Bayesian GLM 기법과 통계학적 Downscaling 기법을 활용한 미래 재해위험도 변동성 평가기법 개발

Development of a Future Disaster Risk Assessment Model for Climate Change Using Bayesian GLM and Statistical Downscaling Model

권현한* · 명수정**

Kwon, Hyun Han · Myeong, Soojeong

Abstract

This study aims to develop a future disaster risk assessment model for climate change. A linear regression model, which has been widely used in previous studies, has limitations such as (1) the underlying probability distribution assumes to be Gaussian and (2) predicted values from the model are often given in negative numbers that are not appropriate in our case. In the present study, a Bayesian GLM-based disaster risk assessment model is introduced in conjunction with relevant predictors. Predictors were initially derived from daily precipitation data. The data that were finally put in the model as main predictors were (1) the number of consecutive cases of over 80 mm precipitation that lasted less than 10 days, and (2) the heaviest rainfall of the year. A nonstationary Markov chain downscaling model using KMA A2 climate change scenario as inputs was adopted to construct future rainfall scenarios in Gangwon, Seoul, Chungnam and Jeju, and the required rainfall predictors were extracted from the constructed scenarios. It was found that the proposed model could predict 90% of the disaster risks in the flood-prone areas such as Gangwon, Seoul and Chungnam. However, the proposed model failed to predict for the nonflood-prone area of Jeju Island. Future disaster risk variability was assessed using probability density function. The probability density function shifted toward the upper tail for all the areas, meaning increased disaster risks under climate change. Based on the results, this study claims that the proposed Bayesian GLM model is able to take into consideration the increased variability associated with climate change and thus can be effectively used in estimating future disaster risks.

Key words : Damage cost, Climate change, Bayesian GLM, Precipitation

요 지

본 연구에서는 미래 기후변화로 기인하는 재해피해액의 변동성을 평가하기 위해서 재해피해액 평가 모형을 개발하였다. 기존 선형회귀분석 방법에서 나타나는 정규분포 가정의 문제 및 예측값이 음의 값을 갖는 등의 문제점을 개선할 수 있는 Bayesian GLM 방법을 도입하였다. 재해피해액을 예측하기 위한 인자들을 강수량으로부터 추출하였으며 최종적으로 80 mm이상 강우 사상이 연속적으로 10일 이내에 발생한 횟수와 연최대강수량이 모형에 입력 자료로서 이용되었다. 미래의 기후변화 영향을 평가하기 위해서 기상청 A2 시나리오와 비정상성 Markov Chain Downscaling 기법이 적용되었으며 강원도, 서울, 충청남도 및 제주도에 대해서 상세강우시나리오를 생산하였다. 생산된 기후변화 상세강우시나리오로부터 동일하게 예측인자들을 추출하였다. 추출된 기상인자와 Bayesian GLM 방법을 통해서 재해피해액을 예측한 결과 상대적으로 강우에 대한 반응이 크게 나타나는 강원도, 서울 및 충청남도에서는 0.9 정도의 상관성을 갖는 정도 높은 모형 구축이 가능하였다. 그러나 상대적으로 강우로 인한 피해가 적게 나타나는 제주도의 경우 효과적인 모형을 구축하기 어려웠다. 미래 기후변화 시나리오부터 추출된 기상인자를 활용하여 미래 재해위험도의 변동성을 확률밀도함수를 통해서 평가한 결과 모든 지역에서 재해위험도가 미래에는 커질 것으로 전망되었다. 본 연구에서 제안된 Bayesian GLM 모형 기반 재해피해액 예측 모형은 기후변화로 기인하는 기상학적 변동성을 고려할 수 있는 모형으로서 미래의 재해위험도를 전망하는 도구로서 활용이 가능할 것으로 판단된다.

핵심용어 : 재해피해액, 기후변화, Bayesian GLM, 강수량

*정회원 · 전북대학교 토목공학과 조교수(E-mail : hkwon@jbnu.ac.kr)

**정회원 · 한국환경정책 평가연구원 부연구위원(교신저자)

1. 서론

도시화와 산업화는 일반적으로 지표의 불투수층을 증가시켜 하천에서 강우의 유출 특성을 변화시키며 동일한 강수량이라 할지라도 홍수 피해를 증가시키는데, 실제 근래의 홍수 피해액은 과거에 비해 증가추세에 있다. 이러한 도시화 및 산업화 영향에 더하여 최근에 나타나고 있는 기후변화는 기상변동성을 증가시키고 있어 자연재해 노출과 그 위험도가 커지고 있다. 우리나라의 대표적인 자연재해 유형은 호우와 태풍인데, 기후변화의 영향으로 전 세계적으로 호우와 태풍의 강도가 증가하는 추세에 있으며 이에 따라 매년 발생하는 피해의 규모가 커지고 있다. 이러한 이상기후는 일시적 현상이 아니라 기후변화의 영향으로 미래에도 빈번히 발생될 것으로 전망되어 그 피해가 우려되고 있다. 특히 인구가 밀집된 도시지역의 경우, 강수량의 증가로 설계기준을 초과하는 홍수 사상이 빈번하게 발생하면서 인명과 재산피해는 물론 피해액이 증가하는 추세이며 특히, 최근에는 소하천 및 도시하천에서 국지성 호우로 인하여 단시간에 범람이 발생하는 패턴이 빈번하게 나타나고 있다.

이처럼 증가하는 재해위험에 대응하기 위해서는 노출된 재해의 특성의 평가와 더불어 이를 구체적으로 모형화하는 과정이 선행되어야 한다. 그러나 동일한 유형의 재해라 하더라도 각 경우에 따라 이를 일으키는 원인과 발생 메커니즘이 매우 상이하기 때문에 실제로 이를 구체적으로 규명하고 모형을 구축하는 데는 어려움이 따른다. 앞서 언급되었듯이 우리나라의 재해피해는 강우에 기인하는 경우가 많은데, 본 연구에서는 이에 착안하여 강수량 자료로부터 재해피해를 효과적으로 설명할 수 있는 인자들을 추출하고 이를 활용하여 간단하면서도 설명력이 큰 해석 모형을 구축하고자 한다.

본 연구는 재해피해액 중 사회기반시설의 피해액을 추출하였으며 이를 예측하기 위한 모형을 구축하고 이를 활용하여 미래의 기후변화로 인한 재해피해액의 변동성을 평가하고자 한다. 이를 위해서 회귀분석에 근거한 재해피해액 예측모형을 구축하고자 하며, Bayesian 기반의 Generalized Linear Model (GLM)을 도입하였다. 본 연구에서 사용된 Bayesian 모형은 비정상성 빈도해석 모형으로도 활용가능하며, 평균적인 재해피해액을 예측할 수 있을 뿐만 아니라 재해 피해액의 극치값 추정도 가능하다. Bayesian 모형의 도입은 모형 및 매개변수에 대한 불확실성을 정량화 할 수 있을 뿐만 아니라 동시에 매개변수를 최적화 할 수 있는 장점이 있다. 또한 기존 회귀분석에서는 재해피해액 예측시 음의 값을 가지는 문제가 발생하기 때문에 이를 효과적으로 처리할 수 있어야 하는데, 이러한 점에서 Bayesian GLM 모형이 매우 유리하다. 본 연구에서는 강수량으로부터 사회기반시설 피해액과 관련이 높은 인자들을 우선 추출하고 이를 활용하여 Bayesian GLM을 구축하였다.

미래의 기후변화를 전망하기 위해서는 여러 단계에서 높은 수준의 해석 과정이 요구된다. 기후변화에 대한 영향을 평가하는 일반적인 방법은 전지구기후모형 (global circulation model, GCM) 또는 GCM을 기본 자료로 구축되는 지역기후

모형(regional climate model, RCM)을 활용하는 것이다. GCM의 경우 공간적 해상도가 200~400km에 이르며 시간적 해상도는 월단위 이상에서 신뢰성을 갖는 것으로 알려져 있다(권현한 등, 2008; Kwon et al., 2011). 그러나 모형 구축을 위해서 극치자료들이 일강수량 자료로부터 추출되어야 하기 때문에 GCM 및 RCM 기후변화 시나리오를 직접 사용하는 것은 실제 적용이 어렵다. 이러한 이유로 상세기상시나리오를 생산하는 과정이 요구되는데, 본 연구에서는 기상청의 RCM 기후변화 시나리오를 입력자료로 활용할 수 있는 비정상성 Markov Chain 모형을 이용하여 일강수량을 모의한 후 이들 자료로부터 모형 구축당시 사용했던 설명인자를 동일하게 추출하였다. 그 다음으로 이를 본 연구에서 구축한 Bayesian GLM 모형에 입력하여 미래의 재해피해액을 평가하였다.

국내에서 재해피해액을 예측하는 연구는 아직 미미한 실정이며, 특히 기후변화에 기인하는 미래 재해피해액을 전망한 사례는 매우 드물다. 전환돈 등(2008)은 지자체를 지역특성에 따라 크게 6개로 구분하고 호우와 태풍에 의한 재해 발생횟수와 총 피해금액을 지자체별로 산정하여 지자체를 4등급으로 분류하였으며 이를 지역특성과 연계하여 호우와 태풍에 의한 피해정도를 분석한 연구를 수행한 바 있으며, 장옥재와 김영오(2009)는 지역회귀분석을 이용한 홍수피해위험도를 산정한 사례가 있다. 수문학적 예측모형으로서의 Bayesian 모형의 경우 다양한 응용 연구사례가 있다. 권현한과 문영일(2007)은 극치강수량의 발생빈도를 예측한 바 있으며, Kwon et al.(2008)은 미국 Montana 지역의 홍수량을 Bayesian 모형을 통해서 예측하였다. 또, Kwon et al.(2009)는 중국 삼협댐 수위관측 지점에 여름 평균유출량 및 홍수량을 Bayesian 기반의 비정상성 빈도해석을 통해 예측하기도 하였다.

본 논문의 2장에서는 대상 자료 및 예측인자 추출에 대해서 살펴보고, 3장에서는 Bayesian 모형 구축과정 및 미래 기후변화 시나리오 생산과정에 대해서 서술하였다. 4장에서는 우리나라의 사례 지역에 대해 앞서 구축한 모형을 적용하고 모형의 적합성과 활용성을 평가하였다.

2. 대상 자료 및 기상인자 추출

사회기반시설의 피해는 강우뿐만 아니라 바람, 화재, 태풍 등 다양한 원인에 의해서 발생한다. 그러나 본 연구는 사회기반시설의 피해액을 예측하는 과정에서 사용되는 예측인자 자료로는 강수량만을 고려하였다. 기후변화 시나리오부터 얻어질 수 있는 일반적인 정보는 강수량과 온도인데, 본 연구는 기후변화에 따른 미래의 재해 피해액을 전망하고자 하였으며, 수문기상학적 변량 중에서 기후변화 시나리오부터 얻을 수 있는 정보가 강수량이기 때문이다.

본 연구에서는 기후변화에 의한 자연재해 취약성을 정량적으로 분석하기 위하여 기상인자와 재해로 인한 피해액의 상관관계를 이용하였다. 재해 피해액은 1994년부터 2008년까지 15년간 전국 시군별로 피해액을 집계한 자료를 이용하였으며,

Table 1. Weather Stations used in this study

ID	Station Name	Start	End	District	ID	Station Name	Start	End	District
90	Sokcho	1968	2008	Gangwon	108	Seoul	1961	2008	Seoul
100	Daegwallyeong	1972	2008		129	Seosan	1968	2008	Chungnam
101	Chuncheon	1968	2008		232	Cheonan	1973	2008	
105	Gangneung	1961	2008		235	Boryeong	1973	2008	
114	Wonju	1973	2008		236	Buyeo	1973	2008	
211	Inje	1973	2008		238	Geumsan	1973	2008	
212	Hongcheon	1973	2008		184	Jeju	1961	2008	Jeju
					189	Seogwipo	1961	2008	

Table 2. Correlation analysis between damage costs and extreme indices

District	Extreme Factor ①	Extreme Factor ②	Extreme Factor ③	Extreme Factor ④
Gangwon	0.3781	0.6475	0.3463	0.8743
Seoul	0.3002	0.4905	0.3125	0.8296
Chungnam	0.5233	0.7337	0.6811	0.9408
Jeju	0.2818	0.3063	0.2464	0.4391

Extreme Factor : ① 80 mm이상 일강수량 발생일수, ② 80 mm 이상 일강수량의 합, ③ 80 mm이상 강우의 발생 간격이 10일 이하인 횟수, ④ 연최대강수량

우리나라 58개 강우관측소의 일강수량 자료를 이용하여 재해에 영향을 줄 수 있는 네 가지 인자를 추출하였고, 연도별 태풍 발생횟수도 기상인자로 고려하였다. 일강수량으로부터 추출한 네 가지 영향인자들은 ①80 mm이상 일강수량 발생일수, ②80 mm이상 일강수량의 합, ③80 mm이상 강우가 연속적으로 10일 이내에 발생한 횟수 및 ④연최대강수량이다.

이 네 영향인자를 고려한 이유를 차례로 기술하면, 첫째, 80 mm이상 강수량은 기상청에서 호우주의보 기준으로서 극치강수량의 발생빈도를 표현하기 위함이고, 둘째, 80 mm이상 강우사상의 합은 앞서 추출한 80 mm이상의 강우빈도와 더불어 특정 연도에 호의주의보 이상의 크기를 갖는 강수량의 총량을 표현하기 위함이다. 셋째, 80 mm이상 강우가 연속적으로 10일 이내에 발생한 횟수는 극치강수량이 특정시간에 연속해서 발생하는 경우로서, 산사태 등과 같이 연속강우로 인한 피해가 원인이 되는 경우와 연속강우로 인한 재해 피해증가를 검토하기 위한 대리변수로 활용될 수 있기 때문이다. 넷째, 매년 발생하는 일최대강수량으로 호우경보 이상의 연최대강수량이 발생하는 경우 재산피해액의 변동성을 평가하기 위함이다.

본 연구는 서울, 강원도, 충청남도, 제주도를 대상으로 하였으며, 각 지역에 해당하는 관측소로부터 추출된 인자들의 지역 평균값을 이용하였다. 본 연구는 미래의 재해위험도 전망모형을 구축하고, 모형의 적용성을 평가하기 위해 기상청에서 제공하는 기후변화시나리오 중 비관적 시나리오로 주로 활용되고 있는 A2 시나리오를 대상으로 연구를 진행하였다. 미래 사회기반시설 피해액을 예측하기 위한 4가지 영향인자는 국립기상연구소의 A2시나리오를 통계학적 Downscaling을 실시

해서 얻은 일강수량으로부터 동일하게 추출하였다. Table 1은 연구에 이용된 관측소의 자료기간과 연구대상지역을 분류한 결과이다. 네 영향인자들과 재해피해액의 상관관계를 분석한 결과는 대부분의 지역에서 연최대강수량이 재해피해액과 가장 높은 상관성을 갖는 것으로 나타났다(Table 2).

3. 기후변화에 의한 피해액 예측 모형

본 연구의 피해액 예측 모형은 Bayesian GLM 방법을 예측모형을 활용하였다. Bayesian GLM 방법은 일반화된 선형 회귀모형으로서 기존 선형회귀방정식과 구별된다. 기존 선형회귀방정식은 정규분포를 가정, 최소자승법(least square method)이 기본적으로 이용되는 반면에, Bayesian GLM은 정규분포를 포함해서 다양한 확률분포의 적용이 가능하며 또, Bayesian 기법을 통해서 매개변수의 사후분포가 추정되며 사후분포를 통해서 매개변수의 불확실성 정량화가 가능하다. GLM 방법은 3단계의 중요 요소로 구분되는데, 첫째, 시계열이 확률분포로 정의되며 둘째, 각 시간에서 확률분포의 평균 $\mu(t)$ 은 매개변수 α 와 예측인자 X , 즉 αX 로 정의된다. 마지막으로 연결함수(link function)에 의해 예측인자 αX 와 종속변수 Y 의 기대치 $\mu(t)$ 를 통합시킨다.

본 연구에서는 시작적인 검토 및 적합성 평가를 통해서 2변수 대수정규분포 (2-parameter lognormal distribution)를 이용하였다. Gamma 분포의 경우에도 적합성 평가를 통과하나 일반적으로 Bayesian Gamma GLM 모형에서 매개변수가 발산하는 문제점이 있어 본 연구에서는 2변수 대수정규분포를 사용하였다. 대수정규분포형의 확률밀도함수와 누가확률밀도함수는 다음 식 (1)~(2)와 같다. 그림 1은 재해피해액에 대한 확률밀도함수와 정규 확률치를 통한 자료의 분포상태를 검증한 결과를 나타낸다. 본 연구에서는 대수정규분포를 적용하므로 그림 1에서 확률밀도함수와 확률치를 통한 검증에 앞서 자료를 Log 변환하여 수행하였다.

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0 \quad (1)$$

$$F(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{e^{-\frac{(\ln(t) - \mu)^2}{2\sigma^2}}}{t} dt \quad (2)$$

여기서 μ 는 위치매개변수(location parameter), σ 는 규모매개

Fig. 1. Probability density function and probability plot for damage costs

변수(scale parameter)를 갖는 함수이다. 대수정규분포를 이용한 이유는 피해액이 음(-)의 값을 갖는 것을 방지하기 위해서인데, 재해 피해액이 매우 적은 경우도 있기 때문에 기상인자를 활용한 회귀분석 과정에서 음의 값을 가질 수 있어 이를 고려한 것이다. 재해로 인한 피해액 Y 가 앞서 제시한 대수정규분포를 따른다고 가정할 때, 대수정규분포함수의 위치 매개변수와 규모매개변수는 식 (3)과 같이 시간에 따라 다른 값을 갖는 비정상성 모형으로 구성할 수 있다.

$$Y(\text{damage costs}) = \text{lognormal}(\mu(t), \sigma) \quad (3)$$

Bayesian GLM 모형에서, 식 (3)의 모든 매개변수들은 확률분포를 가지게 된다. 즉, $\mu(t)$ 와 σ 가 사전분포를 가지게 되며 각 매개변수는 앞에서 추출한 네 가지 기상인자들 중 최종적으로 선정된 두 개의 기상인자들의 함수로서 다음의 식 (4) 및 식 (5)와 같이 회귀관계를 갖는다고 가정한다. 본 연구에서는 통계적인 검토를 통해 각각의 인자에 대해서 2차 회귀식(quadratic regression) 형태를 취하여 재해피해액을 평가하였다.

$$p(\mu(t)) \sim \alpha_0 + \alpha_1 X_3 + \alpha_2 X_3^2 + \alpha_3 X_4 + \alpha_4 X_4^2 \quad (4)$$

$$p(\sigma) \sim G(k, s) \quad (5)$$

여기서, t 는 시간을 의미하며, X_3 과 X_4 는 네 개의 예측인자 중 연속된 극치강수량 횟수와 연최대일강수량이다. 네 인자 중 80 mm이상 강수량 발생횟수와 이들 80 mm이상 강수량의 총합 등 2개의 인자는 3번째 인자인 연속된 극치강수량 횟수와 4번째 인자인 연최대일강수량과 서로 상관되어 있기 때문에 이처럼 동일한 성격을 갖는 인자는 모형의 신뢰성 확보를 위해 배제하였다.

α 는 예측인자와 매개변수 μ 의 사전분포를 연계시키는 회귀모형의 매개변수를 나타낸다. 회귀방정식의 매개변수는 일반적으로 정규분포로 가정된다(Gelman 등, 2004). 정규분포로 가정한 이유는 증가경향과 감소경향에 의한 변화를 모두

고려하기 위함이다. 회귀계수를 정규분포로 가정하더라도, 추정된 결합확률분포(joint distribution)는 대수정규분포를 따르게 된다. α 의 분산 σ_α 는 양의 값을 가져야 하므로 Gamma 분포로 가정하였으며 이를 식 (6)~(7)과 같이 나타낼 수 있다. 결과적으로 6개의 매개변수에 대한 결합확률(joint probability)은 식 (8)과 같다.

$$\alpha_n \sim N(\mu_{\alpha_n}, \sigma_{\alpha_n}^2) \quad n = 0, \dots, 4 \quad (6)$$

$$\sigma_\alpha \sim G(k_\alpha, s_\alpha) \quad (7)$$

$$p(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \sigma_\alpha) \propto 1 \quad (8)$$

Bayes 정리에 의해서 매개변수의 사후분포는 다음과 같이 전개될 수 있다.

$$p(\Theta|y) = \frac{p(\Theta, y)}{p(y)} \propto p(y|\Theta) \cdot p(\Theta) \quad (9)$$

여기서 $\Theta = [\mu(t), \sigma, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \sigma_\alpha]$ 는 전체매개변수들의 집합을 나타내며, $p(\Theta)$ 는 매개변수들의 사전분포를, $p(y|\Theta)$ 는 y 의 우도함수(likelihood function)를 의미한다.

$$p(y|\Theta) = \prod_{t=1}^T \text{lognormal}(y_t | \mu(t), \sigma) \quad (10)$$

여기서, T 는 자료연수를 나타낸다. 식 (4), 식 (5), 식 (8), 식 (10)에서 정의되는 매개변수들의 사전분포들을 식 (11)에 대입시킴으로써 매개변수들의 사후분포를 식 (11)을 통해 추정할 수 있다. 식 (11)에서는 대수정규분포의 위치 및 규모 매개변수와 회귀매개변수들을 위한 비공액사전분포(non-conjugate prior)를 이용하여 모델 매개변수를 추정한다.

$$p(y|\Theta) \propto \prod_{t=1}^T \text{lognormal}(y_t | \mu(t), \tau) \cdot N(\mu(t) | \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4, \sigma_\alpha^2) \cdot G(\tau | k, s) \quad (11)$$

식 (11)은 모든 매개변수에 대한 적분을 통해 직접적으로 추정하는 것은 불가능하므로, 본 연구는 Markov Chain Monte Carlo(MCMC) 기법을 도입하여 매개변수들의 사후분포를 추정하였다. 본 연구는 MCMC 방법 중 Gibbs 표본법을 이용하여 회귀매개변수들을 추정하였다. Gibbs 표본법은 원하는 다변량 확률분포에서 iid 표본을 추출하는 것이 복잡하거나 난해한 경우 이용 가능한 방법으로서 2개의 변수를 갖는 다변량 확률분포를 이용하여 설명하면 다음과 같다. 첫째, 두 변수에 대한 초기 값($\theta_1^{(0)}, \theta_2^{(0)}$)을 부여한다. 둘째, i 번째 난수 벡터 ($\theta_1^{(i)}, \theta_2^{(i)}$)가 주어졌을 때 $i+1$ 번째 난수 벡터를 $\theta_1^{i+1} \sim \pi(\theta_1 | \theta_2^{(i)}), \theta_2^{i+1} \sim \pi(\theta_2 | \theta_1^{i+1})$ 의 조건부 분포에서 추출한다. 마지막으로 위의 과정을 충분히 반복한 후 초기의 일부분 난수를 제거한 이후의 난수들을 이용하게 된다. 조건부 분포에서 조건으로 주어지는 변수들의 값은 정확하게 바로 직전의 단계에서 주어진 값이 사용되게 되며, 따라서 조건부 분포에서 추출된 난수가 안정 상태에 도달하는 것이 주어진 다변량 확률분포를 정확히 따르는 난수가 되는 척도가 되며 Gibbs 표본법을 구현하는데 가장 중요한 부분이 된다.

추정된 사후확률분포를 이용한 미래 재해피해액은 식 (12)의 Quantile 함수를 이용하여 구할 수 있다. 여기서 p 는 식 (2)의 누가확률밀도함수를 나타낸다.

$$Y(p) = F^{-1}(p | \mu, \sigma) = \{x : F(x | \mu, \sigma) = p\} \quad (12)$$

4. 모형 적용 및 결과

본 연구에서 제안된 Bayesian GLM 모형을 사회기반시설 피해액과 추출된 기상인자들을 대상으로 적용하기에 앞서서 기존 선형회귀분석 모형을 적용해 보았다. 그림 2는 관측 값으로서의 재해피해액과 기상인자들을 이용한 선형회귀분석 모형에 의해 예측된 값을 비교한 그림으로, 강원도, 서울 및 충청남도의 경우 피해액이 강우인자들에 의해 정도 높게 모의되고 있는 것을 확인할 수 있다. 이러한 지역의 경우 극치 강수에 의한 홍수피해가 재해의 주원인인 지역으로 해석할 수 있다. 그러나 선형회귀분석의 경우 피해액의 분포가 $-\infty < X < \infty$ 의 범위를 갖는 정규분포로 가정되기 때문에 모형 예측 값이 종종 음의 값을 갖는 문제점이 있으며 이에 대한 처리가 필요하다.

그림 2에 도시한 제주도의 경우 강우인자에 의한 예측 값이 피해액을 명확히 반영하지는 않는 것을 볼 수 있다. 이러한 지역의 경우 재해 발생에 있어 강우로 인한 피해발생 이외에도 다른 주요 원인이 있는 지역으로 해석할 수 있다. 실제로 피해액이 강수량에 의한 예측 값보다 높게 나타난 2002~2003년의 경우 태풍 루사와 매미로 인한 풍해 피해가 컸던 연도라는 것을 알 수 있다.

그림 3은 본 연구에서 제안하는 Bayesian GLM 모형의 적합결과인데, 앞서 도시한 그림 2와 같이 동일한 기상인자를 사용하여 모형을 적합시켰다. 기존 선형회귀방정식은 사회

Fig. 2. Model-fitting results using simple linear regression model

기반시설피해액을 예측하는데 있어 음의 값으로 추정되는 문제점과 자료 자체의 분포특성을 정규분포에 한정하는 문제점이 있다. 이러한 문제점은 본 연구에서 제시한 Bayesian GLM 방법을 통해서 개선될 수 있다. 즉 자료의 분포특성을 정규분포가 아닌 대수정규분포로 가정하며 분포의 특성을 결정짓는 분포의 평균값은 기상인자의 함수로 정의된다. 그림 3에서 보면 추정된 값이 그림 2와는 다르게 항상 양의 값을 가지는 것을 확인할 수 있으며 Bayesian 모형으로부터 계산된 사후분포로부터 매개변수의 불확실성도 정량적으로 해석이 가능하다.

다음으로 기존 선형회귀분석과 본 연구에서 제안한 Bayesian GLM방법을 통한 결과를 통계치를 통해서 평가하였다. 일반적으로 모형의 정확성을 평가하기 위해서 상관계수(correlation coefficient, CC), Nash-Sutcliffe 계수(N-S), 일치계수(Index of agreement, IoA)가 이용되며 본 연구는 이 세 통계치를 사용하여 모형의 적합성을 평가하였다. Nash-Sutcliffe 계수는 Nash와 Sutcliffe(1970)가 제안한 통계적 기준으로 편의를 줄일 수 있는 무차원계수이다. 예측값과 실측값이 잘 일치할수록 1에 가까워지는 성질이 있다. 이 기준은 무차원양으로서 자료의 개수에 관계없이 절대적 평가기준이 될 수 있다.

$$ME = \frac{F_0^2 - F^2}{F_0^2} \quad (13)$$

$$F^2 = \sum_{i=1}^n [q_0(t) - q_s(t)]_i^2 \quad (14)$$

$$F_0^2 = \sum_{i=1}^n [q_0(t) - \bar{q}]_i^2 \quad (15)$$

여기서, q_0 은 실측값을 q_s 는 예측값을 나타낸다.

상관계수는 모형의 효율성을 나타내는 무차원 기준으로 1에 근접할수록 모형의 모의능력이 우수함을 뜻한다.

$$CC = \frac{\sum_{t=1}^n (q_o(t) - \bar{q}_o)(q_s(t) - \bar{q}_s)}{(n-1)} \quad (16)$$

Willmott(1981)은 평균과 분산사이의 차이를 추정할 때 상관계수에 근거한 통계치의 민감성을 보완하고자 일치계수를 개발하였다. 일치계수는 식 (17)과 같다. 일치계수는 0부터 1의 범위를 가지며 1에 가까울수록 모형의 예측능력이 우수함을 뜻한다.

$$IoA = 1 - \frac{\sum_{t=1}^n (q_o(t) - q_s(t))^2}{\sum_{t=1}^n (|q_s(t) - \bar{q}_o| + |q_o(t) - \bar{q}_o|)^2} \quad (17)$$

표 3은 모형 적용 결과를 3가지 통계치에 대해서 비교한 결과이다. 선형회귀분석 결과는 음의 값을 허용하는 경우이므로 모형을 적합시키는데 있어서 좀 더 자유롭다. 이에 반해 Bayesian GLM 모형에서는 음의 값을 제한하게 되므로 일반적으로 기존 선형회귀분석보다 큰 적합성을 기대하기 어렵다. 그럼에도 불구하고 표 3의 경우 Bayesian GLM 방법이 전

Fig. 3. Model-fitting results using Bayesian GLM model

Table 3. Comparison of model performance with selected statistics

Statistics	Gangwon		Seoul		Chungnam		Jeju	
	Linear Model	Bayesian GLM	Linear Model	Bayesian GLM	Linear Model	Bayesian GLM	Linear Model	Bayesian GLM
R	0.93	0.93	0.90	0.88	0.98	0.98	0.48	0.38
N-S	0.87	0.87	0.81	0.77	0.96	0.96	0.23	0.12
IoA	0.96	0.96	0.95	0.94	0.99	0.99	0.58	0.35

체적으로 선형회귀분석과 유사한 모형의 적합성을 갖는 것으로 확인되었다. 특히, 강원도나 충청도와 같이 강우인자에 대한 기반시설피해액에 대한 반응이 큰 경우는 기존 선형회귀분석과 큰 차이를 나타내지 않고 있다. 반면, 제주도의 경우 Bayesian GLM 방법이 선형회귀분석에 비해 정도가 낮은 결과를 주고 있다. 이는 제주도의 경우 강우인자에 대한 재해 피해액의 반응이 크지 않기 때문에 선형회귀분석 시에 과적합(overfitting)으로 기인한 결과로 보인다.

4. 미래 재해피해액 변동성 모의

미래 재해피해액에 대한 변동성을 평가하기 위해서는 모형을 적합시킬 때 사용된 극치강우인자들을 미래에 대해서도 동일하게 구축하는 것이 필요하다. 이를 위해 본 연구에서는 SRES A2시나리오를 바탕으로 미래의 강우변동성을 우선적으로 모의하였는데, 기상청에서 제공하는 지역기후모델 기반 A2시나리오를 이용하였다. 일반적으로 동역학적 모형을 통한 강우시나리오는 극치값을 모의하는데 한계가 있으며 통계학

적 Downscaling 기법을 적용하여 이러한 문제점을 보완하고 있다(권현한 등, 2009). 본 연구에서는 권현한과 김병식(2009)이 제안한 비정상성 Markov Chain 모형을 이용하여 미래기후변화 강우시나리오를 구축하였다. 비정상성 Markov Chain 모형에 대해서 간략히 요약하면 다음과 같다.

비정상성 Markov Chain 모형은 기본적으로 Markov Chain을 이용하여 일강수량을 모의하게 되며 기후변화 시나리오를 반영하기 위해서 Markov Chain 모형의 4개의 매개변수와 외부인자를 연결하기 위한 회귀분석 모형이 도입된다. 비정상성 Markov Chain 모형에서는 무강수에서 강수로 진행하는 천이확률(p_{01}), 강수에서 강수로 진행되는 천이확률(p_{11}), 강우량을 모의하기 위한 Gamma 확률분포의 축척매개변수(α)와 형상매개변수(β)는 입력되는 기후변화 시나리오에 따라 매년 동적으로 변동되면서 추정되게 된다. 이렇게 추정된 미래 2010년부터 2100년까지 모형의 매개변수는 미래 일강수량을 모의하는데 이용된다.

본 연구는 비정상성 Markov Chain 모형을 통해 강우지점별로 일강수량을 모의하고 이로부터 극치강우인자들을 모두

Fig. 4. Future projection of damage cost using simple linear regression model

추출하였으며, 이를 앞서 구축된 모형의 입력 자료로 활용하였다. 먼저 기존 선형회귀분석방법을 통해서 미래 재해위험도의 변동성을 평가하였으며, 그 결과는 그림 4와 같다. 기존 선형회귀분석방법을 이용한 결과를 보면 미래의 재해위험도가 음의 값을 가지고 전망되고 있는 문제점을 확인할 수 있

다. 이러한 문제점은 모형 적합과정보보다 더욱 심하게 나타나고 있으며 이는 극치강수량의 변동성이 과거에 비해 커짐으로서 상대적으로 음의 값이 출현빈도가 잦아지는 것으로 보인다.

그림 5는 본 연구가 제안하는 Bayesian GLM 모형을 통

Fig. 5. Future projection of damage cost using Bayesian GLM model

Fig. 6. Comparison of probability density function of damage cost

해서 전망된 재해피해액으로 기존 방법에서 나타나는 음의 값의 발생 문제점이 개선된 결과를 보여주고 있으며, 예측사후분포(predictive posterior)로부터 불확실성 구간이 적절히 추정되고 있음을 확인할 수 있다. 본 연구는 기후변화로 인한 미래의 재해피해액의 정확한 예측이 아니라 전반적인 재해위험도의 변동성을 전망하고자 하였으므로, 최종적으로 재해피해액의 확률밀도함수를 비교하여 미래 재해위험도를 평가하였다. 이 결과는 그림 6에 나타내었으며, 핵밀도함수방법(kernel density estimation method)을 통해 확률밀도함수를 추정하여 비교하였다. 그림 6에 나타난 결과에 의하면 전체적으로 재해위험도가 증가하는 것으로 전망된다. 모든 지점에서 확률밀도함수의 중심이 오른쪽으로 이동하고 있으며, 이는 미래에는 지금보다 더 큰 재해피해가 발생할 수 있다는 것을 의미한다. 특히, 확률밀도함수의 오른쪽 tail이 과거에 비해서 크게 두꺼워지고 있어 기후변화로 인한 재해위험도가 상승할 것으로 전망된다.

본 연구는 미래 발생 가능한 재해의 여러 원인 인자를 모두 고려하지는 않았고, 기후변화 시나리오에 의해 추출 가능한 강수량에 의한 예측을 중심으로 수행하였으며, 추후 모형을 구성하는 인자들을 적절히 선택함으로써 다른 원인에 의한 피해 예측도 가능할 것이다.

5. 결 론

본 연구는 미래의 기후변화로 기인하는 재해피해액의 변동성을 평가하기 위한 모형을 개발하였다. 재해를 일으키는 다양한 요인이 존재하지만 본 연구는 강수량을 중심으로 연구를 진행하였다. 강수량에 국한하여 재해피해액 평가모형을 구축한 이유는 최종적인 연구목표가 기후변화 영향을 검토하는데 있기 때문이며, 현재까지 기후변화 분석을 위한 신뢰성 있는 미래 기후변화 시나리오 구축의 가능 여부를 고려하였다. 이러한 점에서 미래 기후변화에 따른 강우변동성에 대한 연구가 많이 진행되어 왔으며, 재해피해액 분석을 위해 요구되는 시공간적스케일의 자료 취득이 가능한 강우에 국한하여 연구를 수행하였다.

본 연구의 주안점은 기존 회귀분석방정식에 문제점을 개선하는 것과 불확실성이 동시에 고려될 수 있는 방법론의 개발이었다. 이를 위해 Bayesian GLM모형을 이용한 재해피해액 평가모형을 개발하여 상대적으로 극치강우로 인한 피해가 크다고 생각되는 강원도, 서울, 충청남도를 적용대상으로 하였으며, 이와 더불어 강우로 인한 피해가 상대적으로 적은 제주도를 평가 대상지에 추가하여 앞의 3개 지역과 비교하였다. 본 연구를 통해서 얻은 결론과 주요 토의 사항은 다음과 같다.

첫째, 일반적으로 사용되는 선형회귀방정식의 경우 재해피해액 모형으로 사용하는데 있어서 한계점이 존재하였다. 재해피해액의 경우 확률분포의 형태로 볼 때 극치값으로 왜곡된 형태를 가진다. 즉, 재해피해액이 적을 때는 거의 0의 가깝지만 일단 재해가 발생하면 매우 큰 피해액을 나타내게 된다. 이러한 특징으로 기존 회귀분석을 적용하였을 경우 예측값이

음의 값을 갖는 문제점이 나타났다.

둘째, 이러한 문제점을 개선하기 위해 사용된 Bayesian GLM 모형의 경우 자료의 분포특성을 고려한 확률밀도함수의 적용이 가능하기 때문에 상대적으로 신뢰성 있는 해석이 가능하였다. 또한 연결함수를 통해서 음의 값이 추정되는 문제점을 개선함과 동시에 재해요인과 재해피해액의 분포 특성을 연동시켜 해석할 수 있었다.

셋째, 기후변화 연구에서 가장 중요한 요소 중 하나는 불확실성을 고려하는 것으로, 본 연구에서는 Bayesian 기반의 해석모형으로부터 추정되는 사후분포를 통해 불확실성을 정량화하여 표출할 수 있었다.

넷째, 개발된 모형을 통한 미래 재해위험도를 확률밀도함수 형태로 비교한 결과 현재에 비해 미래의 재해피해액의 중심이 오른쪽으로 이동하는 특징이 나타나고 있으며, 전체적으로 극치사상의 증가로 기인하는 재해피해액 증가가 전망되었다.

본 연구는 매우 제한적인 형태의 접근 방법으로 강수량만을 재해 요소로서 고려하였으며, 본 연구의 접근법은 기후변화 전망을 위한 간단한 모형으로서 활용이 가능할 것으로 판단된다. 그러나 좀 더 개선된 해석모형을 위해서는 강수량 이외에 재해를 유발하는 다양한 인자에 대한 고려해야 할 것이며, 이를 재해 요소들 간의 상관관계 등을 종합적으로 반영할 수 있는 모형 개발이 이루어져야 할 것이다.

사 사

본 연구는 한국환경정책·평가연구원의 녹색성장 정책연구과제(2010-13)의 일환으로 수행되었습니다.

참고문헌

- 권현한, 김병식 (2009) 비정상성 Markov Chain Model을 이용한 통계학적 Downscaling 기법 개발, **한국수자원학회 논문집**, 제42권, 제3호, pp. 213-225.
- 권현한, 김병식, 김보경 (2009) 기후변화에 따른 수자원 영향 평가를 위한 Regional Climate Model 강수 계열의 특성 분석, **대한토목학회 논문집**, 제28권, 제5B호, pp. 525-533.
- 권현한, 문영일 (2007) 기상정보 및 태풍특성을 고려한 계절 강수량의 확률론적 모형 구축, **대한토목학회 논문집**, 제27권, 제1B호, pp. 45-52.
- 전환돈 박무종 김근영 김진옥 (2008) 지자체별 특성을 고려한 자연재해에 따른 피해유형 분석, **한국방재학회 논문집**, 제8권, 제2호, pp. 75-82.
- 장옥재, 김영오 (2009) 지역회귀분석을 이용한 홍수피해위험도 산정, **한국방재학회 논문집**, 제9권, 제4호, pp. 71-80.
- Kwon, H.-H., Brown, C. and Lall, U. (2008) Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling, *Geophysical Research Letters*, Vol. 35, L05404, doi:10.1029/2007GL032220, 2008.
- Kwon, H.-H., Brown, C., Xu, K. and Lall, U. (2009) Seasonal and annual maximum streamflow forecasting using climate information: application to the Three Gorges Dam in the Yangtze River basin, China, *Hydrological Sciences*, Vol. 54, No. 3, pp. 582-595.
- Kwon, H.-H., Sivakumar, B., Moon, Y.-I. and Kim, B.-S. (2011) Assessment of change in design flood frequency under climate

change using a multivariate downscaling model and a precipitation-runoff model, *Stochastic Environmental Research and Risk Assessment*, Vol. 25, No. 4, 567-581, DOI: 10.1007/s00477-010-0422-z.

Nash, J.E. and Sutcliffe, J.V. (1970) "River Flow Forecasting Conceptual Models Part I - A Discussion of Principles", *J. Hydrol.*, Vol. 10, pp. 282-290.
Willmott, C.J. (1981) On the validation of models. *Phys. Geog.*, 2, 184-194.

© 논문접수일 : 11년 03월 25일
© 심사의뢰일 : 11년 04월 04일
© 심사완료일 : 11년 10월 12일