교통사고 위험에 대한 공간의 불확실성 파악: 일반화 가법 혼합 모형 기반의 계층적 베이지안 접근법을 활용하여

임영빈*, 김기정**
*국민대학교 행정학과
**두원공과대학교 전기자동차과
e-mail:yblym0207@kookmin.ac.kr*, kimkj@doowon.ac.kr**

Assessing Spatial Uncertainty in Traffic Accident Risk: A Hierarchical Bayesian Approach Based on Generalized Additive Mixed Models

Youngbin Lym*, Ki-Jung Kim**
*Dept. of Public Administration, Kookmin University
**Dept. of Electric Vehicle, Doowon Technical University

요 약

본 연구는 인명피해 교통사고의 공간적 분포와 그 불확실성을 체계적으로 분석하기 위해, 일반화 가법 혼합 모형에 기반한 계층적 베이지안 분석틀을 적용하였다. 연구 대상은 미국 오하이오주 클리블랜드-엘리리아 광역권에서 2015년부터 2019년까지 발생한 인명피해 사고이며, 사고 데이터를 센서스 블록그룹 단위(소지역)로 집계하여 사고위험에 대한 공간의 영향을 파악하고자 하였다. 설명변수로는 사회인구학적 특성, 물리적 도로환경, 교통 및 토지이용 지표 등을 포함하였으며, 이 중 인구밀도와 빈집비율은 2차 랜덤 워크(Random Walk of Order 2)를 적용하여 비선형적 영향을 추정하였다. 또한, BYM2 모형을 통해 설명되지 않는 잔차의 공간적 상관성과 구조화된 불확실성을 함께 고려하였다.

분석 결과에 따르면 저임금 노동자 비율, 차량 미소유 가구 비율, 활동 밀도, 토지이용 다양성, 통행 균형, 자동차 중심 교차로 밀도 등은 인명피해 사고 위험을 유의하게 증가시키는 요인으로 나타났다. 인구밀도와 빈집비율의 영향은 선형으로 작용하는 것이 아닌 일부 구간에서는 사고 위험을 증가시키거나 약화시키는 등 비선형적 특성을 지닌 것으로 파악되었다. 그리고 도심 지역의 사고위험이 상대적으로 높게 나타났는데, 공간적 구조화 효과가 전체 잔차 변동의 약 78.5%를 설명함으로써 교통사고 위험의 공간적 불균형이 통계적으로 유의미함을 확인하였다. 본 연구는 공간 기반의 사고 위험 예측과 지역 맞춤형 교통안전 정책 수립에 실증적 근거를 제공할 수 있는 분석틀을 제시한다는 점에서 의의가 있다.

키워드: 일반화 가법 혼합 모형, 계층적 베이지안 분석, 소지역, 인명피해사고, 사고위험, 공간적 불평등

1. 서론

교통사고는 전 세계적으로 중대한 공중보건 문제로 간주되며, 그 발생 원인에 대한 학문적 탐구는 주로 운전자의 과실, 차량의 기계적 결함, 기상 조건 등 물리적·환경적 요인에 초점을 맞추어 이루어져 왔다. 그러나 최근에는 지리정보시스템(GIS) 및 지리공간과학의 발전에 힘입어, 교통사고 발생에 대한 공간적 요인의 중요성이 부각되고 있다. 특정 지역에서 사고가 반복적으로 발생하거나 사고의 중증도가 공간적으로 편중되어 나타나는 현상은, 사고의 분포가 무작위적이지 않으며 공간적 불균등성을 내포하고 있음을 시사한다. 이는 도로 인프라의 공간적 불균형, 도시계획의 구조적 제약, 교통정책의 지역적 편차 등 다양한 공간 기반 요인이 사고 발생에 체계적으로 작용하고 있음을 반영한다.

이러한 공간적 이질성과 상관성을 고려한 교통사고 분석을 위해, 본 연구는 계층적 베이지안 모형(Bayesian hierarchical model)을 적용하고자 한다. 본 모형은 사고 발생의 공간적 패턴을 통계적으로 정교하게 추정할 수 있는 유연성을 지니며, 고위험 지역의 식별 및 사고 예측에 있어 실증적 근거를 제공한다. 특히 본 연구는 임영빈(2024)의 선행연구를 확장하여, 미국 오하이오주 북동부에 위치한 클리블랜드-엘리리아 핵심기반 통계영역(Cleveland-Elyria Core-Based Statistical Area)을 대상으로 2015년부터 2019년까지 발생한 인명피해사고 데이터를 분석한다. 분석은 물리적·환경적·사회인구학적 변수들의 고정효과와 함께, 공간적 구조에 기인한 관측 불가능한 확률효과를 통합적으로 고려하여 수행되며, 이를 통해 데이터 기반의 교통안전 정책 수립에 기여할 수 있는 실증적 토대를 마련하고자 한다.

2. 연구방법

2.1 분석데이터

본 연구는 임영빈(2024)의 데이터 수집 및 변수 선정 절차를 준용하여 수행되었다. 교통사고 데이터는 오하이오주 교통국(Ohio Department of Transportation, ODOT)이 제공한 2015년부터 2019년까지의 사고기록 중, 광역 클리블랜드(Cleveland-Elyria Core-Based Statistical Area)에서 발생한 162,579건을 활용하였다. 이들 사고는 지오코딩을 통해 1,691개의 센서스 블록그룹(Census Block Groups, CBGs) 단위로 집계되었으며, 각 블록그룹의 5년간 누적 사고 건수를 산출하였다. 사고는 단순 물적 피해사고(116,566건)와 인명피해 사고(46,013건)로 구분되며, 본 연구에서는 후자에 해당하는 인명피해 사고만을 분석에 포함하였다.

분석에는 사회인구학적, 환경적, 물리적 요인을 포함한 다양한 설명변수를 통합하였다. 관련 데이터는 미국 환경보호국의 Smart Location Database (v3.0), 미국 인구조사국의 American Community Survey (2014-2018 5-Year Estimates), 및 미국 교통국의 Local Area Transportation Characteristics for Households (LATCH, 2021)에서 수집하였다. 주요 변수로는 인구밀도(PopDen), 활동밀도(ActDen), 저임금 노동자 비율(LowWage), 차량 미소유 가구 비율(NoCar), 빈집 비율(Vacancy) 등의 사회인구학적 지표와, 토지이용 다양성(Landuse), 고용·주거 혼합도(WorkEmpMix), 통행 균형지수(TripEq), 도로 밀도(RoadNetwork), 보행자 중심 및 자동차 중심 교차로 밀도(Street, Intersection), 차량주행거리(VMT)가 포함되었다. 아울러, 지역의 도시화 수준을 반영하기 위해 농촌(Rural), 교외(Suburban), 도시(Urban) 구분의 범주형 변수도 도입하였다.

2.2 연구방법

본 연구에서는 계층적 베이지안 모형을 기반으로 사고의 영향변수인 빈집비율과 인구밀도의 비선형적 영향을 파악하기 위해 일반화 가법 모형(Generalized Additive Model, GAM)을 고려하였다. 아울러 설명변수들에 의한 영향이 반영된 후 잔차(residual)에 존재하는 공간의 불확실성을 설명하기 위해 Besag-York-Mollie (BYM) 모형(Besag et al., 1991)의 개선된 형태인 BYM2 모형(Simpson et al., 2017; Lym et al., 2022)을 적용하여 일반화 가법 혼합 모형을 구축하고 이를 분석하였다.

3. 분석결과

분석에 앞서 R의 scale 함수를 활용하여 설명변수들의 표준화가 이루어졌다. 차량주행거리 VMT는 교통량의 지역적 차이를 반영하기 위해 오프셋(offset) 또는 교통사고 위험에 대한 노출도(exposure)로 가정하여 사고의 상대위험도를 추정하는데 활용하였다. 분석결과는 [표 1]과 같이 제시하였는데, 이를 구체적으로 살펴보면 크게 설명변수들에 의한 고정효과와 인구밀도 및 빈집비율의 비선형적 영향 및 공간의 불확실성을 나타내는 확률효과(random effects)로 구성된다.

고정효과를 살펴보면, 인명피해사고의 위험을 높이는 인자로 저임금 노동자의 비율, 차량을 소유하지 않은 가구의 비율, 활동밀도 및 토지이용비율과 통행발생 및 통행유발 균형지수, 고용 및 주거의 다양성 지수, 자동차 중심 교차로 밀도가 통계적으로 유의미한 영향을 미치는 것으로 나타났다. 또한, 농촌지역에 비해 교외 및 도시화 지역의 인명피해 위험이 높은 것으로 추정되었다.

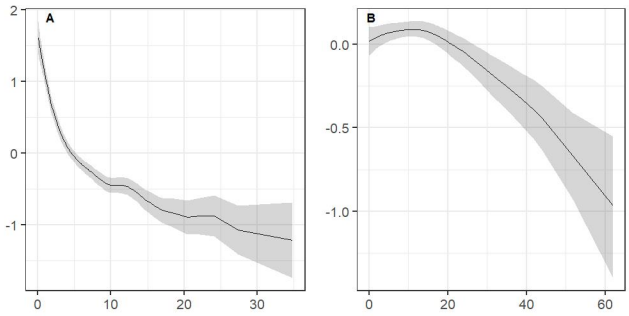
한편, 확률효과에 대해 살펴보면, 인구밀도와 빈집비율의 경우 교통사고에 미치는 영향이 비선형적으로 나타나고 있는데, 데이터를 부드럽게 연결하는 piecewise 다항 함수인 스플라인(spline)의 이산적(difference-based 또는 discretized) 베이지안 해석인 2차 랜덤워크(random walk of order 2)를 적용하여 분석하였다.

[표 1] 일반화 가법 혼합모형의 분석 결과

Variables	Mean	S.D	0.025Q	0.975Q
Fixed effects				
Intercept	−0.778	0.084	−0.943	−0.612
LowWage	0.179	0.032	0.117	0.241
NoCar	0.111	0.025	0.062	0.161
ActDen	0.115	0.02	0.077	0.153
Landuse	0.262	0.021	0.221	0.303
TripEq	0.078	0.021	0.037	0.119
WorkEmpMix	0.152	0.021	0.11	0.194
RoadNetwork	0.092	0.051	−0.007	0.191
Street	−0.017	0.037	−0.089	0.055
Intersection	0.16	0.02	0.12	0.2
Urban	0.457	0.103	0.254	0.659
Suburban	0.36	0.089	0.185	0.535
Random effects				
σ_{Space}	0.828	0.029	0.77	0.882
ϕ	0.616	0.059	0.488	0.717
σ_{PopDen}	0.722	0.201	0.399	1.182
$\sigma_{Vacancy}$	0.119	0.059	0.038	0.264

Goodness of fit	Value
DIC	11070.10
WAIC	10878.06
Log-likelihood	−6846.53

[그림 1]에 제시된 바와 같이, 인구밀도 변수([그림 1]의 A)의 경우 전반적으로 인명피해 위험을 낮추는 것으로 나타났는데, 일부 구간에서 그 위험이 높아지는 경향을 보였고, 빈집비율([그림 1]의 B)은 빈집비율이 낮은 구간에서 사고위험이 높아졌다가 일정 지점을 지나 그 비율이 높아질수록 인명피해 사고의 상대적 위험이 낮아지는 것으로 나타났다.

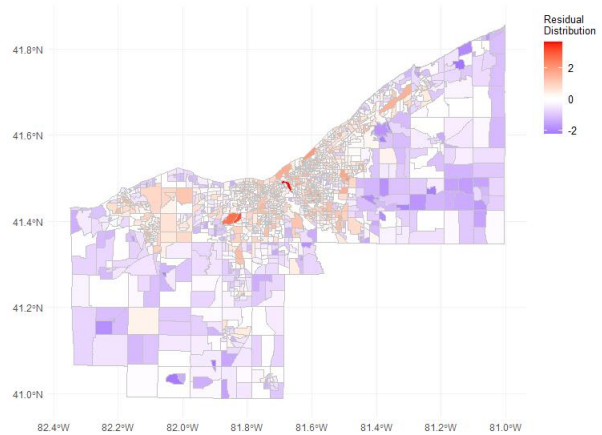


A. 인구밀도 B. 빈집비율

[그림 1] 교통사고 인명피해 위험에 대한 인구밀도(A) 및 빈집비율(B) 변수의 비선형적 영향

즉, 본 연구에서는 인명피해 사고 위험에 있어서의 인구밀도와 빈집비율의 영향이 선형으로 나타나지 않으며, 일부 구간에서는 음(−) 또는 양(+)의 영향을 보이는 것으로 파악되어, 그 변화가 부드럽게 일어나는 비선형적 관계를 가지고 있음을 규명하였다. 아울러 설명변수들의 영향을 반영한 후 잔차의 변동성을 설명하기 위해 적용한 BYM2 모형의 결과에 따르면, 공간의 구조화된 확률효과의 상대적 영향이 $\sqrt{0.616} = 0.785$로 나타나, 잔차의 78.5%가 공간의 구조화된 확률효과에 의해 설명되어지고 있음을 알 수 있었다.

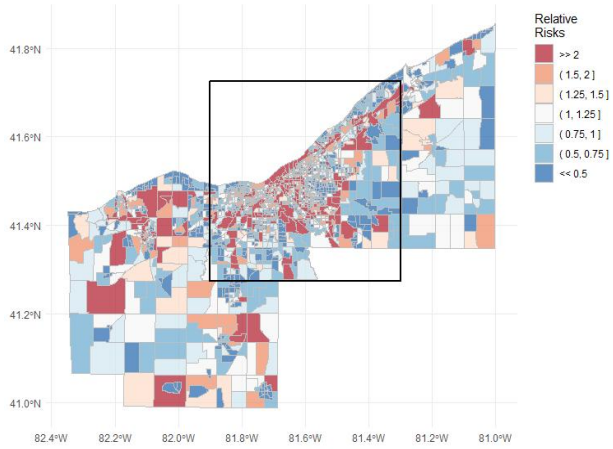
[그림 2]는 이러한 잔차의 공간적 분포를 시각화한 것으로, 설명변수들의 효과를 반영한 이후 인명피해 사고의 상대위험도가 무작위적이 아닌 지역에 따라 크게 차이가 있음을 나타낸다. 즉, 특정 센서스블록그룹의 사고위험이 타 블록그룹에 비해 사고의 상대위험도가 높게 나타나고, 해당 블록그룹에 인접한 곳들 역시 유사한 속성을 지니는 공간적 종속성 및 공간 이질성이 이러한 현상을 설명하며, 결론적으로 해당 지역에서의 인명피해 사고의 상대위험도는 공간적 불평등을 내포하고 있음을 시사한다.



[그림 2] 잔차의 공간적 분포

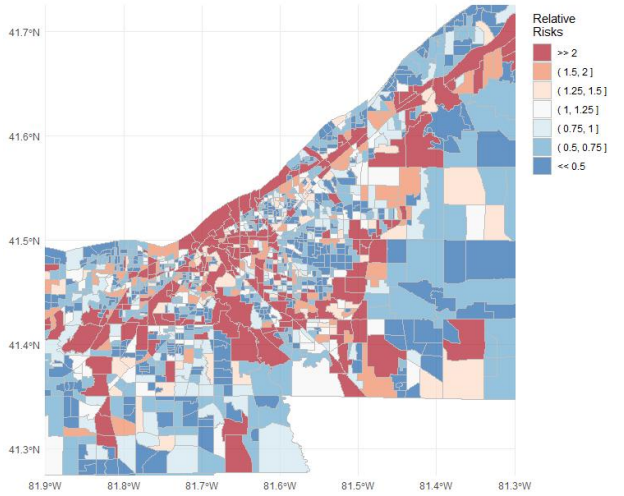
4. 논의 및 결론

이러한 연구결과를 종합하여, [그림 3]과 같이 광역클리블랜드 지역에서 2015−2019년 사이에 발생한 인명피해 교통사고의 (누적)상대위험도의 공간적 분포를 시각화하였다. [그림 3]은 단순히 인명피해 교통사고의 누적발생건수를 지도에 제시한 것이 아닌, 본 연구에서 활용한 일반화 가법 혼합 모형을 통해 도출한 사고의 상대 위험도를 나타낸 것이다. 설명변수들에 의한 고정효과 및 인구밀도와 빈집비율에 의한 비선형의 영향과 공간종속성 및 이질성에 의한 확률효과를 모두 반영하여 상대위험도의 사후평균(posterior mean)을 산출한 후 그 공간적 분포를 제시하였다.



[그림 3] 인명피해 사고 상대위험도 분포

[그림 3]을 살펴보면, 광역클리블랜드 지역 내 1691개 센서스 블록그룹의 인명피해사고 위험은 지역에 따라 다른 양상을 보인다. 범례와 같이 상대위험도(Relative risk)가 1 이상인 블록그룹들은 전체 평균위험 대비 사고의 위험이 높은 곳을 의미하며, 상대 위험도가 2 이상인 곳을 붉은색으로 나타내었다. 동일한 관점에서 사고의 위험이 전체 평균에 비해 상대적으로 낮은 곳들은 파란색으로 표현하여 교통안전 정책 마련을 위한 객관적 근거를 제공하고자 하였다. 센서스 블록그룹이 많은 도심지역([그림 3]의 격자부분)에 대해서는 [그림 4]와 같이 확대하여 도심 지역 내 인명피해사고의 상대위험도의 공간적 분포를 구체적으로 파악해보았다.



[그림 4] Zoom-in map

[그림 4]를 구체적으로 살펴보면, 도심지역의 사고 위험도는 타 외곽지역에 비해 상대적으로 높은 것으로 나타났다. 이는 도심지역이 토지이용의 집약도, 활동 밀도, 통행발생 및 통행유발 균형지수, 고용 및 주거의 다양성 지수, 자동차 중심 교차로 밀도가 타 지역에 비해 상대적으로 높으며, 본 연구에서 제안한 일반화 가법 혼합 모형을 통해 이러한 변수들의 영향이 사고의 위험을 높이는 것으로 추정된 점이 반영되어 있음을 시사한다. 아울러 도심지역 내 사고의 고위험 센서스 블록그룹의 경우 인접한 센서스 블록그룹과 유사하게 사고의 위험이 높은 것으로 나타나고 있어 교통사고 위험을 이해하는 데 있어 공간 종속성 및 이질성을 고려하는 것이 타당하다는 본 연구의 연구 질문에 대한 객관적 검증이 이루어졌다고 판단된다.

참고문헌

[1] 임영빈, “데이터마이닝과 베이지안 기법을 활용한 소규모 지역의 부주의 운전에 의한 교통사고 위험 모형 구축 -미국 오하이오주 북부 지역을 대상으로-”, Crisisonomy, 제 20권 3호, pp. 106-116, 3월, 2024년.

[2] Y. Lym, S. Kim, K.-J. Kim, “Identifying regions of excess injury risks associated with distracted driving: A case study in Central Ohio, USA,” *SSM - Population Health*, vol. 20, p. 101293, Dec. 2022.

[3] J. Besag, Y. Jereny, M. Annie, “Bayesian Image Restoration, with Two Applications in Spatial Statistics.” *Annals of the Institute of Statistical Mathematics*, vol. 42, no. 1, pp. 1-20, 1991.

[4] D. Simpson, R. Havard, R. Andrea, M. Thiago, S. Sigrunn, “Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors.” *Statistical Science*, vol. 32, no. 1, pp. 1-28, 2017.