김오진(제주 추자중) | 김종근(동북아역사재단) |
| | 김진석(전북발전연구원) | 김태호(제주대) | 노영순(한국문화관광연구원) |
| | 류연택(충북대) | 문미성(경기연구원) | 박숙희(한국여성지리학회) |
| | 박지만(국토정보공사) | 백선혜(서울연구원) | 서봉만(인천연구원) |
| | 안영진(전남대) | 오정준(대구대) | 윤영수(한울엠플러스) |
| | 윤정현(태릉고) | 이간용(공주교대) | 이경한(전주교대) |
| | 이상율(대구가톨릭대) | 이상일(서울대) | 이승욱(카이스트) |
| | 이용균(전남대) | 이우평(인천 부광고) | 이의한(강원대) |
| | 이지선(한국교통연구원) | 임은선(국토연구원) | 장영진(한국교원대) |
| | 정치영(한국학중앙연구원) | 조남제(옥션) | 조대현(가톨릭관동대) |
| | 조성욱(전북대) | 조철기(경북대) | 조혜영(한국산업단지공단) |
| | 최영은(건국대) | 최은영(한국도시연구소) | 홍금수(고려대) |
| | 황상일(경북대) | | |
| 고 문 | 권용우(전 성신여대) | 권혁재(전 고려대) | 김일기(전 한국교원대) |
| | 류우익(전 서울대) | 박삼옥(전 서울대) | 박영한(전 서울대) |
| | 박찬석(전 경북대) | 서찬기(전 경북대) | 손 일(전 부산대) |
| | 양보경(성신여대) | 이기석(전 서울대) | 이민부(한국교원대) |
| | 이승호(건국대) | 이정록(전남대) | 이철우(경북대) |
| | 이현영(전 건국대) | 장보용(전 전남대) | 최형규(성지문화사) |
| | 허우궁(전 서울대) | 형기주(전 동국대) | |
| 사 무 장 | 김지영 | | |

학회비 납부 및 회원 가입 안내

본 학회의 원활한 운영을 위해 학회비를 납부하지 않은 회원님들께서는 빠른 시일 내에 납부하여 주시면 감사하겠습니다. 회비납부 계좌는 다음과 같습니다.

국민은행 : 841-25-0006-810 [예금주 : (사)대한지리학회]

본 학회의 회원가입을 희망하시는 분은 홈페이지에 탑재되어 있는 회원가입신청서를 작성하여 대한지리학회 사무실(geography77@hanmail.net)로 보내주시기 바랍니다.

편집위원회

| | | | |
|------|---|---|----------|
| 위원장 | 임석희(대구대) | | |
| 부위원장 | 이자원(성신여대) | 오충원(남서울대) | 임은진(공주대) |
| | 류호상(전북대) | | |
| 편집위원 | 구동회(부산대) | 김숙진(건국대) | 김영훈(교원대) |
| | 백선혜(서울연구원) | 신혜란(서울대) | 이광률(경북대) |
| | 이용균(전남대) | 이정훈(경기연구원) | 장동호(공주대) |
| | 정성훈(강원대) | 정현주(서울대) | 최광용(제주대) |
| | 홍금수(고려대) | Heejun Chang(Portland State University) | |
| | Robert Hassink(Kiel University, Germany) | | |
| | Jeong Chang Seong(University of West Georgia) | | |

대한지리학회지

제54권 제5호(2019년 10월 31일)

발행인 : 이용우

발행처 : 사단법인 대한지리학회

© 04376 서울시 용산구 새창로 213-12, 1413호

전화 : (02) 875-1463, 팩스 : (02) 876-2853

e-mail : geography77@hanmail.net

http://www.kgeography.or.kr

조판·인쇄 : (주)푸른길

전화 : (02) 6942-9570~2, 523-2907 팩스 : (02) 523-2951

Copyright © 2019, 사단법인 대한지리학회

본지에 게재된 논문, 기사, 도표 등의 저작권은 본 학회가 갖는다.

• 이 학술지는 2019년도 정부재원(교육부)으로 한국연구재단의 지원을 받아 출판되었음.