

통행발생예측의 오차 보정을 위한 모형 개발*

Developing Models for Adjusting the Error Rates of Trip Generation Forecasts

이향숙** · 추상호*** · 신현준****

Lee, Hyang-Sook · Choo, Sang-Ho · Shin, Hyun-Joon

Abstract

The paper analyzed error rates between trip generation forecasts and measurement values, and developed regression models for finding a forecast adjustment method. We estimated general linear regression models as well as spatial regression models that consider spatial autocorrelation among analysis regions. Regression models by total, commute, business and other purposes for intra-metropolitan trips and regression models by total, business and other purposes for inter-regional trips were developed, respectively. As results, intra-metropolitan trips showed a higher spatial autocorrelation than inter-regional trips. Spatial error(SE) model was chosen for intra-metropolitan commute trips, while ordinary least square(OLS) model was selected for the others. In cases of inter-regional trips, spatial error(SE) model or spatial lagged(SL) model was suitable for all purpose trips. Based on these models, trip generation forecasts were updated, showing improved RMSE%. The research could contribute to reducing forecast errors and increasing data reliability.

키 워 드 ▪ 통행발생, 통행예측보정, 오차율, 회귀모형, 공간자기상관

Keywords ▪ Trip generation, Trip forecast adjustment, Error rate, Regression model, Spatial autocorrelation

I. 서 론

통행발생의 예측은 일반적으로 4~5년 단위로 실시되는 대규모 조사자료 및 장래 사회경제적 여건 변화 전망을 토대로 구축한 수리적 모형에 기반하고 있다. 그러나 예산의 제약으로 인해 조사수행 주기가 매우 길다는 근본적 한계점이 존재하며, 장래 여건변화가 기준년도 예측과 차이를 보일 경우 통행수요 오차가 발생하게 된다. 실제로 2006년에 예측되었던 2010년 통행수요와 2010년 조사자료를 이용하여 전수화된 통행수요를 비교해 보면 지역별,

목적별로 상당한 오차가 발생한 것을 알 수 있다.

통행수요예측의 오차를 감소시키기 위해서는 예측단계에서 신뢰성 있는 통행발생모형을 추정하는 것이 가장 이상적이나, 활용 가능한 장래 사회경제 지표의 종류가 매우 한정적이어서 정확한 예측이 어려운 실정이다. 따라서 적절한 보정기법의 개발을 통해 예측자료의 정확성을 높일 필요가 있다. 이를 위해서는 우선 과거시점에 예측된 장래년도 통행량과 차후 그 장래년도를 기준년도로 하여 전수화된 통행량 간의 오차율을 분석하고, 오차율에 영향을 미치는 요인을 규명해볼 수 있다. 이를 토대로 과

* 본 논문은 2013년 한국ITS학회 춘계학술대회에서 발표한 논문의 내용을 수정 및 보완한 것입니다.

** 홍익대학교 환경개발연구원 연구교수 (제1저자: hslee3060@gmail.com)

*** 홍익대학교 도시공학과 부교수 (교신저자: shchoo@hongik.ac.kr)

**** 홍익대학교 도시공학과 학사과정 (shout7760@nate.com)

거시점과 현재시점 간의 사회경제적 여건의 변화, 고령자와 1인 가구 증가 등 사회구조의 변화를 고려하여 예측치를 보정할 수 있을 것이다. 추상호 외(2012)는 카테고리법을 이용하여 교통존의 사회경제적 특성에 따른 조정계수를 산정하고, 이를 기존 예측치에 반영함으로써 오차율 감소효과를 검증한 바 있다. 그러나 오차율에 대한 설명변수를 연속함수가 아닌 급간으로 적용하여 정교성이나 해석상의 한계가 존재하였다.

이러한 문제점을 보완하기 위해 본 연구에서는 다양한 사회, 경제, 토지이용 관련 지표를 이용한 회귀모형을 통해 오차율을 추정하고, 예측치를 보정하는 방법을 연구하고자 한다. 김광구(2003)에 따르면 보통 교통존과 같이 공간을 준거로 하는 데이터들은 인접한 지역에서 서로의 값들에 대해 영향을 주고받는 상호의존적인 특성, 즉 공간자기상관성을 가지게 된다. 따라서 일반적으로 이용되는 선형회귀 모형과 오차율의 지역적 연관성을 고려한 공간회귀 모형을 함께 추정하여 예측오차의 감소효과를 비교 분석해 보고자 한다. 이를 통해 적절한 모형을 최종 선택하여 장래 통행수요를 갱신 및 보완함으로써 빈번히 발생하는 예측오차를 감소시키고, 자료의 신뢰성을 향상시킬 수 있을 것으로 기대된다. 선행 연구는 통행발생예측모형을 구축하고, 예측오차, 검증방법 등을 분석하는데 집중되어 있는 반면 본 연구는 오차율에 영향을 미치는 다양한 요인을 규명하고, 모형구축을 통해 예측오차를 줄였다는 점에서 차별성을 가진다.

II. 선행연구 고찰

국·내외 관련 선행연구를 살펴보면 대부분 교통수요예측과정에서 발생하는 오차의 유형 및 영향을 정의하거나, 실측치와 예측치를 비교하여 수요예측

의 정확성을 분석하였다. 또한 예측단계별 검증 또는 개선 방안을 논의한 연구사례도 다수 존재하였다. 그러나 예측자료의 오차율을 감소시키기 위한 연구는 거의 이루어지지 않았다.

다음의 연구들은 수요예측시의 발생하는 오차의 원인을 유형화하여 분석하였다. 정성봉·장수는(2007)은 교통수요추정 4단계의 각 단계별 오차유형과 그 영향을 분석하였다. CSI(2010)는 흔히 발생하는 오차의 종류로 모형구조오차, 모형집계오차, 모형추정자료오차, 입력자료오차, 정산자료오차를 정의하였다. Flyvbjerg(2005)는 수요추정시 발생하는 다양한 오차의 측정방법을 제시하였으며, Bovy·Jansen(1983)은 교통존의 크기와 네트워크의 상세도에 의해 발생하는 오차가 통행배정 결과에 미치는 영향을 규명하였다.

통행수요의 실측치와 예측치를 비교·분석한 연구는 다음과 같다. Skamris·Flyvbjerg(1997)는 대규모 사업을 대상으로 예측 교통량 및 건설비용을 건설 후 실측교통량, 실제건설비용과 비교하였다. Flyvbjerg 외(2005)는 다양한 통계적기법을 활용하여 여러나라의 인프라구축사업에서 통행수요가 과다하게 추정되고 있음을 밝혀냈다. 김강수·조혜진(2008)은 건설전 도로의 예측교통량과 건설 후 실제 교통량을 비교하여 도로교통수요추정의 위험도를 분석하였다.

또한 수요예측의 정확성 제고를 위해 다양한 가이드라인을 제시한 연구들이 있었다. 윤하중 외(2004)는 수요예측의 신뢰성을 향상시키기 위한 교통수요 예측기준 및 지침을 제시하였다. 정일호·오성호(2005)는 자료의 불확실성을 추정과정에 반영하기 위해 구간추정 예측법을 연구하고, 예측시 적정 신뢰수준을 제시하였다. 박민철 외(2010)는 기존 점 통행수요예측에서 발생하는 각 단계별 오차에 대한 검증방법을 논하였다. 정성봉(2011)은 각 단계에서 발생할 수 있는 오차요인을 전문가 합의법

(Delphi Technique)을 통해 도출하고, 개선방안을 제시하였다. Wegmann·Everett(2011)는 대도시권의 현재와 장래 교통수요모형에 대한 정산 및 검증방법을 연구하였으며, Metropolitan Council(2011)은 지역 대중교통을 중심으로 통행수요추정의 검증방안을 제시하였다. 김강수·정성봉(2005)은 여객교통수요 검증을 위한 기초연구를 실시하였다. 최기주 외(2000)는 수요예측오차를 분석하고, 이를 GIS 공간데이터를 활용하여 개선하는 방안을 제시하였다.

한편 추상호 외(2012)는 예측수요의 오차율을 감소시키기 위해 2002년 기준으로 예측된 2006년 통행량과 2006년 전수화된 통행량 간의 오차율을 산출하고, 교통존 특성에 따른 조정계수를 산정하여 예측치를 보정함으로써 예측치의 신뢰도를 향상시켰다.

선행연구에서 회귀모형을 통해 오차율을 추정할 경우는 거의 없었지만, 통행발생량을 추정한 경우는 다수 존재하였다. 이 경우 일반적으로 선형회귀모형이 고려되었으며, 공간회귀모형을 통해 인접지역의 특성을 반영한 경우는 미미하였다. 진창종 외(2012)의 연구가 통행발생모형으로 선형회귀모형과 공간회귀모형을 모두 추정하여 그 결과를 비교·분석한 바 있다.

본 연구는 오차율을 이용하여 예측수요를 보정하는데 그 목적이 있다. 추상호 외(2012)가 카테고리 분석을 통해 교통존별로 적용가능한 조정계수를 산정하였다면, 본 연구는 회귀모형을 통해 오차율을 추정하여 예측통행량을 보정하고자 한다. 이때 오차율을 종속변수로 하는 선형회귀모형과 공간회귀모형을 함께 추정하여 오차율을 더 많이 감소시키는 보정방안을 최종 선택하게 된다.

Ⅲ. 연구의 범위 및 방법

본 연구는 지역적 범위에 따라 두 가지 자료를 이용하였다. 첫째, 수도권자료는 2010년 수도권 가구통행실태조사자료를 토대로 전수화된 목적별 통행발생·도착량(실측치) 및 2006년에 예측한 2011년, 2016년 통행발생·도착량을 이용하여 산출된 2010년 통행발생·도착량(예측치)을 이용하였다. 둘째, 전국자료는 2010년 전국 가구통행실태조사로 토대로 전수화된 목적별 통행발생·도착량(실측치) 및 2006년에 예측한 2011년, 2016년 통행발생·도착량을 이용하여 산출된 2010년 통행발생·도착량(예측치)을 이용하였다. 본 연구는 기준년도에 전수화된 통행발생·도착량, 즉 실측치가 참값이라고 가정하였다. 단, 실측치는 전수조사가 아닌 2~3%의 표본자료를 전수화해서 구축된 자료이므로 표본오차를 포함하고 있다는 한계점이 있다.

통행목적은 총통행과 주요 목적통행인 출근통행, 업무통행, 기타통행에 대해 분석하였다. 서울특별시 및 광역시의 분석단위는 '구'로, 기타지역의 분석단위는 '시·군'으로 설정하였다. 수도권은 총 65개, 전국은 총 229개 교통존에 대한 분석을 수행하였다.

분석은 다음의 단계를 통해 진행되었다. 첫째, 통행발행량의 실측치와 예측치간 오차율을 산정하였다. 둘째, 오차율을 종속변수로 하고, 통행량에 영향을 줄 것으로 예상되는 다양한 사회, 경제, 토지이용 지표를 독립변수로 하는 선형회귀모형을 추정하였다. 이때, 단계적선택방법(stepwise)을 적용하며, 변수 간 다중공선성이 있는 경우(VIF검증 값이 10 이상) 해당변수를 제외시켰다. 셋째, Moran's I 분석으로 종속변수의 공간자기상관을 탐색하여 공간회귀분석의 적용여부를 판별하고, 공간회귀모형을 추정하였다. 넷째, 선형회귀모형과 공간회귀모형에서 얻어진 오차율을 이용하여 기존 예측치를 보정하고, 기존 예측치와 보정된 예측치의 RMSE%를 비교함으로써 결과를 검증하였다.

선형회귀분석을 위해 SPSS 19.0을 이용하였으며,

공간자기상관의 탐색 및 공간회귀분석을 위해 OpenGeoDa 1.0.1을 활용하였다.

IV. 모형의 정립

1. 오차율 분석

총통행량 및 목적별 통행량의 2010년 실측치와 예측치 간 오차율을 KDI 예비타당성 조사의 적정오차수준인 ±15%를 기준으로 하여 다음과 같이 3개 범주로 나누어 분석하였다.

- 과대추정: 15% 이상
- 적정추정: -15% ~ 15%
- 과소추정: -15% 이하

본 연구에서 사용한 오차율은 다음과 같이 정의하였다.

$$(1) \quad \text{오차율}(\%) = \frac{\text{예측치} - \text{실측치}}{\text{실측치}} \times 100$$

여기서 실측치는 2010년을 기준년도로 전수화된 통행량이고, 예측치는 2006년을 기준년도로 하여 장래년도를 예측한 것으로, 기준년도의 차이로 인해 조사방법, 조사지역 등에 차이가 발생하여 자료의 일관성이 다소 결여된다는 한계점이 존재한다.

수도권의 오차율 분석결과, 총통행 발생과 도차는 대부분 적정 범위에 들었지만, 경기외곽지역에서 일부 과소 또는 과대추정을 보인 것으로 분석되었다. 출근통행은 적정범위내 들어가는 지역이 총통행에 비해 줄어드는 경향을 보여 정확성이 다소 떨어지는 것으로 나타났다. 한편 업무통행과 기타통행은 전반적으로 과대추정을 보였다.

목적통행의 오차율은 전반적으로 과대 또는 과소 현상을 보이는 경향이 있어 결과적으로 이들의 합인 총통행에서 적정범위에 들어가는 지역이 많아지

는 원인이 될 수 있다.

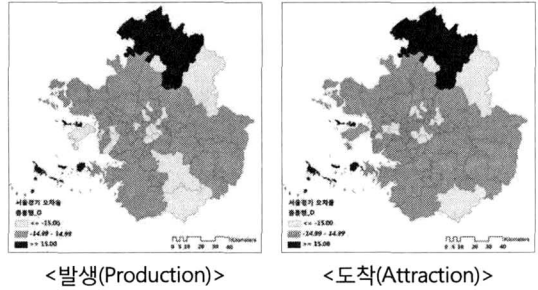


그림 1. 총통행 발생/도착 오차율 (수도권)
Figure 1. Error Rates of Total Trip Production/Attraction(Seoul Metropolitan Area)

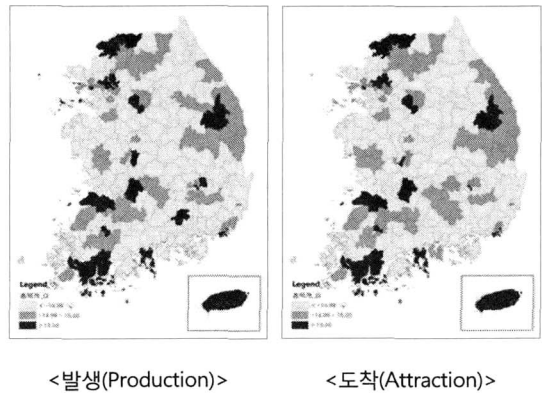


그림 2. 총통행 발생/도착 오차율 (전국)
Figure 2. Error Rates of Total Trip Production/Attraction(Nation)

전국의 오차율을 살펴보면 총통행 발생과 도착 모두 대부분 지역에서 과소추정을 보였다. 업무통행의 경우 총통행에 비해 과소경향이 더 강하게 나타났다. 반면, 기타통행은 3개 범주가 혼재되어 나타나는 형태를 보였다.

2. 모형의 정립

목적별 통행량 및 발생량의 오차율에 대한 선형 회귀모형을 우선 구축하고, 공간자기상관이 있는 경우 공간오차모형과 공간시차모형을 추가로 구

축하여 비교·분석하는 과정을 거치게 된다.

선형회귀모형(Ordinary Least Square Model, OLS모형)의 식은 다음과 같다.

$$y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \dots + \epsilon \quad (2)$$

y : 종속변수, X : 독립변수
 δ : 추정계수, ϵ : 오차항

공간회귀모형은 일반적으로 공간오차모형(Spatial Error Model, SE모형)과 공간시차모형(Spatial Lagged Model, SL모형)을 이용한다. 모형추정에 앞서 Moran's I 분석을 통해 각 목적별 통행발생량에 공간자기상관이 탐색되었다면 공간회귀모형을 통해 이를 통제하게 된다.

SE모형은 오차항의 자기상관을 통제하는 모형으로 다음과 같이 표현된다.

$$y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \dots + u \\ u = \lambda Wu + e \quad (3)$$

y : 종속변수, X : 독립변수
 u : 공간자기상관을 갖는 오차항
 W : 공간가중치 행렬
 δ, λ : 추정계수
 e : 공간자기상관이 없는 오차항

SL모형은 종속변수와 오차항의 자기상관을 통제하기 위한 모형으로 다음과 같이 표현된다.

$$y = \rho Wy + \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \dots + e \quad (4)$$

y : 종속변수, X : 독립변수
 W : 공간가중치 행렬
 δ, ρ : 추정계수
 e : 오차항

GeoDa프로그램의 회귀모형 선정과정은 <그림 3>에 나타나 있다. 먼저 변수의 다중공선성 및 이분산성을 판별하고, Moran's I 공간자기상관 분석, LM 진단 등을 순차적으로 수행하여 최종적으로 적합한 모형을 선택한다.

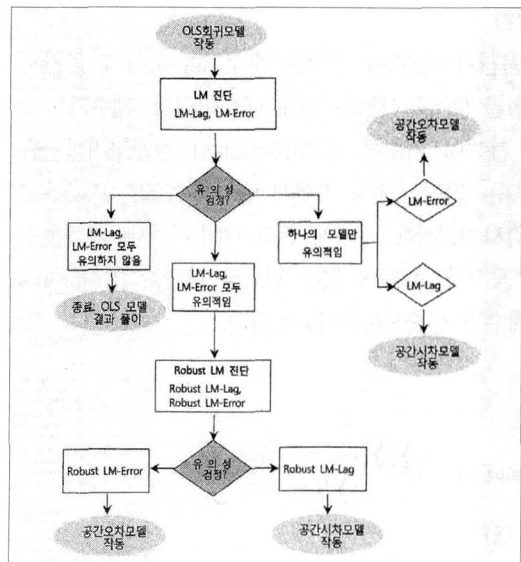


그림 3. 공간회귀모형 선택 과정
 Figure 3. Selection Process of Spatial Regression Model

공간자기상관 검정이 유의하지 않을 경우 OLS모형을 선택하고, 유의할 경우 SL모형과 SE모형 중 더 유의한 모형을 선택하게 된다. 모형선택을 위해 LM-Lag 및 LM-Error 검정을 수행하며, LM-Lag 값을 유의한 경우 SL모형을, LM-Error 값이 유의

한 경우 SE모형을 사용하게 된다.

3. 공간자기상관의 탐색

공간상의 실체들은 공간에 무작위적으로 분포하지 않고 서로 영향을 주고받으며, 가까운 것이 먼 것보다 더 밀접한 관련을 갖는다(Tobler, 1970; 김광구, 2003). 행정구역과 같이 면(Polygon)단위의 교통존을 분석단위로 할 경우 데이터들은 공간단위와 공간현상의 불일치가 발생할 가능성이 있어 올바른 공간정보를 반영하지 못하기 때문에 공간자기상관을 일으키게 된다(변필성, 2009; 진창종 외, 2012).

공간자기상관의 탐색은 주로 Moran's I 공간자기상관계수의 산정을 통해 이루어진다. 계수가 -1에 가까이 갈수록 음의(Negative) 공간자기상관을 가지는 반면, 1에 가까이 갈수록 양의(Positive) 공간자기상관을 가지게 된다. 계수가 0에 가까울수록 공간자기상관이 없는 것으로 판단한다. Moran's I 계수의 산출식은 다음과 같다.

$$Moran's I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2}$$

(5)

x_i : 공간 i 의 값, \bar{x} : x 의 평균값
 n : 개체수, W : 공간가중치 행렬

여기서 공간가중치 행렬로 공간적 인접성(Spatial contiguity)을 정의하는 것이다. 본 연구는 인접기반방식 중 하나인 Queen방식(Queen-Based Contiguity)을 선택하였다. Queen방식은 선과 점을 공유하는 면에 대해 가중치를 부여하는 방법인데,

본 연구의 공간단위인 행정구역은 다양한 크기와 형태를 가지고 있기 때문에 인접기반방식의 가중치 설정이 타당할 것으로 판단된다.

수도권 및 전국 목적별 통행발생·도착량 오차율의 Moran's I 값은 <표 1>과 같다. 수도권의 Moran's I 계수 산정결과, 모두 정적 공간자기상관을 보이는 것으로 나타났다. 그러나 Moran's I 계수가 0.1 미만인 경우가 많아 수도권 지역들의 공간자기상관이 전반적으로 낮음을 알 수 있다.

표 1. 수도권 Moran's I 공간자기상관 계수
 Table 1. Moran's I Spatial Autocorrelation Coefficient in Seoul Metropolitan Area(SMA)

목적(발생) Purpose (Production)	Moran's I	목적(도착) Purpose (Attraction)	Moran's I
총통행 Total	0.0618	총통행 Total	0.1109
출근통행 Commute	0.0376	출근통행 Commute	0.0127
업무통행 Business	0.0720	업무통행 Business	0.1441
기타통행 Other	0.1077	기타통행 Other	0.0233

표 2. 전국 Moran's I 공간자기상관 계수
 Table 2. Moran's I Spatial Autocorrelation Coefficient in Nation

목적(발생) Purpose (Production)	Moran's I	목적(도착) Purpose (Attraction)	Moran's I
총통행 Total	0.4956	총통행 Total	0.4839
출근통행 Commute	0.1150	출근통행 Commute	0.3001
업무통행 Business	0.4714	업무통행 Business	0.2647
기타통행 Other	0.4081	기타통행 Other	0.2946

전국의 경우 마찬가지로 모든 변수가 정적 공간자기상관을 보였다(표 2 참조). 대부분의 Moran's

I 계수가 0.2이상으로 공간자기상관이 비교적 높게 나타났다.

V. 분석결과

1. 수도권 통행량 오차율

위의 분석과정을 통하여 수도권 총통행, 출근통행, 업무통행, 기타통행에 대한 발생·도착량 오차율 모형이 각각 추정되었다. 출근통행 발생 오차율모형에서 SE모형이 선택된 것을 제외하고 모두 OLS모형이 선택되어, 수도권 지역 내 통행시 오차율간의 공간자기상관성이 적음을 시사 하였다.

총통행 발생 오차율모형의 추정 결과, OLS모형이 적합한 것으로 나타났다. 영향을 미치는 변수는 여성비율, 2인 가구 비율, 중학교수, 고속도로연장, 1차종사자수, 공원시설수, 인/가구, 공업지역비율이 있으며, 이 중 2인가구비율과 중학교수, 1차종사자수, 인/가구는 오차율과 양의 상관관계를 보인 반면, 여성비율, 고속도로연장, 공원시설수, 공업지역비율은 오차율과 음의 상관관계를 보였다. 특히, 여성비율과 2인 가구비율이 오차율에 미치는 영향이 크다는 것을 알 수 있었다.

총통행 도착 오차율모형의 경우 발생 오차율모형과 마찬가지로 OLS모형이 선택되었다. 행정구역면적과 1차종사자는 오차율 과대에 기여하는 반면, 여성비율과 고속도로연장은 오차율 과소를 유발하는 것으로 분석되었다.

표 3. 수도권 총통행 발생 오차율모형
Table 3. Model of Error Rates of Total Trip Production in SMA

변수 Variable	OLS	
	계수 Coefficient	P-value
상수 Constant	160.58	0.04*
여성비율(%) Female proportion	-4.64	0.00*
2인가구비율(%) Two-person HH proportion	1.27	0.00*
중학교수(개소) No. of middle schools	0.25	0.05*
고속도로연장(m) Length of highway	-0.00033	0.00*
1차종사자수(인) No. of employees in primary industry	0.019	0.02*
공원시설수(개소) No. of parks	-0.00034	0.02*
가구당인구수(인/가구) Average HH size	13.16	0.08**
공업지역비율(%) Ratio of industrial area	-0.22	0.07**
R ²	0.45	

주: *P-value < 0.05, **P-value < 0.1

표 4. 수도권 총통행 도착 오차율모형
Table 4. Model of Error Rates of Total Trip Attraction in SMA

변수 Variable	OLS	
	계수 Coefficient	P-value
상수 Constant	172.81	0.03*
여성비율(%) Female proportion	-3.61	0.02*
행정구역면적(ha) Administrative area	0.00015	0.01*
1차종사자수(인) No. of employees in primary industry	0.021	0.02*
고속도로연장(m) Length of highways	-0.00028	0.00*
R ²	0.32	

주: *P-value < 0.05

출근통행 발생 오차율모형으로는 SE모형이 선택

되었으며, 상업지역비율, 1인가구비율, 3인가구비율, 3차종사자수, 주택수, 상업지역면적이 주요 영향요인으로 선택되었다. 이 중 상업지역비율 및 면적은 오차율의 과대추정과 관계가 있는 반면, 1인 가구, 3인 가구, 주택수는 과소추정을 야기하는 것으로 나타났다. 특히, 상업지역비율, 1인가구비율, 3인가구비율이 오차율에 미치는 영향이 큰 것으로 분석되었다.

표 5. 수도권 출근통행 발생 오차율모형
Table 5. Model of Error Rates of Commute Trip Production in SMA

변수 Variable	OLS		SEM	
	계수 Coefficient	P-value	계수 Coefficient	P-value
상수 Constant	207.64	0.00*	218.22	0.00*
상업지역비율(% Ratio of commercial area	2.46	0.00*	2.74	0.00*
3인가구비율(% Three-person HH proportion	-7.34	0.00*	-7.67	0.00*
1인가구비율(% Single HH proportion	-2.40	0.00*	-2.57	0.00*
3차종사자수(인) No. of employees in tertiary industry	0.00011	0.00*	0.00011	0.00*
주택수(개소) No. of houses	-0.00016	0.00*	-0.00015	0.00*
상업지역면적(km ²) Commercial area	3.53	0.05*	2.67	0.08**
R ²	0.72		0.74	
Lag coeff.			(Lambda): -0.32, p=0.07**	

주: *P-value < 0.05, **P-value < 0.1

출근통행 도착 오차율모형은 OLS모형이 적합한 것으로 나타났으며, 이 오차율에 영향을 미치는 변수로는 공업지역비율, 고속도로연장, 종사자수/인으로 모두 오차율과 음의 상관관계를 보였다.

표 6. 수도권 출근통행 도착 오차율모형
Table 6. Model of Error Rates of Commute Trip Attraction in SMA

변수 Variable	OLS	
	계수 Coefficient	P-value
상수 Constant	17.33	0.00*
공업지역비율(% Ratio of industrial area	-0.52	0.01*
고속도로연장(m) Length of highways	-0.00037	0.00*
인당종사자수(인/백인) No. of employees per 100 people	-0.35	0.00*
R ²	0.52	

주: *P-value < 0.05

업무통행 발생 오차율모형은 OLS모형이 선택되었으며, 4인가구비율과 대학교수는 오차율의 과대와, 종사자수/인, 고속도로연장은 오차율의 과소와 연관성이 있는 변수로 선정되었다. 특히, 4인가구비율의 영향이 다른 변수에 비해 큰 것으로 나타났다.

표 7. 수도권 업무통행 발생 오차율모형
Table 7. Model of Error Rates of Business Trip Production in SMA

변수 Variable	OLS	
	계수 Coefficient	P-value
상수 Constant	-42.45	0.03*
인당종사자수(인/백인) No. of employees per 100 people	-0.15	0.07**
4인가구비율(% Four-person HH proportion	2.36	0.00*
고속도로연장(m) Length of highways	-0.00066	0.00*
대학교수(개소)	3.93	0.01*
R ²	0.32	

주: *P-value < 0.05, **P-value < 0.1

업무통행 도착 오차율모형의 경우 OLS모형이 적합한 것으로 나타났다. 종사자수/인과 주택수/인이 오차율에 음의 영향을 미치는 것으로 분석되었다.

표 8. 수도권 업무통행 도착 오차율모형
Table 8. Model of Error Rates of Business Trip Attraction in SMA

변수 Variable	OLS	
	계수 Coefficient	P-value
상수 Constant	150.32	0.00*
인당종사자수(인/백인) No. of employees per 100 people	-0.21	0.02*
인당주택수(개소/백인) No. of houses per 100 people	-3.69	0.01*
R ²	0.20	

주: *P-value < 0.05

기타통행 발생 오차율모형 역시 OLS모형이 가장 적합한 것으로 나타났으며, 3인가구비율, 1차종사자수, 주차장수, 종사자수/인이 영향요인으로 선택되었다. 이 중 1차종사자수와 종사자수/인은 오차율과 양의 상관관계를 보이는 반면, 3인가구비율과 주차장수는 음의 상관관계를 가지는 것으로 나타났다. 기타통행 도착과는 달리 3인가구비율의 영향이 매우 크다는 것을 알 수 있었다.

기타통행 도착 오차율모형도 OLS모형이 선택되었다. 승용차/인, 주거지역면적이 오차율 과대에, 고등학교수가 오차율 과소에 영향을 미치는 변수인 것으로 나타났다. 특히, 승용차/인의 영향이 두드러졌다.

표 9. 수도권 기타통행 발생 오차율모형
Table 9. Model of Error Rates of Other Trip Production in SMA

변수 Variable	OLS	
	계수 Coefficient	P-value
상수 Constant	57.22	0.01*
3인가구비율(%) Three-person HH proportion	-2.96	0.00*
1차종사자수(인) No. of employees in primary industry	0.022	0.04*
주차장수(개소) No. of parking lots	-0.00049	0.09**
인당종사자수(인/백인) No. of employees per 100 people	0.20	0.00*
R ²	0.47	

주: *P-value < 0.05, **P-value < 0.1

표 10. 수도권 기타통행 도착 오차율모형
Table 10. Model of Error Rates of Other Trip Attraction in SMA

변수 Variable	OLS	
	계수 Coefficient	P-value
상수 Constant	-23.49	0.06**
고등학교수(개소) No. of high schools	-1.31	0.00*
인당승용차수(대/백인) No. of employees per 100 people	1.04	0.02*
주거지역면적(km ²) Residential area	0.72	0.04*
R ²	0.21	

주: *P-value < 0.05, **P-value < 0.1

2. 전국 지역간 통행량 오차율

전국 지역간 통행의 경우 총통행, 업무통행, 기타통행에 대한 발생·도착량 오차율모형이 추정되었다. 수도권과는 달리 모든 통행에서 공간 회귀모형이 선택되었는데, 이는 수도권에 비해 오차

율의 공간자기상관성이 크다는 것을 의미한다.

표 11. 전국 총통행 발생 오차율모형
Table 11. Model of Error Rates of Total Trip Production in Nation

변수 Variable	OLS		SLM	
	계수 Coefficient	P-value	계수 Coefficient	P-value
상수 Constant	46.27	0.00*	31.94	0.00*
1차종사자수(인) No. of employees in primary industry	0.03	0.00*	0.013	0.12***
공업지역비율(%) Ratio of industrial area	-1.41	0.00*	-0.92	0.00*
녹지지역비율(%) Ratio of green area	-0.61	0.00*	-0.39	0.00*
고속도로연장(m) Length of highways	-0.0007	0.00*	-0.00044	0.00*
R ²	0.30		0.47	
Lag coeff.			(Rho): 0.51, p=0.00*	

주: *P-value < 0.05, ***P-value < 0.2

표 12. 전국 총통행 도착 오차율모형
Table 12. Model of Error Rates of Total Trip Attraction in Nation

변수 Variable	OLS		SLM	
	계수 Coefficient	P-value	계수 Coefficient	P-value
상수 Constant	43.3	0.00*	29.99	0.00*
1차종사자수(인) No. of employees in primary industry	0.032	0.00*	0.014	0.09**
공업지역비율(%) Ratio of industrial area	-1.38	0.00*	-0.94	0.00*
녹지지역비율(%) Ratio of green area	-0.60	0.00*	-0.38	0.00*
고속도로연장(m) Length of highways	-0.00064	0.00*	-0.00042	0.00*
R ²	0.28		0.44	
Lag coeff.			(Rho): 0.47, p=0.00*	

주: *P-value < 0.05, **P-value < 0.1

총통행 발생 오차율모형은 SL모형이 적합한 것

으로 나타났다. 오차율에 영향을 미치는 변수로는 1차종사자수, 공업지역비율, 녹지지역비율, 고속도로연장이 있었으며, 이 중 1차종사자수는 오차율에 양의 상관관계를, 공업지역비율과 녹지지역비율, 고속도로연장은 음의 상관관계를 보이는 것으로 나타났다. 수도권과 비교하여 토지이용특성의 영향이 보다 뚜렷하게 나타나는 것으로 분석되었다.

총통행 도착 오차율모형의 경우 발생과 마찬가지로 SL모형이 선택되었으며, 오차율에 영향을 미치는 변수도 거의 유사한 것으로 나타났다. 주로 공업지역비율, 녹지지역비율, 고속도로연장이 오차율의 과소추정을 유발하는 것으로 분석되었다.

표 13. 전국 업무통행 발생 오차율모형
Table 13. Model of Error Rates of Business Trip Production in Nation

변수 Variable	OLS		SLM	
	계수 Coefficient	P-value	계수 Coefficient	P-value
상수 Constant	120.44	0.00*	94.64	0.00*
5인가구비율(%) Five-person HH proportion	4.44	0.01*	4.13	0.01*
인당사업체수(개소/천인) No of companies per 1000 people	-0.44	0.00*	-0.46	0.00*
주택보급율(%)	-1.67	0.00*	-1.30	0.00*
고속도로연장(m) Length of highways	-0.00056	0.00*	-0.00047	0.00*
일반국도연장(m) Length of national arterial	-0.00031	0.00*	-0.00026	0.00*
R ²	0.496181		0.530679	
Lag coeff.			(Rho): 0.24, p=0.00*	

주: *P-value < 0.05

업무통행 발생 오차율모형으로는 SL모형이 선택되었다. 여러 변수 중 5인가구비율, 사업체수/인, 주택보급율, 고속도로연장, 일반국도연장이 오차율에 영향을 미치는 것으로 나타났다. 과대추정의 영

향요인으로 나타난 5인가구비율을 제외한 나머지 변수들은 과소추정을 야기하였으며, 이 중 5인가구 비율과 사업체수/인의 영향이 두드러졌다.

업무통행 도착 오차율모형은 SE모형이 가장 적합한 것으로 나타났다. 다른 모형들에 비해 많은 변수가 선정되었는데, 이 중 인구밀도와 대학생수는 오차율의 과대를, 승용차/인, 상업지역비율, 고속도로연장, 일반국도연장, 초중고학생수는 과소를 유발하는 것으로 나타났다.

표 14. 전국 업무통행 도착 오차율모형
Table 14. Model of Error Rates of Business Trip Attraction in Nation

변수 Variable	OLS		SEM	
	계수 Coefficient	P-value	계수 Coefficient	P-value
상수 Constant	-14.71	0.27	-14.86	0.29
인당승용차수(대/백인) No. of cars per 100 people	-0.88	0.02*	-1.05	0.01*
인구밀도(인/ha) Population density	0.25	0.00*	0.28	0.00*
상업지역비율(% Ratio of commercial area	-1.81	0.02*	-2.41	0.00*
고속도로연장(m) Length of highways	-0.00054	0.02*	-0.00048	0.03*
일반국도연장(m) Length of national arterial	-0.00028	0.00*	-0.00023	0.02*
대학생수(인) No. of college students	0.0019	0.00*	0.0019	0.00*
초중고학생수(인) No. of elementary, middle, high school students	-0.0004	0.00*	-0.00035	0.01*
R ²	0.42		0.46	
Lag coeff.			(Lambda): 0.295, p=0.00*	

주: *P-value < 0.05, **P-value < 0.1, ***P-value < 0.2

기타통행 발생 오차율모형의 경우 SL모형이 선

택되었으며, 5인가구비율, 1차종사자수, 인구밀도, 행정구역면적은 오차율과 양의 상관관계를, 고속도로연장은 음의 상관관계를 보였다.

표 15. 전국 기타통행 발생 오차율모형
Table 15. Model of Error Rates of Other Trip Production in Nation

변수 Variable	OLS		SLM	
	계수 Coefficient	P-value	계수 Coefficient	P-value
상수 Constant	-35.61	0.16***	-10.93	0.63
5인가구비율(% Five-person HH proportion	7.75	0.01*	3.43	0.18***
1차종사자수(인) No. of employees in primary industry	0.038	0.03*	0.021	0.17***
인구밀도(인/ha) Population density	0.36	0.00*	0.21	0.01*
공업지역비율(% Ratio of industrial area	-1.34	0.01*	-0.84	0.06**
고속도로연장(m) Length of highways	-0.00087	0.00*	-0.00058	0.01*
행정구역면적(ha) Administrative area	0.00036	0.02*	0.00022	0.11**
R ²	0.21		0.34	
Lag coeff.			(Rho): 0.45, p=0.00*	

주: *P-value < 0.05, **P-value < 0.1, ***P-value < 0.2

기타통행 도착 오차율 모형으로는 SL모형이 선택되었다. 사업체수/인, 인구밀도가 오차율 과대에, 2인가구비율과 종사자수/인, 고속도로연장이 과소에 영향을 미치는 것으로 나타났다.

표 16. 전국 기타통행 도착 오차율모형
Table 16. Model of Error Rates of Other Trip Attraction in Nation

변수 Variable	OLS		SLM	
	계수 Coefficient	P-value	계수 Coefficient	P-value
상수 Constant	84.57	0.00*	77.03	0.00
2인가구비율(%) Two-person HH proportion	-2.28	0.00*	-2.03	0.00*
인당사업체수(개소/천인) No. of companies per 1000 people	1.18	0.00*	0.99	0.00*
인당종사자수(인/백인) No. of employees per 100 people	-0.89	0.00*	-0.83	0.00*
인구밀도(인/ha) Population density	0.19	0.02*	0.13	0.12***
고속도로연장(m) Length of highways	-0.00042	0.07**	-0.00037	0.09**
R ²	0.25		0.29	
Lag coeff.			(Rho): 0.26, p=0.00*	

주: *P-value < 0.05, **P-value < 0.1, ***P-value < 0.2

3. 보정결과 검증

선형회귀모형과 공간회귀모형으로부터 도출된 오차율과 2010년 실측치를 이용하여 보정된 예측치를 산출하였다. 이를 위해 오차율을 정의한 식(1)에서 y변수를 오차율에서 예측치로 다음과 같이 변환하였다.

$$\text{보정된 예측치} = \left(1 + \frac{\text{오차율}(\%)}{100}\right) \times \text{실측치} \quad (6)$$

식(6)에 회귀모형을 통해 산정된 오차율과 실측치로 가정한 2010년 전수화된 통행량을 적용함으로써 보정된 예측치를 최종 산출하게 된다. 이를 통해 보정된 예측치를 기존 예측치와 비교하여 자료의 신뢰성을 검증하였

다. 이때 RMSE(Root Mean Square Error)%를 이용하였으며, RMSE%가 낮을수록 모형의 설명력이 높아짐을 의미한다.

$$RMSE = \sqrt{\frac{\sum(T_i^0 - T_i^\beta)^2}{N}}$$

T_i^0 : i 지역의 실측치

T_i^β : i 지역의 추정치

N : 분석대상존의 개수

$$RMSE\% = \frac{RMSE}{\text{실측치 평균}} \times 100$$

수도권 예측치 보정에 대한 검증 결과, 선형회귀 모형 또는 공간회귀모형으로 오차율을 산정함으로써 RMSE%를 감소시킬 수 있는 것으로 나타났다. <그림 4>는 수도권 목적별 발생·도착의 기존 예측치와 OLS모형, SL모형 또는 SE모형에 의해 보정된 예측치의 RMSE%를 비교하여 보여준다. 총통행 발생은 오차율을 선형회귀모형으로 추정함으로써 기존 14.2%에서 12.6%로, 도착은 기존 15.0%에서 12.5%로 약 1.6~2.5%의 감소효과를 가져왔다. 업무통행과 기타통행 역시 RMSE%를 약 1.8~3.6%와 5.0~5.4% 감소시킬 수 있었다. 출근통행 발생을 살펴보면 공간회귀모형(22.9%)으로 인한 RMSE% 감소가 선형회귀모형(23.3%)에 비해 다소 크게 나타났다. 비록 출근통행 도착이 유일하게 개선효과를 보이지 않았지만, 오차율모형 추정으로 인해 수도권 교통수요 예측치의 정확성이 전반적으로 향상되었다고 판단할 수 있을 것이다.

전국 지역간 통행은 수도권 통행과 달리 모든 통행에서 공간회귀모형이 채택되었다. 그러나 채택된 모형이 항상 낮은 RMSE%를 보장하지는 않음이 증명되었다. 예를 들어, 총통행 발생, 업무통행 발생·도착의 경우 SL모형 또는 SE모형으로 결정되

있음에도 불구하고, OLS모형의 RMSE%가 더 낮게 나타났다. 그러나 모든 통행에서 회귀모형을 통해 오차율을 추정함으로써 예측치의 정확성을 향상시킨 것은 분명하다. <그림 5>는 전국 지역간 목적별 발생·도착의 기존 예측치와 OLS모형, SL모형 또는 SE모형에 의해 보정된 예측치의 RMSE%를 나타내고 있다. 목적별로 4.3~33.0%의 감소효과를 보인 것을 알 수 있다.

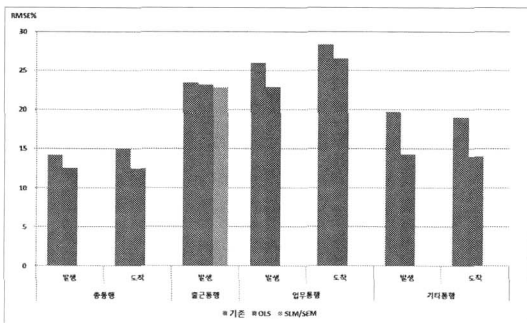


그림 4. 수도권 목적통행의 RMSE%
Figure 4. RMSE% of Purpose Trips in SMA

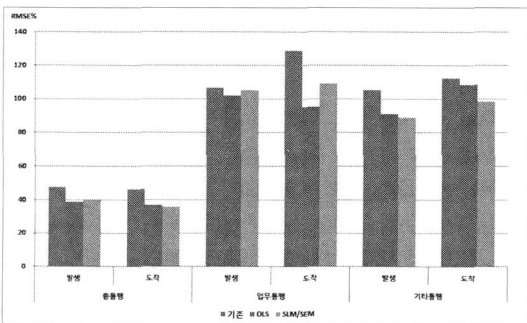


그림 5. 전국 지역간 목적통행의 RMSE%
Figure 5. RMSE% of Purpose Inter-regional Trips in Nation

VI 결론

본 연구는 통행발생예측모형을 보정하는 방법을 개발하기 위해 2010년 수도권 및 전국 지역간 통

행발생·도착량의 예측치와 실측치를 이용하여 오차율을 분석하고, 오차율을 종속변수로 하는 회귀모형을 구축하여 예측치를 보정하였다. 이때 일반적으로 이용하는 선형회귀모형과 분석대상지역의 공간자기상관을 고려하는 공간회귀모형을 함께 구축하여 비교·분석함으로써 예측자료의 오차율을 더욱 감소시키는 방안을 연구하였다.

수도권은 총통행, 출근통행, 업무통행, 기타통행, 전국은 총통행, 업무통행, 기타통행의 발생·도착에 대한 오차율 모형을 추정하였다. 우선 선형회귀모형을 구축하고, 공간자기상관이 있는 경우 공간오차모형과 공간시차모형을 추가로 구축하여 비교하는 과정을 거쳤다. 공간자기상관의 탐색은 Moran's I 공간자기상관계수의 산정을 통해 이루어졌으며, 수도권 통행에 비해 전국 지역간 통행이 높은 상관성을 보이는 것으로 나타났다. 수도권의 경우 출근통행 발생 오차율모형에서 SE모형이 선택된 것을 제외하고 모두 OLS모형이 선택된 반면, 전국에서는 모든 목적통행에서 SL모형 또는 SE모형이 선택되었다.

회귀모형을 통해 산정된 오차율을 이용하여 보정된 예측치와 기존의 예측치를 비교하여 RMSE%를 검증한 결과 한층 개선된 결과가 도출되었으며, 전국 지역간 통행의 감소폭이 훨씬 큰 것으로 나타났다(수도권 1.6~5.4% 감소, 전국 4.3~3.0% 감소). 한가지 흥미로운 점은 채택된 회귀모형(OLS 모형, SL모형 또는 SE모형)이 항상 낮은 RMSE%를 보장하지는 않는다는 것이다. 따라서 우선 프로세스에 따라 모형들을 추정하고 RMSE%를 비교한 후, 가장 낮은 RMSE%를 보이는 모형을 최종 선택함으로써 예측자료의 신뢰성을 더욱 향상시킬 수 있을 것이다.

대부분의 선행연구는 통행발생예측모형을 구축하거나 여객수요 추정시 발생하는 오차의 문제점 및 검증방법을 분석하는데 집중되어 있어 실제로 오차율 개선에 중점을 둔 연구는 미미한 실정이다. 따

라서 본 연구는 오차율에 영향을 미치는 사회, 경제, 토지이용 관련 지표들을 규명하고, 회귀모형을 통해 새로운 오차율을 적용하여 예측치를 산정함으로써 예측오차를 줄였다는 점에서 의의가 있다고 할 수 있다. 또한 실제 오차율에 영향을 주는 주요 변수들을 목적별로 규명하였는데, 이는 향후 통행발생모형 추정에 있어서 이러한 변수를 종속변수로 고려하거나 이들 변수의 장래 예측치를 추정할 경우 보다 신중한 접근방법을 이용해야 한다는 점을 시사하고 있다. 그리고 본 연구에서 지역 간 목적통행의 오차율에 공간자기상관이 존재하는 것으로 나타나 향후 교통수요예측시 통행발생단계에서 공간회귀모형을 고려해야 할 것으로 판단된다.

그러나 본 연구는 한 개 연도를 기준으로 분석되었다는 한계점이 있어 향후 시계열 자료를 이용한 분석을 통해 오차율 및 영향요인의 변화추이를 분석할 필요가 있을 것이다. 또한 오차율의 발생요인을 일차식의 회귀분석모형을 통해 규명하였는데, 이는 목적통행 간의 상관성을 고려할 때 한계가 있는 것으로 판단된다. 따라서 향후 모든 목적통행 오차율간의 인과관계를 종합적으로 규명할 수 있는 연립방정식 형태의 구조방정식 모형을 추정해야 할 것이다.

감사의 글

본 연구는 한국건설교통기술평가원에서 위탁시행한 교통체계효율화사업(09교통체계-지능06)의 지원에 의해 수행되었습니다.

인용문헌

References

1. 김광구, 2003. “공간자기상관의 탐색과 공간회귀분석의 활용”, 『정책분석평가학회보』, 13(1):273-294.

Kim, K., 2003. “Exploration of Spatial Autocorrelation and Application of Spatial Regression Analysis”, *Journal of the Policy Analysis and Evaluation*, 13(1):273-294.

2. 김강수·조혜진, 2008. “도로교통 수요 추정의 위험도 분석 연구”, 『국토계획』, 44(2):179-192.

Kim, K·Cho, H., 2008. “Risk Analysis of Traffic Demand Forecasting”, *Journal of the Korean Planner Association*, 44(2):179-192.

3. 박민철·성홍모·김찬성, 2010. “기종점 통행수요 추정의 단계별 검증방안 개선연구”, 『교통연구』, 17(1):73-85.

Park, M·Sung, H·Kim, C., 2010. “The Improvement of Stepwise Validation Methods for Traffic O/D Estimation”, *Journal of Transport Research*, 17(1):73-85.

4. 변필성, 2009. 공간계량경제모델링 - 공간적 자기상관과 공간회귀모델, 알기 쉬운 국토연구방법론, 경기:국토연구원

Byun, P., 2009. *Spatial Econometric Modeling-Spatial Autocorrelation and Spatial Regression Analysis*, Gyeonggi:KRIHS

5. 윤하중·정일호·김호정·양인석, 2004. 교통수요 예측을 위한 기준 및 절차 지침 연구, 경기:국토연구원.

Yoon, H·Chung, I·Kim, H·Yang, I., 2004. *A Study on The Guideline to Forecast Travel Demand*, Gyeonggi:KRIHS.

6. 정성봉·장수은, 2007. 도로사업의 수요예측 오차발생 원인 및 영향분석, 경기:한국교통연구원.

Chung, S·Chang, S., 2007. *Demand Forecasting Errors in Road Projects: Causes and Effects*, Gyeonggi: KOTI.

7. 정성봉, 2011. “Delphi기법을 통한 교통수요예측 Risk Management 적용 방안”, 『대한안전경영과학학회지』, 13(2): 267-273.

Chung, S., 2011. “Application of Risk Management to Forecasting Transportation Demand by Delphi Technique”, *Journal of the Korea Safety Management & Science*, 13(2): 267-273.

8. 정일호·오성호, 2005. 통행수요 추정 신뢰수준 제고 방안 연구: 구간 통행수요 모형 개발을 중심으로, 경기:국토연구원.
Chung, I·Oh, S., 2005. *Enhancement of Reliability for Traffic Demand Estimation - Focusing on Interval Estimation Model for Traffic Demand*, Gyeonggi:KRISH.
9. 진창중·이향숙·추상호, 2012. "공간회귀분석을 활용한 통행발생모형 추정 연구: 서울특별시를 중심으로", 「국토연구」, 73(1):131-143
Jin, C·Lee, H·Choo, S., 2012. "Developing Trip Generation Models Using Spatial Regression Analysis : A Case for Seoul, Korea", *The Korea Spatial Planning Review*, 73(1):131-143.
10. 최기주·남궁문·윤항목·오승훈, 2000. "교통수요모형의 개관 및개선방안: 모형 및 GIS 측면을 중심으로", 「대한토목학회논문집」, 20(5D): 519-535.
Choi, K·Namkoong, M·Yoon, H·Oh, S., 2000. "An Overview of Travel Demand Modeling and Some Direction of improvements : Model and GIS Aspects", *Journal of the Korean Society of Civil Engineers*, 20(5D): 519-535.
11. 추상호·이향숙·신천규, 2012. "오차율을 이용한 교통존 특성별 통행발생량 조정계수 산정 연구", 「국토계획」, 47(4):253-263.
Choo, S·Lee, H·Shin, C., 2012. "Estimating Adjustment Factors for Trip Generation by Traffic Analysis Zone Characteristics Using the Trip Error Rates", *Journal of the Korean Planner Association*, 47(4):253-263.
12. Bovy, P.H.L. and Jansen, G.R.M., 1983. "Network aggregation effects upon equilibrium assignment outcomes: An empirical investigation", *Transportation Science*, 17(3):240-262.
13. Cambridge Systematics Inc., 2010. "Travel Model Validation and Reasonableness Checking Manual, Second Edition", Travel Model Improvement Program(TMIP), Federal Highway Administration.
14. Flyvbjerg, B., 2005. "Measuring Inaccuracy in Travel Demand Forecasting: Methodological Considerations Regarding Ramp Up and Sampling", *Transportation Research part A*, 39: 522-530.
15. Flyvbjerg, B. Skamris Holm, M.K. and Buhl, S.L., 005. "How (In)accurate Are Demand Forecasts in Public Works Projects?", *Journal of the American Planning Association*, 71(2):131-146.
16. Metropolitan Council, 2011. "Travel Demand Forecasting User Guide: A Supplement to the Regional Transitway Guidelines".
17. Outwater, M.L. and Charlon, B., 2006. "The San Francisco Model in Practice Validation, Testing, and Application", the Innovations in Travel Modeling 2006 Conference, Austin, TX, 21-23.
18. Skamris, M.K. and Flyvbjerg, B., 1997. "Inaccuracy of Traffic Forecasts and Cost Estimates on Large Transport Projects", *Transport Policy*, 4(3):141-146.
19. Tobler, W., 970. "A computer movie simulation urban growth in the Detroit region", *Economic Geography*, 46(1):234-240.
20. Wegmann, F. and Everett, J., 2011. "Minimum Travel Demand Model Calibration and Validation Guidelines for State of Tennessee", Center for Transportation Research, The University of Tennessee.

논 문 투 고 2013-11-12
심 사 완 료 2014-01-10
수 정 일 2014-01-29
계 재 확 정 일 2014-01-29
최 종 본 접 수 2014-02-05