복합포아송 모형 기반의 건물피해액 영향요인 탐색에 관한 연구

Analysis of Influential Factors of Building Damage from Natural Disasters Based on Compound Poisson Model

김택근* · 홍석수** · 허태영*** · 장선명**** · 김학열*****

Kim, Taekgeun*, Hong, Seoksu**, Heo, Tae-Young***, Jang, Sun Myung****, and Kim, Hag-Yeol*****

Abstract

The purpose of this study is to construct statistical models for building damage from natural disasters and analyze the regional pattern of the damage. Both a compound Poisson model (CPM) and a zero-inflated compound Poisson model (ZICPM) have been developed to explain the amount for building damage from natural disasters, while adopting disaster region, disaster duration, and disaster type as explanatory variables. After estimating the above two regression models with the mentioned variables, the ZICPM is verified to be more appropriate than the CPM for building damage data from 1994 to 2015. Results show that amount and probability of building damage for provincial area are much greater than those for the Seoul Metropolitan Area. Additionally, damage from typhoon and heavy rain is statistically proved to be greater than that from strong winds. Accordingly, detailed measures and strategies based on disaster region and type should be prepared to mitigate building damage.

Key words : Natural Disaster, Damage Amount, Compound Poisson, Zero-inflated Compound Poisson

요 지

본 연구는 자연재해로 인한 건물피해에 대한 통계모형을 구축하고 그 피해 패턴을 분석하는 데 목적이 있다. 건물피해액을 설명하기 위하여 재해발생지역, 재해발생기간, 재해종류를 변수로 도입하여 복합포아송(Compound Poisson) 회귀모형과 영과잉 복합포아송(Zero-inflated Compound Poisson)모형을 개발하였다. 1994년부터 2015년 동안 발생한 재해로부터의 건물피해자료와 재해지속기간, 재해지역, 재해종류 자료를 활용하여 모형을 추정한 결과, 영과잉 복합포아송 모형이 보다 적합한 것으로 나타났다. 지방 지역이 서울 지역보다 재해로 인한 피해가 더 크게 나타나고 건물피해 발생확률도 매우 높았으며, 강풍에 비해 태풍과 호우로 인한 피해액이 상대적으로 높았다. 따라서 재해의 발생지역과 재해의 종류에 따라 세부적인 대책을 세우는 것이 건물 피해를 줄이는 주요한 역할을 할 수 있음을 확인하였다.

핵심용어 : 자연재난, 피해액, 복합포아송, 영과잉 복합포아송

*정회원, 충북대학교 정보통계학과 석사 과정(E-mail: nan521@naver.com)

Member, Master Student, Department of Information & Statistics, Chungbuk National University

**충북대학교 정보통계학과 박사 과정(E-mail: blindsage10@gmail.com)

Ph.D. Student, Department of Information & Statistics, Chungbuk National University

***충북대학교 정보통계학과 교수(E-mail: theo@cbnu.ac.kr)

Professor, Department of Information & Statistics, Chungbuk National University

****충북대학교 정보통계학과 박사 수료(E-mail: jsm7653@naver.com)

Ph.D. Candidate, Department of Information & Statistics, Chungbuk National University

*****교신저자, 정회원, 서경대학교 도시공학과 교수(Tel: +82-2-940-7783, Fax: +82-2-940-7616, E-mail: hagkim@skuniv.ac.kr)

Corresponding Author, Member, Professor, Department of Urban Planning and Engineering, Seokyeong University

1. 서 론

전 세계적으로 기후변화가 초래한 기상이변으로 인해 태풍, 홍수 등과 같은 자연재해의 빈도 및 강도가 증가하고 있으며 이에 따른 자연재해의 피해도 증가하고 있는 추세이다(Kim et al., 2017). 국내의 경우도 세계적인 추세와 맞물려 지리적·기후적 특성으로 인해 태풍, 홍수, 장마 등의 자연재해 피해의 규모가 증가하고 있다.

2010년 1월 서울에서 73년 만에 기록적인 강설로 인해 매우 심각한 피해가 발생하였으며 최근인 2017년에 청주를 비롯한 중부내륙 지역에서 발생한 기록적인 폭우로 인하여 약 547억 원의 재산피해와 2,539명의 인명피해와 함께 도로, 농경지, 건물 침수 등의 물적 피해가 발생하였다(Bai, 2014; Song et al., 2016). 따라서 재해의 발생 가능성과 발생 위치, 발생 시기 등을 예측하기 위한 연구가 지속적으로 이뤄지고 있으며, 최근 구글(Google)에서 제안한 홍수 예측 관련 AI 기반 알고리즘 정확도가 90% 수준에 도달했을 정도로, 머지않은 미래에 보다 정확하게 자연재해를 예측할 수 있을 것으로 예상된다.

그러나 자연재해 발생이 예측 정확도가 높아진다 해도 재해 발생 및 이로 인한 재산과 인명 등의 피해를 완벽히 통제하기는 불가능하다. 이에 적절한 대응과 피해 최소화를 위해, 배수펌프, 제방 설치 등과 같은 구조적 대책과 재해 예·경보, 토지이용계획, 대피계획 수립 등과 같은 비구조적 대책의 수립이 필수적이다. 그러나 현재 재해 방지를 위한 다양한 물리적·구조적 대책은 환경단체와 주민들의 반대에 부딪혀 많은 어려움을 겪을 뿐만 아니라, 기후변화와 도시화에 따른 자연재해의 규모와 강도 상승으로 인해 구조적 대책의 한계점이 드러나고 있다(Kim et al., 2007).

이에 비구조적 대책에 관한 관심과 필요성이 높아지고 있으며, 비구조적 대책을 마련하기 위한 수단으로 주로 자연재해 피해발생의 영향인자를 살펴보고자 다양한 통계모형 연구가 시도되었다. 예를 들어, 회귀모형을 이용하여 강우량에 영향을 주는 인자를 확인하는 연구와 함께, 재해가 농축산 시설에 주는 영향 및 재해 피해액의 결정요인을 분석한 연구 등 유사 연구 등이 있다(Davis and Skaggs, 1992; Lee et al., 2016; Kim et al., 2018).

재해연보의 시설별 피해현황에 따르면 지난 10년간(2008년~2017년) 자연재해로 인한 전체 재산피해액 3조 5,324억 중 건물피해액은 1,729억 원으로 전체의 4.9% 정도를 차지하고 있다. 한편 건물 피해액은 재해가 발생하였음에도 건물피해가 발생하지 않아 건물피해액이 0으로 집계되는 이산형 변수와 건물피해가 발생할 경우의 연속형 변수가 혼합되어 있는 특성을 지니기에, 두 특성을 모두 반영할 수 있는 통계모형이 필요하다. 특히 0의 값이 과도하게 포함된 영과잉 자료의 특징을 적절히 반영할 수 있는 영과잉 모형이 고려되어야 한다.

기존의 대부분 영과잉 자료 관련 연구들은 포아송 분포 기반의 모형으로 영과잉 자료를 종속변수로 사용한 선형연구의 경우 주로 영과잉 모형과 포아송 모형을 결합한 영과잉 포아송 모형 또는 음이항 모형을 결합한 영과잉 음이항 모형을 통해 자료를 분석하였다(Kim and Lee, 2008; Chun, 2017). Hwang et al. (2016)의 연구는 정지케도 복합위성의 기상탐재체의 적외선 채널로부터 얻어진 밝기온도와 대류 지수가 강우강도와 어떠한 연관성을 갖는지 알아보기 위해 복합포아송 분포에 기반한 회귀모형을 제안하였으며, Kim et al. (2014)의 연구에서는 서울시 침수저감대책을 수립하기 위해 복합포아송 기반 모형을 활용하였다. 또한, Parveen et al. (2016)의 연구는 0의 값과 양수값이 혼재된 직업훈련비용 자료의 분석에 복합포아송 모형을 활용하여 복합포아송 모형의 활용성을 확인하였다. 한편 영과잉 특성을 반영한 영과잉 복합포아송 모형을 활용한 그 사례를 찾기가 힘들다.

따라서 본 연구에서는 영과잉의 자료 특성을 반영한 건물 피해액 자료를 기반으로 재해발생지역, 재해발생기간, 재해 종류 변수를 활용하여 복합포아송 모형과 영과잉 복합포아송 모형을 적용 후 비교분석을 통해 더 높은 활용성을 가진 모형을 추정함으로써 재해저감대책에 도움이 될 수 있는 설명력 높은 통계모형을 제시하고자 한다.

2. 연구방법

2.1 복합포아송 모형

변수 간의 관계를 알아보는 통계적 방법으로 일반 선형 모형(Generalized linear model, GLM)을 비롯하여 여러 모형이 널리 사용되고 있다. 예를 들어, 재해 분야에서 자연재해가 일어난 횟수를 종속변수로 할 때는 연결함수(link function)를 로그(log)로 취하는 포아송 선형회귀 모형을 기반으로 한다. 하지만 사건의 횟수가 아닌 각 사건에 따른 어떠한 양의 실숫값을 종속변수로 할 때는 포아송 회귀모형을 이용하기에는 적절하지가 않다. 이처럼, 분야별 변수가 가지는 특별한 성질로 인해 모형의 적합도가 현저하게 낮아지거나 모형을 적합 시키지 못하는 경우를 종종 겪는다(Smyth, 1989).

종속변수가 양의 값을 가지는 실수이고 셀 수 있는 사건이 발생함에 따라 종속변수가 얻어지는 경우에 복합포아송 분포(Compound Poisson distribution)를 따르는데, 위의 경우를 예를 들면, 건물피해는 자연재해로 건물에 피해가 일어났을 때 얻을 수 있는 값이며 항상 0 이상의 실수의 값을 가지기 때문에 복합포아송 분포를 이용한 모형을 사용하는 것이 적절할 것이다. 실제로 순보험료, 강우량, 교통사고에 따른 피해액 등을 종속변수로 가지는 경우 복합포아송 분포를 이용한 모형을 실제 응용연구에서 많이 사용하고 있다(Jørgensen and Paes de Souza, 1994).

먼저 복합포아송 모형을 포함하고 있는 지수산포모형(Exponential dispersion model)에 대해 간략한 설명은 다음과

같다. 지수산포모형은 Tweedie (1984)에 처음으로 제안되어 Tweedie 모형으로 불리기도 한다.

Jørgensen (1987)에 의해 연구된 지수산포를 이용한 모형은 여러 가지 통계적 분석에 많이 쓰이는 모형으로 일반적인 지수산포 분포의 확률밀도함수는 Eq. (1)과 같다.

$$f(y|\theta, \phi) = a(y, \phi) \exp\left(\frac{y\theta - \kappa(\theta)}{\phi}\right) \quad (1)$$

여기서 a 는 상수함수이며 κ 는 누율생성함수(cumulant generating function)이며, θ, ϕ 는 모수이다. 특히, ϕ 는 0 이상의 값을 가지는 산포모수(dispersion parameter)이다(Jørgensen, 1987).

Tweedie 분포는 누율생성함수의 고유한 성질을 이용하여 기댓값과 분산인 $E(y) = \kappa'(\theta)$ 와 $Var(y) = \phi\kappa''(\theta)$ 를 쉽게 구할 수 있으며, θ 와 μ 는 일대일 함수임을 통하여 $\kappa''(\theta)$ 를 μ 의 함수 $V(\mu)$ 로 표현할 수 있고 이를 멱분산함수(power variance function)이라고 명명한다(Barndorff-Nielsen, 1978; Kaas, 2005).

멱분산함수의 일반적인 형태는 $V(\mu) = \mu^p$ 로 나타내며, 이때 분산함수의 p 는 지시모수(index parameter)이다. 지수산포 모형에서 멱분산함수는 중요한 성질을 가지는데, 멱분산함수의 형태만 결정된다면 지수산포 모형의 확률분포를 알 수 있다(Jørgensen, 1997).

Tweedie 분포는 지시모수의 형태에 따라 분포가 달라지는 장점이 있다. 지시모수가 $0 < p < 1$ 인 경우에는 확률분포가 존재하지 않으며, $p = 0$ 인 경우에는 정규분포를 따르고, $p = 1$ 일 때는 포아송 분포, $p = 2$ 일 때는 감마분포를 따른다고 알려져 있다. 특히, $1 < p < 2$ 일 때는 복합포아송 분포를 따르며, 복합포아송 분포는 0에서 양의 확률(확률질량함수) 가지는 이산형 변수와 0 이상 양의 실숫값에서는 확률밀도함수를 가지는 연속형 변수가 결합한 분포로 알려져 있다.

그러나 지시모수 p 를 추정하는 데 있어서 복합포아송 분포의 확률밀도함수의 특성상 계산의 어려움으로 인해 다양한 모수 추정 방법들이 제안되었다. Cox and Reid (1987)은 프로파일 가능도(profile likelihood)방법을 제안하여 p 를 추정하였으며, Nelder and Pregibon (1987)은 확장된 준우도(extended quasi-likelihood) 방법을 기반으로 한 방법을 제안하였다. 최근 연구에서는 컴퓨팅의 발전을 통해 더욱 정확한 복합포아송 분포를 구할 수 있다(Dunn and Smyth, 2005, 2008).

복합포아송 모형을 적용하기 위해 본 연구의 관심변수인 건물피해액에 대해 모형식을 적용하면 다음과 같다. 주어진 기간에 대해서 확률변수 N 을 자연재해가 일어난 횟수로 정의한다. 또한 X_i 를 i 번째 자연재해가 발생했을 때의 건물피해액으로 정의하며, 이러한 확률변수들의 가정은 다음의 Eq. (2)와 같다.

$$N \sim Pois(\lambda), X_i \stackrel{iid}{\sim} \Gamma(\alpha, \gamma), N \perp X_i \quad (2)$$

여기서 확률변수 N 은 평균이 λ 인 포아송 분포($Pois$)를 따르며, 확률변수 X_i 는 감마분포(Γ)를 따르고, $N, X_1, X_2, \dots, X_N$ 은 서로 독립임을 가정한다. 이제 확률변수 Y 를 주어진 기간에 대한 누적 건물피해액으로 정의하면 다음의 Eq. (3)과 같다.

$$Y = \sum_{i=1}^N X_i \quad (3)$$

이때 확률변수 Y 를 평균이 λ 인 복합포아송 분포를 따른다고 한다. 즉, 건물피해액 Y 는 X_i 들의 임의의 누적합(random sum)으로 정의되며, 이때 $N = 0$ 인 경우에는 Y 의 값도 0으로 정의한다.

Eq. (4)의 관계식을 통해 건물피해액 Y 의 평균 $\mu = \lambda\alpha\gamma$, 분산 $\phi\mu^p = \lambda\alpha(\alpha+1)\gamma^2$ 을 얻을 수 있다.

$$\lambda = \frac{\mu^{2-p}}{\phi(2-p)}, \alpha = \frac{2-p}{p-1}, \quad (4)$$

$$\gamma = \phi(p-1)\mu^{p-1}$$

복합포아송 분포를 따르는 확률변수 Y 는 Tweedie 분포에서 언급한 바와 같이 이산형과 연속형이 혼합된 분포이므로 확률밀도함수를 다음의 Eq. (5)와 같이 표현할 수 있다.

$$f_{CPM}(y, n|\lambda, \alpha, \gamma) = f(y|n, \alpha, \gamma)f(n|\lambda) \quad (5)$$

$$= \begin{cases} e^{-\lambda}, & y=0 \\ \frac{e^{-y/\gamma} y^{n\alpha-1}}{\Gamma(n\alpha)\gamma^{n\alpha}} \frac{\lambda^n e^{-\lambda}}{n!}, & y>0 \end{cases}$$

2.2 영과잉 복합포아송 모형

전술한 복합포아송 모형(Compound Poisson Model)은 자료가 영(zero)을 과도하게 포함하고 있는 영과잉 자료일 때 자료의 성격상 그 분포를 따르지 못할 때가 있어 분석에 어려움이 있다. 이러한 단점을 보완하기 위하여 분포의 혼합 형태를 가지는 영과잉 모형이 제안되었으며 본 논문에서는 복합포아송 모형과 영과잉 모형을 결합한 영과잉 복합포아송 모형(Zero Inflated Compound Poisson Model)을 활용하여 분석을 시행하였다(Lambert, 1992). 영과잉 복합포아송 분포는 추정된 지시모수 p 가 $1 < p < 2$ 일 때 복합포아송 분포를 따르며 이와 다른 값으로 p 가 추정된 경우에는 다른 분포를 따르기 때문에 영과잉 복합포아송 분포를 적용할 수 없다. 영과잉 복합포아송 분포는 복합포아송 분포와 베르

누이 분포와의 혼합모형(mixed model)으로 확률변수 Y 는 영만 나타나는 상태의 확률값이 따로 정해진다. 정의하면 다음의 Eq. (6)과 같다.

$$Y \sim \begin{cases} 0, \\ CPoi(\lambda, \alpha, \gamma), \end{cases} \quad \begin{matrix} q, \\ 1-q \end{matrix} \quad (6)$$

Eq. (6)에서 $CPoi$ 는 복합포아송분포를 의미하며 $Y=0$ 은 두 가지 확률에서 모두 발생하며 확률밀도함수를 다음 Eq. (7)과 같이 표현할 수 있다.

$$f_{ZICPM}(y, n|\lambda, \alpha, \gamma) \quad (7)$$

$$= \begin{cases} q + (1-q)e^{-\lambda}, & y=0 \\ (1-q)f_{CPM}(y, n|\lambda, \alpha, \gamma), & y>0 \end{cases}$$

Eq. (7)에서의 $f_{ZICPM}(y, n|\lambda, \alpha, \gamma)$ 는 영과잉복합포아송 모형을 의미하며 $f_{CPM}(y, n|\lambda, \alpha, \gamma)$ 는 Eq. (5)의 복합포아송 분포에서의 확률밀도함수와 같다.

2.2.1 건물피해액의 영과잉 복합포아송 분포적용

앞 절에서 건물피해액과 같이 이산형 및 연속형 구조가 혼재된 자료를 분석하기 위해 복합포아송 모형을 사용하는 것이 적절하다. 또한 자료가 영(0)의 값을 과도하게 포함한 경우 영과잉 복합포아송 모형을 적용하는 것이 분석에 적절함을 함께 언급하였다.

Table 1은 본 연구에서 종속변수로 활용한 1994년부터 2015년까지 자연재해에 의한 건물피해액의 빈도표이다. Table 1에서와 같이 총 9,341건의 자연재해 중 건물피해가 발생하지 않은 경우는 4,799건(약 51.3%)으로 자료의 50% 이상이 0의 값을 가지는 영과잉 자료임을 확인할 수 있다.

따라서 복합포아송 분포를 적용하면 건물피해액이 0인 경우 확률을 부여할 수 있는 이산형 변수의 특징과 건물피해액의 값이 0이 아닌 양의 실숫값을 가질 때의 연속형 변수의 특징을 모두 설명할 수 있다는 장점을 가지고 있다.

Table 1. Number of Damage Occurrence and No Occurrence

Type	Total	# of Building Damage occurs (%)	# of Building Damage do not occurs (%)
Building Damage	9341	4542 (48.7)	4799 (51.3)

3. 자료

3.1 자료 설명 및 기초분석

본 분석에서 사용되는 자료는 재해연보를 통해 1994년부

터 2015년까지 한국 전 지역에서 발생한 자연재해에 따른 건물피해액을 중심으로 구축되었다. 종속변수로 사용될 건물피해액은 1994년부터 2015년까지의 자연재해로 인해 발생한 건물손실에 대한 국가지원금액으로써 행정안전부가 수집한 피해 정도를 반영하여 「자연재해 구호 및 복구 비용 부담기준 등에 관한 규정」에 따라 지급된 금액의 합계이다. 본 연구에서는 이를 종속변수의 데이터로 사용하였으며 이는 국내에서 피해 규모를 확인할 수 있는 유일한 공식적인 통계자료이다.

독립변수는 재해기간, 재해지역 그리고 재해종류의 변수로 구성된다. 이 변수들은 건물피해액과 직접적인 연관이 있다고 여겨지는 변수들이며 재해별 건물피해액 차이의 원인을 파악하기 위하여 선정되었다.

첫 번째로 재해기간 변수는 자연재해가 발생한 기간으로 짧게는 하루에서 길게는 22일까지 지속되었다. 자연재해는 평균적으로 5.8일, 약 6일 동안 지속되었으며 재해기간이 길어질수록 건물피해액이 증가하는 양의 상관관계가 존재할 것으로 예상할 수 있다. 두 번째로 재해지역 변수는 종속변수인 건물피해액의 지역별 차이를 확인하기 위하여 독립변수로 사용하였으며 Table 2와 같이 서울특별시와 6개 광역시(부산, 대구, 인천, 광주, 대전, 울산), 세종시, 경기도, 강원도, 충청북도, 충청남도, 전라북도, 전라남도, 경상남도, 제주도로 구분하였다.

마지막은 재해종류 변수이다. 이 변수는 재해유형으로 강풍, 대설, 태풍, 폭풍, 폭풍설, 폭풍우, 호우, 호우태풍, 호우폭풍으로 구분된다. 이 중 폭풍, 폭풍설, 폭풍우, 호우태풍, 호우폭풍은 2001년 2월 이후로 재해 발생 기록이 존재하지 않아 현재 발생하는 자연재해인 강풍, 대설, 태풍, 호우로 자료를 재분류하여 분석에 사용하였다.

Table 3은 자연재해유형별 재해의 원인과 재해기준과 관련 인자를 나타낸 것이다. 위 표에 나타난 자연재해별로 건물에 발생하는 피해액은 각기 다르며 Fig. 1은 연도별 재해로 인한 평균 건물피해액을 나타낸다. 지리적·기후적 특징으로 인해 장마로 인한 호우피해와 태풍에 의한 피해가 가장 크게 나타나는 것을 확인할 수 있으며 특히 2002년과 2003년에는 태풍 루사와 매미의 영향으로 인해 다른 기간보다 훨씬 많은 건물피해가 발생한 것을 확인할 수 있다. 태풍이 발생하지 않은 기간은 태풍이 발생한 기간보다는 건물피해가 적게 발생한다. 한편 호우로 인한 건물피해는 매년 꾸준히 발생하고 있으며 강풍과 폭설로 인한 피해도 지속해서 발생한다.

자연재해로 인한 피해는 재해종류(강풍, 대설, 태풍, 호우)뿐만 아니라 재해가 발생하는 지역에 따라서도 차이를 보인다. Table 4는 지역과 재해종류에 따른 재해발생과 재해피해액에 관한 기초통계를 나타낸 표이다. 0이 과도하게 포함된 자료의 특성상 재해가 발생한 모든 경우의 기초통계량과 재해발생 시 건물피해가 발생한 경우의 기초통계량을 따로 표기하였다.

Table 2. Independent and Dependent Variables

Variable Name	Variable Type	Variable Description
Building Damage	Mixed	Building Damage from natural disasters (compensation amount by government)
Disaster Duration	Continuous	Period of disaster
Disaster Area	Categorical	Seoul (SE), Busan (BS), Daegu (DG), Incheon (IC), Gwangju (GJ), Daejeon (DJ), Ulsan (US), Sejong (SJ), Gyeonggi (GG), Gangwon (GW), Chungbuk (CB), Chungnam (CN), Jeonbuk (JB), Jeonnam (JN), Gyeongbuk (GB), Gyeongnam (GN), Jeju (JJ)
Disaster Type	Categorical	Gale, Heavy snow, Typhoon, Heavy rain, Storm (Typhoon), Snowstorm (Heavy snow), rainstorm (Heavy rain), Typhoon storm (Typhoon), Heavy rain storm (Typhoon)

Table 3. Disaster Causes, Disaster Standards, and Related Factors

Disaster Type	Disaster Causes	Disaster Impact	Disaster Standards (weather warning)	Related Factors
Gale (GA)	Typhoon Gale	Gale	Land: Wind speed 14m/s or more, Instant wind speed 20m/s or more Wold: Wind speed 17m/s or more, Instant wind speed 24m/s or more	Wind speed Instant wind speed
Heavy snow (HS)	Heavy snow	Heavy snow	24 hours Fresh snow cover more than 5cm	Fresh snow cover
Typhoon (TP)	Typhoon, Gale, Wind wave, Heavy rain, Surge, Debris flow	River flooding or flooding	Weather warning standard of Gale, Wind wave, Heavy rain, Flood, and Surge	Flood area, depth, rate, wave Wind speed Wave height
Heavy rain (HR)	Rain Debris flow	River flooding or flooding	12 hours Rainfall more than 80mm	Flood area, depth, rate, wave

Fig. 1. Barplot of Yearly Building Damage Due to Natural Disasters

Table 4. Summary of Building Damage by Natural Disaster

Category		building damage with zero case			building damage without zero case		
		Average Damage	Number of Disaster	Average Period	Average Damage	Number of Damage	Average Period
Region	SE	18495.34	514	4.69	36423.78	253	4.94
	BS	21505.95	375	5.13	38587.23	166	5.90
	DG	18235.02	99	6.51	34061.64	46	8.16
	IC	15690.28	374	5.19	31892.19	190	6.19
	GJ	18296.83	184	6.06	34006.23	85	7.03
	DJ	6969.36	162	6.06	16851.30	95	7.13
	US	25290.96	129	7.21	50977.09	65	8.62
	SJ	74476.05	38	6.50	128640.50	16	8.13
	GG	23942.92	1416	5.15	49566.05	732	5.79
	GW	42785.84	776	6.51	85571.68	388	7.83
	CB	109941.70	517	5.47	238822.90	279	6.86
	CN	100922.20	770	6.01	208898.10	398	6.95
	JB	51553.07	734	6.33	111622.30	395	7.05
	JN	100778.70	1281	5.96	222582.00	701	7.02
	GB	54209.72	1002	6.42	105472.10	487	7.83
	GN	28893.24	897	5.92	60132.79	466	7.06
	JJ	28121.74	73	4.41	57024.64	37	5.61
Type	GA	2837.492	622	3.34	24512.78	72	3.52
	HS	8068.145	1373	4.88	78011.01	142	9.32
	TP	109745.00	2860	5.98	204342.9	1536	6.60
	HR	35666.65	4486	6.34	57306.81	2792	6.89

기초통계 결과 중 먼저 지역별 건물피해액의 차이를 확인하면 서울지역을 포함한 광역시 지역(부산, 대구, 인천 등)보다 지방 지역(충북, 충남, 경북 등)은 자연재해의 지속 기간은 크게 차이가 없으나 평균적인 건물피해액은 큰 것으로 확인되었다. 다음으로 재해종류별 기초통계량을 확인한 결과 가장 많이 발생한 재해는 호우이지만 가장 많은 건물피해를 끼친 재해는 태풍으로 확인되었다. 강풍과 대설재해는 다른 두 재해에 비해 발생빈도 대비 건물피해를 끼칠 확률은 낮으나 대설의 경우 건물피해가 발생할 시 호우보다 오히려 더 큰 건물피해를 끼친다. 또한, 대설은 모든 재해 중 가장 긴 재해 기간을 가지며 이는 강설 기간이 증가할수록 적설량이 늘어나는 대설재해의 특징에 의한 것으로 예상된다.

4. 분석 및 결과

복합포아송 모형과 영과잉 복합포아송 모형을 이용하여 건물피해액에 대하여 어떠한 변수가 유의하게 영향을 미치는지 알아보고자 한다. 분석에 앞서 종속변수인 건물피해액 Y의 값이 매우 크기 때문에 백만원 단위로 변환하였으며,

이는 복합포아송 모형이 Maximum Likelihood (ML) 추정방법을 기반으로 모수를 추정하기 때문에 모수 추정을 위한 계산 과정 중 계산이 수렴되지 않는 문제점을 보완하기 위함이다. 분석 및 모수 추정을 위하여 통계패키지 R을 사용하였으며 모수 추정결과, 복합포아송 모형의 p 와 ϕ 에 대한 모수 추정값은 각각 $\hat{p} = 1.8526$ 과 $\hat{\phi} = 6.1624$ 로 추정되었고, 영과잉 복합포아송의 p 와 ϕ 에 대한 모수 추정값은 각각 $\hat{p} = 1.99$ 과 $\hat{\phi} = 4.6359$ 로 추정되었다.

지시모수인 p 값이 1과 2 사이에 존재하기에 건물피해액에 대한 영향요인을 확인하기 위한 두 모형의 사용은 적합하다고 평가되었다. 한편 두 모형의 비교를 위한 판단 기준으로 AIC를 사용하였으며 AIC는 로그 가능도함수와 모형의 모수의 수를 더한 함수로 모수의 수가 많을수록 적합 결과에 페널티를 주는 방법이다. 따라서 AIC의 크기가 작을수록 제안된 모형이 자료에 잘 적합된 것으로 판단할 수 있다. AIC와 로그 가능도함수의 식은 Eq. (8)과 같다(Akaike, 1973).

$$\begin{aligned}
 -2l(\theta) &= -2\log\left(\prod_{i=1}^n f_i(y_i|\theta)\right) \\
 AIC &= -2l(\theta) + 2|\theta|
 \end{aligned} \tag{8}$$

건물피해액을 종속변수로 하여 모든 변수를 두 모형에 적합 시켜 얻어진 결과는 Table 5에 제시되어있다. Table 5 하단의 복합포아송 모형과 영과잉 복합포아송 모형의 추정결과 AIC를 비교했을 때 복합포아송 모형의 AIC는 -18808.69이며 영과잉 복합포아송 모형의 AIC는 -20465.57이다. 이는 0이 과도하게 포함된 건물피해액과 같은 자료에는 영과잉 복합포아송 모형을 적합한 것이 통계적으로 더 유의하다는 것을 의미한다.

둘째로 영과잉 복합포아송 모형의 결과인 Table 5의 모형 추정 결과에서는 건물피해액이 0이 아닌 자료 즉 건물피해액이 존재할 때의 자료에 대한 영과잉 복합포아송 모형추정결과와 건물피해 발생 여부에 대한 영과잉 복합포아송 모형 추정 결과가 함께 제시되어있다.

건물피해 발생 여부에 대한 영과잉 복합포아송 모형추정결과 중 유의수준 0.05에서 유의한 변수의 회귀계수를 살펴

보면 재해기간(Duration)이 1일 증가 시 건물피해가 발생하지 않을 확률은 평균적으로 93.32% ($e^{-0.0691} = 0.9332$)로 감소한다. 즉 재해기간이 증가할수록 건물피해가 발생할 확률이 증가하는 것으로 분석되었다. 또한, 재해지역의 기준 수준인 서울지역과 유의한 차이를 보인 지역은 대전지역으로 건물피해가 발생하지 않을 확률이 서울지역과 비교하여 47.49% ($e^{0.3886} = 1.4749$) 증가하며 대전지역을 제외한 지역은 건물피해가 발생할 확률이 서울지역과 차이가 나지 않는다. 이는 재해로 인한 건물피해 발생 여부는 지역과는 연관성이 크지 않은 것을 의미한다. 또 재해종류의 기준 수준인 강풍과 유의한 차이를 보이는 재해는 태풍, 호우로 강풍과 비교하여 건물피해가 발생하지 않을 확률이 각각 12.92% ($e^{-2.046} = 0.1292$), 9.24% ($e^{-2.3811} = 0.0924$)로 감소한다. 즉 태풍과 호우 발생 시 강풍과 비교하여 건물피해가 발생할 확률이 급격하게 증가하므로 이에 따른 대비가 필요하다.

Table 5. Estimated Compound Poisson Regression Coefficient and Zero Inflated Compound Poisson Regression Coefficient

		Compound Poisson			Zero inflated Compound Poisson					
		Estimate	Std. Error	Pr (> z)	Logit			Compound Poisson		
					Estimate	Std. Error	Pr (> z)	Estimate	Std. Error	Pr (> z)
(Intercept)		-6.6155	0.3949	<0.0001 ***	2.2820	0.1550	<0.0001 ***	-4.3090	0.2910	<0.0001 ***
Duration		0.0770	0.0125	<0.0001 ***	-0.0691	0.0050	<0.0001 ***	0.0384	0.0057	<0.0001 ***
Disaster region	SE									
	BS	0.1483	0.4295	0.7298	-0.1642	0.1466	0.2627	0.2016	0.2036	0.3220
	DG	-0.6211	0.7304	0.3952	-0.1606	0.2462	0.5140	-0.5196	0.3320	0.1175
	IC	-0.1234	0.4365	0.7773	0.1013	0.1463	0.4888	-0.0368	0.2114	0.8618
	GJ	-0.2018	0.5521	0.7147	-0.0291	0.1864	0.8760	-0.1464	0.2595	0.5726
	DJ	-0.6995	0.5977	0.2419	0.3886	0.1975	0.0492 *	-0.7711	0.3037	0.0111 *
	US	0.2193	0.6119	0.7201	0.1676	0.2154	0.4364	0.5076	0.3053	0.0964
	SJ	1.5249	0.9599	0.1122	-0.4830	0.3888	0.2142	1.1204	0.4850	0.0209 *
	GG	0.0866	0.3301	0.7931	0.0291	0.1116	0.7947	0.2099	0.1597	0.1886
	GW	0.7410	0.3579	0.0385 *	-0.0678	0.1255	0.5892	0.8238	0.1772	<0.0001 ***
	CB	1.3677	0.3831	0.0004 ***	-0.0555	0.1385	0.6886	1.5345	0.1979	<0.0001 ***
	CN	1.3506	0.3534	0.0001 ***	-0.1059	0.1258	0.4001	1.3795	0.1778	<0.0001 ***
	JB	1.1231	0.3572	0.0017 **	0.1477	0.1255	0.2392	1.2635	0.1803	<0.0001 ***
	JN	1.2820	0.3265	0.0001 ***	0.2161	0.1134	0.0567	1.4214	0.1650	<0.0001 ***
	GB	0.9551	0.3402	0.005 ***	-0.1441	0.1191	0.2263	1.0121	0.1698	<0.0001 ***
	GN	0.2130	0.3524	0.5455	0.0040	0.1207	0.9737	0.1726	0.1742	0.3220
	JJ	-0.4666	0.7989	0.5592	-0.0424	0.2658	0.8732	-0.2414	0.3914	0.5374
Disaster type	GA									
	HS	-0.0016	0.3472	0.9962	0.2525	0.1552	0.1038	0.3414	0.3267	0.2959
	TP	3.1342	0.3105	<0.0001 ***	-2.0460	0.1323	<0.0001 ***	1.5513	0.2701	<0.0001 ***
	HR	1.9345	0.3062	<0.0001 ***	-2.3811	0.1301	<0.0001 ***	0.2848	0.2676	0.2872
Model AIC		-18808.69			-20465.57					

(*:p<0.05, **:p<0.01, ***:p<0.005, unit = million)

다음으로 건물피해액이 0이 아닌 값을 가질 때의 영과잉 복합포아송 모형 추정 결과 중 유의수준 0.05에서 유의한 변수의 회귀계수를 살펴보면, 건물피해액은 재해기간이 1일 증가 시 평균적으로 3.91% ($e^{0.0384} = 1.0391$) 증가하는 것으로 분석되었다. 또한, 재해지역의 기준 수준인 서울지역에 비해 대전지역은 46.25% ($e^{-0.7711} = 0.4625$) 낮은 건물피해액을 가질 것으로 분석되며 세종, 강원, 충북, 충남, 전북, 전남, 경북 지역은 각각 206.6% ($e^{1.1204} = 3.066$), 127.91% ($e^{0.8238} = 2.2791$), 363.9% ($e^{1.5345} = 4.639$), 297.29% ($e^{1.3795} = 3.9729$), 253.77% ($e^{1.2635} = 3.5377$), 314.29% ($e^{1.4214} = 4.1429$), 175.13% ($e^{1.0121} = 2.7513$) 높은 건물피해액을 가질 것으로 분석된다. 이는 지방 지역이 재해로 인한 피해가 더 크게 나타나는 것을 의미한다.

분석결과, 대전을 제외한 충청지역(세종, 충북, 충남)은 건물피해 발생확률이 전반적으로 매우 높고, 강원 및 경상지역은 상대적으로 낮은 것으로 나타났는데, 이러한 지역적 차이는 상당히 주목해야 할 현상으로 판단된다. 이와 같은 결과는 소득수준과 자연재해 피해액과의 관계성을 분석한 이전의 연구결과와 같이, 서울지역보다 지방 지역의 재해예방 시설 및 설비 부족, 건물 노후현상에 대한 반증으로 보인다 (Kim et al., 2017).

마지막으로 재해종류의 기준 수준인 강풍에 비해 태풍은 52.97% ($e^{1.5513} = 1.5297$) 높은 건물피해액을 가질 것으로 분석된다. 특이하게 강풍보다 호우로 인해 매우 큰 건물피해액이 발생할 것으로 추정된 복합포아송 모형 추정결과와는 달리, 영과잉 복합포아송 추정결과 건물피해가 발생할 경우 호우는 강풍대비 건물피해액에 유의미한 차이를 보이지 않는다고 분석되었다.

5. 결 론

본 연구는 자료의 특성으로 인해 모형 적합이 어려운 재해피해액 중 건물피해액에 대해 복합포아송 모형과 0의 자료가 과잉된 자료의 특성을 반영할 수 있는 영과잉 복합포아송 모형을 적합시켜, 보다 우수한 모형을 선택하고 해당 모형의 결과를 통해 건물피해액에 영향을 주는 재해종류와 재해지역을 확인하였다. 이를 통해 건물피해를 최소로 할 수 있는 지역별 차등적인 재해대응정책 마련의 필요성을 제시하였다.

본 연구에서는 재해지역, 재해기간, 재해종류를 독립변수로 하여 1994년부터 2015년까지 발생한 자연재해로 인한 건물피해액에 대하여 복합포아송 모형 및 영과잉 복합포아송 모형을 활용하여 각 변수의 계수를 추정하였다. 모형 추정결과 0 과잉의 특성을 반영할 수 있는 영과잉 복합포아송 모형이 더 타당한 모형임을 확인하였다. 세부 결과를 살펴보면, 영과잉 복합포아송 모형의 결과 중 건물피해 발생 여부에 관한 추정결과에서 태풍과 호우는 강풍을 기준으로 매우

높은 확률로 건물피해가 발생하는 것으로 분석되었다. 따라서 태풍과 호우의 경우, 상대적으로 심각한 건물의 피해가 우려되므로 이를 경감할 수 있는 철저한 예방 및 대응 대책 마련이 필요할 것으로 보인다.

아울러 서울지역 보다 지방은 자연재해에 의한 건물피해액이 증가할 것으로 예상되므로 지방의 건물피해 저감방안에 관한 연구사업, 재해예방 정책 및 전략 마련이 필요할 것으로 판단된다. 또한 재해기간이 길어질수록 건물피해액이 증가하므로 재해기간에 따른 대책마련과 함께, 향후 재난 유발요인의 강도와 연계하여 그 피해정도를 예측하는 연구도 필요할 것으로 판단된다.

비록 본 연구가 자연재해 피해액이 지나는 특별한 자료분포 특성을 반영하여 모형을 적합시킴으로써 의미 있는 결과를 얻었으나, 재해연보에 수록된 피해액이 실제 재산피해액과는 상이함에 따른 근본적인 한계가 존재한다. 현재 국가연구개발과제로 실제 피해액을 과학적으로 조사하고 합리적으로 추정하는 기법이 연구 개발되고 있으므로 이 결과를 적용하여 새로운 모형이 추정될 필요가 있다.

또한 본 연구는 분석이 광역시·도 차원에서만 이루어졌다는 것과 건물피해에 중요한 영향을 미치는 강우량, 강우강도, 풍속 등의 기상현상, 건물의 구조 및 노후도 등의 주요 인자가 고려되지 못했다는 한계점이 있다. 이는 연구의 한계점인 동시에, 이를 보완하여 추가적인 연구수행이 필요하다는 것을 시사한다. 아울러 본 연구는 피해액에 관한 설명모형에 국한되어 있으므로, 향후 인과관계 모형으로 발전시키기 위한 지속적인 연구가 필요할 것으로 판단된다.

감사의 글

본 연구는 행정안전부 재난예측및저감연구개발사업의 지원을 받아 수행된 연구입니다[MOIS-재난-2015-05].

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the 2nd International Symposium on Information Theory*, Akademia Kiado, Budapest, Hungary, pp. 267-281.
- Bai, J.S. (2014). *A study on priority of determinants of career decision level and career preparation behavior in high school students based on decision tree analysis*. Ph.D. dissertation, Soonchunhyang University, Korea.
- Barndorff-Nielsen, O. (1978). *Information and exponential families: Statistical theory*. Chichester, UK: Wiley.
- Chun, H.J. (2017). Fit of the number of insurance solicitor's turnovers using zero-inflated negative binomial

- regression. *Journal of the Korean Data and Information Science Society*, Vol. 28, No. 5, pp. 1087-1097.
- Cox, D.R., and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B*, Vol. 49, No. 1, pp. 1-18.
- Davis, S.A., and Skaggs, L.L. (1992). *Catalog of residential depth-damage functions used by the army corps of engineers in flood damage estimation*. IWR Report 92-R-3, U.S. Army Corps of Engineer, Institute for Water Resources, Alexandria, VA, USA.
- Dunn, P.K., and Smyth, G.K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, Vol. 15, No. 4, pp. 267-280.
- Dunn, P.K., and Smyth, G.K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, Vol. 18, No. 1, pp. 73-86.
- Hwang, H.M, Song, E.J, Park, N.U., and Lee, W.J. (2016). Analyzing precipitation data with zeroes using compound poisson distribution. *Journal of the Korean Data Analysis Society*, Vol. 18 No. 1, pp. 129-140.
- Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society B (Methodological)*, Vol. 49, No. 2, pp. 127-162.
- Jørgensen, B. (1997). *Theory of dispersion models*. London, UK: Chapman & Hall.
- Jørgensen B., and Paes de Souza, M.C. (1994). Fitting Tweedie's compound poisson model to insurance claims data. *Scandinavian Actuarial Journal*, Vol. 1994, No. 1, pp. 69-93.
- Kaas, R. (2005). Compound poisson distribution and GLM's-Tweedie's distribution. *Proceedings of the Contact Forum 3rd Actuarial and Financial Mathematics Day*, pp. 3-12.
- Kim, G.Y., Joo, H.T., and Kim, H.J. (2017). Analyzing damage effect factors of agricultural and livestock facilities caused by heavy snowfall disasters. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 4, pp. 65-72.
- Kim, G.Y., Joo, H.T., and Kim, H.J. (2018). The study for damage effect factors of heavy snowfall disasters: Focused on heavy snowfall disasters during the period of 2005 to 2014. *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 19, No. 2, pp. 125-136.
- Kim, H.Y., Kang, I.S., and Heo, T.Y. (2014). Estimation of a compound poisson linear model for non-structural measure to mitigate flood-damaged areas in Seoul Metropolitan area. *Seoul Studies*, Vol. 15, No. 2, pp. 49-59.
- Kim, J.H., Kim, H.J., Lee, S.O., and Cho, Y.S. (2007). Numerical simulation of flood inundation with quadtree grid. *J. Korean Soc. Hazard Mitig.*, Vol. 7, No. 2, pp. 45-52.
- Kim, J.Y., and Lee, S.K. (2008). A case study on the credit scoring model with zero-inflated poisson regression. *Journal of the Korean Data Analysis Society*, Vol. 10, No. 6, pp. 3255-3265.
- Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, Vol. 34, No. 1, pp. 1-14.
- Lee, J.S., Eo, G., Choi, C.H., Jung, J.W., and Kim, H.S. (2016). Development of rainfall-flood damage estimation function using nonlinear regression equation. *Journal of the Korea Society of Disaster Information*, Vol. 12, No. 1, pp. 74-88.
- Parveen, N., Abu Shadeque Mullah, M., and Ahshanullah, M. (2016). Tweedie model for analyzing zero-inflated continuous response: An application to job training data. *British Journal of Economics, Management & Trade*, Vol. 14, No. 3, Article No. BJEMT.26043.
- Nelder, J.A., and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, Vol. 74, No. 2, pp. 221-232.
- Smyth, G.K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society B (Methodological)*, Vol. 51, No. 1, pp 47-60.
- Song, Y.S., Lim, C.H., Joo, J.G., and Park, M.J. (2016). A study on heavy rain forecast evaluation and improvement method. *J. Korean Soc. Hazard Mitig.*, Vol. 16, No. 2, pp. 113-121.
- Tweedie, M.C.K. (1984). An index which distinguishes between some important exponential families. *Statistics: Applications and new directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, Indian Statistical Institute, Calcutta, India, pp. 579-604.

Received	September 16, 2019
Revised	September 16, 2019
Accepted	October 22, 2019