

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 존재된 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_x^2)’라 부르하고자 한다. 이 결정계수는 다시 공통 패턴 요소(C)(보다 정확하게는, 공통 고유벡터)만 투입한 회귀분석을 실행했을 때의 결정계수인 ‘공통 결정계수(R_{xc}^2)’와 특수 패턴 요소(U)(보다 정확하게는 특수 고유벡터)만 투입한 회귀분석을 실행했을 때의 결정계수인 ‘특수 결정계수(R_{xu}^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.

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

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.

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

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


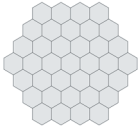


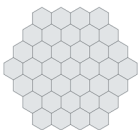


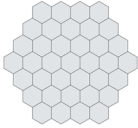


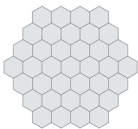


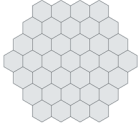




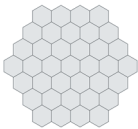



상관의 최소한 25% 수준을 보여주는 고유벡터를 선정 후, 단계적 회귀분석을 실행하여 최종적인 고유벡터를 추출하는 방식에 의거해 이루어진다(Griffith, 2010; Chun and Griffith, 2013; Chun *et al.*, 2016; Griffith, 2017). 둘째, 산출된 공간적 패턴 요소를 이용하여 하위 상관계수를 산출하고, 결정계수의 요소분해를 실행한다. 셋째, 산출된 하위 상관계수와 결정계수를 이용하여 식(9)에 의거한 피어슨 상관계수 요소분해를 실행한다. 각 단계별로 결과에 대한 세세한 해석을 제공함으로써 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^2)	공통 결정계수 (R^2_{cc})	특수 결정계수 (R^2_{uu})	전체 결정계수 (R^2_y)	공통 결정계수 (R^2_{cc})	특수 결정계수 (R^2_{uu})
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 통계량과 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