

# 공간적 자기상관 통계량의 고유벡터 간 비교 연구: 공간근접성행렬의 효과와 공간적 회귀분석에의 함의를 중심으로

이상일\* · 조대현\*\* · 이민과\*\*\*

## Comparing Spatial Eigenvectors Associated with Spatial Autocorrelation Statistics: A Focus on the Effects of Spatial Proximity Matrices and the Implications for Spatial Regression Models

Sang-Il Lee\* · Daeheon Cho\*\* · Minpa Lee\*\*\*

**요약** : 본 연구의 주된 목적은, 상이한 공간적 자기상관 통계량(모런 통계량, 기어리 통계량, S\* 통계량)과 상이한 공간근접성행렬(이항연접성행렬과 행표준화행렬)로부터 추출된 고유벡터의 공간 패턴을 체계적으로 비교함으로써 고유벡터의 다양성에 대한 일반론을 정립하고, 이러한 고유벡터의 다양성이 고유벡터공간필터링 접근에 대해 갖는 함의를 실 데이터를 통해 검토하는 것이다. 고유벡터간 일치도 평가를 위해 일종의 상관관계 매트릭스 그래프가 사용되었고, 대각선성과 대응성이라는 두 가지 기준에 의거해 해석되었다. 이와 관련된 결과를 요약하면 다음과 같다. 첫째, 동일한 공간적 자기상관 통계량에 상이한 공간근접성행렬을 적용한 결과 상당히 이질적인 고유벡터의 세트가 추출되었다. 둘째, 상이한 공간적 자기상관 통계량 쌍 간에 일치성의 정도와 양상에서 상당한 차이가 있는 것으로 드러났고, 공간근접성행렬의 효과도 현저한 것으로 나타났다. 고유벡터의 다양성이 공간적 회귀분석에 가지는 함의를 분석하기 위해 푸에르토리코의 경험 데이터에 대해 6개의 서로 다른 ESF 모형을 실행하였다. 세 가지의 기본적인 사항이 관찰되었다. 첫째, 모형 별로 다양한 개수의 고유벡터가 다양한 순위의 조합으로 선정되어 투입된다. 둘째, 투입된 고유벡터의 종류에 따라 ESF 모형이 잔차의 공간적 자기상관을 제거하는 능력이 달라진다. 셋째, 회귀계수의 크기와 유의성이 모형별로 상당한 차이를 보인다. 이러한 기본적인 결과를 바탕으로 두 가지 함의가 도출되었다. 첫째, 기본적으로 잔차의 공간적 자기상관을 가장 잘 제거하는 모형이 가장 우수하다고 말할 수 있다. 둘째, 회귀계수의 크기와 유의성을 비공간적인 기본 모형과 비교하고, 그것을 바탕으로 상이한 ESF 모형들을 평가하는 것이 가능하다.

**주요어** : 공간적 자기상관 통계량, 공간근접성행렬, 공간적 고유벡터, 고유벡터공간필터링

**Abstract** : The main objective of this study is to elucidate the source and aspects of the variability of eigenvectors by comparing the spatial patterns of different eigenvectors in association with different spatial autocorrelation statistics (Moran's *I*, Geary's *c*, and Lee's *S\**) and/or different spatial proximity matrices (binary contiguity-based and row-standardized) and, based on this, to discuss some potential implications for the

본 연구는 국토교통부 국토공간정보연구사업의 연구비지원(과제번호14NSIP-B080144-01)에 의해 수행되었습니다.

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

\*\* 가톨릭관동대학교 지리교육과 조교수(Assistant Professor, Department of Geography Education, Catholic Kwandong University, dhncho@gmail.com)

\*\*\* (주)망고시스템 기술연구소 연구소장(Director of R&D, Institute of Technology, Mango System Inc., minpa.lee@mango-system.com)

eigenvector spatial filtering modeling. A modified form of the correlation matrix graph is used as a visual analytic and two criteria, diagonality and correspondence, are set to evaluate the degree of coincidence between two sets of eigenvectors. Regarding this, two things are observed: (1) the spatial proximity matrix matters even when the same spatial autocorrelation statistic is concerned; (2) different spatial autocorrelation statistics and different spatial proximity matrices are jointly responsible for the variability of spatial eigenvectors. In order to draw some implications of the variability of spatial eigenvectors for the eigenvector spatial filtering approach, six different ESF models are established for the Puerto Rico agricultural data. Regarding this, three things are observed: (1) different numbers and compositions of eigenvectors are selected for the models; (2) different models due to the different eigenvector input have different ability to control spatial autocorrelation in residuals; (3) the magnitude and significance of the regression coefficients vary among the models. Based on these, two implications are drawn. First, a better model should remove spatial autocorrelation in residuals better. Second, another criterion can be set based on a comparison between ESF models and the basic model in terms of the magnitude and significance of regression coefficients

**Key Words** : spatial autocorrelation statistics, spatial proximity matrix, spatial eigenvectors, eigenvector spatial filtering (ESF)

## 1. 서론

공간적 자기상관을 “관측개체들의 위치 유사성과 속성 유사성 간의 특정한 관련성”으로 정의할 때, 모든 종류의 공간적 자기상관 통계량(spatial autocorrelation statistics, SAS) 혹은 공간연관성통계량(spatial association statistics)은 반드시 “공간단위 간의 지리적 관련성을 정의하는 요소와 공간단위가 보유한 속성값 간의 수치적 유사성을 정의하는 요소로 구성되어야 한다.”(이상일 등, 2015, 330). 따라서 모든 종류의 SAS는 이러한 두 요소를 정식화한 두 매트릭스 간의 ‘교차곱 통계량(cross-product statistics)’의 형태를 띠게 된다(Mantel, 1967; Hubert *et al.*, 1981; Getis, 1991; Anselin, 1995; Lee, 2004; 2009). 이 때 전자의 매트릭스를 ‘공간근접성 행렬(spatial proximity matrix, SPM)’ (Bailey and Gatrell, 1995) 혹은 ‘공간가중행렬(spatial weights matrix)’이라고 부른다. 비록, SPM을 어떻게 정의하느냐가 공간통계분석의 결과에 무시할 수 없는 영향을 끼친다는 사실이 밝혀져 있고(Griffith, 1996a), 매우 다양한 방식의 SPM이 존재할 수 있음이 강조되어 왔지만(Dubin, 2009), SAS를 활용한 공간통계분석에서 SPM은 단순히 주어진 것으로 간주되어 왔으며,

그 자체가 연구의 핵심 사항으로 등장한 경우는 매우 드물었다. 그런데 최근의 공간통계학적 연구를 살펴보면 SPM에 대한 기존의 시각이 바뀔 필요가 있음을 알게 된다.

일단의 연구는 SPM 혹은 ‘조정된(adjusted) SPM’ 으로부터 각 SAS가 가진 통계량으로서의 특성에 대한 중요한 정보를 얻어낼 수 있다는 사실을 보여 주고 있다. 모런 통계량(Moran’s  $I$ )의 조정된 SPM으로부터 정규성(normality) 가정에 의지한 네 개의 중심적률(central moments)을 추출할 수 있을 뿐만 아니라(Henshaw, 1966; 1968; Hepple, 1998; Tiefelsdorf, 2000), 통계량의 ‘가능치 범위(feasible range)’ (de Jong *et al.*, 1984)와 상이한 공간적 자기상관 수준을 보여주는 공간 패턴을 추출할 수 있다는 점이 밝혀졌다(Boots and Tiefelsdorf, 2000; Griffith, 2003; Tiefelsdorf and Griffith, 2007). 이러한 방법은 전역적 통계량 뿐만 아니라 국지적 모런 통계량(local Moran’s  $I_i$ )에도 적용 가능한 것으로 드러났다(Boots and Tiefelsdorf, 2000). 더 나아가, 이상일 등(2015; 2016)의 연구는 SPM을 이용한 이러한 방법론을 일반화함으로써 다른 SPM의 특성을 파악하는 데도 확장 적용될 수 있음을 보여주었다. 또한 이상일 등(2015; 2016)은 기어리 통계량(Geary’s  $d$ )과 이상일의  $S^*$  통계량 역시 교차곱 통계량의 일반 형태를 취

한다는 점에서 일반화된 방법론의 적용가능성을 보여주고, 각 통계량이 고유한 ‘조정된 SPM’을 가진다는 점에서 서로 다른 중심적률, 가능치 범위, 공간 패턴이 도출된다는 점을 보여주었다.

SPM이 핵심적인 역할을 하는 공간통계학적 방법론 상의 또 다른 진전은 고유벡터공간필터링(eigen-vector spatial filtering, ESF) 기법인데, 이는 공간적 회귀분석(spatial regression)의 한 방법론으로 제시된 것이다. 공간적 회귀분석이란 넓게 정의하자면 공간 데이터의 본질을 반영한, 혹은 공간적 효과를 다룰 수 있게 디자인된, 모든 종류의 회귀분석이라 할 수 있지만, 좁게 정의하자면 잔차의 공간적 자기상관의 문제를 해결하고자 하는 모든 종류의 회귀분석을 의미하는 것으로 이해할 수 있다. ESF는 다니엘 그리피스(Daniel A. Griffith) 교수의 선구적인 연구(1996b; 2000; 2003)에 기반하고 있는데, 선형회귀모형(혹은 OLS 회귀분석)의 구조를 그대로 따르는 대신 SPM에서 추출한 고유벡터를 일종의 통제변수(control variables)로 사용함으로써 잔차에 존재하는 공간적 자기상관의 문제를 해결하고자 한다. 다시 말해 고유벡터를 ‘인공(synthetic)’ 변수 혹은 ‘대리(proxy)’ 변수로 사용하여 공간 데이터에 내재되어 있는 공간적 의존성을 걸러냄으로써 공간회귀모형이 본질적으로 겪는 ‘오특정화의 문제(misspecification problem)’를 해결하고자 한다(Griffith, 2003; Thayn, 2017). 그런데 ESF는 거의 예외 없이 SPM 자체 혹은 모런 통계량의 조정된 SPM에서 추출된 고유벡터만을 사용한다.

본 연구는 위에서 살펴본 SPM에 초점을 둔 두 종류의 공간통계학적 방법론을 새로운 연구과제로 연결하여 발전시키고자 한다. Griffith(2003)는 모런 통계량에서 추출한 고유벡터의 공간 패턴과 기어리 통계량에서 추출한 고유벡터의 공간 패턴 사이에 일대일 일치가 존재하지 않음을 보고한 바 있다. 또한 이상일 등(2015)은 세 종류의 SAS로부터 추출된 고유벡터의 공간 패턴 상에 상당한 차이가 있음을 시사한 바 있다. 그러나 이들 연구는 모두 체계적인 연구 디자인을 개발하지 못했고, 따라서 일반화가 가능한 수준의 결론에는 도달하지 못했다. 또한 ESF 연구에서

모런 통계량이 아닌 다른 SAS에서 추출된 고유벡터가 투입된 적은 전무하다. 서로 다른 SAS는 공간적 자기상관을 서로 다른 방식으로 측정하기 때문에, 특수성을 보유한 조정된 SPM으로부터 추출된 고유벡터가 ESF 회귀분석의 결과에 어떠한 차이점을 초래하게 될지를 살펴보는 것은 의미 있는 일이 될 것이다. 더 나아가 동일한 SAS라 하더라도 SPM의 종류(이항근접성행렬 혹은 행표준화행렬)에 따라 추출된 고유벡터가 다를 가능성 역시 존재한다. 따라서 본 연구의 주된 목적은, 상이한 SAS(Moran's  $I$ , Geary's  $c$ , Lee's  $S^*$ )와 상이한 SPM으로부터 추출된 고유벡터의 공간 패턴을 체계적으로 상호 비교함으로써 고유벡터의 다양성에 대한 일반론을 정립하고, 이러한 고유벡터의 다양성이 ESF 접근에 대해 갖는 함의를 실 데이터를 통해 검토하는 것이다.

## 2. 연구 방법론

### 1) ‘조정된 공간근접성행렬’과 고유벡터의 추출

표 1은 본 연구에서 사용되는 세 종류의 SAS에 대한 수식을 보여주고 있다. 매트릭스 표현은 ‘편도 벡터(deviate vector)( $\delta$ )’(원 벡터에서 평균만 뺀 것)를 통해 ‘이차형식의 비(ratio of quadratic forms)’로 각 통계량을 표현한 것이다(Tiefelsdorf, 2000; Lee, 2008). 표 1에 나타나 있는 수식이 SPM에 대한 특정한 사고를 반영한 것이라는 사실에 주목할 필요가 있다. SPM은 다양한 이분법에 의거해 서로 다른 방식으로 정의될 수 있다. 즉, SPM은 연결성-기반 혹은 거리-기반일 수 있고, 이항비표준화 혹은 행표준화일 수 있고, 주대각 요소가 0일 수도 있고 0이 아닐 수도 있다. 이 모든 다양성을 포괄하기 위해 SPM에 대한 기호로  $\mathbf{V}$ 의 사용이 제안되었으며(Tiefelsdorf, 2000; Lee, 2004; 2009), SAS에 본질적인 차이를 발생시키는 것이 SPM의 주대각 요소라는 점을 강조하기 위해  $\mathbf{V}$ 를  $\mathbf{V}^0$ 와  $\mathbf{V}^*$ 로 구분할 것이 제안되었다(이상일 등, 2015; Lee, 2017).

표 1. 일변량 공간적 자기상관 통계량

	합산 표현	매트릭스 표현
모런 통계량	$I = \frac{n \sum_i \sum_j v_{ij}^0 (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j v_{ij}^0 \sum_i (x_i - \bar{x})^2}$	$I = \frac{n}{\mathbf{1}^T \mathbf{V}^0 \mathbf{1}} \frac{\delta^T \mathbf{V}^0 \delta}{\delta^T \delta}$
기어리 통계량	$c = \frac{n-1}{2 \sum_i \sum_j v_{ij}^0} \frac{\sum_i \sum_j v_{ij}^0 (x_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}$	$c = \frac{n-1}{\mathbf{1}^T \mathbf{V}^0 \mathbf{1}} \frac{\delta^T (\mathbf{\Omega} - \mathbf{V}^0) \delta}{\delta^T \delta}$
S*통계량	$S^* = \frac{n}{\sum_i (\sum_j v_{ij}^*)^2} \frac{\sum_i (\sum_j v_{ij}^* (x_i - \bar{x}))^2}{\sum_i (x_i - \bar{x})^2}$	$S^* = \frac{n}{\mathbf{1}^T (\mathbf{V}^{*T} \mathbf{V}^*) \mathbf{1}} \frac{\delta^T (\mathbf{V}^{*T} \mathbf{V}^*) \delta}{\delta^T \delta}$

주: 수식에 대한 자세한 설명은 이상일 등(2015)과 Lee(2004) 참조.

이상일 등(2015)이 보여준 바처럼, 표 1에 나타나 있는 세 SAS는 모두 다음과 같은 이차형식의 비로 표현될 수 있다.

$$\Gamma = \frac{\delta^T \mathbf{T} \delta}{\delta^T \delta} \quad (1)$$

이 때  $\mathbf{T}$ 는 SPM을 전체 공간단위의 수와 공간가중치의 합 등을 통해 재정의한 것으로 일종의 ‘정규화된(normalized)’ SPM으로 간주할 수 있다(Lee, 2009; 이상일 등, 2015).

$$\begin{aligned} \mathbf{T}(I) &\equiv n \frac{\mathbf{V}^0}{\mathbf{1}^T \mathbf{V}^0 \mathbf{1}}, \quad \mathbf{T}(c) \equiv (n-1) \frac{(\mathbf{\Omega} - \mathbf{V}^0)}{\mathbf{1}^T \mathbf{V}^0 \mathbf{1}}, \\ \mathbf{T}(S^*) &\equiv n \frac{\mathbf{V}^{*T} \mathbf{V}^*}{\mathbf{1}^T (\mathbf{V}^{*T} \mathbf{V}^*) \mathbf{1}} \end{aligned} \quad (2)$$

이를 바탕으로 최종적인 조정된 SPM을 의미하는  $\mathbf{K}$  매트릭스를 도출할 수 있다(Tiefelsdorf, 2000).

$$\mathbf{K} \equiv \mathbf{M}_{(1)} \frac{1}{2} (\mathbf{T} + \mathbf{T}^T) \mathbf{M}_{(1)} \quad (3)$$

여기서,  $\mathbf{M}_{(1)} \equiv \mathbf{I} - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T$ 인데,  $\mathbf{M}_{(1)}$ 는 독립변수가 존재하지 않는 상태에서의 ‘사영행렬(projection matrix)’을 의미한다(Tiefelsdorf and Boots, 1995).  $\mathbf{I}$ 는 단위행렬(unit matrix)이고,  $\mathbf{1}$ 은 요소가 모두 1인 열벡터(column vector)이다.

이  $\mathbf{K}$  매트릭스를 분해하면  $n$ 개의 고유치(eigen-values),  $\{\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_n\}$  ( $\lambda_i \geq \lambda_{i+1}$ )와 그에 상응하는  $n$ 개의 고유벡터(eigenvectors),  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_i, \dots, \mathbf{e}_n\}$ 를 추

출할 수 있는데, 각 고유벡터는 고유치에 상응하는 정도의 공간적 자기상관을 보여주는 벡터이다. 여기서 중요한 점은 고유벡터를 지도화하면 해당 고유치가 나타내는 수준의 공간적 자기상관을 가진 공간 패턴이 드러난다는 점이다(Boots and Tiefelsdorf, 2000; Griffith, 2003; Tiefelsdorf and Griffith, 2007; 이상일 등, 2015). 이는 자연스럽게 ESF 접근과 연결되어 있다.

## 2) 고유벡터공간필터링 접근

ESF는 공간적 회귀분석의 하나로 창안된 공간통계학적 모델링 방법론이다. 이것의 특징을 이해하기 위해 우선 가장 기본적인 비공간적 회귀모형에서부터 시작할 필요가 있다. 일반적인 선형회귀모형 혹은 OLS 회귀모형은 다음과 같이 표현된다.

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon} \quad (4)$$

여기서  $\mathbf{y}$ 는 종속변수 벡터,  $\mathbf{X}$ 는 독립변수 매트릭스, 그리고  $\boldsymbol{\varepsilon}$ 는 잔차 벡터이다. 여기서 잔차의 각 요소는 상호독립적인 것으로 간주된다. 그러나 모델에 개입된 변수(특히 종속변수)에 공간적 의존성이 존재하게 되면 이러한 정식화는 더 이상 유효하지 않게 되며, 공간적 자기상관의 문제를 해결하기 위한 다른 방식의 모델 특정화가 시도되어야 한다. 가장 널리 사용되어 온 공간적 회귀모형은 주로 공간계량경

제학(spatial econometrics)에 기반한 다양한 종류의 공간자기회귀모형(spatial autoregressive models)이다(Anselin, 1988; Anselin, 2009; Anselin and Rey, 2014). 이러한 접근은 잔차의 공간적 자기상관이 종속변수의 공간적 의존성 때문에 발생한다고 전제하는 공간지체모형(spatial lag models)과 잔차 자체가 본원적으로 서로 연관되어 있다는 공간오차모형(spatial error models)으로 나뉘어 진다. 그러나 핵심적인 사항은 공간적 자기회귀 계수(spatial autoregressive coefficients)를 추정해냄으로써 공간적 자기상관이 없는 잔차가 도출되도록 한다는 데 있다.

ESF는 공간적 자기상관이 없는 잔차가 도출되도록 하려는 목적을 가졌다는 점에서는 공간자기회귀모형과 동일하지만 기본 개념과 추정 방법에서 상이한 접근법을 취한다. ESF 모형은 다음과 같은 형태를 취한다.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\eta} \quad (5)$$

여기서  $\mathbf{E}$ 는 선택된 고유벡터로 구성된 매트릭스,  $\boldsymbol{\gamma}$ 는 고유벡터 매트릭스에 대한 회귀계수 벡터, 그리고  $\boldsymbol{\eta}$ 는 공간적 자기상관이 존재하지 않는 잔차이다. 식 (4)와 비교해 보면,  $\boldsymbol{\varepsilon} = \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\eta}$ 이 성립함을 알 수 있는데, 이는 ESF의 기본 전제가 잔차는 공간적 자기상관을 가진 부분과 그렇지 않는 부분으로 구성되어 있다는 사실을 잘 보여준다. 따라서 ESF의 기본 전략은 잔차로 갈 공간적 자기상관을 고유벡터가 흡수하게 함으로써 잔차에 공간적 자기상관이 남아 있지 않게 하는 것이다. 고유벡터는 실제로 존재하는 변수가 아니라 만들어진 변수라는 점에서 ‘인공적’이며, 특정한 공간 패턴을 보유한 누락된 독립변수를 대신한다는 의미에서 ‘대리적’이며, 공간적 효과를 배제한 순수한 독립변수의 설명력을 평가할 수 있게 해준다는 의미에서 ‘통계적’이다. 이 모든 논리의 근간에는 놀라운 가정이 숨어 있는데, 개입된 변수의 공간 패턴과 관계 없이, 공간근접성행렬로부터 도출된 고유벡터들은 그 공간구조에 내재되어 있는 본원적 공간 패턴을 대변하고 있다는 사실이다.

ESF 모형은 공간자기회귀모형에 비해 몇 가지 이점이 있다. 첫째, 모델 추정의 편의성이다. ESF는 OLS

추정을 그대로 이용할 수 있다(Griffith, 2010). 둘째, ESF의 기본 공리 하에서 공간지체모형과 공간오차모형을 수용할 수 있다(Tiefelsdorf and Griffith, 2007). 그러나 ESF의 가장 큰 난점은 고유벡터를 선택하는 방식의 문제이다. 기본적으로 고유벡터들은 서로 직교하기 때문에 상관관계가 존재하지 않는다. 이는 다공선성(multicollinearity)의 문제로부터 벗어날 수 있다는 크나큰 이점으로 작용한다. 고유벡터의 순서는 앞에서도 언급한 것처럼 공간적 자기상관의 정도 순이다. 첫 번째 고유벡터는 최대의 (양의) 공간적 자기상관을 보여주는 공간 패턴과 관련되어 있고, 두 번째 고유벡터는 첫 번째 고유벡터와 상관관계가 없으면서 최대의 공간적 자기상관을 보여주는 공간 패턴과 관련되어 있다. 마지막 고유벡터는 가능한 최소의 공간적 자기상관(가장 큰 음의 공간적 자기상관)을 보여주는 공간 패턴과 관련되어 있다. 모든 고유벡터를 회귀식에 투입할 수 없기 때문에 어떠한 방식으로 고유벡터를 선정할 것인가가 핵심적인 방법론적 난제일 수밖에 없다. 가장 널리 사용되고 있는 방식은(Griffith, 2010; Chun and Griffith, 2013; Chun *et al.*, 2016; Griffith, 2017), 우선 최대 공간적 자기상관의 최소한 25% 수준을 보여주는 고유벡터를 선정한 후, 단계적 회귀분석(stepwise regression)을 실행하는 것이다.

### 3) 분석 디자인

본 연구에서 분석은 크게 두 가지 부분으로 나뉘어 진다. 하나는 세 SAS로부터 도출된 고유벡터의 공간 패턴을 상호 비교하는 것이고, 또 다른 하나는 상이한 SAS로부터 도출된 고유벡터를 ESF 회귀분석에 투입했을 때 어떠한 결과가 나타나는지를 분석하는 것이다.

첫 번째 연구 과제는 다시 두 가지로 구분된다. 우선 고유벡터 추출에 대한 상이한 SPM의 효과를 검토한다. 주로 이항비표준화 SPM과 행표준화 SPM 간의 차이에 주목하고자 한다. 이를 위해 네 종류의 SPM을 구성하였는데,  $\mathbf{C}^0$ 는 주대각 요소가 0인 이항 연결성 SPM,  $\mathbf{C}^*$ 은  $\mathbf{C}^0$ 의 주대각선에 1을 넣은 SPM,  $\mathbf{W}^0$

는  $C^0$ 의 행표준화 SPM,  $W^*$ 은  $C^*$ 의 행표준화 SPM이다.  $C^0$ 와  $W^0$ 은 모런과 기어리 통계량에 대해,  $C^*$ 와  $W^*$ 은  $S^*$  통계량에 대해 사용하도록 한다. 예를 들어 모런 통계량에서  $C^0$ 를 사용하여 추출한 고유벡터와  $W^0$ 를 사용하여 추출한 고유벡터 사이에 얼마만큼의 유사성이 있는지 알아본다. 동일 SAS 내에서 상이한 SPM의 효과를 우선 파악하는 것은 그 뒤를 이어 진행될 상이한 SAS 간의 비교분석의 결과를 보다 정확하게 해석하기 위함이다. 즉, 고유벡터 간의 차이가 SAS의 차이에 기인한 것인지 SPM의 차이에 기인한 것인지 정확히 구분할 필요가 있기 때문이다.

첫 번째 연구 과제를 수행하기 위해 육각형 정다각 테셀레이션을 활용한다(이상일 등, 2015, 그림 3 참조). 64개 육각형으로 구성된 테셀레이션을 기본적으로 사용하지만(그림 1(a)), 공간단위 개수의 효과를 검토하기 위해 256개 육각형으로 구성된 테셀레이션도 보조적으로 사용한다. 분석 도구로는 국토교통부 국토공간정보연구 사업의 공간정보 SW활용을 위한 오픈소스 가공기술개발 과제로 개발되고 있는 분석 도구 및 R을 함께 사용하였다(이상일 등, 2015; 2016; 국토교통과학기술진흥원, 2017). 연구 결과를 효과적으로 보여주기 위해 특정한 방식의 시각화 기법이 사용된다. 동일한 SAS에 대해 두 개의 서로 다른 SPM의 결과를 비교하는 것이건, 동일한 SPM에 대한 두 개의 서로 다른 SAS의 결과를 비교하는 것이건, 기본적으로 64개 고유벡터와 64개 고유벡터간의 관계를 살펴보아야 한다. 이를 위해 일종의 ‘상관관계 행렬 시각화(correlation matrix visualization)’ 방법

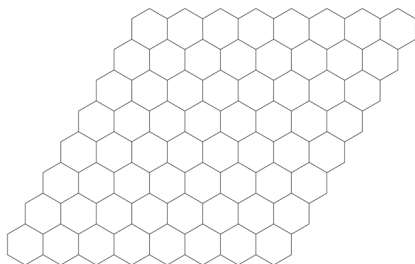
을 사용했다. 일반적인 경우라면 64개의 변수간 관계를 64X64 매트릭스 형태로 표현하는 것이겠지만, 본 연구의 경우는 64개 벡터와 64개 벡터를 서로 교차하여 상관계수를 산출하고 그것을 64x64 매트릭스 형태로 표현한다는데 독특함이 있다. 구체적인 시각화에는 R 환경에서 작동하는 ‘ggcorrplot’ 패키지가 사용되었다(<https://cran.r-project.org/web/packages/ggcorrplot/>).

두 번째 연구 과제를 수행하기 위해 푸에르토리코의 73개 자치시(municipality)에 대한 농업 관련 데이터를 사용한다(그림 1(b) 참조). 이 데이터를 사용하는 것은 모런 통계량의 고유벡터를 사용한 ESF 회귀분석 결과가 이미 출판되어 있기 때문에(Chun and Griffith, 2013; Griffith and Chun, 2014), 다른 SAS에서 추출된 고유벡터를 투입한 ESF 회귀분석의 결과와 비교하기에 용이하기 때문이다.

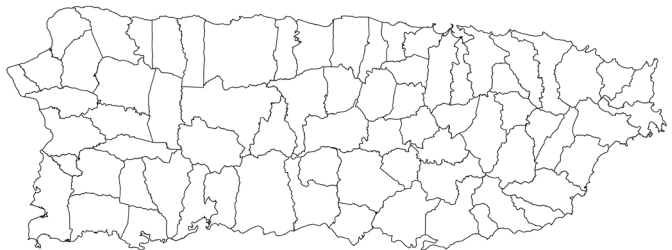
### 3. 연구 결과

#### 1) 고유벡터 간 상관관계 분석

식 (2)와 식(3)에 의거해 각 SAS 별로 고유벡터를 추출할 수 있는데, 행표준화의 효과를 밝혀 내기 위해 모런 통계량과 기어리 통계량에 대해서는  $C^0$ 와  $W^0$ 를,  $S^*$  통계량에 대해서는  $C^*$ 와  $W^*$ 를 적용하였다. 따라서 각 SAS별로 두 세트의 고유벡터가 산출되었다. 여



(a) 육각형 정다각 테셀레이션( $n=64$ )



(b) 푸에르토리코 행정구역( $n=73$ )

그림 1. 육각형 정다각 테셀레이션과 푸에르토리코 행정구역

기서 기어리 통계량의 고유벡터 추출 과정에 대해 언급할 필요가 있다. 주지하는 바처럼 기어리 통계량은 기댓값이 1이고, 0에 가까워질수록 양의 공간적 자기상관이, 2에 가까워질수록 음의 공간적 자기상관이 강하다는 것을 나타낸다. 이는 값이 클수록 양의 공간적 자기상관이 커지는 다른 두 SAS와의 비교분석에 난점을 제공한다. 이를 극복하기 위해  $1-c$ 의 수식을 통해 기어리 통계량을 변형하였다(O'Sullivan and Unwin, 2010, 211). 이렇게 하면 기어리 통계량이 대략 -1에서 1 사이의 값을 갖게 되며(기댓값은 0), 양의 값이 클수록 양의 공간적 자기상관이 커지는 형태가 된다. 이에 따라 고유벡터의 순서도 정반대로 바뀌게 되는데, 첫 번째 고유벡터가 원래 음의 공간적 자기상관이 가장 큰 것에서 양의 공간적 자기상관이 가장 큰 것으로 바뀌게 된다.

#### (1) 공간근접성행렬의 효과

공간근접성행렬의 효과만을 분리하여 살펴보기 위해, 동일한 SAS 내에서 서로 다른 두 개의 SPM을 적용하여 추출된 고유벡터 간의 상관관계를 분석하였다. 그림 2는 이것을 시각화한 것이다. 예를 들어 그림 2(a)를 보면, x-축에는  $C^0$ 가 적용된 모런 통계량의 64개 고유벡터가, y-축에는  $W^0$ 가 적용된 모런 통계량의 64개 고유벡터가 놓이게 된다. 그리고 4,096개 셀(64X64)은 x-축의 특정한 고유벡터와 y-축의 특정한 고유벡터 간의 상관계수를 나타내는데, 각 상관관계의 방향(양의 상관관계는 빨간색, 음의 상관관계는 파란색)과 강도(음영의 정도)에 따라 상이하게 표시된다. 이 그래프를 올바르게 해석하기 위해 두 가지 점이 지적될 필요가 있다. 첫째, 고유벡터의 성격상 상관관계의 방향성은 큰 의미가 없으며 강도가 훨씬 더 큰 의미를 가진다. 왜냐하면 상관관계가 1이거나 -1인 경우 모두 두 고유벡터가 동일하다는 점을 의미한다. 둘째, 한 고유벡터가 둘 이상의 고유벡터와 극단적으로 높은 상관관계(예를 들어, 0.9보다 크거나 -0.9보다 작은 경우)를 가질 가능성은 거의 없다. 왜냐하면 고유벡터의 세트 내에서 각 고유벡터간 상관관계는 0이므로, x-축의 한 고유벡터가 y-축의 한 고유벡터와 극단적으로 높은 상관관계를 보일 경우,

y-축의 또 다른 고유벡터와는 이와 유사한 수준의 상관관계를 보이는 것이 불가능하기 때문이다.

이 두 가지 점을 전제된 상태에서 패턴 해석을 위한 두 가지 개념을 제안하고자 한다. 하나는 '대각선성(diagonality)' 개념이고, 또 다른 하나는 '대응성(correspondence)' 개념이다. 대각선성은 "상대적으로 높은 상관관계를 보이는 셀이 그래프 상의 대각선을 중심으로 형성되는 정도"를 의미한다. 그래프에서 대각선의 형상이 뚜렷하다는 것은 두 고유벡터 세트 사이에서 '서열적 일치성'이 높다는 것을 의미한다. 이에 반해 대응성은 "특정한 고유벡터가 오로지 특정한 고유벡터와만 절대적으로 높은 상관관계를 보이는 정도"를 의미한다. 이는 절대적으로 높은 상관관계가 오로지 특정한 쌍들에서만 배타적으로 나타나고 나머지 셀들은 매우 낮은 상관관계를 나타내는 것으로, 두 고유벡터 세트 사이에서 '일대일 일치성'이 높다는 것을 의미한다. 이 두 가지 개념을 조합하면 그래프의 해석을 좀 더 용이하게 할 수 있다. 예를 들어, 높은 대각선성과 높은 대응성을 보이는 그래프는 대각선 상에만 극단적으로 높은 상관관계가 나타나고 나머지 셀들은 거의 0에 가까운 상관관계를 보이는 경우이다. 이에 반해 낮은 대각선성과 낮은 대응성을 보이는 그래프는 전체 셀들에 서로 다른 수준의 상관관계가 무작위적으로 분포하는 경우이다.

그림 2의 (a)~(c)를 비교해 보면, 상이한 SPM의 효과가 상당함을 알 수 있다. 전체적으로 봤을 때 어느 정도의 대각선성, 즉 대각선 주변에서 높은 상관성을 보이는 셀들이 집중해 있는 경향이 발견되기는 하지만, 일대일 대응성이 현저하다고 말할 수 있는 수준은 아닌 것으로 판단된다. 가장 현저한 양상을 보인 것은 모런 통계량(그림 2(a))으로 최상위에서 중후반 고유벡터 쌍에 이르는 지점까지 대각선성이 현저하다. 하지만 이후에는 대각선성이 사라짐과 동시에 대응성도 현저히 약화된다. 기어리 통계량(그림 2(b))은 최상위 고유벡터 쌍을 제외하고는 대각선성과 일대일 대응성이 모두 약한 것으로 나타났다.  $S^*$  통계량의 경우(그림 2(c))는 두 SAS의 중간 정도 양상을 보이는데, 양 극단 쪽이 가운데에 비해 대각선성과 대응성이 모두 좋은 것으로 나타났다. 그림 2(d)는 2(a)와 동일

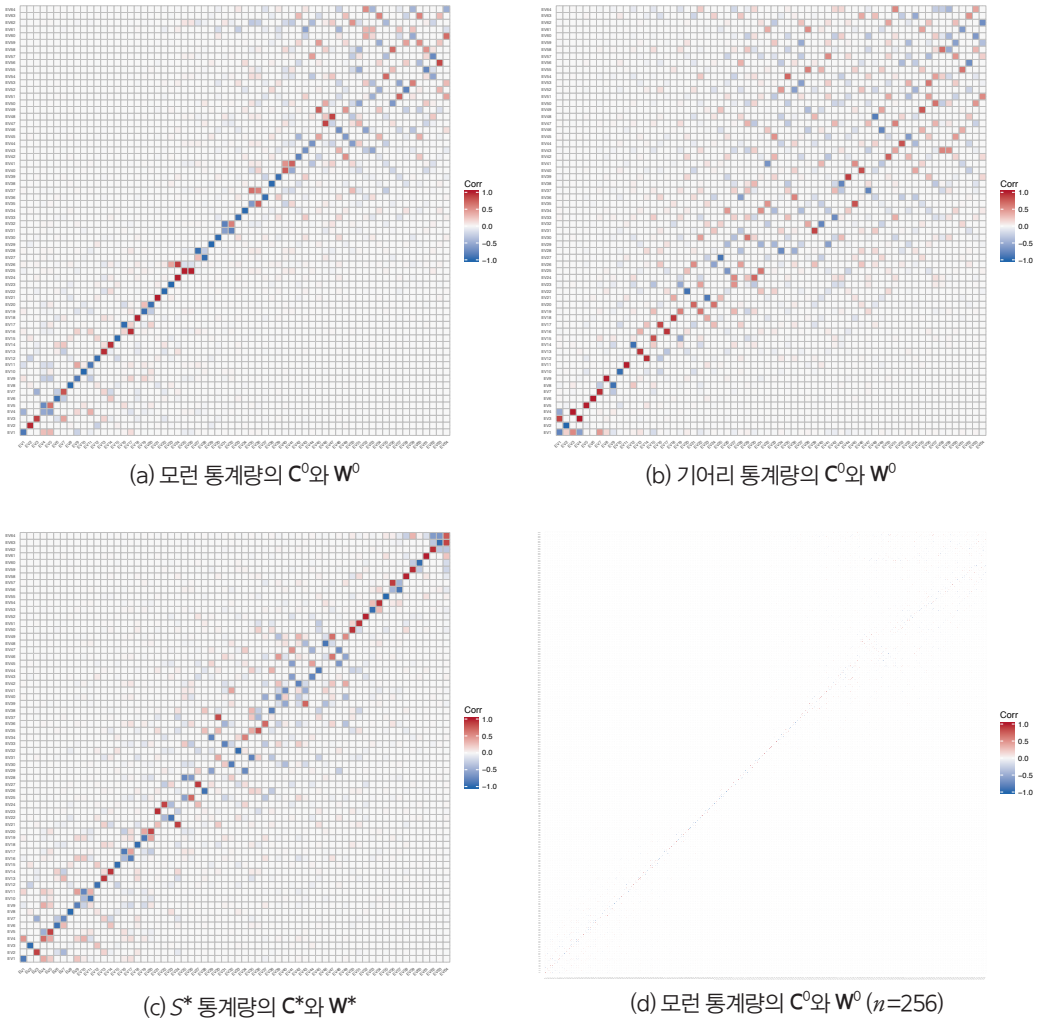


그림 2. 고유벡터 간 상관관계의 시각화: 공간근접성행렬의 효과

한 비교를 256개 육각형 테셀레이션에 적용한 결과를 보여주고 있다. 너무 많은 셀(256×256=65,636)이 개입되어 있어 정확한 분석이 어렵지만 패턴은 비슷한 것으로 판단된다.

표 2는 극단적 일치도를 보이는 고유벡터의 개수를 계산하여 나타낸 것이다. 이는 대응성의 정도를 수치적으로 파악해 보고자 하는 목적에서 시도된 것이다. 그림 2에 대한 분석에서 나타난 것과 마찬가지로, 일대일 대응성은 모런 통계량, S\* 통계량, 기어리 통계량 순이었다. 모런 통계량의 경우 총 64개의 고유

벡터 중 28개의 고유벡터(43.8%)가 극단적 일치도를 보인 반면, 기어리 통계량은 16개(25.0%)에 불과한 것으로 나타났다. 256개 육각형의 경우, 일대일 대응성의 순위는 64개 육각형의 경우와 동일하게 나타났다지만, 전체적으로 극단적 일치도를 보이는 고유벡터의 비중은 현저히 줄어 드는 것으로 나타났다. 특히 기어리 통계량과 S\* 통계량의 경우는 절반 이하의 수준으로 감소했다.



표 2. 극단적 일치도를 보이는 고유벡터의 개수: 공간근접성행렬의 효과

통계량	비교 SPM	극단적 일치도를 보이는 고유벡터의 개수					
		64개 육각형의 경우			256개 육각형의 경우		
		> 0.9	< -0.9	합계	> 0.9	< -0.9	합계
모런	$C^0 - W^0$	9 (14.1)	19 (29.7)	28 (43.8)	41 (16.0)	46 (18.0)	87 (34.0)
기어리	$C^0 - W^0$	11 (17.2)	5 (7.8)	16 (25.0)	14 (5.5)	13 (5.1)	27 (10.5)
$S^*$	$C^* - W^*$	11 (17.2)	13 (20.3)	24 (37.5)	21 (8.2)	24 (9.4)	45 (17.6)

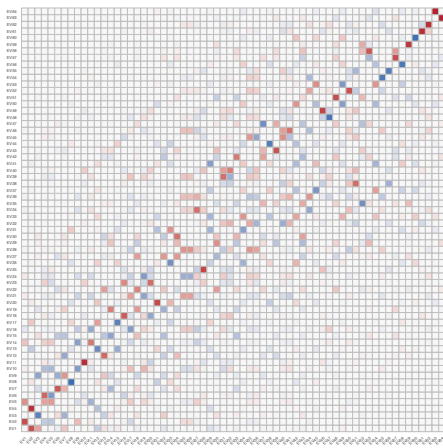
주: 괄호 안의 숫자는 %임.

## (2) 공간적 자기상관 통계량 간의 비교

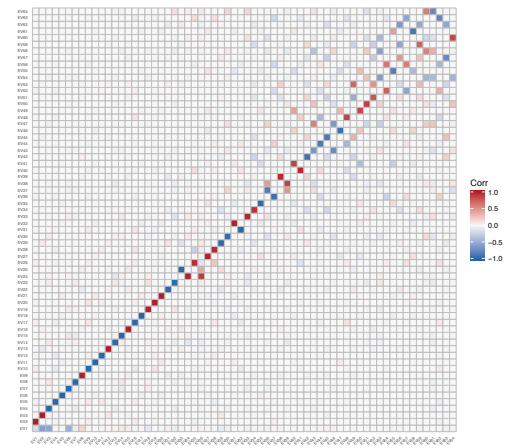
세 SAS에서 추출된 고유벡터 간의 상관관계를 비교하기 위해, 모런 통계량과 기어리 통계량, 모런 통계량과  $S^*$  통계량, 기어리 통계량과  $S^*$  통계량 등 세 쌍의 조합을 설정하였다. SPM의 효과도 동시에 살펴 보기 위해 이항비표준화 SPM과 행표준화 SPM을 비교하였다. 예를 들어 모런 통계량과 기어리 통계량을 비교하는 경우,  $C^0$ 를 적용하여 추출된 고유벡터간의 상관관계를 분석하였을 뿐만 아니라  $W^0$ 를 적용하여 추출된 고유벡터 간의 상관관계도 분석하였다.

그림 3은 SAS 간 상관관계가 매우 다양하게 나타날 뿐만 아니라 SPM의 효과 역시 무시할 수 없다는 것을 뚜렷이 보여주고 있다. 그림 3의 (a)와 (b)를 살펴보면, 다른 두 쌍에 비해 모런 통계량과 기어리 통계량의 상관관계 양상이 그림 2에서 살펴보았던 패턴과 가장 유사함을 알 수 있다. 특히 행표준화 SPM이 사용된 (b)에서 대각선성이 매우 두드러지게 나타남을 알 수 있다. 더구나 대응성도 최상위에서 중간 순위에 이르는 구간에서 상당히 뚜렷하게 나타나고 있다. 이에 반해 그림 3(a)의 대각선성과 대응성은 상대적으로 매우 낮게 나타나고 있는데, 이항비표준화 SPM을 사용하느냐 행표준화 SPM을 사용하느냐의 선택이 고유벡터간 일치도에서 상당한 정도의 차이를 만들어 낼 수 있다는 점을 명확하게 보여주고 있다. 그림 3(c)와 3(d)는 모런 통계량과  $S^*$  통계량간의 상관관계를 보여주고 있는데, 매우 특이한 양상을 보

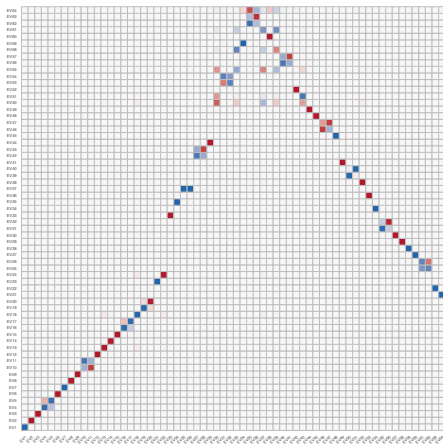
여주고 있다. 특히 3(c)는 매우 흥미로운 패턴을 보여주고 있는데, 대각선성은 매우 낮는데 반해 대응성은 극단적으로 높게 나타나 있다. 이는 이항비표준화 SPM을 적용했을 때 두 SAS에서 추출된 고유벡터의 공간 패턴 상에 호환성이 매우 높다는 것을 의미한다. 낮은 대각선성을 보여주고 있는 것은 두 SAS의 본질적 특성에서의 차이점을 반영하고 있는 것으로 보인다. 모런 통계량은 음의 공간적 자기상관에도 매우 민감하게 반응하며 정밀하게 측정하지만  $S^*$  통계량은 게티스-오드의  $G_i^*$  통계량과 마찬가지로 양의 공간적 자기상관에 집중하며, 음의 공간적 자기상관과 공간적 자기상관의 부재를 잘 구분하여 측정하지 못한다(이상일 등, 2015; 2016). 이러한 차이로 말미암아 전반부에서는 대각선 형상과 함께 최소한 양의 관련성이 나타나고 있는 반면, 후반에서는 음의 관련성이 나타나게 된 것으로 보인다. 또한 모런 통계량과 달리  $S^*$  통계량은 주대각 요소가 0이 아닌 SPM을 사용한다는 사실도 이러한 차이에 일조했을 것으로 추정된다. 그런데 흥미로운 점은 이러한 높은 일치도가 행표준화 SPM을 적용할 경우 상당히 훼손된다는 점이다. 모런 통계량과 기어리 통계량이 행표준화 SPM에서 더 높은 일치도를 보인 것과 대조되는 부분이다. 그림 3의 (e)와 (f)를 살펴보면, 기어리 통계량과  $S^*$  통계량의 상관성은 다른 두 쌍의 상관성에 비해 상대적으로 낮은 것을 알 수 있다. 행표준화 SPM이 적용된 (f)는 (d)와 매우 유사한 패턴을 보이고 있는 반면, (e)에서



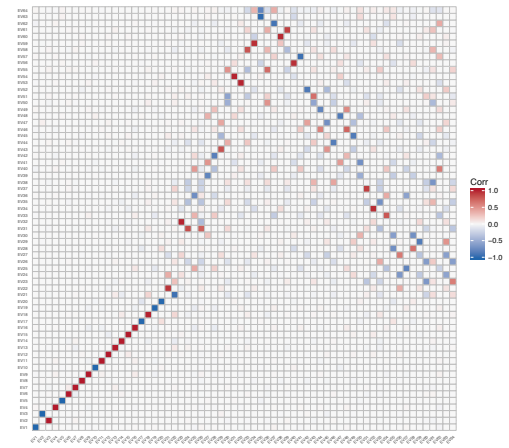
(a) 모런-기어리( $C^0-C^0$ )



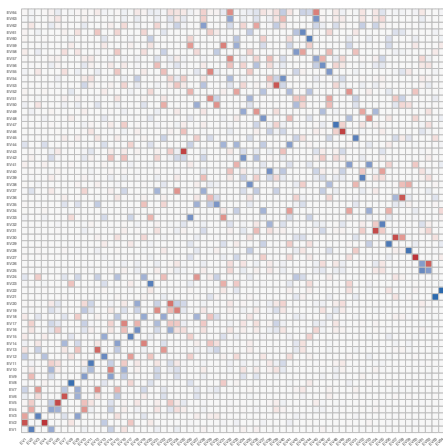
(b) 모런-기어리( $W^0-W^0$ )



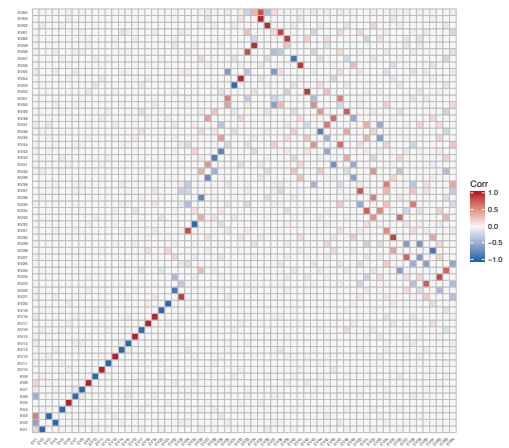
(c) 모런- $S^*(C^0-C^*)$



(d) 모런- $S^*(W^0-W^*)$



(e) 기어리- $S^*(C^0-C^*)$



(f) 기어리- $S^*(W^0-W^*)$

그림 3. 고유벡터 간 상관관계의 시각화: 공간적 자기상관 통계량의 비교

표 3. 극단적 일치도를 보이는 고유벡터의 개수: 공간적 자기상관 통계량 간의 비교

비교 통계량	비교 SPM	극단적 일치도를 보이는 고유벡터의 개수					
		64개 육각형의 경우			256개 육각형의 경우		
		> 0.9	< -0.9	합계	> 0.9	< -0.9	합계
모런-기어리	$C^0-C^0$	7 (10.9)	2 (3.1)	9 (14.1)	12 (4.7)	13 (5.1)	25 (9.8)
	$W^0-W^0$	17 (26.6)	22 (34.4)	39 (60.9)	70 (27.3)	81 (31.6)	151 (59.0)
모런- $S^*$	$C^0-C^*$	24 (37.5)	23 (36.0)	47 (73.4)	86 (33.6)	105 (41.0)	191 (74.6)
	$W^0-W^*$	21 (32.8)	10 (15.6)	31 (48.4)	50 (19.5)	61 (23.8)	111 (43.4)
기어리- $S^*$	$C^0-C^*$	2 (3.1)	4 (6.3)	6 (9.4)	14 (5.5)	13 (5.1)	27 (10.5)
	$W^0-W^*$	13 (20.3)	15 (23.4)	28 (43.8)	56 (21.9)	42 (16.4)	98 (38.3)

주: 괄호 안의 숫자는 %임.

는 어떠한 패턴도 발견되지 않았다.

표 3은 그림 3에 나타난 사항을 요약하고 있다. 가장 높은 일대일 대응성을 보여준 그림 3(c)의 경우 73.4%의 고유벡터가 극단적인 일치도를 보이는 대응 고유벡터를 보유하고 있는 것으로 드러났다. 뒤이어 그림 3(b)도 60.9%의 일치도를 보였다. 이에 반해 가장 낮은 일치도를 보인 그림 3(e)의 경우 그 비율은 9.4%에 불과했다. 256개 육각형에 적용한 경우를 살펴보면, 일치도의 순위뿐만 아니라 극단적 일치도를 보이는 고유벡터의 비중에서도 64개 육각형에 적용한 경우와 거의 유사한 결과를 보여주고 있다. 따라서 SAS 간 고유벡터의 상관성이라는 측면에서는 공간 스케일의 효과가 거의 없는 것으로 보인다.

## 2) 고유벡터공간필터링 회귀분석에의 함의

SAS와 SPM에 따른 고유벡터의 다양성이 공간적 회귀분석에 가지는 함의를 분석하기 위해 사용된 경험 데이터는 푸에르토리코의 73개 자치시에 대한 농업 관련 데이터이다. 구체적으로 종속변수는 2007년 농가밀도(farm density)로 농가수를 전체 면적으로

나누어 산출된 것이다. 이 변수의 분포가 심한 왜도를 보여주기에 때문에 박스-콕스 변환(Box-Cox transformation)을 통해 재계산된 값을 사용했다. 독립변수는 연강수량 하나로 단순회귀모형이 이 실험에 사용되었다(데이터에 대한 자세한 사항은 Griffith and Chun, 2014 참조). 기본 모형과 비교하기 위해 모두 6개의 ESF 모형이 디자인 되었는데, 세 종류의 SAS에 두 종류의 SPM(이항비표준화와 행표준화)을 적용하여 모두 여섯 세트의 고유벡터를 산출하였다. 각 고유벡터의 세트에 대해 가장 큰 고유치의 최소한 25%에 해당하는 크기를 가진 고유벡터만을 선정하여 ESF 회귀분석에 투입하였고, 단계적 선택 옵션을 통해 특정한 기준을 통과한 고유벡터만이 최종적인 회귀식에 포함되도록 하였다.

표 4는 기본 모형과 함께, 여섯 종류의 ESF 모형의 결과가 제시되어 있다. 비공간적인 일반 모형을 보면, 독립변수인 연강수량의 회귀계수가 0.0138이고 1% 유의수준에서 유의한 것으로 나타나 있다. 중요한 사실은 잔차의 모런 통계량 값이 0.4304로 통계적으로 유의한 수준의 양의 공간적 자기상관이 존재한다는 점이다. 이 문제를 해결하기 위해 Griffith and Chun

(2014)이 시도한 ESF 모형은  $C^0$ 를 적용한 모런 통계량으로부터 추출한 고유벡터를 투입한 모형이다. 2단계의 선정 절차를 통해 총 8개의 고유벡터가 ESF 회귀모형 속에 투입된 것으로 드러났다. 독립변수의 회귀계수는 0.0210으로 바뀌었지만 0.1%에서 유의할 정도로 유의성은 더 향상되었다. 가장 중요한 사실은 잔차의 모런 값이  $-0.1643$ 으로 통계적으로 유의한 수준의 공간적 자기상관이 발견되지 않았다는 점이다. 이로서 ESF의 목적인, 종속변수를 독립변수가 구조적으로 설명하는 부분, 공간적으로 구조화된 잔차 부분, 완전히 무작위적인 잔차 부분으로 분해하는 데 성공하였다(이 네 요소에 대한 지도가 Griffith and Chun, 2014, Fig. 75.5에 나타나 있다.)

본 연구의 궁극적인 질문은, 만일 상이한 SAS로부터 상이한 고유벡터의 세트가 산출된다면, 혹은 동일한 SAS라 하더라도 SPM에 따라 상이한 고유벡터가 산출된다면, 그러한 상이한 세트의 고유벡터가 투입되었을 때 ESF 모형의 결론이 어떻게 나타날까이다. 표 4에 나타나 있는 나머지 다섯 개 모형은 바로 그러

한 질문에 답을 구하기 위해 디자인된 것이다. 주목할 만한 몇 가지 사항이 관찰되었다. 첫째, 모형 별로 다양한 개수의 고유벡터가 다양한 순위의 조합을 통해 투입되었다. 예를 들어 모런 통계량( $W^0$ )의 경우는 모두 11개의 고유벡터가 투입된 반면,  $S^*$  통계량( $C^*$ )의 경우는 불과 6개의 고유벡터만이 투입되었다. 모런 통계량( $W^0$ )과  $S^*$  통계량( $C^*$ )의 경우는 고유벡터의 순위와 투입 순서가 다소 뒤죽박죽인 반면, 기어리 통계량( $C^0$ )과  $S^*$  통계량( $W^*$ )의 경우는 상대적으로 조응적이다. 둘째, 모런 통계량과 기어리 통계량을 이용한 모형의 잔차에서는 통계적으로 유의한 수준의 공간적 자기상관이 발견되지 않았지만,  $S^*$  통계량을 사용한 경우에는 양의 공간적 자기상관이 여전히 존재하는 것으로 나타났다. 이는 투입된 고유벡터의 종류에 따라 ESF 모형이 잔차의 공간적 자기상관을 제거하는 능력이 달라진다는 점을 함축하고 있다. 셋째, 모든 모형에서 독립변수는 0.1% 유의수준에서 유의미한 설명 인자인 것으로 판명되었지만, 회귀계수의 크기는  $S^*$  통계량( $W^*$ )의 0.0140에서 모런 통계량( $W^0$ )

표 4. 고유벡터공간필터링 모형 비교

	기본 모형	고유벡터공간필터링 모형					
		모런 통계량		기어리 통계량		$S^*$ 통계량	
		$C^0$	$W^0$	$C^0$	$W^0$	$C^*$	$W^*$
절편	0,6622*	0,1623	-0,2237	0,6841**	0,4902*	0,4484	0,6465**
연강수량	0,0138**	0,0210***	0,0265***	0,0135***	0,0163***	0,0169***	0,0140***
투입된 고유벡터		EV12*** EV04*** EV01*** EV02*** EV18** EV10* EV09 EV15	EV01*** EV05*** EV12*** EV13* EV16** EV15 EV19** EV02** EV06* EV04* EV03	EV02*** EV06*** EV16*** EV18** EV13 EV12* EV11* EV19 EV22	EV02*** EV06*** EV13*** EV16*** EV14* EV12* EV17* EV22 EV15	EV12*** EV04*** EV01*** EV02*** EV10* EV09	EV01*** EV05*** EV12*** EV16** EV13** EV15** EV14
조정 $R^2$	0,1176	0,6895	0,6783	0,6140	0,6410	0,6198	0,6114
잔차의 모런 값	0,4304***	-0,1643	-0,1324	-0,0301	-0,0615	0,1171***	0,1498***

유의수준 \*: 0.05, \*\*: 0.01, \*\*\*: 0.001

의 0.0265에 이르는 범위를 보여 다소간 편차가 존재하는 것으로 드러났다. 또한 절편의 경우 Griffith and Chun(2014)의 모형을 포함한 세 개의 모형에서 유의하지 않은 것으로 나타났지만, 나머지 세 개의 모형에서는 유의한 것으로 나타났다.

표 4가 보여주고 있는 궁극적인 결론은 투입되는 고유벡터가 달라지면 ESF의 결과도 달라진다는 점이다. 이는 6개의 모형의 우열을 객관적으로 평가할 수 있는 토대가 아직은 마련되지 않았다는 점을 함축하고 있다. 그러나 그럼에도 불구하고 SAS와 SPM에 따른 고유벡터의 다양성이 ESF 방법론에 대해 가지는 몇 가지 함의를 표 4의 결과를 바탕으로 제시하고자 한다. 첫째, 기본적으로 잔차의 공간적 자기상관을 가장 잘 제거하는 모형이 가장 우수하다고 말할 수 있다. 그런 의미에서 모런 통계량( $C^0$ )이 가장 우수하고  $S^*$  통계량에 기반한 두 모형이 가장 열등하다. 그러나 잔차의 공간적 자기상관을 평가할 보다 '중립적인' 방식이 고려될 필요는 있어 보인다. 왜냐하면 잔차의 공간적 자기상관을 평가하기 위해 모런 통계량이 사용되고 있기 때문이다. 앞에서 언급한 것처럼, 서로 다른 SAS는 서로 다른 방식으로 공간적 자기상관을 측정한다. 어떠한 관점에서 공간적 자기상관을 보느냐에 따라 잔차에서의 공간적 자기상관의 정도가 달리 평가될 수 있다.

둘째, 회귀계수의 크기와 유의성을 비공간적인 기본 모형과 비교하고, 그것을 바탕으로 상이한 ESF 모형들을 평가하는 것이 가능하다. 예를 들어 기본 모형의 절편과 독립변수는 모두 통계적으로 유의할 뿐만 아니라 대략 0.65와 0.014의 값을 보이고 있다. 이런 측면을 고려하면 모런 통계량의 두 모형은 다소 불만족스러운 것으로 평가될 수도 있다. 두 모형에서는 절편이 통계적으로 유의하지 않을 뿐만 아니라 연강수량의 회귀계수도 0.020을 초과하여 다른 모형과의 편차가 크게 나타난다. 이런 측면에서 보면, 기어리 통계량( $C^0$ )과  $S^*$  통계량( $W^*$ )의 두 모형이 상대적으로 만족스러운 것으로 판단될 수 있다. 두 모형 모두에서 절편과 독립변수가 통계적으로 유의하고, 회귀계수의 크기도 기본 모형과 상대적으로 유사한 것으로 나타나 있기 때문이다. 이런 두 가지 사항을 종합적으로

고려할 때, 표 4에 나타나 있는 6개 모형 중 가장 우수한 것으로 기어리 통계량( $C^0$ )이 선정될 수도 있을 것이다.

#### 4. 결론

본 연구는 조정된 SPM으로부터 추출되는 고유벡터의 세트가 SAS의 종류와 SPM의 종류, 그리고 그 조합에 따라 다양하게 나타나는지 그렇지 않은지를 밝히는 데 주목적이 있었다. 이를 위해 세 종류의 SAS와 네 종류의 SPM이 고려되었으며, 고유벡터간 일치도에 대한 시각적 평가를 위해 일종의 상관관계 매트릭스 시각화 기법이 사용되었다. 각 그래프는 대각선성과 대응성이라는 두 가지 기준에 의거해 해석되었으며, 이를 통해 동일한 SAS 내의 상이한 SPM 간에서, 그리고 동일한 SPM를 가진 상이한 SAS 간에서 고유벡터의 상관성이 얼마나 다양하게 나타날 수 있는지가 분석되었다. 부가적으로 극단적 일치도를 보이는 고유벡터의 개수를 산출함으로써 대응성의 정도를 요약하는 수치로 사용하였다. 이와 관련된 주요 결과를 요약하면 다음과 같다. 첫째, 동일한 SAS에 상이한 SPM을 적용한 결과 SPM의 효과가 상당한 것으로 드러났다. 어느 정도의 대각선성은 나타나고 있지만 일대일 대응성은 높지 않은 것으로 드러났으며, 가장 높은 일치도를 보인 SAS는 모런 통계량이었다. 둘째, 상이한 SAS 쌍 간에 일치성의 정도와 양상에서 상당한 차이가 나타났고 SPM의 효과도 현저했다. 모런-기어리(특히  $W^0-W^0$ )의 경우는 대각선성이 우세한 반면, 모런- $S^*(C^0-C^*)$ 의 경우는 매우 높은 대응성을 나타냈다.

SAS와 SPM에 따른 고유벡터의 다양성이 공간적 회귀분석에 가지는 함의를 분석하기 위해 푸에르토리코의 경험 데이터에 대해 6개의 서로 다른 ESF 모형을 실행하였다. 여기에서 세 가지의 기본적인 사항이 관찰되었다. 첫째, 모형 별로 다양한 개수의 고유벡터가 다양한 순위의 조합을 통해 투입된다. 둘째, 투입된 고유벡터의 종류에 따라 ESF 모형이 잔차의

공간적 자기상관을 제거하는 능력이 달라진다. 셋째, 회귀계수의 크기와 유의성이 모형별로 상당한 차이를 보인다. 이러한 기본적인 결과를 바탕으로 모형의 비교우위에 대해 논의하고 평가하였으며 두 가지 결론에 도달했다. 첫째, 기본적으로 잔차의 공간적 자기상관을 가장 잘 제거하는 모형이 가장 우수하다고 말할 수 있다. 둘째, 회귀계수의 크기와 유의성을 비공간적인 기본 모형과 비교하고, 그것을 바탕으로 상이한 ESF 모형들을 평가하는 것이 가능하다.

본 연구를 확장하기 위한 몇 가지 향후 연구과제를 제시하면 다음과 같다. 첫째, 본 연구에서는 육각형 정다각 테셀레이션만을 사용하였지만, 이를 삼각형과 사각형 테셀레이션으로 확대하여 적용할 필요가 있다. 특히 사각형의 경우 래스터 데이터의 일반 형태를 대변하고 있기 때문에 본 연구의 결과가 그러한 데이터에 대해서도 동일하게 나타나는지를 따져보는 것은 의미 있는 일이다. 둘째, 보다 다양한 형태의 SPM이 가지는 효과에 대한 분석이 이루어져야 한다. 본 연구에서는 연결성에 기반한 이항 SPM과 그것의 행표준화 SPM에만 집중했다. ‘분산-안정화(variance-stabilizing)’ SPM(Tiefelsdorf *et al.*, 1999)이나 거리-기반의 다양한 SPM이 적용되었을 때 고유벡터의 다양성이 어떠한 방식으로 전개될지를 살펴보는 것은 매우 의미 있는 작업이 될 것이다. 셋째, ESF 모형에 대한 보다 일반적인 함의를 이끌어 내기 위한 추가적인 연구가 뒤따를 필요가 있다. 우선 고유벡터의 초기 선정 기준의 변화가 결과에 어떠한 영향을 미치는지에 대해 살펴볼 필요가 있다. 본 연구에서는 최소 25% 기준을 사용했지만 Chun *et al.*(2016)이 제시한 함수를 사용하거나, 음의 공간적 자기상관을 보이는 고유벡터도 함께 투입하는 방안을 모색할 필요가 있다. 또한 보다 다양한 경험 데이터를 활용하거나 보다 복잡한 모형을 수립하는 방식을 통해 ESF 모형 간 비교분석의 폭을 넓히고 양을 늘리는 작업이 뒤따라야 할 것이다.

## 사사

본 논문 속에 포함되어 있는 아이디어를 발전시키는 데 있어 그 크기를 계측할 수 없을 정도의 영감과 도움을 준 미국 텍사스대학교(달러스) 지리공간정보과학 프로그램의 Daniel A. Griffith, Yongwan Chun, Michael Tiefelsdorf 교수에게 감사의 마음을 전합니다.

## 참고문헌

- 국토교통과학기술진흥원, 2017, 공간정보 SW활용을 위한 오픈소스 가공기술개발 4차년도 연차실적계획서(내부자료).
- 이상일·조대현·이민파, 2015, “일변량 공간연관성통계량에 대한 비교 연구 (I): 전역적 S 통계량을 중심으로,” 한국지리학회지, 4(2), 329-345.
- 이상일·조대현·이민파, 2016, “일변량 공간연관성통계량에 대한 비교 연구 (II): 국지적 S 통계량을 중심으로,” 한국지리학회지, 5(3), 375-396.
- Anselin, L., 1988, *Spatial Econometrics: Methods and Models*, Boston: Kluwer Academic Publishers.
- Anselin, L., 1995, Local indicators of spatial association: LISA, *Geographical Analysis*, 27(2), 93-115.
- Anselin, L., 2009, Spatial regression, in Fotheringham, A. S. and Rogerson, P., eds., *The SAGE Handbook of Spatial Analysis*, London: SAGE, 255-275.
- Anselin, L. and Rey, S. J., 2014, *Modern Spatial Econometrics in Practice*, Chicago: GeoDa Press.
- Bailey, T. C. and Gatrell, A. C., 1995, *Interactive Spatial Data Analysis*, Harlow, Essex: Longman.
- Boots, B. and Tiefelsdorf, M., 2000, Global and local spatial autocorrelation in bounded regular tessellations, *Journal of Geographical Systems*, 2(4), 319-348.
- Chun, Y. and Griffith, D. A., 2013, *Spatial Statistics & Geostatistics*, Los Angeles: SAGE,
- 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.

- de Jong, P. D., Sprenger, C., and Veen, F. V., 1984, On extreme values of Moran's  $I$  and Geary's  $c$ , *Geographical Analysis*, 16(1), 17-24.
- Dubin, R., 2009, Spatial weights, in Fotheringham, A. S. and Rogerson, P., eds., *The SAGE Handbook of Spatial Analysis*, London: SAGE, 125-157.
- Getis, A., 1991, Spatial interaction and spatial autocorrelation: a cross-product approach, *Environment and Planning A*, 23(9), 1269-1277.
- Griffith, D. A., 1996a, Some guidelines for specifying the geographic weights matrix contained in spatial statistical models, in Arlinghaus, S. L., ed., *Practical Handbook of Spatial Statistics*, Boca Raton, FL: CRC, 65-82.
- Griffith, D. A., 1996b, 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*, Berlin: Springer.
- Griffith, D. A., 2010, Spatial filtering, in Fischer, M. M. and Getis, A., eds., *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, New York: Springer, 301-318.
- Griffith, D. A., 2017, Spatial filtering, in Shekhar, S., Xiona, H., and Zhou, X., eds., *Encyclopedia of GIS*, Volume 3, 2nd edition, New York: Springer, 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*, Berlin: Springer-Verlag, 1477-1507.
- Henshaw, R. C., Jr., 1966, Testing single-equation least squares regression models for autocorrelated disturbances, *Econometrica: Journal of the Econometric Society*, 34(3), 646-660.
- Henshaw, R. C., Jr., 1968, Errata: Testing single-equation least squares regression models for autocorrelated disturbances, *Econometrica: Journal of the Econometric Society*, 36(3/4), 646-660.
- Hepple, L. W., 1998, Exact testing for spatial autocorrelation among regression residuals, *Environment and Planning A*, 30(1), 85-107.
- Hubert, L. J., Golledge, R. G., and Costanzo, C. M., 1981, Generalized procedures for evaluating spatial autocorrelation, *Geographical Analysis*, 13(3), 224-233.
- 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., 2008, A generalized procedure to extract higher order moments of univariate spatial association measures for statistical testing under the normality assumption, *Journal of the Korean Geographical Society*, 43(2), 253-262.
- 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, New York: Springer, 360-368.
- Mantel, N., 1967, The detection of disease clustering and a generalized regression approach, *Cancer Research*, 27(2 Part 1), 209-220.
- O'Sullivan, D. and Unwin, D. J., 2010, *Geographic Information Analysis*, 2nd edition, Hoboken, NJ: John Wiley & Sons.
- 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, New York: Springer, 511-522.
- Tiefelsdorf, M., 2000, *Modelling Spatial Processes: The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's  $I$* , New York: Springer.
- Tiefelsdorf, M. and Boots, B., 1995, The exact distribution of Moran's  $I$ , *Environment and Planning A*, 27(6), 985-999.
- Tiefelsdorf, M. and Griffith, D. A., 2007, Semiparametric filtering of spatial autocorrelation: The eigenvector

이상일·조대현·이민파

tor approach, *Environment and Planning A*, 39(5), 1193-1221.

Tiefelsdorf, M., Griffith, D. A., and Boots, B., 1999, A variance-stabilizing coding scheme for spatial link matrices, *Environment and Planning A*, 31(1), 165-180.

교신: 이상일, 08826, 서울특별시 관악구 관악로 1, 서울대학교 사범대학 지리교육과 (이메일: si\_lee@snu.ac.kr)

Correspondence: Sang-Il Lee, 1 Gwanak-ro, Gwanak-gu, Department of Geography Education, College of Education, Seoul National University, Seoul 08826, South Korea (Email: si\_lee@snu.ac.kr)

최초투고일 2017. 9. 25

수정일 2017. 10. 3

최종접수일 2017. 10. 12