

■ 論 文 ■

## 공간가중회귀분석을 이용한 통행발생모형

Trip Generation Model based on Geographically Weighted Regression

**김 진 희**

(연세대학교 도시공학과 박사과정)

**박 일 섭**

((주)한국스마트카드 사원)

**정 진 혁**

(연세대학교 도시공학과 교수)

### 목 차

- |   |   |
|---|---|
| <p>I. 서론</p> <p>1. 연구의 배경 및 목적</p> <p>2. 연구의 범위</p> <p>II. 이론적 고찰</p> <p>1. 전역적/국지적 공간상관성</p> <p>2. 공간가중회귀분석(GWR)</p> <p>III. 모형의 추정 및 결과분석</p> | <p>1. 모형의 정립</p> <p>2. GWR 모형의 계수 추정결과</p> <p>3. OLS 모형과 GWR모형의 결과비교</p> <p>4. 국지적 공간상관성 분석</p> <p>IV. 결론</p> <p>참고문헌</p> |
|---|---|

**Key Words :** 공간가중회귀분석, 공간상관성, 국지적 공간상관성지표, 공간의존성, 통행발생모형  
 Geographically Weighted Regression, Spatial Autocorrelation, Local Indicator of Spatial Association, Spatial Dependence, Trip Generation Model

### 요 약

대다수의 현대 도시들은 집적의 이익을 극대화하기 위해 군집을 형성하고 각 지역 간에 다양한 공간적 영향을 주고받는다. 그러나 전통적 4단계 수요예측방법의 첫 단계인 통행발생단계에서 주로 적용되는 선형회귀분석모형은 공간적 영향을 반영할 수 없다는 단점이 있다. 이러한 문제를 해결하기 위해서 공간적 상관성을 반영할 수 있는 통행발생모형을 구축하는 것이 필요하다.

본 연구에서는 공간적 상관성을 고려할 수 있는 통행발생모형으로 공간가중회귀모형(Geographically Weighted Regression)을 제안한다. 공간가중회귀모형은 공간적 상관성을 고려할 수 있는 가중치 행렬을 추정하고 이를 이용하여 회귀식의 계수를 각 존별로 추정하는 것이다. 본 연구에서는 대구광역시 통행자료를 이용하여 공간가중회귀모형을 적용하였다. 공간가중회귀모형의 우수성을 평가하기 위하여 일반적인 회귀모형과 적합도, RMSE 등을 비교분석하였다. 또한 국지적 공간상관성을 측정하는 척도인 LISA(Local Indicator of Spatial Association) 지표를 각 모형별로 산출하였다. LISA 지표를 통하여 현재 분석대상지역은 국지적 공간상관성이 존재함을 확인할 수 있으며 공간가중회귀모형을 적용함으로써 공간상관성으로 인한 오차가 크게 개선됨을 확인할 수 있다.

In most of the urbanized cities, socio-economic attributes tend to cluster as patterns of similarity in space, namely spatial autocorrelation, by agglomeration forces. The classical linear regression model, the most frequently adopted in the trip generation step, cannot sufficiently represent this effect. In order to take into account the effect properly, we need a model which adequately deals with the spatial dependence patterns.

In this study, the Geographically Weighted Regression (GWR) model is adopted as an alternative method for the local analysis of relationships in multivariate data sets: that is GWR extends this traditional regression framework by estimating local rather than global parameters. This study shows the existence of spatial effects in the production and attraction of home base/non-home based trips through the GWR model using travel data collected in Daegu metropolitan area. Furthermore, LISA is employed to verify the fact that the local spatial autocorrelation exists.

## 1. 서론

### 1. 연구의 배경 및 목적

현대 도시들은 집적의 이익을 극대화하기 위해 군집을 형성하고 다양한 공간적 영향을 주고받는다. 공간적 영향의 한 예로는 업무·상업시설들이 밀집되어 발생하는 집적의 효과(Agglomeration force)를 들 수 있다. 집적의 효과는 업무 및 상업시설에 대한 접근성이 높아지는 것을 의미할 수 있으며, 이러한 사실은 사람들이 더 많은 사회·경제적 활동을 유발할 수 있다는 것을 의미한다. 통행은 사회·경제적 활동에서 파생되는 것이기 때문에 공간적 영향과 통행발생은 밀접한 관계를 가지게 된다.

집계적 관점에서 최소의 분석단위인 교통분석존(Traffic Analysis Zone, TAZ)은 사회·경제적 군집현상을 반영하기 보다는 행정구역단위에 의해 결정된다. 이는 교통수요 예측과 관련된 사회·경제지표 수집이 용이하기 때문이다. 하나의 행정구역단위 안에는 동질성이 존재한다고 볼 수 있으나, 인접한 행정구역과의 관계가 독립적이라고 판단하기에는 어려움이 있다.

전통적 4단계 수요예측방법의 첫 단계인 통행발생단계에서 주로 적용되는 선형회귀모형은 일반적으로 일반화최소자승법(Ordinary least square method, 이하 OLS)에 의해 추정된다. OLS는 종속변수인 각 존의 통행발생량이 서로 독립적임을 가정한다. 따라서 현재의 통행발생모형은 앞서 설명한 공간적 영향을 반영할 수 없고, 만약 공간적 영향이 존재함에도 불구하고 이를 고려하지 않는다면 OLS로 추정한 통행발생모형은 왜곡된 결과를 나타낸다. 따라서 통행발생모형 정립 시, 현재 대상지역의 통행발생량이 공간적 영향이 존재하는지를 우선적으로 파악한 후, 공간적 영향을 고려할 수 있는 통행발생모형을 구축하여야 한다.

통행발생량에 대한 공간적 영향은 전역적(Global)으로 존재할 수 있고, 국지적(Local)으로 존재할 수 있다. 기존의 연구들은 주로 전역적인 공간적 영향을 분석하였다(Kim et al, 2009; 박일섭 외, 2009). 그러나 국지적으로만 공간적 영향이 강하게 존재함에도 불구하고 전역적인 공간적 영향을 분석한다면, 국지적인 현상이 전역적인 현상인 것으로 오인될 우려가 존재한다. 이러한 현상은 직관적으로 유추될 수 있는데, 앞에서 설명한 집적의 효과는 도시 전체의 전반적인 현상이라기보다는 도심 또는 부도심과 같은 특정지역에 나타나는 국지적인 현

상이라고 보는 것이 일반적이다. 따라서 공간적 영향을 국지적으로 고려할 수 있는 통행발생모형이 필요하다.

본 연구에서는 국지적인 공간적 영향이 존 간 통행발생량에 존재함을 확인하고 이를 고려할 수 있는 모형인 공간가중회귀모형(Geographically Weighted Regression, 이하 GWR)을 적용하는 것이 본 연구의 목적이다.

### 2. 연구의 범위 및 방법

본 연구의 공간적 범위는 대구광역시와 인접 8개 시군을 포함한 대구광역권의 257개 동을 대상으로 하였다. 본 분석에 사용한 자료는 2006년 한국교통연구원 국가교통DB센터에서 실시한 5대 광역권 가구통행실태 조사를 바탕으로 전수화 과정을 통해 구축된 목적별 기중점통행량을 활용하였다. 통행목적은 PA기반으로 구분하였으며 그 중 가정기반통행과 비가정기반통행의 기타통행만을 대상으로 분석을 수행하였다. 출근과 등교는 필수통행으로 공간적 영향에 의해 통행빈도수가 달라지지 않을 것으로 예상되기 때문에 출근과 등교를 제외하고 쇼핑, 여가 등 나머지 목적을 기타통행으로 분류하여 분석하였다.

분석대상이 되는 대구광역권의 가정기반 기타통행은 총 2,721,045통행/일 이고, 비가정기반 기타통행은 831,501통행/일 이다. 총 인구는 3,531,706명으로 가



〈그림 1〉 공간적 범위(대구광역권)

정기반 기타통행은 1인당 0.77통행/일 이며, 비가정기 반 기타통행은 1인당 0.24통행/일로 나타났다.

본 연구의 방법은 우선적으로 대구광역시권 통행자료를 이용하여 각 통행목적별로 일반적인 선형회귀모형을 구축하고 선형회귀모형과 동일한 독립변수와 종속변수로 GWR모형을 구축한다. 구축된 모형을 추정하고 두 모형의 추정된 parameter 및 모형 적합도의 차이를 비교하여 두 모형의 차이를 분석한다. 그 후 추정된 선형회귀모형의 잔차와 추정된 GWR모형의 잔차를 이용하여 공간상관성을 분석한다. 공간상관성은 전역적 공간상관성과 국지적 공간상관성으로 구분하고 각각의 크기를 나타낼 수 있는 적합한 지표로써 분석한다(“II. 이론적 고찰 - 1. 전역적/국지적 공간상관성” 참고). 두 모형의 공간상관성 지표 값의 차이를 통해 GWR모형으로 인하여 공간상관성이 고려되었음을 확인한다.

## II. 이론적 고찰

### 1. 전역적/국지적 공간상관성

서론에서 언급한 공간적 영향은 지리학 및 계량경제학 등 다양한 분야에서 연구되어왔는데 공간의존성(Spatial dependence) 또는 공간상관성(Spatial association or Spatial autocorrelation)이라는 용어로 정의되고 있다. 본 연구에서는 공간적 영향을 “공간상관성”으로 정의하여 사용하겠다. 공간상관성은 어떤 현상이 멀리 떨어진 공간보다 인접한 공간의 현상과 유사한 패턴을 나타내는 것을 말한다. 이와 같은 공간상관성은 전역적인 상관성과 국지적인 상관성으로 구분할 수 있는데 전역적 공간상관성을 나타내는 지표로는 Moran’s I가 대표적이며, 국지적 공간상관성을 나타내는 지표로는 Local Moran’s I(Cliff and Ord, 1981; Anselin, 1995)가 대표적이다.

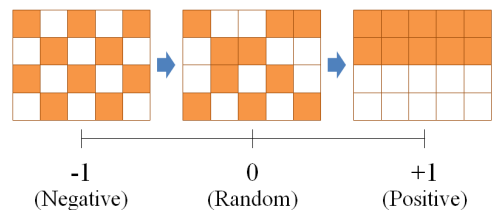
$$Moran's I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \hat{y}_i)(y_j - \hat{y}_j)}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$Local Moran's I_i = (y_i - \hat{y}_i) \sum_{j=1}^n w_{ij} (y_j - \hat{y}_j) \quad (2)$$

여기서  $y_i$ 는 존 i의 관측된 종속변수를 의미하고,  $\hat{y}_i$ 는 모형에 의해 추정된 존 i의 종속변수 값을 의미한다.  $w_{ij}$ 은 존 i와 존 j의 공간적 상관관계를 정의하는 값으로 공간가중치(Spatial weight)라고 한다. 즉, 총 n개의 존에 대하여 (n×n) 공간가중치 행렬을 구성할 수 있으며 공간가중치 값은 연구자가 정의하는 방법론에 따라 그 값이 결정된다. 일반적으로 인접하면 1, 아니면 0으로 정의하는 경우와 거리에 반비례하게 정의하는 경우 등이 있다.

Moran’s I 값은 관측치와 모형추정치 차, 즉 잔차의 크기가 공간적으로 유사하게 분포되어 있는 정도를 의미한다. 즉, 잔차의 크기가 유사한 존들이 인접하여 분포되어 있으면 Moran’s I 값은 1에 가깝게 산출되며, 무작위 패턴을 가지고 있으면 0에 가깝게 산출된다. 또한 -1에 가까울수록 부(negative)의 공간적 상관성을 갖는 것으로 해석할 수 있다(Anselin and Bera, 1998). <그림 2>는 Moran’s I 값과 공간적 분포의 개념도로서 부의 상관관계일수록 서로 다른 값들이 공간적으로 인접하게 위치하는 양상으로 분포되고 양의 상관관계일수록 동일한 값이 인접하게 분포되는 것을 의미한다.

Local Moran’s I 값은 식(2)에서 볼 수 있듯이 각 존마다 산출된다. 즉, 한 존의 잔차와 인접한 존의 잔차가 유사하게 분포되어 있으면 positive 자기상관으로, 반대의 경우에는 negative 자기상관으로 판정한다. 따라서 각 존마다 주변지역과의 공간상관성을 분석하므로 국지적인 공간상관성을 파악할 수 있다. Local Moran’s I 값이 통계적으로 유의하지 않은 경우 해당 존은 인접한 존들과 공간상관성이 존재하지 않는 것으로 판단할 수 있다. Local Moran’s I의 통계적 검정은 “공간상관성이 존재하지 않는다(즉,  $E[I_i] = 0$ ).”는 귀무가설을 검정하는 것으로 검정통계량은 z-value를 이용한다(Local Moran’s I 값의 통계적 분포를 증명한 것과 가설검정 및 관련 실험에 대한 상세내용은 Anselin(1995)의 연구 pp.108~110을 참고).



<그림 2> Moran’s I 값과 공간적 분포 개념도

Moran's I와 Local Moran's I은 관측치와 모형 추정치의 차인 잔차(즉,  $y_i - \hat{y}_i$ )를 이용하여 분석한다. 따라서 현재 공간상관성이 존재함에도 불구하고 모형이 공간상관성을 제대로 고려하지 못한다면 Moran's I와 Local Moran's I의 값은 크게 산출될 것이다. 반대로 공간상관성을 고려한 모형은 공간상관성 지표 값이 작게 산출될 것이다. 이와 같은 현상은 식(1)과 식(2)에서 도출할 수 있는데 두 지표 값은 잔차의 값이 클수록, 공간가중치가 클수록 증가하게 된다. 잔차의 값이 크다는 것은 해당 지점의 종속변수에 대해 모형이 설명하지 못한 부분이 크다는 것을 의미하고, 공간가중치가 크다는 것은 공간적으로 서로 인접하다는 것을 의미한다. 따라서 모형이 설명하지 못한 부분이 큰 지점들이 인접하게 있을 경우 두 지표 값이 가장 크게 산출되게 된다.

## 2. 공간가중회귀분석(GWR)

GWR은 공간상 서로 다른 위치에 존재하는 데이터들에 대하여 서로 다른 회귀모형을 적용함으로써 비정상성(Non-stationarity)을 탐색하는 방법이다. 일반적으로 통행발생단계에서 사용되는 선형회귀분석은 분석대상지역 전체에서 나타나는 평균적이고 전역적인 효과를 전제로 하지만, GWR은 좀 더 낮은 분석단위를 적용해서 국지적인 패턴을 반영한 모형이다. 이는 분석대상지역 내에서 개별 관측치에 대해 회귀분석 기준점을 선정하여 그 중심점으로부터 기준거리 내에 있는 관측치들을 대상으로 회귀분석을 실시한다. 따라서 GWR을 사용하여 통행발생모형을 구축할 경우 각 Zone의 개수만큼의 회귀식이 산출된다. 또한 GWR은 회귀분석 기준점에서 멀어질수록 낮아지는 차별적 가중치를 독립변수에 적용하기 때문에 지리적으로 먼 거리에 있는 변수보다 가까운 거리에 있는 변수가 더 큰 영향을 미치게 된다.

식(3)은 GWR모형을 나타낸다. GWR은 다중선형회귀분석에 좌표를 부여함으로써 모형을 지역적인 구조로 확장시킨다.

$$y_i = \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (3)$$

여기서  $(u_i, v_i)$ 는 존 i의 중심점의 좌표를 의미하며,  $\beta_k(u_i, v_i)$ 는 존 i의 k번째 독립변수의 계수를 의미한다. 따라서 각 독립변수 별로 모든 회귀분석의 중심점에 대

하여 각각 다른 계수들이 추정되며, 추정치는 식(4)에 의해서 계산된다.

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad (4)$$

여기서,

$$W(u_i, v_i) = \begin{bmatrix} W_{i1} & 0 & \dots & 0 \\ 0 & W_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{in} \end{bmatrix}$$

여기서  $X$ 는  $(n \times k)$  행렬,  $Y$ 는  $(n \times 1)$  벡터이고  $W(u_i, v_i)$ 는  $(n \times n)$  행렬이다.  $W(u_i, v_i)$ 의 대각선에 존재하는 값들은 좌표  $(u_i, v_i)$ 를 갖는 존 i에 대해 주변 n개 관측치의 지리적 가중치를 의미한다. Moran's I와 Local Moran's I을 산출하기 위해 적용한 공간가중치 행렬과 유사한 개념이지만 행렬의 형태가 다를 수 유의하여야 한다. 또한 분석 시, 공간상관성을 분석하기 위한 공간가중치 행렬은 연구자가 외생적으로 정의하는 것이지만, GWR에서의 공간가중치 행렬은 내생적으로 추정된다.

$W_{ij}$ 는 식(5)와 같이 Kernel function으로 정의할 수 있다.

$$W_{ij} = \exp(-\gamma d_{ij}^2) \quad (5)$$

여기서  $d_{ij}$ 는 존 i의 중심점에서 존 j의 중심점까지의 거리를 의미하고,  $\gamma$ 는 Kernel bandwidth로 가중치를 적용하는 기초가 되며 연구지역을 분할하는 기준이 된다. 즉,  $\gamma$  값이 추정이 필요한 parameter로 이를 추정하는 방법론은 대표적으로 AICc(Akaike Information Criterion) 방법론과 CV(Cross-Validation) 방법론이 있다. AICc 방법론은 모형의 적합도를 나타내는 AIC 값을 최소화 시키는 bandwidth를 선택하는 것이다. CV 방법론은 관측치와 추정치의 차이의 제곱합을 CV로 정의하고, bandwidth를 변화시킬때 CV가 감소하다가 증가하는 변곡점에서 bandwidth를 결정하게 된다(Fotheringham et al., 2002).

## III. 모형의 추정 및 결과분석

### 1. 모형의 정립

본 연구에서는 대구광역권을 대상으로 기타통행에 대

〈표 1〉 모형의 정립

| 종속변수     |           | HBO                     |            | NHBO |    |
|----------|-----------|-------------------------|------------|------|----|
|          |           | 유출                      | 유입         | 유출   | 유입 |
| 독립 변수    | $\beta_0$ | 상수항                     |            |      |    |
|          | $\beta_1$ | 거주인구                    |            |      |    |
|          | $\beta_2$ | -                       | 3차 산업 종사자수 |      |    |
|          | $\beta_3$ | -                       | 초중고 수용학생수  |      |    |
|          | $\beta_4$ | -                       | 대학생 수용학생수  |      |    |
| 공간가중치 행렬 |           | Binary Contiguity-based |            |      |    |

해 가정기반 기타통행(Home-based Other trip, 이하 HBO)과 비가정기반 기타통행(Non-home-based Other trip, 이하 NHBO)으로 목적을 구분하고 유출(Production)과 유입(Attraction)으로 나누어 분석하였다. 각 종속변수를 설명하는 독립변수는 〈표 1〉과 같다. 이와 같은 변수를 이용하여 일반적인 선형회귀모형 및 GWR모형을 구축하였다.

또한 공간상관성을 분석하기 위한 공간가중치행렬은 Binary Contiguity-based weights matrix로 적용하였다. 이것은 존의 경계를 공유하는 지에 따라  $w_{ij}$ 에 1 또는 0을 부여하는 방법이다. 이러한 공간가중치행렬을 이용하여 모형의 전역적/국지적 공간상관성을 분석하였다.

## 2. GWR 모형의 계수 추정결과

앞서 설명한 것과 같이 GWR모형의 계수는 Zone의

개수만큼 추정되므로 추정된 모든 계수값을 제시하는 것 보다는 지역적인 차이를 비교하는 것이 효과적이다. 〈그림 3〉은 통행목적별 GWR모형에 의해 추정된 계수값을 도시한 것이다. 각 독립변수별로 추정된 계수의 평균과 표준편차 값을 이용하여 표준화하여 1-std 간격으로 표현하였다.

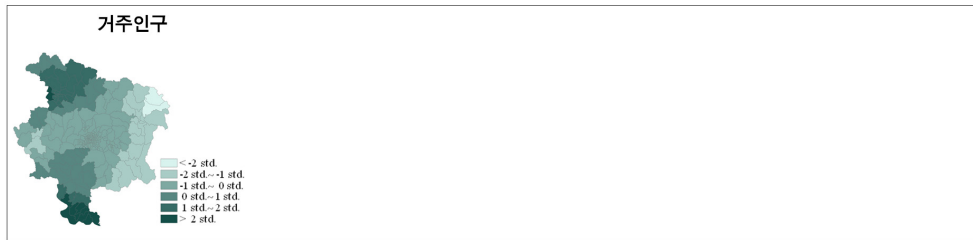
즉, 한 Zone에서 추정된 계수값이 전체적 지역의 평균에 비해 클수록 진하게 표현된다. 결과에서 볼 수 있듯이 추정된 계수값의 크기는 지역별로 차이가 존재하며, 계수값의 크기가 유사지역끼리 군집을 이루는 현상을 확인할 수 있다.

본 연구에서 설정한 모형의 종속변수가 통행발생량인 점을 고려했을 때, 각 독립변수의 계수값은 해당 독립변수에 대한 통행발생원단위의 개념으로 해석이 가능하다. GWR 모형에 의해 지역적으로 상이한 계수가 산출되었다는 것은 해당 변수에 의한 통행발생원단위가 지역적으로 서로 다르다는 것을 의미한다. HBO 유출통행의 경우, 대구광역시 중심부에 비해서 남쪽지역과 북쪽지역이 거주인구당 통행발생원단위가 큰 것을 알 수 있다. 그에 반해 NHBO 유출통행의 경우 대구광역시 중심부가 기타지역에 비해 거주인구당 통행발생원단위가 더 크게 나타났다. 이와 같이 GWR모형을 적용하면 분석대상지역 내에서 지역적 특성에 의해 통행발생원단위가 서로 다른 현상을 분석할 수 있다.

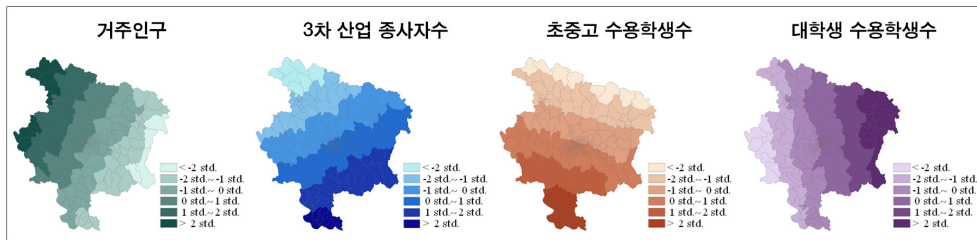
〈표 2〉의 HBO 유입통행의 추정결과에서 확인할 수 있듯이 추정된 계수의 표준편차가 "0"에 가까움에도 불

〈표 2〉 OLS 및 GWR 모형추정결과

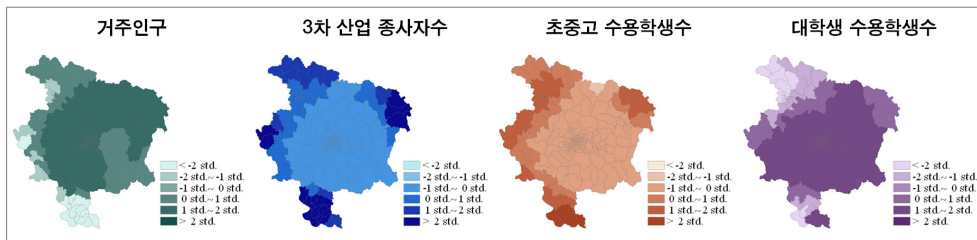
| 통행목적 | 변수 | OLS Model |          |                |       |           | GWR Model |          |                |       |           |        |
|------|----|-----------|----------|----------------|-------|-----------|-----------|----------|----------------|-------|-----------|--------|
|      |    | 추정치       | t-static | R <sup>2</sup> | AIC   | Moran's I | Average   | Std.     | R <sup>2</sup> | AIC   | Moran's I |        |
| HBO  | 유출 | b0        | 91.817   | 0.475          | 0.866 | 4635.0    | 0.1194    | -188.923 | 405.097        | 0.884 | 4624.5    | 0.0487 |
|      |    | b1        | 0.438    | 40.537         |       |           |           | 0.453    | 0.037          |       |           |        |
|      | 유입 | b0        | 677.627  | 1.923          | 0.626 | 4909.3    | 0.0575    | 677.483  | 0.091          | 0.626 | 4911.6    | 0.0578 |
|      |    | b1        | 0.134    | 4.319          |       |           |           | 0.134    | 0.000          |       |           |        |
|      |    | b2        | 0.878    | 8.342          |       |           |           | 0.878    | 0.000          |       |           |        |
|      |    | b3        | 0.527    | 4.397          |       |           |           | 0.527    | 0.000          |       |           |        |
|      | b4 | -0.084    | -2.017   | -0.084         | 0.000 |           |           |          |                |       |           |        |
| NHBO | 유출 | b0        | 188.048  | 0.982          | 0.646 | 4596.0    | 0.2176    | 164.316  | 404.838        | 0.739 | 4563.3    | 0.0506 |
|      |    | b1        | 0.109    | 6.461          |       |           |           | 0.094    | 0.029          |       |           |        |
|      |    | b2        | 0.368    | 6.430          |       |           |           | 0.460    | 0.236          |       |           |        |
|      |    | b3        | 0.162    | 2.487          |       |           |           | 0.205    | 0.150          |       |           |        |
|      |    | b4        | 0.100    | 4.439          |       |           |           | 0.064    | 0.106          |       |           |        |
|      | 유입 | b0        | 162.980  | 0.735          | 0.558 | 4671.5    | 0.2089    | 103.324  | 607.438        | 0.660 | 4647.1    | 0.0842 |
|      |    | b1        | 0.084    | 4.298          |       |           |           | 0.077    | 0.053          |       |           |        |
|      |    | b2        | 0.524    | 7.907          |       |           |           | 0.589    | 0.181          |       |           |        |
|      |    | b3        | 0.176    | 2.330          |       |           |           | 0.203    | 0.233          |       |           |        |
|      |    | b4        | 0.002    | 0.069          |       |           |           | -0.035   | 0.090          |       |           |        |



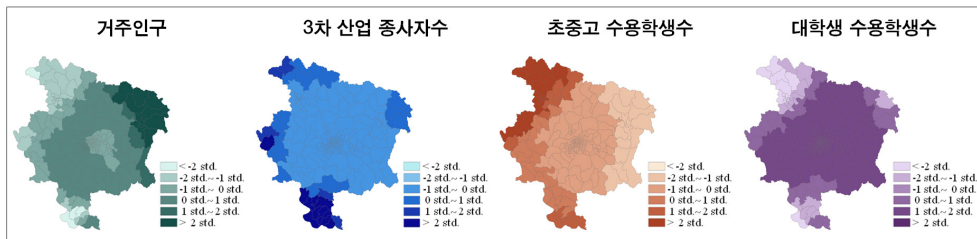
(a) 가정기반 기타통행 유출(HBO Production)



(b) 가정기반 기타통행 유입(HBO Attraction)



(c) 비가정기반 기타통행 유출(NHBO Production)



(d) 비가정기반 기타통행 유입(NHBO Attraction)

〈그림 3〉 GWR모형에 의해 추정된 Parameter 값의 분포도

구하고 (즉, 지역적으로 parameter의 차이가 없음에도 불구하고), 〈그림 3(b)〉의 HBO 유입통행의 경우 추정된 계수값의 분포가 일정한 방향성을 갖는 것으로 보인다. 그러나 이것은 본 연구에서 계수값의 분포를 나타내는 기준을 1-std 간격으로 표현하였기 때문에 나타난 왜

곡된 결과이다. 계수값의 크기가 매우 유사함에도 표준편차가 "0"에 가깝기 때문에 1-std 이상 차이 나는 것으로 판정하고, 이들이 서로 다른 범주에 포함되어 나타난 현상이다.

HBO 유입통행은 공간적 상관성이 존재하지 않기 때

문에 GWR 모형에 의해 추정된 각 Zone의 계수값이 서로 유사하다. 따라서 GWR 모형을 적용할 경우, 공간적 상관성이 존재하지 않는다면 일반적인 회귀모형과 유사한 결과를 산출한다는 것을 알 수 있다.

### 3. OLS모형과 GWR모형의 결과비교

OLS 및 GWR 모형의 추정결과는 <표 2>와 같다. OLS 모형 추정결과, HBO 유출통행, NHBO 유출 및 유입통행의 상수항과 NHBO 유입통행의 대학생 수확생수를 제외한 모든 변수의 계수 값이 95% 신뢰수준에서 통계적으로 유의한 것으로 나타났다. GWR 모형은 각 존마다 개별적으로 계수를 추정하므로 계수 값의 분포를 결과로 산출한다. 모든 목적통행의 유의한 변수에 대해서 OLS 모형의 추정된 계수 값과 GWR 모형의 추정된 계수의 평균값이 유사한 것으로 나타났다.

모형의 적합도는 R2와 AIC로 분석하였는데 R2 값은 1에 가까울수록, AIC 값은 작을수록 모형의 적합도가 더 높다. 모형의 적합도를 비교해 보면, HBO 유입통행을 제외한 모든 목적통행에서 GWR 모형의 적합도가 더 높은 것으로 나타났다. 이와 같은 현상은 전역적 공간상관성을 나타내는 Moran's I의 결과를 통해 의미를 찾을 수 있다.

이론적 고찰에서 설명한 것처럼 Moran's I는 -1에서 1 사이의 값을 갖으며 0에서 멀어질수록 Negative 또는 Positive 자기상관을 갖는 것으로 판단할 수 있다. 그러나 공간상관성의 존재 여부를 판단하는 임계치(threshold)는 존재하지 않으며 분석가의 기준에 따라 판단할 수 있다. <표 2>에 나타난 OLS 모형의 Moran's I의 결과는 NHBO 유출(I=0.2176)의 경우 공간상관성이 다른 목적통행발생량에 비해 강하게 존재한다고 볼 수 있으며 NHBO 유입 통행(I=0.2089)도 비교적 강한 공간상관성이 존재한다고 판단할 수 있다. HBO통행의 경우 유출(I=0.1194)과 유입(I=0.0575) 모두 비교적 약한 공간

상관성이 존재한다고 판단할 수 있다. 특히 HBO 유입통행의 경우 그 값이 "0"에 매우 가까워 다른 목적통행발생량에 비해 공간상관성이 거의 존재하지 않는 것으로 판단된다.

반면 GWR 모형의 Moran's I의 값은 HBO 유입통행을 제외한 다른 목적통행에서 OLS 모형에 비해 그 값이 크게 작아진 것을 확인할 수 있다. 즉, GWR 모형을 적용함으로써 공간상관성에 의한 편이(bias)에 대한 보정 효과가 나타난 것을 의미한다. 앞서 설명한 것과 같이 이들 목적통행에 대한 GWR 모형의 적합도 OLS 모형에 비해 더 높다는 것을 고려할 때, 공간상관성이 비교적 강하게 존재하는 상황에서는 일반적인 OLS 모형에 비해 GWR 모형이 더 설명력이 있다는 것을 의미한다. 그러나 HBO 유입통행과 같이 공간상관성이 거의 존재하지 않는 상황에서는 GWR 모형은 일반적인 OLS 모형과 큰 차이가 없음을 의미한다.

### 4. 국지적 공간상관성 분석

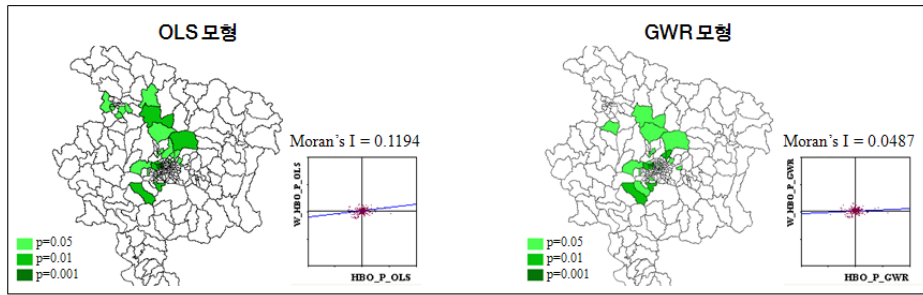
<그림 4>는 각 존별로 산출되는 국지적 공간상관성 지표인 Local Moran's I 값이 유의한 존을 표시한 것이다. 즉 공간상관성이 대상지역 전체에 고루 분포하는 것이 아니라 특정지역에 국지적으로 나타남을 나타내고 있다. GWR 모형의 경우 OLS 모형에 비해 국지적 공간상관성이 유의한 지역이 줄어들어든 것을 볼 수 있는데 이것은 GWR 모형이 국지적 공간상관성을 반영하여 개선하는데 효과적임을 보여주는 것이다.

<표 3>은 <그림 4>의 결과를 정리한 것으로 공간상관성이 존재하지 않는 HBO 유입통행을 제외한 모든 통행에서 국지적 공간상관성 지표의 평균값이 GWR 모형을 적용함으로써 현저히 줄어든 것을 볼 수 있다. 또한 국지적 공간상관성이 GWR 모형에 반영되었기 때문에 결과적으로 국지적 공간상관성이 유의하게 나타나는 존의 개수가 감소한 것을 확인할 수 있다.

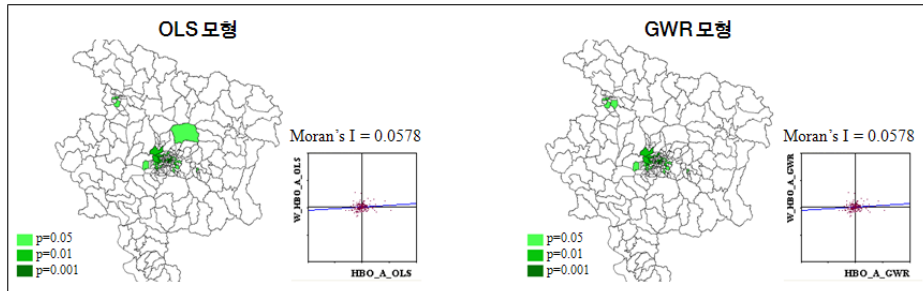
<표 3> 목적통행별/모형별 국지적 공간상관성 지표(Local Moran's I) 값 차이 비교

| 통행목적             | 모형 | 평균  | 표준편차  | 최소값   | 최대값    | 유의한 존의 개수 |           |
|------------------|----|-----|-------|-------|--------|-----------|-----------|
| H<br>B<br>O      | 유출 | OLS | 0.119 | 0.551 | -2.732 | 2.834     | 38(14.8%) |
|                  |    | GWR | 0.049 | 0.495 | -2.324 | 2.655     | 29(11.3%) |
|                  | 유입 | OLS | 0.058 | 0.649 | -3.273 | 5.037     | 42(16.3%) |
|                  |    | GWR | 0.058 | 0.649 | -3.273 | 5.037     | 43(16.7%) |
| N<br>H<br>B<br>O | 유출 | OLS | 0.217 | 1.020 | -1.390 | 9.371     | 51(19.8%) |
|                  |    | GWR | 0.050 | 0.549 | -1.442 | 4.073     | 34(13.2%) |
|                  | 유입 | OLS | 0.209 | 0.996 | -1.710 | 8.088     | 47(18.3%) |
|                  |    | GWR | 0.084 | 0.660 | -1.945 | 5.127     | 36(14.0%) |

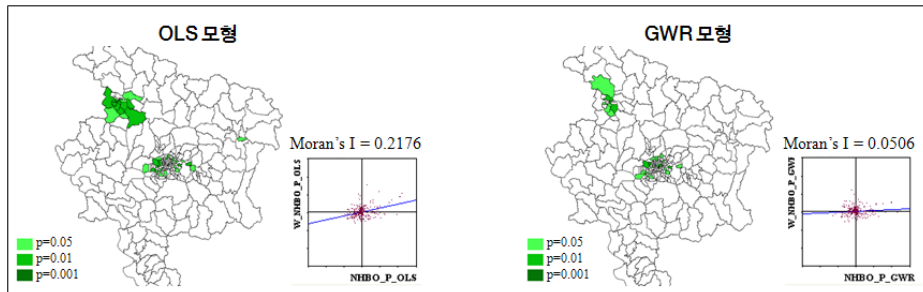




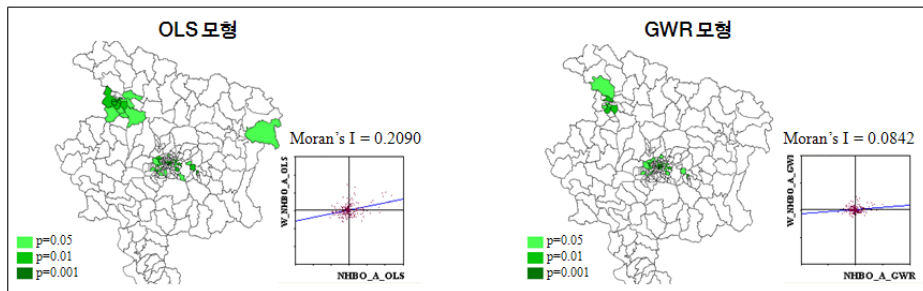
(a) 가정기반 기타통행 유출(HBO Production)



(b) 가정기반 기타통행 유입(HBO Attraction)



(c) 비가정기반 기타통행 유출(NHBO Production)



(d) 비가정기반 기타통행 유입(NHBO Attraction)

〈그림 4〉 목적통행별/모형별 공간상관성 지표의 유의성 비교

#### IV. 결론

본 연구에서는 인간의 사회·경제적인 활동으로 인하여 필연적으로 발생하는 공간적 상관관계가 통행발생에

영향을 미치는 것을 확인하였다. 이러한 공간적 상관관계는 전역적인 현상으로 확인할 수 있고 국지적인 현상으로 확인할 수 있다. 본 연구에서 제시한 모형은 국지적 공간상관성으로 인하여 일반적인 통행발생모형의 결과가



왜곡되는 것을 개선하는 모형이다. 이와 같은 개선효과를 국지적 공간상관성 지표인 Local Moran's I의 변화로 확인 할 수 있다.

분석대상지역에서 분석하고자 하는 통행발생량이 공간적 상관관계가 존재할 때는 본 연구에서 제시한 모형이 효과적이다. 그러나 공간적 상관관계가 크게 존재하지 않을 경우에는 일반적인 선형회귀모형과 GWR 모형은 유사한 결과를 도출하는 것을 확인하였다. 따라서 통행발생모형 구축 시에는 우선적으로 본 연구에서 제시한 공간상관성 지표를 이용하여 공간상관성 존재여부를 확인한 후 적절한 모형을 선택하여 적용하는 것이 합리적이다.

본 연구는 국지적 공간상관성을 확인하고 이를 반영할 수 있는 모형을 제안하는 것에 초점을 맞추었기 때문에 GWR 모형의 산출결과를 충분히 다루지 못한 한계가 존재한다. 존 별로 각각 다른 계수가 추정되기 때문에 이에 대한 해석은 분석대상지역의 통행발생 특성을 이해하기에 매우 효과적이다. 따라서 향후에는 이와 같은 부분에 대한 상세한 연구가 필요한 것으로 생각된다. 또한 공간상관성 지표에 적용한 공간가중치와 GWR 모형에 적용한 공간가중치가 상이하다는 한계가 존재한다. 공간상관성 지표의 공간가중치는 외생적으로 정의가 가능하기 때문에 GWR 모형에서 추정된 공간가중치를 적용하여 공간상관성 지표를 산출하는 것이 가능할 것으로 예상된다. 이와 같이 공간가중치를 공통적으로 적용하여 GWR 모형의 효과를 더욱 명확히 분석하는 연구도 필요하다.

알림: 본 논문은 대한교통학회 제63회 학술발표회 (2010. 10. 29)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

## 참고문헌

1. 박일섭·정진혁(2009), "공간상관성을 고려한 통행발생모형", 제61회 학술발표회, 대한교통학회, pp.467~472.
2. Anselin L.(1995), "Local Indicator of Spatial Association - LISA", Geographical Analysis, Vol.27, No.2.
3. Anselin L. and A. K. Bera(1998), "Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics", In: A. Ullah and D.E.A Giles, Eds., Handbook of Applied Economic Statistics, New York: Marcel Dekker, pp.237~289.
4. Cliff A. D. and Ord J. K.(1981), "Spatial Processes : Models & Applications", Pion Ltd., 207 Brondesbury Park, London.
5. Fotheringham A. S., Brunson C. and Charlton M.(2002), "Geographically Weighted Regression: the analysis of spatially varying relationships", John Wiley & Sons LTD, University of Newcastle, UK.
6. Kim J. H., Chung J. H., Kim I. K. and Lim U.(2009), "Spatial Dependence in Trip Generation Models: A Spatial Econometric Perspective", Transportation Research Board 88th Annual Meeting, Washington, D.C.

- ♣ 주 작성자 : 김진희
- ♣ 교신저자 : 정진혁
- ♣ 논문투고일 : 2010. 11. 25
- ♣ 논문심사일 : 2010. 12. 31 (1차)  
2011. 3. 3 (2차)  
2011. 3. 7 (3차)
- ♣ 심사판정일 : 2011. 3. 7
- ♣ 반론접수기한 : 2011. 8. 31
- ♣ 3인 익명 심사필
- ♣ 1인 abstract 교정필