

최단경로 기반 교통량 공간 예측에 관한 연구

허태영¹⁾ 박만식²⁾ 엄진기³⁾ 오주삼⁴⁾

요약

본 연구에서는 연평균일교통량 예측을 위한 공간회귀모형을 제시하였다. 비록 공간 분석을 위하여 조사지점들 간의 유클리디안 거리가 일반적으로 사용되고 있지만, 조사되지 않는 도로의 교통량 예측을 위하여 교통량 조사지점들 간의 최단경로를 이용한 공간회귀모형을 새롭게 시도하였다. 공간예측방법으로는 일반크리깅을 사용하였으며 교차검증을 통하여 정량적으로 최단경로 기반의 교통량 공간예측모형의 타당성을 제시하였다.

주요용어: 교차검증, 세미베리오그램, 연평균 일교통량, 크리깅.

1. 서론

연평균 일교통량 (Annual Average Daily Traffic: AADT)이란 교통량을 측정하는 도로의 한 지점에서 1년간 조사한 24시간 교통량의 평균이다. 연평균 일교통량은 교통량의 증가나 변환 등에 관하여 시간적, 공간적 비교를 용이하게 하며 도로를 계획하고 설계하는데 있어 매우 중요한 기초자료로 활용된다. 이는 연평균 일교통량을 기준으로 해당 도로의 유지 또는 보수와 같은 서비스를 예측하며, 신설 및 확장될 도로의 기하구조가 결정하기 때문에 조사지점에서의 연평균 일교통량이 정확히 산출되어야 도로를 경제적으로 건설할 수 있으며 적당한 도로의 서비스 수준을 유지하면서 교통소통을 원활히 할 수 있다 (오주삼 등, 2002). 당해 년도 1월부터 12월까지 각 지점에 설치된 조사장비를 통하여 365일동안 상시 조사지점에서 교통량 조사가 실시되므로 쉽게 연평균 일교통량을 구할 수 있지만 교통량 조사 예산의 제약으로 모든 도로에 대하여 연평균 일교통량을 측정하기는 현실적으로 불가능하며 상시조사가 이루어지는 조사지점에서도 교통량 조사장비의 불량 및 고장으로 정확한 연평균 일교통량을 얻지 못하는 경우가 존재한다.

환경자료나 지리자료의 공간적 분석에 대한 연구는 활발히 이루어지고 있으나 도로에서 조사되는 교통량 자료에 대한 공간적 분석에 대한 연구는 미미한 실정이다. 조재영 등

- 1) (606-791) 부산광역시 영도구 동삼동 1번지, 한국해양대학교 나노데이터시스템학부 데이터정보전공, 전임강사
E-mail: heoty@hhu.ac.kr
- 2) (136-701) 서울시 성북구 안암동, 고려대학교 의과대학원 유전체및단백체환경독성센터, 연구조교수
E-mail: bayesia@korea.ac.kr
- 3) (437-757) 경기도 의왕시 월암동 360-1, 한국철도기술연구원 철도정책물류연구본부, 선임연구원(Ph.D)
E-mail: jkom00@krri.re.kr
- 4) (411-712) 경기도 고양시 일산구 대화동 2311번지, 한국건설기술연구원 첨단도로교통연구실, 선임연구원(Ph.D)
E-mail: jusam@kict.re.kr

(2001)는 환경자료에 대하여 미지의 위치에 대한 예측력을 알아보고자 공간 통계학의 범용 크리깅 (Ordinary kriging) 방법과 일반 통계학 방법 중 비모수 방법인 국소적 회귀분석방법, 그리고 일반화 가법모형을 비교하여 공간 분석법이 적절함을 보였으며, 유성모와 엄익현 (1999)은 강우강도 자료를 통하여 세미베리오그램 (Semi-variogram)의 모수추정과 공간 이상치 자료에 대하여 연구하였다. 김선우 등 (2005)은 공간 상관이 존재하는 지리통계 자료에 대하여 일반적으로 널리 사용되는 크리깅모형과 공간 모형중의 하나인 지리적 가중 회귀 모형을 고려하고 미지의 위치에 대한 예측력을 비교하였다. 본 연구에서는 교통량 조사 지점의 연평균 일교통량 추정을 위한 공간통계학적 적용 방안에 대해 다루고자 하며 공간자료 분석에서 가장 널리 사용되고 있는 세미베리오그램 모형을 통한 공간예측을 실시하였다.

앞서 언급했듯이 연평균 일교통량 예측을 위한 연구는 기본적인 통계적 방법만을 사용하고 있다. 특히, 기존의 일교통량이나 시간교통량을 기반으로 한 연평균 일교통량을 예측하는 통계적 방법들은 활발히 연구되고 있으나 해당 조사지점의 도로 정보, 인구통계적 정보, 그리고 지리적 정보를 활용하여 연평균 일교통량을 예측하는 연구는 극히 적은 실정이다. 오주삼 등 (2002)은 일교통량의 변동계수, 시간교통량, 그리고 일교통량을 이용하여 회귀 분석을 통한 연평균 일교통량을 추정하는 방법과 신경망 모형을 통하여 연평균 일교통량을 추정하는 방법을 비교하였다. Cheng (1992), Xia 등 (1999), 그리고 Zhao와 Chung (2001)은 연평균 일교통량의 추정을 위하여 일반선형회귀분석을 실시하였으나 Cheng (1992) 등이 제안한 일반선형회귀분석은 교통량 자료의 상관성을 고려하지 않았다는 단점을 가지고 있다. 또한 기존의 대부분의 연평균 일교통량의 예측에 관한 연구 (오주삼 등, 2002, Lam과 Xu, 2000; Tang 등, 2003)는 수시교통량 (예산상의 제약으로 일년 365일간의 교통량을 조사하지 못하고 2-3일간의 교통량만을 조사하는)을 통하여 연평균 일교통량을 추정하는 연구에 국한되어 있으며, 교통량이 조사 되지 않은 지점의 예측에 대한 연구는 단순히 다중 회귀분석에 의한 연구들 (Cheng, 1992; Xia 등, 1999; Zhao와 Chung, 2001)이외의 다른 연구들이 진행되지 않았기 때문에 본 연구에서는 선형회귀분석과 공간선형회귀분석과의 비교만을 통하여 공간상관성을 반영한 공간회귀모형이 우수성을 입증하였다.

최단경로 (Shortest travel path)를 활용한 교통량 연구로 Heo와 Hughes-Oliver (2005)는 미국 노스캐롤라이나 주의 20000 여개의 교통량 조사지점간의 최단경로를 활용하여 공간 상관성을 고려한 공간모형을 통하여 관측된 연평균 일교통량의 신뢰구간을 통하여 이상치를 탐색하는 연구를 실시하였으며, Rathbun (1998)은 강하구 (Estuary)의 염도 (Salinity)와 용존산소 (Dissolved Oxygen)의 공간예측을 위하여 유클리디안 기반 모형과 수거리 (Water distance) 기반 공간 모형을 비교하였다.

본 연구에서는 교통량 예측을 위하여 여러 가지 도로 특성, 인구통계적 특성, 지리 특성과 교통량 조사지점의 거리에 대한 공간정보를 활용하여 교통량 예측 모형을 제시하였다. 특히, 본 연구에서는 Eom 등 (2006)가 제시한 유클리디안 거리 (Euclidean distance)를 기반으로 공간상관성을 반영한 연평균 일교통량 예측에 관한 연구의 확장으로 기존의 대부분의 공간자료 분석에서 유클리디안 거리를 공간거리로 사용한것과 달리 본 연구에서는 조사지점 사이의 최단경로를 활용한 네트워크기반 공간분석을 시도하였다.

조사지점들간의 최단거리를 얻기 위하여 지리정보시스템 (Geographic information system, GIS)을 이용하였다. 지리정보시스템과 공간분석과는 불가분의 관계에 있으며 지리정보시스템에서 교통정보 체계는 교통계획 및 교통영향평가 등의 교통문제 전반에 관한 관리 및 분석을 위한 정보로 활용될 수 있다. 일반적으로 공간분석을 위하여 사용되는 유클

리디안 거리는 공간분석을 하기에 적합하지 않은 경우가 있다. 따라서 도로상에 설치되어 있는 교통량 조사지점간의 거리를 일반적으로 공간통계학에서 사용되는 유클리디안 거리가 아닌 최단경로를 이용하여 얻은 정확한 실제 공간거리를 이용하여 공간분석을 실시하였다. 이 실제 공간거리를 구하기 위하여 GIS 기능을 갖춘 교통분석 프로그램인 트랜스 캐드 (TransCAD) 소프트웨어를 활용하였다.

기존의 대부분의 공간자료 연구에서는 베리오그램 (혹은 세미베리오그램)의 모수 추정을 위해 최소제곱법 (Least squares estimation method)과 최대우도추정법 (Maximum likelihood estimation method)을 사용하였다. 최소제곱법은 분포의 가정없이 구할 수 있다는 것과 대용량의 공간자료 분석에도 용이하다는 장점을 가지고 있는 반면에 모수추정을 위해서 래그 (Lag)로 그룹핑하여 표본 베리오그램 값을 필요로 한다는 단점을 가지고 있다. 베리오그램의 모수 추정에 있어서 대표적인 방법인 최소제곱법 이외에 모수 추정시 표본 베리오그램값을 사용하지 않고 모든 공간자료를 사용하여 직접적으로 모수 추정값을 구하는 최대우도추정법은 대용량의 자료일 때 분산-공분산행렬 (Variance-covariance matrix)의 계산에 많은 시간이 소요된다는 단점을 가지고 있다.

본 연구에서는 모형선택의 기준으로 최소제곱법은 잔차 제곱합 (Residual sum of squares, RSS)을 사용하였으며 최대우도추정법은 우도함수값 (Likelihood)을 사용하였다. 또한 예측 모형의 타당성을 정량적으로 제시하기 위하여 최소제곱방법과 최대우도방법의 모수 추정값으로 계산된 공간 예측성능을 평가하기 위한 기준으로 교차검증방법 (Cross validation)을 사용하였다. 본 논문에서는 2장에서 미국 노스캐롤라이나 주에 있는 Wake county 의 200개 지점의 교통량 조사지점에서 얻어진 교통량 자료에 대한 소개와 최단경로에 대한 간단한 예 및 기본적인 공간통계적 분석 방법에 관하여 설명을 하였으며, 3장에서는 세미베리오그램의 간단한 소개와 최소제곱법과 우도함수추정법을 제시하였으며, 4장에서는 공간 예측 방법으로서 일반 크리깅 (Universal kriging)에 관하여 간략히 소개하였다. 5장에서는 실증 분석으로서 교통량 자료를 통하여 최단경로에 기반한 공간회귀모형을 구축하고 모형의 모수를 추정하고 공간 예측력을 비교하였다. 끝으로 6장에서는 결론 및 향후 연구 과제에 대하여 논하였다.

2. 교통량자료에 대한 공간통계적 분석

공간통계 분석기법은 공간적으로 변동하는 정보를 이용해서 이들의 분포특성 및 상관관계를 분석하고 신뢰할만한 정보를 예측하는 기법으로 광산학, 수리지질학, 기상학, 환경공학, 석유공학, 지리정보시스템, 수산학 등 여러 다양한 분야에서 적용되고 있다 (Journel과 Huijbregts, 1978; Isaaks와 Srivastava, 1989; Goovaerts, 1997; Kitanidis, 1997). 공간통계기법의 핵심은 공간자료가 무작위하게 분포하는 것이 아니라 특정한 공간적 상호관계와 연속성을 가진 공간변이도로 표현된다는 것이다. 공간적 상호관계는 베리오그램으로 정량화 되는데 이것은 일정한 거리만큼 떨어진 자료들의 유사성 정도를 나타내는 지표로 공분산의 계산과 크리깅 가중치를 계산하기 위하여 필요하다.

베리오그램 ($2\gamma(h)$)은 확률장에서 존재하는 공간적 관련에 대한 성질과 구조 등의 공간 정보를 전달하는 구조적인 도구라고 할 수 있으며 n 개의 공간 자료 $Z(s_i)$, $i = 1, 2, \dots, n$ 에 대하여 다음과 같이 정의된다.

$$2\gamma(h) = \text{Var}(Z(s+h) - Z(s)). \quad (2.1)$$

여기서 베리오그램의 반에 해당하는 값인 $\gamma(h)$ 는 세미베리오그램이라 한다. 공간상에 분포하는 표본자료의 특성은 식 (2.1)과 같이 베리오그램 (혹은 세미베리오그램)에 의해서 나타나는데, 베리오그램은 공간상에 분포하는 임의의 두 자료 값의 차이의 분산으로 구해지므로 일정한 거리에 있는 자료들의 유사성을 나타내는 척도로 이용된다. 여기서 h 는 지연거리 (lag distance)라 하여 두 지점간의 거리를 나타낸다. 따라서 베리오그램은 지연거리 h 만큼 떨어져 있는 두 지점 사이의 자료값 차이를 제공한 것에 대한 기대값이라 할 수 있다.

공간자료에 대한 세미베리오그램 분석은 관측된 자료를 통하여 공간적 변동을 알아보기 위하여 널리 사용되고 있으며 공간 예측을 위한 크리깅 방법에서 아주 중요한 역할을 하고 있다. 식 (2.1)에서 정의된 세미베리오그램에 대한 적률추정량은 아래의 식 (2.2)같이 표현된다 (Matheron, 1963).

$$\hat{\gamma}(h_k) = \frac{1}{2|N(h_k)|} \sum_{(s_i, s_j) \in N(h_k)} (Z(s_i) - Z(s_j))^2. \quad (2.2)$$

여기서 $N(h_k) = \{(s_i, s_j) : \|s_i - s_j\| \in I_k\}$, $k = 1, \dots, K$, $|N(h)|$ 는 $N(h)$ 집합안의 원소수를 나타낸다. 그리고 $0 < h_1 < \dots < h_K$ 를 만족하는 구간 I_k 는 $I_1 = (0, h_1), \dots, I_k = (h_{k-1}, h_k), \dots, I_K = (h_{K-1}, h_K)$ 이다.

이론적 세미베리오그램은 대부분 멍치 (Nugget), 부분문턱 (Partial sill)과 범위 (Range)와 같은 모수들을 가지고 있다. 베리오그램 모형의 모수 중 멍치모수는 두 공간위치간의 거리 h 가 0으로 가까이 갈때 0이 되기 직전의 베리오그램 값으로 정의하는데, 보통 측정오차를 의미한다. 범위모수는 공간적으로 더이상 상관이 존재하지 않는 최소거리를 나타내고, 자료가 범위 이상의 거리만큼 떨어져 있을 때 베리오그램은 일정한 값을 가지는데 이 값을 문턱 (Sill)이라고 한다 (문턱은 멍치와 부분문턱을 합한 것이다). 결과적으로 세미베리오

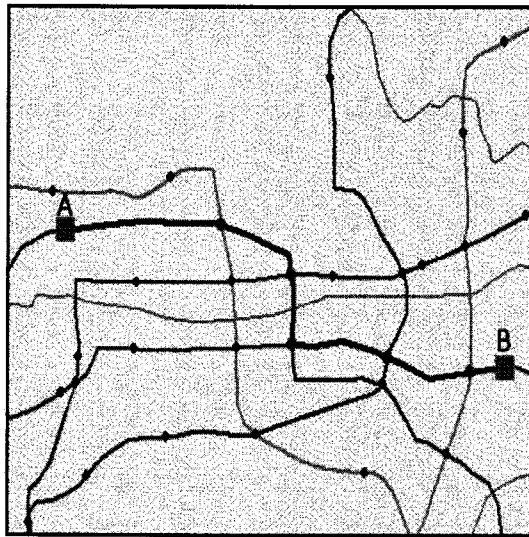


그림 2.1: 트랜스 캐들을 이용한 최단경로 계산의 예 (A 조사지점과 B 조사지점 사이의 수많은 조합의 경로 중에 최단경로는 굵은 선으로 표시되어 있다.)

그램의 추정은 세미베리오그램이 가지고 있는 모수의 추정이며, 적절한 이론적 세미베리오그램의 모수들을 가지고 크리깅 예측의 가중치를 결정하기 때문에 이론적 베리오그램의 모형적합은 공간 예측의 단계에서 매우 중요하다.

베리오그램의 모수를 추정을 위하여 본 연구에서는 최단경로를 기반으로 가중최소제곱법 (Weighted least squares estimation method)과 최대우도추정법 (Maximum likelihood estimation method)을 사용하였다. 각 추정방법 하에서 여러 세미베리오그램 모형들을 적합시킨 다음 최적의 모형을 선택하였다. 이 때 모수 추정값을 이용하여 공간 예측값들을 공간상관성을 고려하지 않은 다중선형회귀모형의 예측값들과 비교하였다. 그림 2.1은 교통분석 프로그램인 트랜스 카드를 이용하여 계산된 최단경로의 예를 보여주고 있다.

3. 추정방법

관측된 자료로부터 계산된 세미베리오그램은 일반적으로 경험적 세미베리오그램이라 하는데, 이는 관측된 자료들만을 대상으로 계산되기 때문에 공간보간 (Spatial interpolation)을 위해서는 관측되지 않은 지점에 대한 세미베리오그램의 계산이 필요하다. 따라서 경험적 세미베리오그램을 함수화한 이론적 세미베리오그램을 추정하게 된다.

공간분석에서 자주 사용되는 이론적 세미베리오그램 모형에는 가우시안모형 (Gaussian model), 지수모형 (Exponential model), 구형모형 (Spherical model), 파워모형 (Power model) 등이 있다 (Cressie, 1993). 본 연구에서는 가우시안모형과 지수모형을 사용하고자 한다. 지수 모형은

$$\gamma(h; \Theta) = \begin{cases} \frac{1}{2} \left[\theta_0 + \theta_1 \left\{ 1 - \exp\left(-\frac{h}{\theta_2}\right) \right\} \right], & \text{if } h > 0, \\ 0, & \text{if } h = 0, \end{cases}$$

가우시안 모형은

$$\gamma(h; \Theta) = \begin{cases} \frac{1}{2} \left[\theta_0 + \theta_1 \left\{ 1 - \exp\left(-\left(\frac{h}{\theta_2}\right)^2\right) \right\} \right], & \text{if } h > 0, \\ 0, & \text{if } h = 0 \end{cases}$$

이다. 여기서 $\Theta = (\theta_0, \theta_1, \theta_2)^T$ 이며 θ_0 는 문치모수, θ_1 은 부분문턱, 그리고 θ_2 는 범위모수이다.

최소제곱법에 의해 세미베리오그램의 모수를 추정하는 방법은 다음과 같다. 첫째, 자료로부터 세미베리오그램의 표본 추정치들을 도출하고, 둘째, 표본 추정치들을 이론적 세미베리오그램 모형에 적합시킨다. 다시 말해서, 주어진 래그 $h_k, 1 \leq k \leq d$ 는 두 지점간의 가능한 최대 거리 (d)에 대하여 표본 세미베리오그램의 값들 $\{\hat{\gamma}(h_1), \hat{\gamma}(h_2), \dots, \hat{\gamma}(h_k)\}$ 과 이론적 세미베리오그램 사이의 제곱합을 최소화시키는 모수 벡터 $\Theta = (\theta_0, \theta_1, \theta_2)^T$ 를 아래와 같은 식으로 구할 수 있다.

$$\min_{\Theta} \sum_{i=1}^k (\hat{\gamma}(h_i) - \gamma(h_i; \Theta))^2.$$

최소제곱법에서는 자료들이 서로 다른 래그에 중복적으로 사용되기 때문에, 세미베리오그램의 표본 추정치들이 서로 상관되어 있다. 따라서 이 단점을 극복한 추정방법으로 일반

화 최소제곱법의 하나인 가중최소제곱법이 있다. 가중최소제곱법에 의한 추정량은 다음과 같이 구해진다. $\hat{\gamma}(h)$ 의 공분산 행렬을 Σ 라 놓으면, 일반화 최소제곱 추정량은 아래의 식 (3.1)을 최소화시킴으로써 구할 수 있다.

$$(\hat{\gamma}(h) - \gamma(h; \Theta))^T \Sigma^{-1} (\hat{\gamma}(h) - \gamma(h; \Theta)). \quad (3.1)$$

여기서 $\hat{\gamma}$ 는 경험적 세미베리오그램이다. Cressie (1993)는 세미베리오그램의 표본 추정치의 분산이 근사적으로 식 (3.2)를 따르는 것을 보였다.

$$\text{Var}(\hat{\gamma}(h)) \simeq \frac{2\gamma^2(h)}{|N(h)|}. \quad (3.2)$$

가중최소제곱법은 Σ 를 $\hat{\gamma}(h)$ 의 분산 근사값으로 만들어진 대각행렬 V 로 대체하는 방법이다. 따라서 세미베리오그램의 모수에 대한 가중최소제곱추정량은 식 (3.3)과 같은 제곱합을 최소화하여 구할 수 있다.

$$(\hat{\gamma}(h) - \gamma(h; \Theta))^T V^{-1} (\hat{\gamma}(h) - \gamma(h; \Theta)). \quad (3.3)$$

여기서 V 의 대각원소는

$$\frac{2\gamma^2(h_1)}{|N(h_1)|}, \frac{2\gamma^2(h_2)}{|N(h_2)|}, \dots, \frac{2\gamma^2(h_k)}{|N(h_k)|}$$

와 같다. 가중최소제곱법을 포함한 일반적인 최소제곱법의 단점은 경험적 세미베리오그램의 값을 구하기 위하여 사용자 주관에 따라 래그의 개수를 선택하고, 이에 선택된 표본 세미베리오그램의 값을 반응변수로 사용한다는 것이다. 따라서 선택된 래그의 개수에 따라 적합한 추정량을 제공할 수도 있지만 그렇지 않은 경우도 있으므로 사용시 주의를 하여야 한다.

세미베리오그램의 추정에 있어서 최소제곱법과 더불어 최대우도추정법이 사용된다. 최대우도추정법은 모수의 추정을 위하여 표본 세미베리오그램 대신에 모든 자료를 사용하여 모수를 추정하는 방법이다. 공간변수 $\mathbf{z} = (Z(s_1), Z(s_2), \dots, Z(s_n))^T$ 가 n -변량 정규분포를 따르며 정상성조건(Stationary condition)하에서 평균 $E(\mathbf{z}) = \boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$ 과 분산 $\text{Var}(\mathbf{z}) = \boldsymbol{\Sigma}_\Theta$ 를 가진다고 가정하자. 그러면 Θ 의 최대우도추정량(Maximum likelihood estimator; MLE)은 식 (3.4)와 같은 대수우도함수(Log likelihood function)를 최소화시킴으로써 얻어질 수 있다.

$$\log L(\boldsymbol{\mu}, \Theta; \mathbf{z}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_\Theta| - \frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_\Theta^{-1} (\mathbf{z} - \boldsymbol{\mu}). \quad (3.4)$$

다음장에서는 가중최소제곱방법과 최대우도추정법으로 얻어진 모수 추정치를 이용하여 공간예측의 한 기법인 일반크리깅 방법으로 조사되지 않은 지점의 연평균 일교통량을 예측하였다.

4. 공간예측을 위한 크리깅

공간 분석에서 가장 중요하게 다루는 문제는 관측된 지점의 자료를 근거로 하여 관측되지 않은 지점의 값을 예측하는 것이다. 공간상의 상관관계를 고려하는 예측 방법은 공간 통

계학에서 크리깅으로 알려져 있다. 크리깅의 종류에는 단순 크리깅 (Simple kriging), 범용 크리깅 (Ordinary kriging), 일반 크리깅 (Universal kriging), 지시 크리깅 (Indicator kriging) 등이 있다 (Isaaks와 Srivastava, 1989). 본 연구에서 고려한 일반 크리깅은 주어진 공간 영역에서 평균이 일정하다고 보기 어려운 경우에 적합한 예측방법으로서 설명변수를 고려한 방법이다.

일반 크리깅에서는 평균이 갖는 공간적 추세를 제거한 잔차를 이용하여 세미베리오그램을 추정한 후 미관측 지점의 자료값을 예측하게 된다. 따라서 일반 크리깅에서 자료는 식 (4.1)과 같은 모형으로 표현할 수 있다.

$$Z(s) = \mu(s) + \epsilon(s). \tag{4.1}$$

여기서 $\epsilon(s)$ 는 평균이 0이고 정상성을 만족하는 확률변수 (오차항)로서 공분산 행렬 Σ_Θ 와 세미베리오그램 $\gamma(h; \Theta)$ 를 가진다. 평균함수 $\mu(s)$ 는 일반적으로 설명변수들의 함수와 같은 변동을 의미하며 아래의 식 (4.2)와 같은 p 개의 설명변수($X_0(s), X_1(s), \dots, X_p(s)$)의 선형 결합으로 나타낼 수 있다.

$$\mu(s) = \beta_0 X_0(s) + \beta_1 X_1(s) + \dots + \beta_p X_p(s). \tag{4.2}$$

여기서 $\beta_0, \beta_1, \dots, \beta_p$ 는 회귀계수이다. 식 (4.1)은 식 (4.2)와 같은 설명변수들과 공간의존적 구조를 가지고 있는 오차항이 있기 때문에, 이와 같은 모형을 공간회귀모형 (Spatial regression model)이라고 부른다. 크리깅에서 관심있는 새로운 지점 s_0 에서의 확률과정 $Z(s_0)$ 는

$$Z(s_0) = \mathbf{x}^T(s_0)\boldsymbol{\beta} + \epsilon(s_0) \tag{4.3}$$

이고, $\mathbf{x}(s_0) = (X_0(s_0), X_1(s_0), \dots, X_p(s_0))^T$ 는 미지의 지점 s_0 에서의 설명변수의 값들을 나타내고 $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ 이다. 일반크리깅에 의한 예측값 $\hat{Z}(s_0)$ 은 n 개의 지점에서 관측된 $Z(s_i)$ 의 선형 결합으로 아래와 같이 표현될 수 있다.

$$\hat{Z}(s_0) = \sum_{i=1}^n w_i Z(s_i).$$

여기서 가중치인 선형 계수인 $w_i, i = 1, \dots, n$ 는 균일적 불편 추정량의 조건 $w^T X = X^T(s_0)$ 와 s_0 지점에서 최소 크리깅 분산 $\sigma^2(s_0)$ 이라는 제약 조건인 $\sigma^2(s_0) = E(Z(s_0) - \hat{Z}(s_0))^2$ 에서 계산된다.

본 연구에서는 첫번째 추정방법으로 평균함수를 제거한 후 잔차들만을 이용하여 가중 최소제곱방법으로 모수의 추정값을 얻은 후 크리깅을 실시하였으며, 두번째 추정방법으로는 우도함수를 이용하여 평균함수의 회귀계수와 공분산함수의 모수들을 동시에 추정하는 최대우도추정법을 사용하여 모수를 추정한 후 크리깅을 실시하였다. 보다 자세한 이론적 설명은 Cressie (1993)를 참조하기 바란다.

5. 실증분석 및 결과

본 실증분석에서는 트랜스 캐드를 통하여 얻어진 교통량 조사지점 사이의 최단경로를 기반으로 세미베리오그램의 모수 추정을 위하여 공간분석에서 많이 활용되고 있는 가중최

소제곱방법과 최대우도방법을 사용하여 모수들을 추정하고 결과값들을 공간 정보를 포함하고 있지 않은 기존의 다중선형회귀모형 (Multiple linear regression model)의 결과와 비교하였다. 분석에 사용된 교통량자료는 미국 노스캐롤라이나주의 교통국에서 2001년에 시행된 상시조사지점들 중에서 Wake County내에서 조사된 총 1154개 지점에서 200개 지점을 선택하였다. 지점 선택을 위해서 지리정보시스템인 ArcGIS를 이용하여 일정한 간격의 격자 (Grid system)를 구축한 뒤, 계통임의추출법 (Systematic random sampling method)을 통해 200개의 지점을 지역적으로 고르게 추출하였다. 그림 5.1은 모형 구축을 위하여 선택된 200개의 교통량 조사지점들을 나타낸다. 총 1154개의 교통량 조사지점에서 관측된 설명변수는 교통량지점을 나타내는 위치자료, 도로의 특성과 관련된 5개 변수들, 그리고 조사지점과 인접한 지역의 2000년도 인구조사관련 변수 6개 등으로 구성되어 있다 (표 5.1). 도로의 특성을 나타내는 설명변수들은 조사지점이 위치한 곳의 토지이용형태, 도로의 차선수, 제한속도, 도로의 기능, 단위거리당 신호등 밀도이며, 이들은 Wake County 교통수요모형 자료에서 획득하였다. 인구관련자료는 GIS로 구축된 인구센서스자료를 ArcGIS를 이용하여 조사지점의 위치정보와 결합한 후, 조사지점과 연관된 지역의 인구정보를 통해 구축하였다. 본 연구에서 사용된 교통량 자료, 도로의 특성 자료 그리고 인구관련자료는 노스캐롤라이나 주립대학 교통공학연구소 (Institute for Transportation Research and Education, ITRE)에서 제공된 자료이다. 이상의 구축된 설명변수들은 표 5.1에 제시되었다.

모형적합에 사용하게 되는 200개의 자료에 있어서 종속변수의 정규성가정을 위하여 Box-Cox변환방법을 통한 교통량 자료의 변환후 공간분석을 실시하였다. 고려된 설명변수들은 단계별 변수선택 (Stepwise variable selection)방법을 통하여 위치정보 (위도, 경도)를

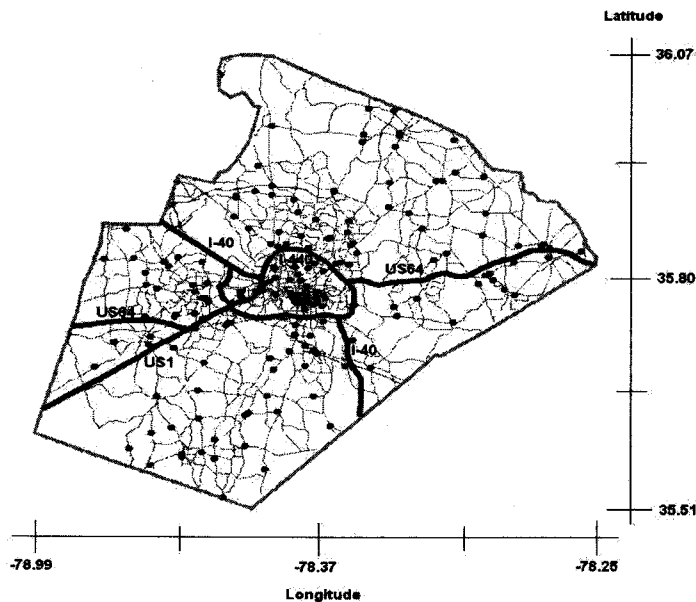


그림 5.1: 공간모형 구축을 위하여 선택된 200개의 교통량 조사지점

표 5.1: 연평균 일교통량 예측을 위한 조사지점에서의 설명변수들

변수	변수 설명
위치 설명변수	경도
	위도
도로 설명변수	토지이용행태 (도시; 준도시; 시골)
	차선의 수
	제한 속도
	도로의 기능
	(주요간선도로; 보조간선도로; 지방부고속도로; 국지도로; 접속도로) 1마일당 신호등의 밀도
센서스 설명변수	교통량 조사지역의 인구 (센서스 구역상)
	전체 가구 수
	총 가구 수입
	6세이하의 자녀를 둔 가구 수
	교통량 조사지역 내부에서 일하는 근로자의 수 (센서스 구역상) 교통량 조사지역 외부에서 일하는 근로자의 수 (센서스 구역상)

포함한 상태에서 통계적으로 유의한 변수들을 선택하였으며 최종 선택된 설명변수들은 표 5.2와 같다. 이들 변수들은 교통량 예측을 위하여 식 (5.1)과 같은 평균함수의 형태로 활용되었다.

$$Z_*(s) = \beta_0 + X_1(s)\beta_1 + X_2(s)\beta_2 + \dots + X_p(s)\beta_p + \epsilon(s). \tag{5.1}$$

여기서 $Z_*(s) = (Z(s))^{0.25}$ 는 변환된 교통량 자료이다. 연평균 일교통량 예측을 위하여 공간의존적 구조없이 설명변수들만을 이용한 다중선형회귀모형을 구축한 결과 결정계수 값(R^2)이 0.65로 나타났다

표 5.2에서 제시된 바와 같이 모형에 기본적으로 포함시킨 위치좌표인 위도와 경도 이외에 도로의 특성을 나타내는 변수인 도로의 제한속도, 차선의 수, 토지이용행태, 도로의 기능과 센서스 자료중 총 가구 수입이 최종모형의 설명변수로 선택되었다. 표 5.2로부터 알 수 있듯이 도로의 제한 속도가 높은 지역은 원활한 교통의 흐름이 전제되기 때문에 교통량이 많으며, 토지이용행태에 의하면 도시지역과 준도시지역이 시골지역에 비해 상대적으로 교통량이 많다. 그리고 도로의 기능의 관점에서는 지방부 고속도로는 도로의 설계속도가 높고 도시 간을 연결함으로써 다른 형태의 도로보다 교통량이 많음을 알 수 있다. 특히, 총 가구수입이 높은 지역일수록 그 지역에 위치한 조사지점에서 관측한 교통량이 상대적으로 많음을 알 수 있다. 이는 교통량과 여가활용을 위한 차량의 이용빈도가 밀접한 관계가 있다고 하겠다.

이러한 설명변수들의 영향을 제거하고 순수한 공간적 상관관계를 알아보기 위하여 평균함수 (5.1)를 제거한 잔차들만을 가지고 세미베리오그램 분석을 실시하였으며 교통량 조사지점의 최단경로를 기반으로 여러 가지 공간 모형을 고려하였다. 모형 선택 기준으로 가중최소제곱법은 잔차제곱합을, 최대우도추정법은 로그우도함수를 사용하였다.

표 5.2: 단계적 변수선택방법에 의해 최종 선택된 설명변수들과 회귀계수의 추정값

선택된 변수	가중최소제곱법			최대우도추정법		
	계수	표준오차	t-값	계수	표준오차	t-값
절편	88.405	142.4	0.621	26.750	213.5	0.125
경도	0.5080	1.253	0.405	-0.153	1.921	-0.080
위도	-1.361	1.631	-0.835	-1.041	2.441	-0.426
제한속도	0.0899	0.173	5.193	0.0584	0.025	2.336
총 가구수입	1.9E - 5	7.2E - 6	2.639	1.6E - 5	1.0E - 5	1.153
차선의 수	0.6931	0.134	5.186	0.7026	0.177	3.970
도시	1.9272	0.495	3.894	1.9001	0.665	2.857
준도시	1.4762	0.385	3.838	1.3553	0.523	2.591
주요간선도로	3.5253	0.451	7.817	3.4124	0.663	5.147
보조간선도로	2.0363	0.460	4.428	1.8070	0.660	2.738
지방부고속도로	7.3683	1.146	6.432	7.1394	1.534	4.654
국지도로	1.6541	0.403	4.103	1.6405	0.521	3.149

표 5.3: 최종으로 적합된 세미베리오그램 모수들의 추정값

방법	모형	모수			기준	
		뭉치	문턱	범위 (mile)	RSS	log L
가중최소제곱법 (WLS)	지수	0.5918	2.3556	1.0728	0.5552	.
	가우시안	1.3577	2.3493	2.0865	0.5541	.
최대우도추정법 (ML)	지수	2.2269	4.4792	1.2758	.	-212.5
	가우시안	2.7010	4.4189	1.2907	.	-212.7

공간회귀모형을 구축을 위하여 식 (5.1)의 오차항에 대한 분산-공분산행렬을 구축할 때 최단경로를 이용한 실제거리를 고려하였으며 이 행렬은 식 (2.1)의 변환을 통해서 쉽게 생성되어진다. 회귀모수 뿐만 아니라 오차항의 공간모형에 관계된 모수의 추정을 위해서 R 2.2.1의 nlminb함수와 geoR패키지를 이용하였으며 이론적 세미베리오그램의 추정에 있어서 식 (2.2)에 있는 유틸리티안거리가 아닌 도로상의 실제 거리를 이용하였다.

표 5.3에는 각 방법으로 적합된 모형들의 모수 추정값들과 잔차제곱합, 우도함수값을 제시하고 있다. 가중최소제곱법으로는 잔차제곱합을 최소화하는 가우시안 모형이, 최대우도추정법으로는 로그우도함수를 최대화하는 지수모형이 최종 선택되었다. 표 5.3에서 최대우도추정법 하에서 얻은 뭉치와 문턱 (= 뭉치 + 부분문턱)의 추정값이 모형에 관계없이 가중최소제곱법 하에서의 추정값보다 훨씬 크다는 것을 알 수 있다. 또한 추정된 범위는 대략 1-2마일로 상당히 작게 추정되었지만 상관관계가 거의 미미하게 되는 유효범위 (Effective range)는 3-6마일이다.

5.1. 크리깅과 예측 성능의 비교

최단경로 기반으로 가중최소제곱법과 최대우도방법으로 선택된 두 공간모형을 가지고 크리깅을 실시하였으며 각 방법별로 예측력을 비교하였다. 일반 크리깅을 사용하여 교통량이 관측되지 않은 도로 공간상에 예측값을 구하였다. 예측 성능을 비교하는 도구로서 한 자료씩을 제거해 나가는 교차 검증방법 (Leave-one-out cross validation)을 고려하였다. 또한 추정된 세미베리오그램 모형의 예측력 비교 기준으로서 다음과 같이 정의되는 예측잔차제곱합 (Predicted residual sum of squares, PRESS) 척도를 사용하였다.

$$PRESS = \frac{1}{n_s} \sum_{i=1}^{n_s} \left(Z_*(s_i) - \hat{Z}_*(s_{-i}) \right)^2. \quad (5.2)$$

여기서 $n_s = 200$ 는 교통량 관측지점의 수를 나타내며 $Z_*(s_i)$ 는 i 번째 위치한 관측값이며 $\hat{Z}_*(s_{-i})$ 는 i 번째를 제외한 나머지 관측값들을 이용하여 i 번째 위치의 교통량을 예측한 값이다. 또한 모형적합에 사용되지 않은 교통량 조사지점의 예측값을 실제값과 비교, 평가하기 위하여 평균제곱예측오차 (Mean squared prediction error, MSPE) 척도를 사용하였다.

$$MSPE = \frac{1}{n_{ns}} \sum_{j=1}^{n_{ns}} \left(Z_*(s_j) - \hat{Z}_*(s_j) \right)^2. \quad (5.3)$$

여기서 $n_{ns} = 954$ 는 모형적합에 사용되지 않은, 그러나 실제 교통량은 관측한 지점의 수이고, $\hat{Z}_*(s_j)$ 는 n_s 개의 모형적합에 사용한 관측자료를 이용하여 구한 j 번째 예측지점의 예측값이다. 예측값 성능의 지표로 사용되는 포함확률 (Coverage probability)을 계산하기 위하여 우선 교차검증방법을 이용하여 i 번째 위치의 관측값을 제외한 모형으로 i 번째 관측값의 95% 신뢰구간을 구한다음 i 번째 관측값이 그 신뢰구간안에 포함되는지의 여부를 살펴본다. 이 과정을 모든 관측값들에 적용하여 전체 관측값들 중 몇 %가 신뢰구간에 포함되는지를 보여주는 지표가 포함확률이다.

각 모수 추정방법에 따른 예측잔차제곱합, 평균제곱예측오차 그리고 포함확률이 표 5.4에 있다. 예측잔차제곱합과 평균제곱예측오차의 관점에서는 항상 최대우도추정법으로 얻은 지수모형이 다른 모형에 비해 우수하였고 특히 오차항이 등분산이고 독립이라는 가정하에서의 다중선형회귀모형이 가중최소제곱법에서 얻은 가우시안모형보다 훨씬 더 작은 예측잔차제곱합을 가진다. 하지만 모형적합에 사용되지 않은, 그러나 실제로 관측된 지점의 관측값들을 이용한 평균제곱예측오차의 관점에서는 가우시안모형이 회귀모형보다 우수하다. 포함확률의 경우, 최대우도추정법에 의한 모형과 회귀모형이 유사함을 알 수 있다.

표 5.4: 선택된 모형들의 교차검증 결과

방법	관측 지점		예측 지점	
	예측잔차제곱합	포함확률	평균제곱예측오차	포함확률
OLS	2.8528	93.0%	3.1818	92.8%
WLS	10.131	89.0%	2.9926	97.7%
ML	2.4396	98.0%	2.8657	97.5%

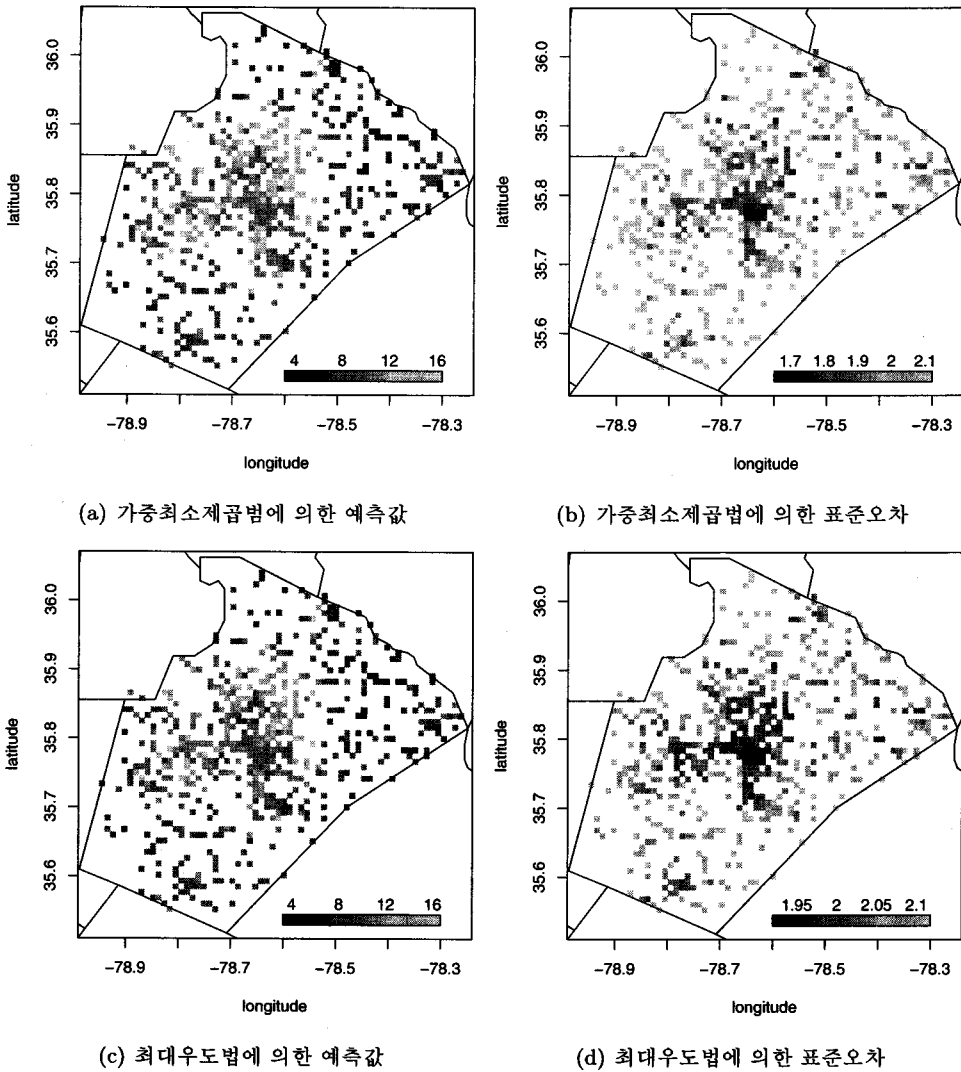


그림 5.2: 크리깅방법으로 도출된 예측값과 표준편차

그림 5.2에서는 일반크리깅방법으로 구한 예측값 ($n_{ns} = 954$)과 표준편차를 표시하였다. 추정방법에 관계없이 예측값은 큰 차이가 없으며 대도시인 Raleigh (지도의 가운데 부분)에서 가중최소제곱법에 의한 크리깅 표준편차가 최대우도추정법에 비해 다소 작음을 알 수 있다. 물론 그림 5.2에서 제시된 값들은 Box-Cox 방법에 의해 변환된 값의 예측값들이므로 원자료로 재변환하여 비교하는 작업이 필요하겠다.

제 5장에서는 가중최소제곱법과 최대우도추정법으로 얻은 세미베리오그램의 모수들을 통하여 미지의 위치에 대한 모형 예측력 연구에 이용하였다. 최단경로를 기반으로 한 최대우도추정법의 공간회귀모형이 가중최소제곱법의 공간회귀모형보다 교통량 예측 측면에서

상대적으로 좋은 결과가 나왔다. 이는 공간 분석에서 최대우도방법의 장점인 원자료를 래깅이라는 부가적인 과정없이 모수 추정에 사용한다는 점을 우도함수 추정값들이 잘 반영하고 교통량 자료의 특성상 조사지점 간의 최단경로가 교통량 예측을 위하여 타당한 거리정보임을 보여주고 있다. 그러나 공간정보를 반영하지 않은 일반 다중회귀분석 방법이 가중최소제곱법보다 관측지점의 예측잔차제곱합 관점에서 성능이 좋은 것으로 나타났다. 그림 5.2는 가중최소제곱법과 최대우도추정법으로부터 얻은 두 가지 모형을 일반크리깅을 통하여 예측지점에서의 예측된 일교통량과 표준오차를 보여주고 있다.

6. 결론

본 연구에서는 공간 분석의 응용에서 중요하게 사용되는 세미베리오그램의 모수 추정 방법을 간략히 제시하였으며 교통량 자료를 통하여 공간회귀분석을 활용한 교통량 예측 모형을 구축하였다. 도로상의 교통량 조사지점의 최단경로를 계산하여 공간분석의 거리로 사용하였으며, 최단경로를 활용하여 추정된 모수를 통한 일반크리깅을 통해 공간 예측력을 비교하였다. 비교 결과로는 최대우도방법의 공간회귀모형이 가중최소제곱법의 공간회귀모형보다 뛰어난 예측값을 제공하는 것으로 나타났다. 또한 교차검증을 이용하여 최단경로를 기반으로 한 공간분석이 교통량 자료에 적합함을 공간 상관성을 고려하지 않은 회귀분석과의 비교를 통하여 제시하였다.

본 연구에서 새롭게 제시된 최단경로기반 공간회귀모형은 도로위에 설치되어 있는 조사장비의 불량 및 오류로 인한 오류교통량에 대한 평가나 조사장비가 설치되어 있지 않은 도로상의 교통량에 대한 올바른 예측에 아주 중요한 모형이 될 것이다. 향후 연구 주제로서 교통량 예측을 위하여 공간뿐만 아니라 시간을 고려한 시공간 모형을 통하여 좀더 정확한 시공간 예측을 할 수 있는 모형을 만들 수 있도록 하겠다.

참고문헌

- 김선우, 정애란, 이성덕 (2005). 공간자료를 이용한 세미베리오그램 추정, <응용통계연구>, 18, 271-280.
- 유성모, 엄익현 (1999). 강우강도데이터를 이용한 세미베리오그램의 추정과 공간이상치에 관한 연구, <응용통계연구>, 12, 125-141.
- 오주삼, 조윤호, 변상철 (2002). 연 평균 일교통량 추정 방법론에 관한 연구, <대한토목학회논문집>, 22, 59-69.
- 조재영, 최승배, 김규곤 (2001). 일반통계에 대한 공간통계 방법의 예측성능에 관한 연구, *Journal of the Korean Data Analysis Society*, 3, 41-49.
- Cheng, C. (1992). Optimum Sampling for Traffic Volume Estimation, Ph.D. Dissertation, University of Minnesota, Carlson School of Management, Minnesota.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*, John Wiley & Sons, New York.
- Eom, J. K., Park, M. S., Heo, T. Y. and Huntsunger, L. F. (2006). Improved prediction of annual average daily traffic for non-freeway facilities by applying spatial statistical method, *Proceedings of the 85th Annual Meeting of the Transportation Research Board*, Washington DC, and under consideration for publication in Transport Policy.

- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York.
- Heo, T-Y. and Hughes-Oliver, J. (2005). A Spatial Editing and Validation Process for Short-Count Traffic Data, Technical report, North Carolina State University.
- Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, New York.
- Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*, Academic Press, New York.
- Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications in Hydrogeology*, Cambridge University Press, New York.
- Lam, W. H. and Xu, J. (2000). Estimation of AADT from short period counts in Hong Kong: a comparison between neural network method and regression analysis, *Journal of Advanced Transportation*, **34**, 249-268.
- Matheron, G. (1963). Principles of geostatistics, *Economic Geology*, **58**, 1246-1266.
- Rathbun, S. L. (1998). Spatial modelling in irregularly shaped regions: Kriging estuaries, *Environmetrics*, **9**, 109-129.
- Tang, Y. F., Lam, W. H. K. and Pan, L. P. (2003). Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong, *Journal of Transportation Engineering*, **129**, 271-277.
- Xia, Q., Zhao, F., Chen, Z., Shen, L. D. and Ospina, D. (1999). Development of a Regression Model for Estimating AADT in a Florida County, In *Transportation Research Record 1660*, Transportation Research Board, National Research Council, Washington, D.C.
- Zhao, F. and Chung, S. (2001). Estimation of Annual Daily Traffic in a Florida County Using GIS and Regression, *Transportation Research Board, 80th annual meeting*, Washington, D.C.

[2006년 11월 접수, 2007년 5월 채택]

A Study on the Prediction of Traffic Counts Based on Shortest Travel Path

Tae-Young Heo¹⁾ Man Sik Park²⁾ Jinki Eom³⁾ Ju sam Oh⁴⁾

ABSTRACT

In this paper, we suggest a spatial regression model to predict AADT. Although Euclidian distances between one monitoring site and its neighboring sites were usually used in the many analysis, we consider the shortest travel path between monitoring sites to predict AADT for unmonitoring site using spatial regression model. We used universal Kriging method for prediction and found that the overall predictive capability of the spatial regression model based on shortest travel path is better than that of the model based on multiple regression by cross validation.

Keywords: Cross-validation, semivariogram, annual average daily traffic, Kriging.

-
- 1) Full-time Lecturer, Division of Nano Data System, Korea Maritime University, Busan, Korea
E-mail: heoty@hhu.ac.kr
 - 2) Research professor, Department Preventive Medicine, Korea University, Korea
E-mail: bayesia@korea.ac.kr
 - 3) Senior researcher, Korea Railroad Research Institute, Korea
E-mail: jkom00@krri.re.kr
 - 4) Senior researcher, Korea Institute of Construction Technology, Korea
E-mail: jusam@kict.re.kr