

## A Comparative Analysis of Applying Traffic Accident Models with Spatial Effects

Bo Mi Jeong<sup>1#</sup>, Jae Kyeong Lee<sup>2</sup>, Yong Un Ban<sup>3</sup>, Tae Young Heo<sup>1+</sup>

<sup>1</sup> Department of Information and Statistics, Chungbuk National University, 1, Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Korea

<sup>2</sup> Idea Commercialization Center, Korea Institute of Science Technology Information, 66, Hoegi-ro, Dongdaemun-gu, Seoul, Korea

<sup>3</sup> Department of Urban Engineering, Chungbuk National University, 1, Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Korea

### Abstract

The purpose of this study is to develop traffic accident models based on spatial effects and traffic accident data characteristics. Statistical models are constructed with the number of traffic injuries observed from 277 segments of interstate highway I-5 in Washington for 9 years from 1999 to 2007 as a response variable. A significance test for Moran's Index is conducted to determine whether or not spatial autocorrelation exists among the segments. Spatial zero Inflated model and spatial hurdle model are used to reflect excessive zeros in the response variable as well as spatial effects. Those models are found to be appropriate to handle the data according to DIC. Therefore, we propose these models as an optimal traffic accident model, which allows us to determine the effect of major factors of traffic accidents by investigating the influence of each explanatory variable in the models.

**Key words:** traffic accident, spatial correlation, zero inflated model, hurdle model

### 1. 서론

교통안전 분야에서의 주요 목적은 교통사고 발생에 영향을 미치는 요인 탐색에 있다. 따라서 자료의 특성을 정확히 파악하여 그에 맞는 최적의 모형을 찾는 것은 매우 중요하다. 기본적으로 교통사고 발생건수는 가산 자료로서 교통사고 부상자 수 분석을 일반화된 포아송 모형을 활용하여 실시한 Wulu, *et. al.*(2002)와 같이 포아송 분포를 가정하고 모형을 개발한다.

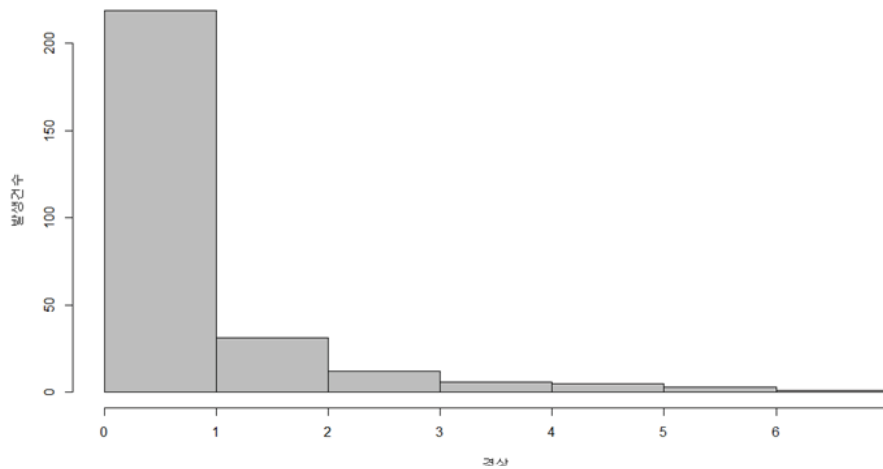
포아송 분포는 단위 구간(발생 위치 또는 지역)에 평균과 분산이 동일해야 한다는 가정이 존재한다. 하지만

본 연구에서 활용한 교통사고 발생건수는 평균보다 분산이 더 큰 값을 가지며 0값이 과잉된 자료이다. 따라서 이러한 자료의 특성을 반영한 모형 개발이 필요하다. Famoye, *et. al.*(2002)은 자료에 0이 많아 과대산포가 발생하는 경우에 영 과잉 일반화된 포아송 모형을 적용하여 적합하였으며 Rui Fang(2013)은 평균과 분산이 다른 생물의학 분야의 자료의 특성을 반영하여 영 과잉 음이항 모형을 연구하였다.

또한 교통사고 발생건수는 발생 위치에 따른 공간적인 영향을 받을 수 있는 자료이므로 공간적 영향력을 고려하여 더 적합성이 더 높은 모형을 개발할 수 있다.

# The 1st author: Bo Mi Jeong, Tel. +82-10-4161-9103, e-mail, qhal0324@hanmail.net

+ Corresponding author: Tae Young Heo, Tel. +82-43-261-3741, e-mail, theo@chungbuk.ac.kr



※ Source: Washington State Department of Transportation.

Figure 1. The number of evidence injury of traffic accidents of Interstate highway I-5

Black & Thomas(1998)는 Moran's I를 확인하여 분할된 도로의 공간상관성을 연구하였으며 Choi, *et. al.* (2010)은 포아송 모형과 음이항 모형, 일반화된 포아송 모형에 공간영향력을 반영하여 서울시 범죄 영향요인을 도출하였다.

따라서 본 연구에서는 자료의 특성을 반영하여 통계적 가정을 위배하지 않는 최적의 교통사고 모형을 제안하고자 한다. 먼저 가산자료의 특성을 반영하여 포아송 모형과 음이항 모형을 고려하였으며 자료의 과잉된 0값을 반영할 수 있는 영 과잉 모형(Zero Inflated Model)과 허들 모형(Hurdle model)을 고려하였다. 또한 Moran 지수(Moran's I)를 사용하여 자료의 공간상관성을 확인한 후 과잉된 0값과 공간영향력을 동시에 반영한 공간 영 과잉 모형(Spatial Zero Inflated Model)과 공간 허들 모형(Spatial Hurdle Model)을 적합하여 공간상관성을 반영하지 않은 기존의 영 과잉 모형과의 비교를 통해 공간모형의 우수성을 확인하였다.

## II. 분석대상 및 자료

본 연구는 미국 I-5 고속도로에서 1999년부터 2007년까지 발생한 교통사고 자료를 이용하였으며, 각 년도 별로 하행선을 277개의 동질한 구간으로 분할하여 관측된 교통사고 발생건수를 종속변수로 사용하였다. 본

연구에서 사용된 설명변수로는 각 구간에서 나들목(Interchange) 존재여부, 왼쪽 길어깨, 오른쪽 길어깨, 차선 수, 도로길이, 지방에 속하는 구간과 나들목과의 거리, 도심에 속하는 구간과 나들목과의 거리, 일 교통량을 사용하였다. 특히, 나들목 존재여부 변수는 각 구간에서 나들목의 존재 여부를 나타내는 변수로 나들목이 있을 경우 0, 없는 경우 1을 나타낸다. 왼쪽 길어깨와 오른쪽 길어깨는 도로 옆 쪽에 있는 갓길의 길이를 나타내며, 일교통량 변수는 각 구간별 교통량으로 로그 함수를 통해 변환된 값을 사용하였다.

〈Figure 1〉에서와 같이 본 연구에서 사용한 종속변수인 교통사고 발생 건수는 교통사고가 발생하지 않은 0값이 과대하게 관측된 영 과잉 자료의 특성을 가지고 있는 것으로 확인되어 영 과잉 기반의 모형들을 통해 교통사고 발생에 영향을 미치는 영향요인을 확인하고자 한다.

## III. 분석 모형

본 장에서는 다양한 가산자료 모형 및 공간모형에 대해 간략히 기술하고자 한다. 본 연구에서는 0이 팽창된 자료의 특성을 반영한 모형으로 영 과잉 모형과 허들모형을 활용하고 포아송 모형과 음이항 모형 기반의 모형을 적합하고자 한다.

1. 영 과잉 모형

가산자료의 분석에서 주로 사용되는 포아송 모형은 분산이 평균보다 큰 과대산포의 문제가 발생하게 될 경우 적합한 모형이라 할 수 없으며, 이러한 문제가 발생하는 원인 중 하나는 관측 값에 0값이 과대하게 관측되는 경우이다. 따라서 0의 팽창을 반영하기 위한 방법으로 영 과잉 모형과 허들 모형이 개발되었다.

영 과잉 모형은  $w$ 의 발생확률을 가지는 0값과  $1-w$ 의 발생확률로 0값을 포함한 건수 값으로 나누어 분석하는 모형으로 확률밀도함수는 식 (1)과 같다.

$$f(Y=y) = \begin{cases} w + (1-w)f(y), & y=0 \\ (1-w)f(y), & y=1,2,\dots \end{cases} \quad (1)$$

영 과잉 모형에서 0값을 포함한 가산형 자료를 분석하기 위해 고려하는 모형 중 포아송 분포를 기반으로 제안된 모형이 영 과잉 포아송 모형(Zero Inflated Poisson Model, ZIP)이며 확률밀도함수는 다음 식 (2)와 같다.

$$f(Y=y) = \begin{cases} w + (1-w)e^{-\lambda}, & y=0 \\ (1-w)\frac{e^{-\lambda}\lambda^y}{y!}, & y=1,2,\dots \end{cases} \quad (2)$$

여기서  $w(0 \leq w \leq 1)$ 는 0사건 발생 확률을 나타내며, 영 과잉 포아송 모형의 평균과 분산은 식 (3)과 같다.

$$\begin{aligned} E(Y=y) &= (1-w)\lambda \\ Var(Y=y) &= (1-w)(\lambda + w\lambda^2) \end{aligned} \quad (3)$$

모수  $\lambda$ 와  $w$ 의 정준연결함수는 로그함수와 로짓모형으로 표현되며 식 (4)와 같이 설명변수들의 선형결합으로 표현된다.

$$\begin{aligned} \log(\lambda) &= X\beta \\ \log\left(\frac{w}{1-w}\right) &= \text{logit}(w) = Z\alpha \end{aligned} \quad (4)$$

영 과잉 모형에서 0값을 제외한 가산형 자료를 분석

하기 위해 고려하는 모형 중 음이항 분포를 기반으로 제안된 모형이 영 과잉 음이항 모형(Zero Inflated Negative Binomial Model, ZINB)이며 확률밀도함수는 식 (5)와 같다.

이 때,  $0 \leq w \leq 1$ ,  $\lambda \geq 0$ 이며,  $\theta$ 는 과산포 모수로서 양수 값을 가진다. 영 과잉 음이항 모형의 평균과 분산은 식 (6)과 같다.

$$f(Y=y) = \begin{cases} w + (1-w)\frac{1}{(1+\theta\lambda)^{1/\theta}}, & y=0 \\ (1-w)\frac{\Gamma(1/\theta+y)}{y!\Gamma(1/\theta)}\left(\frac{\theta\lambda}{1+\theta\lambda}\right)^y\left(\frac{1}{1+\theta\lambda}\right)^{1/\theta}, & y=1,2,\dots \end{cases} \quad (5)$$

$$\begin{aligned} E(Y=y) &= (1-w)\lambda \\ Var(Y=y) &= (1-w)(\lambda + \theta\lambda^2 + w\lambda^2) \end{aligned} \quad (6)$$

2. 허들 모형

0이 팽창된 가산자료를 분석하는 방법으로 제안된 허들 모형은  $p'$ 의 발생확률을 가지는 0값과  $1-p'$ 의 발생확률을 가지는 0값을 제외한 1 이상의 값으로 나누어 분석하는 모형으로 이에 대한 확률은 식(7)과 같으며 확률밀도함수는 식 (8)과 같다. 영 과잉 모형과 달리 허들모형은 0의 값과 0의 값을 제외한 절단된(truncated) 값으로 구분되어 있는 것을 알 수 있다.

$$Y \sim \begin{cases} 0, & p=p' \\ f(Y), & p=1-p' \end{cases} \quad (7)$$

$$f(Y=y) = \begin{cases} p', & y=0 \\ (1-p')\frac{f(y)}{1-f(0)}, & y=1,2,\dots \end{cases} \quad (8)$$

허들 포아송 모형의 분포는 0값과 0을 제외한 1 이상의 값으로 나누어 나타내며 식 (8)에서 포아송 분포를 혼합하여 식(9)와 같이 표현된다.

$$f(Y=y) = \begin{cases} p', & y=0 \\ (1-p')\frac{e^{-\lambda}\lambda^y}{(1-e^{-\lambda})y!}, & y=1,2,\dots \end{cases} \quad (9)$$

이 때,  $0 \leq p' \leq 1$ ,  $\lambda \geq 0$ 이며,  $\lambda$ 와  $p'$ 의 정준연결함수는 영 과잉 모형의 형태와 같다.

허들 음이항 모형의 분포는 다음 식 (10)과 같다.

$$f(Y=y) = \begin{cases} p' & , y=0 \\ (1-p') \frac{\frac{\Gamma(\theta^{-1}+y)}{\Gamma(\theta^{-1})\Gamma(y+1)} \left(\frac{\theta\lambda}{1+\theta\lambda}\right)^y \left(\frac{1}{1+\theta\lambda}\right)^{1/\theta}}{1-\left(\frac{1}{1+\theta\lambda}\right)^{1/\theta}} & , y=1,2,\dots \end{cases} \quad (10)$$

### 3. 공간모형

공간효과를 나타내는  $\delta_i$ 는 조건부분포를 따르며 그 분포는 다음 식 (11)과 같다.

$$\delta_i | \delta_{-i} \sim N\left(\mu_i + \sum_{j=1}^n c_{ij}(\delta_j - \mu_j), \tau_i^2\right) \quad (11)$$

$\{Z(A_i) : A_i \in (A_1, \dots, A_n)\}$ 이 격자  $D$ 를 형성하는 정규화률과정(Gaussian random process)을 따르며  $Z(A_i)$ 가 지역  $A_i$ 에서의 관측된 값이라고 하고  $\{A_1, \dots, A_n\}$ 이 격자  $D$ 를 분할한다고 가정한다. 그러면 조건부 자기회귀모형은 식 (12)와 같이 표현된다.(Jung Soon, Choi *et. al.*, 2010)

$$Z(A_i) | Z(A_{-i}) \sim N\left(\mu_i + \sum_{j=1}^n c_{ij}(Z(A_j) - \mu_j), \tau_i^2\right) \quad (12)$$

여기서  $Z(A_{-i}) = \{Z(A_j) : j \neq i\}$ 를 뜻하며  $E(Z(A_i)) = \mu_i$ ,  $\tau_i^2$ 는 조건부 분산을 나타낸다. 이를  $n$ 이 유한인 경우 행렬  $C = (c_{ij})$ ,  $T = \text{diag}\{\tau_1^2, \dots, \tau_n^2\}$ 을 나타내며, 인수분해정리를 통해 식 (13)과 같이 표현된다.

$$Z \sim N(\mu, (I_n - C)^{-1}T) \quad (13)$$

행렬  $(I_n - C)^{-1}T$ 는 대칭이며 양정치행렬(Positive definite matrix)이다. 행렬  $C$ 는 가중행렬  $W$ 를 통해 변환된 행렬이며 일반적으로 지역  $i$ 와  $j$ 가 인접하면  $w_{ij} = 1$ 이고 인접하지 않으면  $w_{ij} = 0$ 으로 나타낸다.  $c_{ij} = w_{ij}/w_{i+}$ 로 정의하고,  $\tau_i^2 = \sigma_c^2/w_{i+}$ 로 가정하며, 여기서  $\sigma_c^2$ 는 공간 상관의 평활성을 나타내는 모수를 나타낸다.

### 4. 공간 영 과잉 모형

공간상관성이 존재하는 경우 공간영향력을 반영하여 공간모형을 통해 분석을 진행하여야 한다. 본 연구

에서 활용한 영 과잉 모형은 영이 과잉된 자료에서 공간적 영향력이 반영된 공간자료로 확장하여 공간 영 과잉 모형으로 나타낼 수 있으며 일반 영 과잉 모형과 동일하게 공간 영 과잉 포아송 모형과 공간 영 과잉 음이항 모형으로 각각 생각해 볼 수 있다.

공간 영 과잉 모형의 발생확률은 식 (1)과 동일하며 이에 따른 확률밀도함수도 식 (2)와 동일하다. 공간 영 과잉 모형에 포아송 분포와 음이항 분포를 각각 반영한 모형에서 모수  $\lambda_i$ 와  $w$ 의 정준연결함수는 0값과 건수값으로 나누어 표현되며 로그연결과 로짓연결 함수들이 설명변수들의 선형결합으로 표현된다. 이 때 공간 영향력인  $\delta_i$ 를 반영하여 나타내면 공간 영 과잉 모형은 다음 (14)와 같이 표현된다.

$$\begin{aligned} \log(\lambda_i) &= X\beta + \delta_i \\ \log\left(\frac{w}{1-w}\right) &= \text{logit}(w) = Z\alpha + \delta_i \end{aligned} \quad (14)$$

$\delta_i$ 는  $i$ 지역에서 공간영향력을 나타내며, 임의효과인 조건부자기회귀모형으로 표현된다.

허들 모형도 영 과잉 모형과 같이 영이 과잉된 자료에서 공간적 영향력이 반영된 공간자료로 확장하여 공간 허들 모형으로 나타낼 수 있다. 일반 허들 모형과 동일하게 공간 허들 포아송 모형과 공간 허들 음이항 모형으로 나타낼 수 있다.

### 5. 모란지수

본 연구에서는 I-5번 고속도로 하행선 277개 구간에 대한 교통사고 발생건수 자료를 활용하여 교통사고 모형을 개발하였고 교통사고 발생건수에 영향을 미치는 다양한 요인들의 영향력을 파악하였다. 본 연구의 목적에 따라 분석하기 전 자료가 가지고 있는 공간상관성을 진단하여 공간상관성이 존재하지 않는다고 판단되면

독립이라고 가정하여 분석을 진행하고 공간상관성이 존재한다고 판단되면 그 영향력을 반영한 공간모형을 통해 분석하게 된다. 따라서 분석에 앞서 자료의 공간상관성에 대한 유의성 검정이 먼저 진행되어야 한다. 이 때, 공간상관성을 진단하는 척도로 Moran's I가 주로 사용되고 있다.

Moran's I는 공간상관성을 진단하기 위해 일반적으로 쓰이는 척도이며 이는 -1부터 1까지 값을 가진다. -1에 가까울수록 음의 공간상관성을 가지며 1에 가까울수록 양의 공간상관성을 가진다. Moran's I는 하나의 변수에 대해서 해당 지역과 그 주변지역의 값을 이용하여 계산하며 그 계산식은 다음 식 (15)와 같다.

$$I = \frac{N \sum_i \sum_j w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_i \sum_j w_{ij}) \sum_i (Y_i - \bar{Y})^2} \quad (15)$$

여기서,  $N$ 은 관측치의 수,  $Y_i$ 와  $Y_j$ 는 각 지역의 관측치,  $\bar{Y}$ 는 관측 값들의 평균값이고  $w_{ij}$ 는 열의 합을 1로 하는 공간가중치 행렬( $W$ )의  $(i, j)$ 번째 원소값이다.

Moran's I의 유의성 검정은 정규분포를 기반으로 하며, 그에 대한 검정통계량  $Z$ 의 계산식은 다음 식 (16)과 같다(Cliff & Ord, 1972).

$$Z = \frac{I - E(I)}{S.E(I)} \quad (16)$$

여기서  $E(I)$ 는 Moran's I의 평균이고  $S.E(I)$ 는 Moran's I의 표준오차이다.

유의성 검정에 대한 귀무가설은 공간상관성이 존재하지 않음을 뜻하고 대립가설은 공간상관성이 존재함을 의미한다. 이에 따른 가설 식은 다음 식 (17)과 같다.

$$H_0: I=0 \quad vs \quad H_1 = I > 0 \quad (17)$$

검정통계량  $Z$ 가 유의수준 5%하에서 임계치 1.96보다 크면 공간상관성이 존재한다는 근거가 된다.

#### IV. 분석 결과

##### 1. 공간상관성 확인

먼저 1999년부터 2007년까지 9개년도의 교통사고 자료를 이용하여 공간상관성의 여부를 진단하기 위해 Moran's I 통계량을 계산하였고 그 결과는 <Table 1>과 같다. 9개년도 모두에서 P-값이 <.0001로 매우 작아 유의수준 5%에서 통계적으로 유의하기에 공간상관성이 존재함을 알 수 있다. 또한 Moran's I 통계량이 양의 값을 가지기 때문에 양의 상관성이 존재하여 공간적 상관성을 반영하지 않은 모형보다 공간적 상관성을 반영한 공간모형이 더 적합함을 알 수 있다.

##### 2. 모형비교

본 연구에서 활용한 미국 교통사고 경상 자료에 0값이 과대하게 존재함을 확인하였고 이를 반영한 영과잉 모형을 활용하여 적합하였으며 공간상관성이 존재함을 확인하였으므로 공간효과를 반영한 공간모형이 더 우수함

Table 1. Moran's I for 9 years from 1999 to 2007

Year	Moran's I	P-value
1999	0.3057	<.0001
2000	0.3277	<.0001
2001	0.3354	<.0001
2002	0.3295	<.0001
2003	0.4430	<.0001
2004	0.3910	<.0001
2005	0.4647	<.0001
2006	0.4361	<.0001
2007	0.4071	<.0001

을 확인하고자 한다. 본 연구에서 사용한 모형비교 척도 중 베이저안 기법의 모형 비교 척도에 관해서는 DIC(Deviance information criterion)를 사용하였으며 식 (18)과 같이 표현한다(Spiegelhalter, *et. al.*, 2002).

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \quad (18)$$

여기서,  $p_D = \bar{D} - D(\bar{\theta})$ 이며,  $D(\bar{\theta})$ 은 로그가능도의 함수로 이루어진 이탈도(Deviance)를 의미하고  $\bar{\theta}$ 은 모수  $\theta$ 의 사후평균을 의미한다.  $\bar{D}$ 는 이탈도의 사후평균,  $p_D$ 는 모형을 이루는 효과적인 모수의 개수를 나타낸다. DIC는 모형의 적합성과 복잡성을 동시에 고려할 수 있는 척도로서 AIC(Akaike Information Criteria)와 비슷하며, DIC의 값이 작은 모형일수록 적합한 모형이라 할 수 있다.

〈Table 2〉는 1999년부터 2007년까지 미국 교통사고 경상 자료를 영 과잉 모형과 허들 모형에 공간효과를 반영하지 않은 경우와 공간효과를 반영한 경우를 비교하기 위하여 DIC값을 구한 결과이다. 2002년과 2004년을 제외한 7개년도에서 공간효과를 반영한 공간모형의 DIC값이 작으므로 공간모형이 더 우수하다는 것을 확인 할 수 있었다. 1999년, 2001년, 2003년, 2005년은 공간 허들 포아송 모형이 가장 우수하다는 것을 알 수 있었으며 2006년, 2007년은 공간 영 과잉 포아송 모형이 가장 우수하다는 것을 알 수 있었다. 2000년은 공간 영 과잉 포아송 모형과 공간 허들 포아송 모형이 우수하며 2006년은 공간 영 과잉 포아송 모형과 공간

영 과잉 음이항 모형이 우수함을 알 수 있다. 따라서 자료에 공간상관성이 존재하는 경우 공간영향력을 반영한 모형이 그렇지 않은 모형보다 더 적합한 모형이라 판단할 수 있다.

### 3. 교통사고 모형 추정 결과

모란지수를 통하여 공간상관성이 존재함을 확인하였고 〈Table 2〉를 통해 공간상관성을 반영한 공간모형의 우수함 또한 확인하였다. 본 연구에서는 공간 영 과잉 포아송 모형, 공간 영 과잉 음이항 모형, 공간 허들 포아송 모형, 공간 허들 음이항 모형을 적합하였으며 그중 공간 영 과잉 포아송 모형, 공간 허들 포아송 모형이 우수함을 확인하였다.

본 연구에서 제안한 모형의 모수추정 방법으로 베이저안 기법을 이용하였다. 각 모수에 대한 사전분포(Prior Distribution)는 다음 식 (19)와 같다.

$$\begin{aligned} \alpha_j &\sim Normal(0,1), j = 0,1,2,\dots,k \\ \beta_j &\sim Normal(0,1), j = 0,1,2,\dots,k \\ \tau &\sim \Gamma(0.1,0.1) \end{aligned} \quad (19)$$

모수에 대한 추론은 사후분포를 통해 이루어지며 모형의 모수를 추정하기 위해 마코프 연쇄 몬테카를로(MCMC; Markov Chain Monte Carlo)방법을 사용하였다. 모수추정은 세 개의 초기값(Initial Value)을 사용하여 30,000번 반복 시행 후 처음 500번을 제거한 나머지 결과로 모수 추정 값은 분석결과에 제시하였다(Choi, *et. al.*, 2010).

Table 2. DIC of Zero Inflated Model and Spatial Zero Inflated Model for 9 years from 1999 to 2007

Year	Wtiout Spatial Effects				With Spatial Effects				
	ZIP	ZINB	H-P	H-NB	ZIP	ZINB	H-P	H-NB	
1999	DIC	5540790	5540770	5540780	5540770	5540750	5540760	5540740	5540760
2000	DIC	5540780	5540760	5540790	5540770	5540740	5540750	5540740	5540760
2001	DIC	5540780	5540770	5540770	5540770	5540760	5540770	5540750	5540760
2002	DIC	5540740	5540740	5540700	5540750	5540730	5540730	5540740	5540750
2003	DIC	5540700	5540700	5540700	5540700	5540700	5540700	5540690	5540700
2004	DIC	5540690	5540690	5540700	5540700	5540700	5540700	5540700	5540710
2005	DIC	5540710	5540690	5540700	5540690	5540680	5540680	5540670	5540690
2006	DIC	5540720	5540700	5540720	5540710	5540680	5540680	5540680	5540690
2007	DIC	5540650	5540650	5540660	5540660	5540630	5540640	5540640	5540650

Table 3. Estimate and Confidence Interval of Poisson Zero Inflated Model for 9 years from 1999 to 2007

Year	Variable	Esimate	std	2.50%	97.50%
1999	intercept	-0.1343	0.1451	-0.4217	0.1450
	LENGTH	1.0170	0.1669	0.6888	1.3470
	ADT	0.4188	0.2239	-0.0223	0.8565
	LSHW	-0.0737	0.0832	-0.2376	0.0890
	RSHW	0.0231	0.0764	-0.1237	0.1758
	NLN	0.0826	0.1639	-0.2463	0.3980
	RURAL	-0.5972	0.1833	-0.9553	-0.2377
	URBAN	-0.4009	0.1478	-0.6892	-0.1100
2000	INTORNOT	0.0707	0.2626	-0.4335	0.5974
	intercept	-0.1537	0.1619	-0.4744	0.1600
	LENGTH	1.2950	0.1678	0.9669	1.6260
	ADT	0.3404	0.2343	-0.1226	0.7971
	LSHW	-0.1008	0.0795	-0.2555	0.0571
	RSHW	-0.0708	0.0703	-0.2080	0.0681
	NLN	-0.0265	0.1549	-0.3315	0.2760
	RURAL	-0.9088	0.1768	-1.2530	-0.5587
2001	URBAN	-0.6399	0.1376	-0.9087	-0.3683
	INTORNOT	0.5528	0.2479	0.0661	1.0380
	intercept	-0.0862	0.1574	-0.3929	0.2239
	LENGTH	0.9825	0.1636	0.6635	1.3040
	ADT	0.4150	0.2056	0.0094	0.8225
	LSHW	-0.0371	0.0777	-0.1901	0.1146
	RSHW	-0.1125	0.0627	-0.2359	0.0097
	NLN	0.2693	0.1468	-0.0211	0.5547
2002	RURAL	-0.5743	0.1748	-0.9147	-0.2314
	URBAN	-0.3791	0.1358	-0.6421	-0.1124
	INTORNOT	0.2314	0.2397	-0.2411	0.6966
	intercept	-0.1015	0.1670	-0.4320	0.2210
	LENGTH	0.9459	0.1666	0.6217	1.2750
	ADT	0.3338	0.2229	-0.1085	0.7637
	LSHW	-0.0601	0.0741	-0.2057	0.0853
	RSHW	-0.1123	0.0700	-0.2500	0.0241
2003	NLN	0.2232	0.1455	-0.0618	0.5113
	RURAL	-0.5423	0.1737	-0.8841	-0.2033
	URBAN	-0.2911	0.1352	-0.5567	-0.0282
	INTORNOT	0.1659	0.2362	-0.2928	0.6314
	intercept	-0.2280	0.1837	-0.5863	0.1349
	LENGTH	1.0320	0.1801	0.6733	1.3810
	ADT	0.4716	0.2115	0.0600	0.8874
	LSHW	-0.0348	0.0826	-0.1946	0.1283
2004	RSHW	0.0143	0.0785	-0.1380	0.1692
	NLN	0.1632	0.1473	-0.1233	0.4530
	RURAL	-0.6446	0.1891	-1.0130	-0.2693
	URBAN	-0.3992	0.1461	-0.6843	-0.1084
	INTORNOT	0.3668	0.2830	-0.1912	0.9238
	intercept	-0.2280	0.1837	-0.5863	0.1349
	LENGTH	1.0320	0.1801	0.6733	1.3810
	ADT	0.4716	0.2115	0.0600	0.8874
2005	LSHW	-0.0348	0.0826	-0.1946	0.1283
	RSHW	0.0143	0.0785	-0.1380	0.1692
	NLN	0.1632	0.1473	-0.1233	0.4530
	RURAL	-0.6446	0.1891	-1.0130	-0.2693
	URBAN	-0.3992	0.1461	-0.6843	-0.1084
	INTORNOT	0.3668	0.2830	-0.1912	0.9238
	intercept	-0.2280	0.1837	-0.5863	0.1349
	LENGTH	1.0320	0.1801	0.6733	1.3810

2005	intercept	-0.3437	0.1819	-0.6983	0.0147
	LENGTH	0.9038	0.1950	0.5153	1.2830
	ADT	0.4410	0.2194	0.0214	0.8757
	LSHW	-0.0431	0.0884	-0.2178	0.1299
	RSHW	-0.0489	0.0764	-0.1977	0.1024
	NLN	0.0566	0.1659	-0.2648	0.3895
	RURAL	-0.3933	0.2026	-0.7863	0.0122
	URBAN	-0.2573	0.1575	-0.5629	0.0569
	INTORNOT	0.1790	0.2685	-0.3499	0.7075
2006	intercept	-0.3930	0.1838	-0.7582	-0.0390
	LENGTH	0.7547	0.1943	0.3679	1.1300
	ADT	0.8260	0.2596	0.3227	1.3450
	LSHW	-0.0479	0.0831	-0.2092	0.1167
	RSHW	-0.1249	0.0732	-0.2711	0.0169
	NLN	0.3232	0.1565	0.0135	0.6311
	RURAL	-0.3357	0.2049	-0.7282	0.0735
	URBAN	-0.3089	0.1606	-0.6162	0.0106
	INTORNOT	0.3050	0.2835	-0.2588	0.8541
2007	intercept	-0.2362	0.1853	-0.6026	0.1203
	LENGTH	0.6199	0.2172	0.1887	1.0380
	ADT	0.2599	0.2327	-0.1904	0.7199
	LSHW	0.0968	0.0916	-0.0853	0.2761
	RSHW	0.0279	0.0861	-0.1371	0.2017
	NLN	-0.2231	0.1806	-0.5803	0.1304
	RURAL	-0.2968	0.2292	-0.7398	0.1571
	URBAN	-0.0912	0.1738	-0.4283	0.2536
	INTORNOT	-0.0516	0.2865	-0.6149	0.5080

※ LENGTH : 도로길이, ADT : 일 교통량, LSHW : 왼쪽 길어깨, RSHW : 오른쪽 길어깨, NLN : 차선 수, RURAL : 지방에 속하는 구간과 나들목과의 거리, URBAN : 도심에 속하는 구간과 나들목과의 거리, INTORNOT : 나들목의 존재 여부

본 연구에서 최적모형이라 판단된 모형 중 <Table 3>은 공간 영 과잉 포아송 모형의 모수 추정 및 신뢰구간 결과이다. <Table 3>을 통해 9개년도 모두에서 도로 길이(LENGTH)가 교통사고 발생건수에 유의한 영향을 미친다는 것을 알 수 있으며 2001년과 2003년부터 2006년까지 5개년도에서 교통량(ADT)이 교통사고 발생건수에 유의한 영향을 미친다는 것을 알 수 있다. 또한 RURAL과 URBAN은 1999년부터 2004년까지 6개년도에서 교통사고 발생건수에 유의한 영향을 미친다는 것을 알 수 있다. 이를 통해 도로길이가 교통량의 모수 추정량이 양의 값을 가지므로 도로길이가 교통량이 늘어나면 교통사고 발생건수가 증가한다고 판단 할 수 있으며 도심과 지방 모두에서 구간의 중심과 인터체인지까지의 거리가 멀어질수록 교통사고 발생건수가 감소한다고 판단 할 수 있다.

## V. 결론

본 연구에서는 미국 고속도로 277개 구간의 교통사고 자료를 활용하여 교통사고 모형을 개발하였다. 교통사고 자료 중 경상 자료에 과대하게 포함된 0값을 확인하였으며 이를 반영하기 위해 영 과잉 모형을 활용하였다. 또한 고속도로의 특성상 사고 발생 지점의 공간적 영향력을 받을 것으로 예상되어 각 구간의 공간영향력을 확인하였으며 이를 반영한 공간 영 과잉 모형을 통해 최적의 교통사고 모형을 개발하였다.

본 연구에서 제시한 공간 영 과잉 모형의 추정 결과 도로길이가 길고 교통량이 많을수록 교통사고 발생건수가 증가하며 도심과 지방 모두 각 구간과 인터체인지와 거리가 멀수록 교통사고 발생건수가 감소한다는 것을 알 수 있었다.



이러한 분석 결과 고속도로 교통사고 예방을 위해 적합한 교통사고 모형이 필요하며 이를 통한 고속도로 교통안전정책의 수립 및 관리체계 확립이 이루어져야 할 것이다.

본 연구에서 활용한 자료는 각 구간별 공간영향력뿐 아니라 1999년부터 2007년에 해당하는 시간영향력도 존재한다는 것을 알 수 있다. 이를 반영하여 기존 공간모형에서 확장한 시공간모형(Spatial-Temporal Model)을 고려한다면 보다 우수한 모형이 될 것이라 생각된다. 또한, 본 연구에서 제안한 공간모형을 다양한 건수자료에 활용할 수 있다. 예를 들어 전염병 발생건수와 같은 질병자료는 더 많은 공간영향력이 존재하므로 이를 활용하여 다양한 예측 및 보다 정확한 질병의 영향요인을 확립 할 수 있을 것이라 생각된다.

## 감사의 글

이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5B6914198).

## References

- Black, W. R. and I. Thomas. 1998. Accidents on Belgium's Motorways: a Network Autocorrelation Analysis. *Journal of Transport Geography*. 6(1): 23-31.
- Choi, Jung Soon, Man Sik Park, Yu Bok Won, Hag Yeol Kim, and Tae Young Heo. 2010. Analysis of Total Crime Count Data Based on Spatial Association Structure. *The Korean Journal of Applied Statistics*. 23(2): 335-344.
- Cliff, A. and J. Ord. 1972. Testing for Spatial Autocorrelation among Regression Residual. *Geographic Analysis*. 4: 267-284.
- Famoye, F., J. T. Wulu, and K. P. Singh. 2004. On the Generalized Poisson Regression Model with an Application to Accident Data. *Journal of Data Science*. 2: 287-295.
- Kim, Hag Yeol and Tae Young Heo. 2012. A Development of Traffic Accident Model Using Generalized Poisson Regression Model. *Journal of Korea Planning Association*. 47(1): 129-138.
- Park, Byoung Ho and Yun Hwan Kim. 2007. Analyzing the Relationships Between Geometric Structure and Traffic Accident of Intersection in the Case of Korea Expressway: Using the Multi-Regression, Poisson regression and Ridge Regression Model. *Journal of the Institute of Construction Technology*. 26(1): 43-58.
- Rui, Fang. 2013. Zero-Inflated Negative Binomial(ZINB) Regression Model for Over-Dispersed Count Data with Excess Zeros and Repeated Measures, an Application to Human Microbiota Sequence Data. Master's Thesis. University of Colorado.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and van der A. Linde. 2002. Bayesian Measures of Model Complexity and Fit(with discussion). *Journal of the Royal Statistical Society*. 64: 583-639.
- Wulu, J. T., K. P. Singh, F. Famoye, T. N. Thomas, and G. McGwin. 2002. Regression Analysis of Count Data. *Journal of the Indian Society of Agricultural Statistics*. 55: 220-231.
- Korean References Translated from the English*
- 김학열, 허태영. 2011. 일반화포아송회귀모형을 이용한 교통사고모형개발. 국토계획. 47(1): 129-138.
- 박병호, 김윤환. 2007. 고속도로 교차점의 기하구조와 교통사고의 관계분석: 다중회귀분석, 포아송 회귀분석, 능형회귀분석 모형을 중심으로. 건설기술연구소 논문집. 26(1): 43-58.
- 최정순, 박만식, 원유복, 김학열, 허태영. 2010. 공간적 연관구조를 고려한 총범죄 자료 분석. 응용통계연구. 23(2): 335-344.

Received: Jun. 1, 2018 / Revised: Jul. 17, 2018 / Accepted: Jul. 23, 2018

## 공간효과를 반영한 교통사고 모형 비교 및 응용

**국문초록** 본 연구에서는 교통사고 자료의 특성 및 공간효과를 반영한 교통사고 모형을 개발하였다. 미국 I-5 고속도로를 277개의 동일한 구간으로 구분하여 1999년부터 2007년까지 발생한 교통사고 자료를 활용하였다. 공간모형의 활용성 여부를 판단하기 위해 모란지수(Moran's I)를 통하여 구간별 공간상관성이 존재하는지 검정하였다. 교통사고 발생건수 자료에서 팽창된 0값을 확인하였으며 이러한 특성을 반영하기 위해 기존의 가산자료 모형이 아닌 영 과잉 모형(Zero Inflated Model)과 허들 모형(Hurdle Model)을 활용하였다. 본 연구에서는 공간효과를 반영한 공간 영 과잉 모형(Spatial Zero Inflated Model)과 공간 허들 모형(Spatial Hurdle Model)을 적합하였다. 분석 결과, DIC를 통해 팽창된 0의 특성을 반영한 공간 영 과잉 모형과 공간 허들 모형의 우수함을 확인하였다. 따라서 이를 최적의 교통사고 모형으로 제안하고 각 설명변수의 영향력을 통해 교통사고 발생에 영향을 주는 주요 요인들의 효과를 정확하게 확인할 수 있다.

주제어 : 교통사고, 공간상관성, 영 과잉 모형, 허들 모형

**Profiles** **Bo Mi Jeong** : She is the Ph. D. candidate in the Department of information and Statistics, Chungbuk National University. Her current research interests are environmental statistics and drug discovery(qhal0324@hanmail.net).

**Jae Kyeong Lee** : He received the B. S. and M. S. in Statistics from the Chungbuk National University. Now he is researcher in the Korea Institute of Science Technology Information(KISTI). His current research interests are development of business decision support model for start-up culture diffusion based on weblog and social big data(gngee@naver.com).

**Yong Un Ban** : He received his M. A. from Yonsei University and Ph. D. from University of Pennsylvania. He is a professor of the Department of Urban Engineering at Chungbuk National University. His research interests include sustainable development, urban spatial configuration, LID(Low Impact Development), EIP(Eco Industrial Park), climate change, and environmental justice. He has published 79 articles in journals and written 5 co-author books(bybyyu@chungbuk.ac.kr).

**Tae Young Heo** : He received the B. S. in Statistics from the Chungbuk National University and the M. S. and Ph. D. in Statistics from the North Carolina State University. Now he is a professor in department of Information and Statistics, Chungbuk National University. His current research interests include spatial modeling, transportation and environmental statistics based on various spatial methodologies(theo@cbnu.ac.kr).