

시공간 베이저안 모형 기반의 대전광역시 근린지역 교통사고 위험 분석

임영빈
국민대학교 행정학과

Examining Neighborhood-Level Vehicle Crash Risks in Daejeon Metropolitan City using Spatiotemporal Bayesian Models

Youngbin Lym
Department of Public Administration, Kookmin University

요약 이 연구는 대전광역시에서 2015~2020년 기간 발생한 교통사고를 대상으로 한 것으로, 개별 행정동을 분석단위로 삼아 교통사고에 영향을 미치는 요인들을 탐색하고자 하였다. 이를 위해 시공간의 영향인 확률효과와 설명변수에 의한 고정효과를 모두 고려할 수 있는 일반화선형혼합모형 기반의 시공간 베이저안 모형을 활용하였다. 연구의 결과에 따르면 행정동 단위 면적 기준 교차로 밀도 및 횡단보도 밀도가 커지면, 그 사후평균이 각각 0.328, 0.711로 제시되어, 교통사고의 상대위험도가 증가하는 것으로 나타났다. 한편, 인구밀도가 높아지게 되면 그 사후평균이 -0.351로 추정되어, 상대위험도에 음(-)의 영향을 미치는 것으로 나타났다. 확률효과 중 시간종속성의 모수인 ρ_T 의 사후평균이 0.971로 추정되어, 공간종속성의 모수인 λ_S 의 사후평균 0.476에 비해 상대적으로 높게 나타나, 시계열적 자기상관이 확률효과 중 큰 부분을 차지하고 있음을 알 수 있었다. 연구의 결과를 통해, 대전광역시 내 소규모 지역 단위에서의 교통 안전 정책 수립을 위한 객관적 근거자료로 활용할 수 있으며, 또한, 본 연구에서 제시한 연구기법의 타 지자체 적용을 통해 보다 일반화된 결과를 얻을 수 있을 것이다.

Abstract This study examined vehicle crashes between 2015 and 2020 in Daejeon Metropolitan City, utilizing the administrative dong (a neighborhood-level division) as the unit of analysis to explore influential factors in these crashes. This study used generalized linear mixed models capable of addressing fixed and random effects within a spatiotemporal Bayesian framework. The results estimated that the posterior mean of the variables high intersection density and high pedestrian crossing density were 0.328 and 0.711, respectively, suggesting an increase in the relative risks of crashes. Conversely, the posterior mean of population density was estimated to be -0.351, implying a decrease in crash risks. Regarding random effects, the study estimates temporal autocorrelation at 0.971 and spatial autocorrelation at 0.476, pointing to the dominance of temporal factors in the overall random effects. These findings provide objective evidence for formulating transportation safety policies at the localized level. Furthermore, the methodologies applied can be utilized to analyze crashes in other municipalities, thereby enhancing the understanding of the associated risks.

Keywords : Crash Risks, Daejeon Metropolitan City, Neighborhood-Level, Random Effects, Spatiotemporal Bayesian

본 논문은 국민대학교 2023년도 신입교원 논문지원사업의 지원을 받아 수행되었음(C2023-0023).

*Corresponding Author : Youngbin Lym(Kookmin University)

email: yblym0207@kookmin.ac.kr

Received February 26, 2024

Revised March 21, 2024

Accepted May 3, 2024

Published May 31, 2024

1. 서론

1.1 연구배경

우리나라의 교통정책은 산업화 과정에 따른 급격한 교통 수요 증가로 인해 안전보다는 교통 인프라의 공급과 이를 통한 사람 및 물류의 원활한 이동에 중점을 두고 이루어졌다[1]. 도로교통공단에 따르면, 우리나라는 1990년 OECD 국가 33개국 중 상위 5번째로 많은 교통사고가 발생했으며, 2020년 인구 10만 명당 교통사고 발생건수는 OECD 회원국 평균인 162.6건의 약 2.5배인 404.5건으로 나타나, 여전히 OECD 회원국 중 하위 수준에 머물러 있다[1-3].

이에 대응하여 교통안전 수준 향상을 위해 정부 및 지자체와 도로교통공단 등 관계기관의 노력이 있어왔고, 전국 기초자치단체를 대상으로 교통안전지수 등을 산출하여, 교통안전에 대한 지역의 관심과 경쟁을 유도하였다[4]. 하지만, 교통안전지수의 경우 기초자치단체인 시군구를 대상으로 하고 있기 때문에, 국가 전체의 교통 안전 수준의 공간적 분포를 파악하기에는 용이하나, 소규모 지역인 읍면동 수준에서의 교통사고 위험에 대한 파악이 어려운 단점이 있다.

한편, 교통사고는 차량 파손 및 사망과 부상 초래와 같은 직접적인 피해 이외에도, 사고로 인한 의료비 및 수리비, 장애인 지원 등 사회·경제적 비용을 발생시킨다. 또한, 사고로 인한 교통체증 및 혼잡, 그리고 사회적 안전감 저하와 같은 심리적인 피해를 주고 있으므로, 도시 및 지역 안전정책의 주요 관심사 중 하나로 간주해 왔다[5,6].

지역의 교통안전 정책 수립에 있어 필수적인 요소로는 사고 영향요인의 파악에 있다. 교통사고의 사고 빈도를 줄이거나 사고의 심각도를 낮추기 위한 다양한 영향요인 파악을 위해 데이터 및 통계모형 기반의 연구가 국내외에서 다수 이루어졌다[7-9]. 최근에는 빅데이터 및 기계학습, 심층학습의 여러 알고리즘 활용을 통해 데이터 기반의 교통사고 연구가 활발하게 이루어지고 있다[10-13].

이러한 데이터 기반 연구의 학문적 움직임과 궤적을 같이 하여, 본 연구에서는 현재 국내 연구에서 제한적으로 다루어진 시공간 베이지안 모형(spatiotemporal Bayesian model)을 활용하여 소규모 근린지역 단위에서의 교통사고 영향요인을 탐색하고자 한다[14,15]. 최근 기계학습 또는 심층학습 기반의 연구기법을 통해 교통사고의 영향요인을 파악하고 있으나, 시간 및 공간에 의한 불확실성과 같은 관측되지 않은 변동에 대한 이해

또는 설명에는 한계를 보인다. 한편, 베이지안 기법의 경우, 그 모형의 유연성으로 인해, 설명변수들에 의한 영향 요인과 공간 단위로 수집되는 자료의 분석에서 발생할 수 있는 공간적 이질성에 의한 불확실성과 반복에 따른 시계열적 특성에 의한 영향을 모두 파악할 수 있는 장점이 있다.

이 연구는 대전광역시를 실증 분석의 대상으로 선정하여, 대전광역시 내 79개 행정동에서 2015~2020년 동안에 발생한 교통사고에 대한 분석을 수행하고자 한다. 연구의 대상 지역으로 대전광역시를 선정한 이유는 2019년 기준, 특·광역시 중 대전광역시가 인구 10만 명당 교통사고 사망자 수 5.2명으로 가장 높은 반면, 타 특·광역시와 비교하여 대전광역시를 대상으로 한 연구가 부족하기 때문이다[16]. 아울러 이 연구의 주된 관심사인 소규모 근린지역인 대전광역시 내 행정동 단위에서의 사고위험과 그 영향요인 파악에 관한 연구는 기존에 수행되지 않았다. 또한, 대부분의 국내 소규모 지역(행정동) 교통사고 분석 연구가 단일 시점에서의 공간의 영향을 고려하고 있는 점과는 다르게, 이 연구는 6년 동안 대전광역시 내 79개 행정동에서 발생한 교통사고건수를 반복적으로 고려하고 있다는 점에서 다른 연구와 차이점이 있다.

행정동별 교통사고 위험의 공간적 분포와 위험의 시간에 따른 변화를 고려하여 그 영향요인을 데이터 및 모형을 통해 파악해보는 것은 소규모 행정구역 단위에서의 교통안전 정책 수립에 있어 객관적인 근거를 제시할 수 있다는 점에서 연구의 의의가 있다 하겠다.

1.2 연구목적

일반적으로 교통사고에 대한 분석은 사고 빈도(crash frequency)와 사고 심각도(crash severity)로 구분되며, 연구의 관심사가 사고의 빈도 또는 상대위험도(relative risks)에 대해 규명하고자 하는지 혹은 사고 심각도에 대한 분류인지의 여부에 따라 그 접근 방식이 달라진다[17]. 이 연구는 개별 사고의 심각도에 미치는 영향요인에 대한 탐색과 같은 미시적(microscopic) 분석이 아닌, 소규모 근린지역인 행정동을 연구의 분석단위 혹은 주제로 보고, 개별 행정동의 지역 특성이 연도별 교통사고의 사고 빈도(상대위험도)에 어떠한 영향을 미치는지 파악해 보기 위해 거시적(macrosopic)인 접근을 고려하였다. 이를 위해 다양한 공공데이터를 수집하여 교통사고 분석을 위한 데이터베이스를 구축한 후, 통계모형을 활용하여 다음과 같은 연구 질문에 대한 해답을 모색하고자 하였다.

- 1) 대전광역시 내 소규모 지역의 교통사고의 상대위험도에 미치는 공간의 영향은 존재하는가?
- 2) 이러한 교통사고의 상대위험도는 시간에 따라 변화하는가?
- 3) 시간 및 공간에 의한 확률효과 이외의 행정동 별 교통사고의 상대위험도에 미치는 근린환경 및 사회인구학적 특성은 무엇인가?

이 연구에서는 관측되지 않는 특성인 시간 및 공간의 확률효과와 파악을 위해 일반화선형혼합모형 기반의 시공간 베이지안 모형을 채택하고, 시뮬레이션을 활용하여 분석을 수행하였다.

2. 연구방법

2.1 연구대상 지역 및 분석자료 구축

국내 다수의 교통사고 연구는 한국교통안전공단에서 제공하는 공공데이터인 교통사고분석시스템(TAAS: Traffic Accident Analysis System, 이하 TAAS)의 사고데이터를 주로 이용하여 수행되어 왔다[18]. TAAS 데이터에는 발생한 개별 사고의 속성정보로 사고내용(경상, 부상신고, 증상, 사망사고 등 사고의 심각도)과 사고유형, 법규위반, 노면 상태, 사고 발생 시의 기상상태, 가해자 및 피해 운전자의 정보, 사고 발생일시, 그리고 공간정보인 시도, 시군구, 법정동에 관한 정보를 제공하고 있다. 하지만, 사고 발생지점에 대한 좌표정보를 제공하지 않기에, 공간정보를 활용한 교통사고 연구는 제한적으로 이루어졌다.

이러한 한계를 극복하기 위해 이 연구에서는 웹 페이지에서 데이터를 추출하는 기법인 웹 스크레이핑(web scraping)을 활용하여 교통사고가 발생한 지점의 좌표를 수집한 후, 이 좌표를 TAAS의 원데이터와 결합하여 새로운 데이터베이스를 구축하였다. 그리고 지리정보시스템(GIS: Geographic Information System, 이하 GIS)의 공간결합(spatial join) 기법을 통해 TAAS의 개별교통사고에 대전광역시 행정동의 공간정보를 포함할 수 있도록 데이터베이스를 확장하였다. 이 연구의 주된 관심사가 대전광역시 내 행정동의 교통사고 위험에 대한 시공간적 분석이기 때문에, 연구의 목적에 부합하도록 데이터베이스의 적절한 변환이 이루어졌다.

수집된 TAAS 데이터를 분석해 보면, 2007~2021년 기간 대전광역시에는 총 98,205건의 교통사고가 발생하

였다. 이 중 본 연구는 2015~2020년에 발생한 45,207건을 대상으로 하고 있는데, 그 이유는 해당 기간 대전광역시 행정동이 79개로 변화가 없었기에, 소규모 지역 사고위험의 시공간적 분석이 가능하기 때문이다. 아울러 교통사고 위험의 영향요인으로 고려할 수 있는 사회인구학적 지표 및 도로 및 버스정류장 밀도와 같은 근린환경에 대한 변수 수집 또한 이를 기반으로 이루어졌다. 이와 같은 분석데이터 구축, 시공간 베이지안 모형 개발 및 분석 등 데이터마이닝 기법의 전반적인 과정은 Fig. 1에 제시하였다.

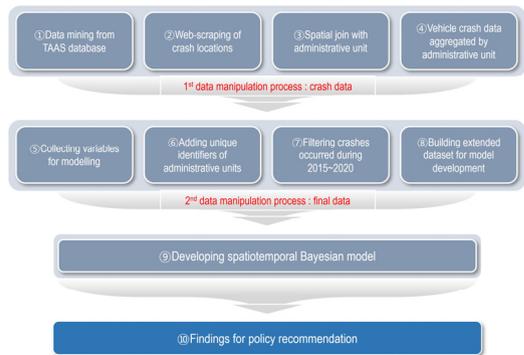


Fig. 1. Flow chart of the study process

2.2 변수선정

이 연구는 대전광역시 내 79개 행정동을 분석단위로 보고, 2015~2020년 6년 동안의 자료들을 수집한 총 $79 \times 6 = 474$ 개 관측치를 대상으로 하고 있다. 연구의 종속변수는 행정동 단위로 집계한 교통사고의 발생빈도이며, 이에 영향을 미칠 수 있는 사고 영향요인으로 인구, 인구밀도, 인당 차량등록 건수, 도로면적률, 교차로 밀도, 버스정류장 밀도, 횡단보도 밀도를 고려하였다. 이때 인구는 연도별 주민등록 연앙인구를 수집하였으며, 행정동별 사고의 기대교통사고 건수 계산 및 인당 차량등록 건수를 산출하는 데 활용하였다. 이 연구의 근린환경은 주로 도로환경을 고려하였는데, 행정동 내 도로면적과 버스정류장 수, 교차로의 개수, 그리고 횡단보도의 개수와 같은 데이터가 수집되었다. 수집된 자료의 기술 통계량은 Table 1에 제시되어 있다.

연구의 종속변수인 행정동 별 교통사고 빈도는 평균이 95.3, 분산의 경우 5,665.6(표준편차인 75.27의 제곱)으로 나타났다. 이는 대전광역시 행정동별 교통사고 발생 빈도가 평균과 비교하면 분산이 큰 우측으로 치우친 분포를 따르고 있으며, 과대 산포(overdispersion)의 문제

를 야기할 수 있음을 의미한다. 한편, 연구에서 활용된 설명변수들은 Table 1의 기술 통계량에서 나타난 것과 같이 그 척도(measurement scale)가 서로 다르기 때문에, 인당 차량등록 대수, 단위 면적 당 버스 정류장 수, 단위 면적당 교차로 수 등으로 변환하여, 변수 간의 상관관계를 파악하고자 하였다.

Table 1. A brief summary of data used for this study

Variables	Mean	S.d	Min	Max
Dependent variable				
Number of crashes	95.3	75.27	9	549
Independent variables				
<i>Sociodemographic</i>				
Population	18,912	9,392.9	2,491	51,122
Population density	10.802	7,973.9	39	36,771
Vehicle registration	8,385	4,602.7	1,223	25,733
Area(km ²)	6.83	12.79	0.46	63.58
<i>Road environment</i>				
Road area(km ²)	0.575	0.479	0.104	2.662
No. of bus stops	18.1	15.83	0	76
No. of intersections	383	394	7	2,529
No. of crossroad	158	111.35	26	593

Note: 1) Each variable was measured by the administrative unit, Dong.
2) Variables were further converted to meet the needs of statistical investigation

Number of observations: 79(dongs) × 6(years) = 474

2.3 시공간 베이지안 모형

대전광역시 내 행정동(i) 및 특정 시점(t)에서 관측된 교통사고 발생 건수를 $Y_{i,t} = y_{i,t}$ 라고 할 때, 이 값은 음의 값을 가지지 않는 가산자료(countable data)라 볼 수 있다. 가산자료의 분석은 일반적으로 포아송(Poisson) 모형을 통해 이루어지는데, 모수(parameter)인 $\mu_{i,t}$ 즉, 평균 사고 건수를 따른다고 가정하자. 모수는 다음과 같이 $\mu_{i,t} = O_{i,t} \times \theta_{i,t}$ 로 분해(decomposition)할 수 있으며, 이 때, $O_{i,t}$ 는 기대교통사고 건수 또는 오프셋(offset)을 의미하고, $\theta_{i,t}$ 는 상대위험도를 나타낸다[19,20]. 이를 수식으로 표현하면, 다음과 같다.

$$Y_{i,t} \mid \theta_{i,t} \sim \text{Poisson}(O_{i,t} \times \theta_{i,t}) \quad (1)$$

이 연구의 관심인 상대위험도($\theta_{i,t}$)는 포아송 모형의 정준연결함수(canonical link function)인 로그함수

(logarithmic function)를 이용하여 다음과 같이 나타낼 수 있다.

$$\log(\theta_{i,t}) = \eta_{i,t} \quad (2)$$

$$\eta_{i,t} = \sum_{j=0}^J \beta_j X_{i,t,j} + \psi_{i,t} \quad (3)$$

로그 상대위험도는 Eq. (2)와 같이 선형 예측자(linear predictor)인 $\eta_{i,t}$ 로 나타난다. 구체적으로는 Eq. (3)에서 제시된 것처럼, 절편 β_0 를 포함한 계수값 벡터인 $\beta = (\beta_0, \beta_1, \dots, \beta_{J-1}, \beta_J)^T$ 와 독립변수인 $X_i = (x_{i,t,0}, x_{i,t,1}, \dots, x_{i,t,J-1}, x_{i,t,J})^T$ 벡터의 선형결합에 의한 고정효과(fixed effect) 부분과 확률효과(random effect)를 나타내는 $\psi_{i,t}$ 로 이루어져 있다. 관련하여 다양한 시공간 베이지안 확률효과 모형이 연구되어 왔고, 그 중 이 연구에서는 이 연구의 목적인 시공간의 확률효과를 규명하기 위해 Rushworth et al.[21]이 제안한 1차 시공간 자기 회귀 모형(AR1: first order spatiotemporal autoregressive model, 이하 AR1 모형)을 활용하여 분석을 수행하였다.

$$\psi_{i,t} = \phi_{i,t} \quad (4)$$

$$\phi_t \mid \phi_{t-1} \sim \text{Normal}(\rho_T \phi_{t-1}, \sigma^2 \mathbf{Q}(\mathbf{W}, \lambda_S)^{-1})$$

$$\phi_1 \sim \text{Normal}(0, \sigma^2 \mathbf{Q}(\mathbf{W}, \lambda_S)^{-1})$$

$$\sigma^2 \sim \text{Inverse-Gamma}(a, b)$$

$$\lambda_S \sim \text{Uniform}(0, 1)$$

$$\rho_T \sim \text{Uniform}(0, 1)$$

이 때, $t = 1, 2, \dots, 6$, $i = 1, 2, \dots, 79$ 이며, ρ_T 는 1차 시계열 모형의 자기상관 계수를, λ_S 는 공간 종속성을 나타내는 모수를 의미한다. 이 연구에서는 공간의 확률효과에 대하여 Leroux et al.[22]이 제안한 모형을 활용하였는데, Eq. (5)와 같다.

$$\mathbf{Q}(\mathbf{W}, \lambda_S)^{-1} = [\lambda_S(\mathbf{D}_W - \mathbf{W}) + (1 - \lambda_S)\mathbf{I}]^{-1} \quad (5)$$

이 때, \mathbf{W} 는 공간가중치 행렬을 의미하며, 어떤 두 행정동(i)와 행정동(k)이 인접하고 있는 경우 $w_{i,k} = 1$, 그

렇지 않은 경우 $w_{i,k} = 0$ 으로 구분되며, 자기 자신인 경우 $w_{i,i} = 0$ 과 같이 공간의 영향을 반영하였다. 대각행렬 (diagonal matrix)인 D_w 는 공간가중치 행렬 W 의 개별 행을 통해 구축되며, W 의 각 행에서 0이 아닌 값들의 개수를 더하여 대각행렬 D_w 의 i 번째 대각 성분으로 반영하였다. 그리고 I 는 항등행렬(identity matrix)을 나타내며, 공간의 영향을 나타내는 모수인 λ_S 는 $0 \leq \lambda_S \leq 1$ 과 같이 제시하였다.

Eq. (5)를 구체적으로 살펴보면, 공간의 영향을 나타내는 $D_w - W$ 행렬 부분과 영향이 없는 부분인 I 행렬의 선형결합으로 나타나며, 공간 종속성을 나타내는 모수 λ_S 가 가중치의 역할을 하고 있음을 알 수 있다. 만일 λ_S 의 추정치가 1에 가까우면, 그 확률효과가 매우 강한 공간 상관성을 띠고 있음을 의미하며, 0인 경우 공간의 영향이 없음을 나타낸다고 하겠다. 즉, 공간효과의 지역적 차이 반영을 위해 공분산인 Q^{-1} 에 Eq. (5)와 같은 구조를 구축하였다고 볼 수 있다. 한편, σ^2 은 AR1 모형의 확률효과의 크기를 나타내는 모수로 Eq. (4)와 같이 시간의 확률효과 및 공간의 확률효과의 변동성이 하나의 모수를 통해 나타나고 있음을 알 수 있다.

3. 분석 결과 및 고찰

3.1 상관분석

2.3절의 계층적 베이지안 접근 방식 기반의 AR1 모형의 구축에 앞서, 투입변수(설명변수) 간 상관 정도를 파악하기 위해 상관분석을 수행하였다. 설명변수들 사이에 상관관계가 강하게 나타나는 경우 혹은 다중공선성 (multicollinearity)이 존재하는 경우 회귀모형의 계수 추정치의 변동성이 커져, 잘못된 결론을 내릴 수 있는 우려가 있기 때문이다. Fig. 2에서 나타나는 바와 같이, 이 연구에서 고려하는 사회인구학적 변수들과 도로환경 변수들 사이에는 큰 상관관계가 존재한다.

Fig. 2를 구체적으로 살펴보면, 1인당 차량등록대수 (CarPerPerson)와 트럭의 비율(TruckRate) 사이에는 0.335로 유의미한 양(+)의 상관관계가 나타난다. 도로율 (Road ratio)과 단위면적당 버스정류장 수(Bus Stops), 단위면적당 교차로 수(Intersections), 그리고 단위면적당 횡단보도의 수(Crossings) 사이에는 각각 0.746, 0.624, 0.850으로 매우 큰 양(+)의 상관관계가 존재하고

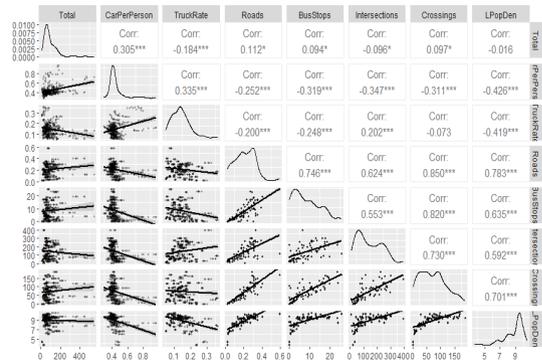


Fig. 2. Pairwise correlation plot among variables

있음을 알 수 있다. 또한, 도로율, 버스정류장 수, 교차로 수, 횡단보도의 수 변수들은 인구밀도(Population Density)와도 매우 큰 양(+)의 상관관계가 있는 것으로 나타났다.

이러한 설명변수 간 다중공선성 문제를 우회하기 위해 트럭의 등록비율 변수는 배제하고, 교통사고 위험에 영향을 미칠 수 있는 중요한 변수라 볼 수 있는 인구밀도, 행정동 단위 면적 당 버스 정류장의 수, 교차로의 수 및 횡단보도 개수의 경우는 범주형 변수로 변환하였다. 이러한 과정을 통해 최종적으로 얻은 본 연구의 활용 변수들은 다음과 같다.

- 1) 종속변수: 행정동 별 교통사고 발생건수
- 2) 오프셋: 행정동별 인구
- 3) 설명변수(연속형): 1인당 차량등록대수, 도로율
- 4) 설명변수(범주형): 인구밀도(낮음, 중간, 높음), 단위면적당 버스정류장 수(낮음, 중간, 높음), 단위면적당 교차로의 수(낮음, 중간, 높음), 단위면적당 횡단보도의 수(낮음, 중간, 높음)

3.2 AR1 모형 분석 결과

이 연구는 대전광역시 79개 행정동을 대상으로 2015~2020년의 연간 발생교통 건수를 종속변수로 두고, 이러한 교통사고에 영향을 미치는 요인을 파악하고자 하였다. 교통사고 발생빈도는 교통량에 대한 노출이나 인구 등에 따라 지역에 따른 차이가 있다. 따라서 이러한 지역적 차이를 반영하기 위해 인구를 바탕으로 오프셋/기대교통사고 건수($O_{i,t}$)를 산출하고, Eq. (2)와 같이 교통사고의 상대위험도를 추정하였다.

AR1 모형의 추정을 위해 이 연구에서는 마르코프 체인 몬테카를로(MCMC: Markove Chain Monte Carlo, 이하 MCMC) 기법을 활용하였다[23,24]. 독립된(independent)

MCMC 체인 3개 각각에 초기 번인(burn-in) 샘플 200,000개, MCMC 샘플 5,200,000개 및 마르코프 체인 샘플의 자기상관을 줄이기 위해 매 1,000번째의 샘플을 가져올 수 있도록 thin = 1,000으로 지정해주어, MCMC 체인 1개당 5,000개씩 총 5,000×3 = 15,000개의 MCMC 샘플을 도출하도록 하였다.

Table 2는 MCMC를 통해 도출된 결과를 정리한 것으로, 연구관심사인 설명변수의 고정효과와 시공간의 확률효과의 모수 추정값을 보여주고 있다. 여기에서는 모수의 점추정값인 사후평균(posterior mean)과 이를 중심으로 한 95% 신용구간(credible interval)을 통해 변수들의 영향력을 파악하였다.

우선 고정효과에 대해 살펴보면, 행정동별 인당 차량등록 대수(Car-registered per person)의 경우 사후평균 2.379로 추정되었으며, 95% 신용 구간이 0을 포함하고 있지 않아 교통사고의 (로그) 상대위험도에 유의미한 영향을 끼치는 것으로 나타났다. 즉, 행정동의 인당 차량등록 대수가 한 단위 높아질수록 교통사고의 상대위험이 증가한다고 하겠다. 도로 비율(Road ratio)의 경우 사후평균은 -0.708로 추정되었으나 신용 구간이 (-1.7, 0.283)으로 구간 내 0이 포함되어 있어, 사고위험에 유의미한 영향을 미치지 않는 것으로 추정되었다.

Table 2. MCMC results

	Mean	2.5%	97.5%	Effective samples
Intercept	-1.202	-1.510	-0.893	5963.1
Car-registered per person	2.379	1.824	2.929	3413.2
Road ratio	-0.708	-1.700	0.283	1172.2
Population density _M	-0.242	-0.364	-0.117	1591.3
Population density _H	-0.351	-0.507	-0.197	1397.5
Bus stops _M	0.063	-0.115	0.239	3009.5
Bus stops _H	-0.024	-0.237	0.190	6612.8
Intersections _M	0.229	0.057	0.405	8465.0
Intersections _H	0.328	0.123	0.540	8830.8
Pedestrian crossings _M	0.136	-0.076	0.352	3995.3
Pedestrian crossings _H	0.711	0.421	0.995	2226.1
σ^2	0.127	0.090	0.169	7283.0
λ_S	0.476	0.243	0.730	6653.1
ρ_T	0.971	0.924	0.999	15044.9

- 1) DIC = 3,708.3, p.d = 333.23, LMPL = -1,933.2
- 2) Reference categories: Population density_L(Low), Bus stops_L, Intersection_L, Pedestrian crossings_L
- 3) All values of PSRF of parameters are less than 1.1, suggesting good convergence of MCMC chains.

한편, Fig. 2의 상관분석 결과에 따라 인구밀도, 단위 면적 당 버스정류장의 수, 교차로의 수, 및 횡단보도의 수는 범주형 변수의 형태로 분석이 수행되었다. 우선 인구밀도 변수(Population density)는 낮음(L) 대비(reference), 중간(M), 높음(H)의 범주에서 음(-)의 유의미한 영향을 미치고 있다고 추정되었다. 즉, 인구의 크기를 오피셋으로 반영하고, 다른 변수들의 효과들을 고정했을 때, 인구밀도가 높아지면, 인구밀도가 낮은 곳에 비해 교통사고의 상대위험도가 낮아짐을 의미한다고 볼 수 있다. 인구밀도가 높은 지역이 교통법규가 강하거나 단속 등 교통안전을 위한 대응이 상대적으로 강하기 때문이라 하겠다.

근린환경을 나타내는 행정동 면적 당 버스정류장의 수(Bus stops) 변수는 음(-)의 영향을 보여주는 것으로 추정되어 졌으나, 그 변동성으로 인해 통계적으로 유의미한 결과를 보여주지는 못했다. 교차로 변수(Intersections)의 경우 교차로 밀도가 낮은 지역(L)과 비교하여, 중간(M), 높음(H) 모두 양(+)의 유의미한 영향을 미치는 것으로 나타났다. 마지막으로 횡단보도(Pedestrian crossings) 밀도 변수의 경우 밀도가 낮은 지역(L) 대비, 높음(H)에서만 통계적으로 유의미하게 사고의 상대적 위험이 증가하는 것으로 추정되었다.

Eq. (4)와 (5)에서 제시된 시공간의 확률효과에 대한 MCMC 결과를 살펴보면, 시간의 효과를 나타내는 모수인 ρ_T 의 경우 사후평균이 0.971로 나타나 매우 강한 시간 종속성을 보임을 알 수 있다. 한편, 공간 종속성을 나타내는 λ_S 의 사후평균은 0.476으로 나타나 시간의 영향에 비해 상대적으로 낮은 것으로 나타났다. 즉, 대전광역시 교통사고의 상대위험도에는 79개 행정동 사이에 존재하는 공간의 확률효과에 비해, 연도별로 반복되어 관측되는 시간 종속성의 영향이 더 크다고 볼 수 있다 하겠다. 마지막으로 확률효과의 변동성의 크기를 나타내는 분산 σ^2 의 사후평균은 0.127로 추정되었는데, 이는 모형에서 고정효과를 반영하고 남은 잔차(residual)의 분포를 설명하고 있다 하겠다.

Fig. 3은 연구에서 활용된 일부 설명변수들의 MCMC 시뮬레이션 결과를 보여준다. Fig. 3의 가장 윗부분은 행정동의 인당 차량등록 대수의 영향을 보여주고 있는데, MCMC 체인 3개가 잘 혼합되어(mixed) 있으며 대부분의 추정치가 1.5에서 3.5 사이에 위치하고 있어, 그 영향이 양(+)임을 의미한다고 볼 수 있다. 두 번째 그림은 도로 비율의 영향을 담고 있는데, MCMC 샘플의 값이 -2

와 1 사이에 분포하고 있어 0을 포함하고 있으므로 그 영향이 유의미하지 않음을 알 수 있다. 세 번째와 네 번째 그림은 각각 교차로 밀도 높음(H)과 횡단보도 밀도 높음(H) 변수에 대한 것으로 MCMC 분포가 모두 0에서 벗어난 양(+)의 범위에 있어, 교통사고 위험을 증가시키는 영향요인이라 볼 수 있다.

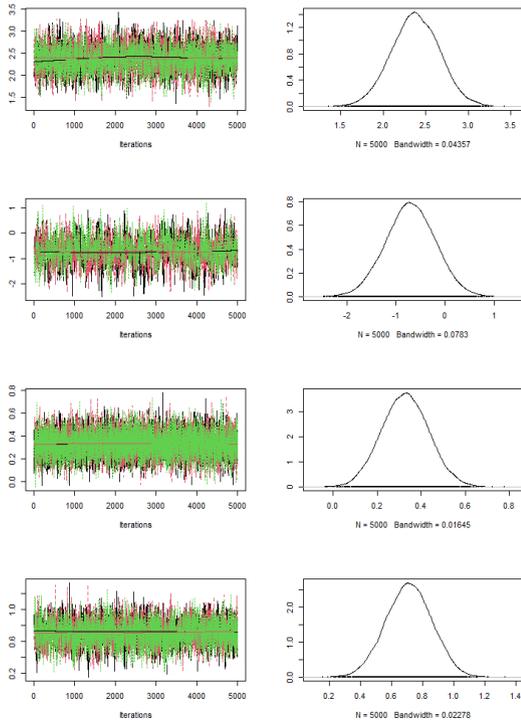


Fig. 3. MCMC plots for selected variables: Car-registered per person (Top), Road ratio (Upper middle), IntersectionsH (lower middle), Pedestrian crossingsH (Bottom)

이러한 연구 결과를 종합하여, 2015~2020년 6년 동안의 대전광역시 교통사고의 상대위험도 경향을 Fig. 4와 같이 제시하였다. 이 결과는 MCMC 시뮬레이션을 통해 적합된(fitted) 값을 활용하였으며, 사고위험의 중간값(median)과 95% 신용 구간을 나타내고 있다. 대전광역시 내 평균 교통사고 위험은 2015년부터 2018년까지 대체로 증가하였으며, 2019년도에 가장 높은 위험도를 보였다. 이는 TAAS 웹사이트 내 교통사고정보 인포그래픽에서 제시된 우리나라의 최근 10년간 교통사고 추세에서도 드러난다[25]. 우리나라의 교통사고는 2015년 이후 감소하다가 2019년에 정점을 찍고, 다시 감소하는 패턴을 보이고 있으며, 대전광역시 역시 2016년 이후 2019

년까지 지속적으로 증가하다가 2020년에 감소하였다. 본 연구에서 제시한 모형이 교통사고 변화에 대한 이와 같은 추세를 잘 반영하고 있다 하겠다.

아울러 도로교통공단의 2019 및 2020년도 전국 기초자치단체별 교통안전지수 보고서에 따르면, 대전광역시 5개구의 교통안전지수는 대덕구(C → B), 동구(C → C), 서구(E → E), 유성구(C → B), 중구(D → C)로 2019년에 비해 2020년의 등급이 상승하였다[26]. 이는 교통안전에 대한 정부 및 지방자치단체의 지속적 노력과 시민들의 의식개선이 2020년의 교통안전수준 향상과 그에 따른 사고위험의 감소에까지 이어졌으리라 생각된다.

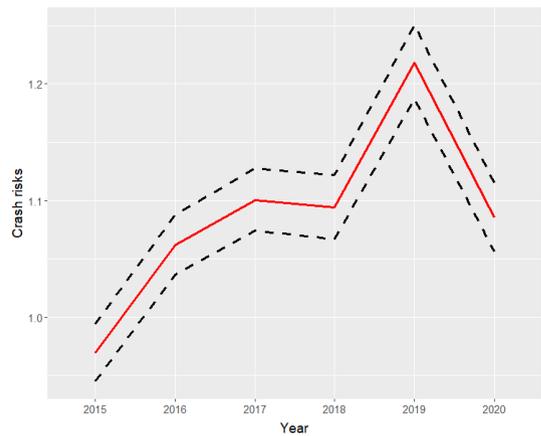


Fig. 4. Posterior median and 95% credible interval for the temporal trend in crash risk over 2015~2020

4. 결론 및 시사점

이 연구는 2015~2020년의 6년 기간 동안 대전광역시에서 발생한 교통사고를 대상으로, 웹스크레이핑을 활용한 데이터 수집과 베이지안 모형을 적용한, 데이터마닝 기반의 교통사고 요인의 영향 분석에 관한 것이라 볼 수 있다.

이 연구에서 활용된 시공간 베이지안 모형의 추정은 MCMC 시뮬레이션에 의해 이루어졌다. 이 모형은 고정효과와 확률효과를 함께 고려한 일반화선형혼합모형(generalized linear mixed model) 기반이며, 설명변수에 의한 고정효과영역의 영향력은 다음과 같다.

- 1) 유의미한 양(+)의 영향 : 행정동 1인당 차량등록 대수, 교차로 밀도 중간(M) 및 높음(H), 횡단보도 밀도 높음(H)

2) 유의미한 음(-)의 영향 : 인구밀도 중간(M) 및 높음(H)

한편, 시공간에 의한 확률효과 역시 추정되었는데, 시간 종속성을 나타내는 자기 상관계수의 사후평균이 0.971로 나타나 매우 강한 시간 종속성을 보임을 알 수 있었다. 공간 종속성의 경우 모수의 사후평균 추정치가 0.476으로 나타나 시간의 확률효과와 비교하여 상대적으로 낮은 것으로 나타났다. 이는 대전광역시 교통사고의 상대위험도가 행정동 간의 공간의 영향에 비해, 동일 행정동 내에서 반복적으로 작용하는 시계열적 자기상관의 영향이 크다는 것을 의미한다고 하겠다.

이러한 분석 결과를 종합하여, 모형 기반의 대전광역시 교통사고의 상대적 위험도 변화 추이를 살펴보았다. 대전광역시 내 평균 교통사고 위험은 2015년부터 증가하다가 2019년도에 가장 높았던 반면 2020년에는 감소하였다.

이 연구는 대전광역시를 대상으로 하고 있는데, 대전광역시의 경우 타 특·광역시에 비해 실증 데이터를 활용한 교통사고 분석이 제한적으로 이루어졌으며, 대전광역시 및 다른 지자체를 대상으로 한 시공간 베이지안 모형 기반의 교통사고 연구는 현재 존재하지 않기에 연구의 독창성이 확보된다고 하겠다. 따라서 이 연구에서 활용한 기법을 다른 지역에도 적용한다면, 지방자치단체 간 교통위험의 차이에 대한 비교와 함께, 소규모 행정구역 단위에서의 교통안전 정책 수립을 위한 객관적인 근거를 제시할 수 있다는 점에서 연구의 의의를 찾을 수 있을 것이다.

References

- [1] Traffic Accident Analysis System (TAAS). Traffic Accident Statistical Report 2021 and 2022 [Internet]. Traffic Accident Analysis System, c2022 [cited 2022 October 13], Available From: https://taas.koroad.or.kr/web/bdm/srs/selectStaticReportsList.do?menuId=WEB_KMP_IDA_SRS_TAA (accessed Feb. 23, 2024)
- [2] International Transport Forum. Road Safety Annual Report 2021: The Impact of Covid-19 [Internet]. International Transport Forum / OECD, c2021 [cited 2021 December 22], Available from: <https://www.itf-oecd.org/road-safety-annual-report-2021-impact-covid-19> (accessed Feb. 24, 2024)
- [3] H. Kang, M. Noh, "Classifying the severity of pedestrian accidents using ensemble machine learning algorithms: A case study of Daejeon City," *Journal of Digital Convergence*, vol. 20, no. 5, pp.39-46, May. 2022. DOI: <https://doi.org/10.14400/JDC.2022.20.5.039>
- [4] Traffic Accident Analysis System (TAAS). Statistical Report: Transportation Safety Level Index of the year 2016 [Internet]. Traffic Accident Analysis System, c2017 [cited 2017 December 22], Available From: https://taas.koroad.or.kr/web/bdm/srs/selectStaticReportsList.do?menuId=WEB_KMP_IDA_SRS_TSI (accessed Feb. 23, 2024)
- [5] S. Pervaz, T. Bhowmik, N. Eluru, "Integrating macro and micro level crash frequency models considering spatial heterogeneity and random effects," *Analytic Methods in Accident Research*, vol. 36, p. 100238, Dec. 2022. DOI: <https://doi.org/10.1016/j.amar.2022.100238>
- [6] 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. DOI: <https://doi.org/10.1016/j.ssmph.2022.101293>
- [7] B. M. Jeong, J. K. Lee, Y. U. Ban, T. Y. Heo, "A Comparative Analysis of Applying Traffic Accident Models with Spatial Effects," *Crisis and Emergency Management: Theory and Praxis*, vol. 14, no. 7, pp. 131-140, Jun. 2018. DOI: <https://doi.org/10.14251/crisisonomy.2018.14.7.131>
- [8] J. H. Na, Y. Kang, J. H. Kim, H.-J. Oh, "Real Time Black Ice Warning Alert Service based on Accident Risk Index," *The Journal of Korean Institute of Information Technology*, vol. 20, no. 2, pp. 11-19, Feb. 2022. DOI: <https://doi.org/10.14801/jkiit.2022.20.2.11>
- [9] J. Jang, J. Park, "Examining Factors of Severity of Traffic Accidents of Senior Drivers Using Machine Learning," *International Journal of Highway Engineering*, vol. 24, no. 5, pp. 109-115, Oct. 2022. DOI: <https://doi.org/10.7855/ijhe.2022.24.5.109>
- [10] H. Lee, S. Lee, "Comparative Analysis of Machine Learning Models for the Prediction of Pedestrian Crash Severity: Focused on Balancing Pedestrian Crash Dataset," *Journal of Korean Society for Geospatial Information Science*, vol. 29, no. 2, pp. 3-15, Jun. 2021. DOI: <https://doi.org/10.7319/kogsis.2021.29.2.003>
- [11] C. Kwon, H. Chang, "Comparative Analysis of Traffic Accident Severity of Two-Wheeled Vehicles Using XGBoost," *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 20, no. 4, pp. 1-12, Aug. 2021. DOI: <https://doi.org/10.12815/kits.2021.20.4.1>
- [12] M. Ijaz, L. Ian, M. Zahid, A. Jamal, "A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw," *Accident Analysis & Prevention*, vol. 154, p. 106094, May 2021. DOI: <https://doi.org/10.1016/j.aap.2021.106094>
- [13] S. Kim, "Elderly Driver-involved Crash Analysis and Crash Data Policy," *The Journal of The Korea Institute of*

- Intelligent Transport Systems*, vol. 21, no. 5, pp. 90-102, Oct. 2022.
DOI: <https://doi.org/10.12815/kits.2022.21.5.90>
- [14] P. Moraga, *Spatial Statistics for Data Science*. CRC Press, pp. 65-100, 2023.
- [15] A. B. Lawson, *Using R for Bayesian Spatial and Spatio-Temporal Health Modeling*. CRC Press, pp. 67-86, 2021.
- [16] Daily Truth News 24(dtnews24). Daejeon, the city with the highest number of traffic accidents, to invest 1.033 trillion won in improving traffic environment [Internet]. dtnews24, c2020 [cited 2020 December 28], Available From: <https://www.dtnews24.com/news/articleView.html?idxno=595598> (accessed Feb. 23, 2024)
- [17] Z. Chen, Y. Lym, "The influence of built environment on distracted driving related crashes in Ohio," *Transport Policy*, vol. 101, pp. 34-45, Feb. 2021.
DOI: <https://doi.org/10.1016/j.tranpol.2020.11.011>
- [18] Traffic Accident Analysis System (TAAS). TAAS GIS [Internet]. Traffic Accident Analysis System, c2024 [cited 2024 January 5], Available From: https://taas.koroad.or.kr/gis/mcm/mcl/initMap.do?menuId=GIS_GMP_STS_RSN (accessed Feb. 23, 2024)
- [19] V. Gomez-Rubio, *Bayesian inference with INLA*. CRC Press, pp. 150-185, 2020.
- [20] S. Sahu, *Bayesian Modeling of Spatio-Temporal Data with R*. CRC Press, pp.301-355, 2022.
- [21] A. Rushworth, D. Lee, R. Mitchell, "A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London," *Spatial and Spatio-temporal Epidemiology*, vol. 10, pp. 29-38, Jul. 2014.
DOI: <https://doi.org/10.1016/i.sste.2014.05.001>
- [22] B. G. Leroux, X. Lei, N. Breslow, Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence, *Statistical models in epidemiology, the Environment, and Clinical trials.*, vol. 116. New York, New York, United States: Springer New York, pp.179-191, 2000.
- [23] D. Lee, A. Rushworth, G. Napier, "Spatio-Temporal Areal Unit Modeling in R with Conditional Autoregressive Priors Using the CARBayesST Package," *Journal of Statistical Software*, vol. 84, no. 9, 2018.
DOI: <https://doi.org/10.18637/jss.v084.i09>
- [24] D. Lee, "A tutorial on spatio-temporal disease risk modelling in R using Markov chain Monte Carlo simulation and the CARBayesST package," *Spatial and Spatio-temporal Epidemiology*, vol. 34, p. 100353, Aug. 2020.
DOI: <https://doi.org/10.1016/i.sste.2020.100353>
- [25] Traffic Accident Analysis System (TAAS). Traffic accident information Infographic for 2022 [Internet]. Traffic Accident Analysis System, c2024 [cited 2024 March 20], Available From: https://taas.koroad.or.kr/sta/acs/gus/selectStaInfoGraph.do?menuId=WEB_KMP_IDA_TAI (accessed Mar. 21, 2024)
- [26] Traffic Accident Analysis System (TAAS). Statistical Report: Transportation Safety Level Index of the year 2022 [Internet]. Traffic Accident Analysis System, c2023 [cited 2023 December 06], Available From: https://taas.koroad.or.kr/web/bdm/srs/selectStaticalReportsList.do?menuId=WEB_KMP_IDA_SRS_TSI (accessed Mar. 21, 2024)

임 영 빈(Youngbin Lym)

[정회원]



- 2020년 8월 : 오하이오주립대학교 도시 및 지역계획학과 (도시 및 지역계획학 박사)
- 2021년 9월 ~ 2022년 8월 : 충남대학교 자연과학연구소 박사후연구원
- 2022년 9월 ~ 현재 : 국민대학교 행정학과 조교수

<관심분야>

베이저안 통계학, 데이터 과학, 도시 및 지역안전, 공간정보 활용