로지스틱 분포를 이용한 태풍피해예측함수 개발

Development of Typhoon Damage Prediction Function Using a Logistic Distribution

김종성* · 최창현** · 김태균***

Kim, Jongsung*, Choi, Changhyun**, and Kim, Taegyun***

Abstract

The purpose of this study is to develop a typhoon damage prediction function considering the correlation between damage amount due to typhoons and hydrometeorological factors estimated from disaster statistics in coastal districts. As a typhoon approaches, heavy rain, gale, and storms jointly and simultaneously cause severe damage. A total of 32 variables were selected to derive the function, such as factors that affect rainfall-runoff, maximum wind velocity, and central typhoon pressure. As observed, typhoon damages have a wide range-from hundreds of thousands of KRW to more than KRW 1 trillion-and accordingly have been classified into large and small groups depending on the amount of damage; thereafter, the typhoon damage prediction function was developed for each group. The logistic regression analysis was used for group classification of typhoon damages. Furthermore, the principal component regression analysis was employed for the typhoon damage prediction function. The principal component analysis effectively solved the problem of multicollinearity due to high correlation between 32-dimensional independent variables and rainfall-runoff related variables. According to the results of the typhoon damage prediction function, the Normalized Root Mean Squared Error (NRMSE) of the large group was 12.4%, while that of the small group was 18.5%.

Key words : Typhoon Damage Prediction Function, Logistic Regression, Principle Component Analysis, Coastal District

요 지

본 연구에서는 연안지역에 위치한 시군구를 대상으로 재해 통계자료를 이용하여 태풍으로 인한 피해액과 수문기상학적 인자의 관계를 유도하여, 태풍피해를 예측할 수 있는 태풍피해예측함수를 개발하였다. 태풍은 호우, 강풍, 풍랑에 의한 피해가 동시에 발생하므로, 강우-유출관계에 영향을 미치는 인자와 최대풍속, 태풍중심기압 등 총 32개 변수를 독립변수로 선정하였다. 관측된 태풍피해액은 수십만원에서 1조원 이상의 넓은 범위를 가지므로, 피해액이 큰 집단과 작은 집단으로 구분하고, 각각의 집단에 대하여 태풍피해예측함수를 개발하였다. 집단 구분은 로지스틱회귀분석을 이용하고, 태풍피해예측함수는 주성분회귀분석을 이용하였다. 주성분분석은 32개의 독립변수로 인한 고차원 문제와 강우-유출 관련 변수의 높은 상관관계에 의한 다중공선성 문제를 해결할 수 있었다. 두 집단에 대한 태풍피해예측함수 적용 결과 큰 집단의 NRMSE는 12.4%, 작은 집단은 18.5%로 나타났다.

핵심용어 : 태풍피해예측함수, 로지스틱회귀분석, 주성분분석, 연안지역 시군구

1. 서 론

태풍은 호우로 인한 홍수, 침수, 산사태 등의 피해, 강풍에

의한 피해, 파랑으로 인한 피해가 복합적으로 일어나는 재해이다. 2017년 기준으로 과거 10년간의 피해 현황을 살펴보면 태풍이 가장 많은 피해를 야기하였고, 거의 같은 수준으로

*정회원, 인하대학교 토목공학과 박사과정(E-mail: kjs0308@naver.com)

Member, Doctor's Course, Department of Civil Engineering, Inha University

**교신저자, 정회원, 인하대학교 수자원시스템연구소 박사후 연구원(Tel: +82-32-874-0069, Fax: +82-32-876-9787, E-mail: karesma0cch@naver.com)

Corresponding Author, Member, Post-Doctoral Researcher, Institute of Water Resources System, Inha University

***정회원, 경남과학기술대학교 조경학과 교수

Member, Professor, Department of Landscape Architecture, Gyeongnam National University of Science and Technology

호우에 의한 피해가 발생하였다(MOIS, 2018). 따라서 태풍에 의한 피해를 미리 파악하고 대비하는 것은 효율적인 방재업무를 위하여 대단히 중요한 일이다. 태풍으로 인한 지역별 피해양상이 다르게 나타나는데, Jun et al. (2008)은 내륙, 해안, 산지, 평지 등 지형적 특성에 따라 시군구를 구분하고, 태풍으로 인한 피해횟수와 피해액의 크기에 따라 분류하였다. 해안지역에 위치한 시군구가 많은 횟수를 나타내었고, 피해액은 백두대간에 위치한 산지 시군구와 해안지역에 위치한 시군구가 가장 큰 피해액을 나타낸 것으로 분석하였다. Lee et al. (2016)은 태풍 내습시 강풍, 호우 등에 따른 지역별 피해액의 지역적 특성과의 관계를 연구하였다.

자연재해로 인한 피해예측방법에는 손실·손상함수를 이용하여 피해를 예측하는 방법(Kim et al., 2014; Lee et al., 2017)과 수문기상학적 인자와 피해액 사이의 상관관계를 이용하여 산정하는 방법(Kim et al., 2014; Kwon and Chung, 2017; Choi et al., 2018; Choo et al., 2018; Y. Kim et al., 2018)이 있다.

수문기상학적 인자와 피해액의 관계를 유도하기 위한 방법으로 다양한 회귀분석방법이 이용되었다. Kwon and Chung (2017), Oh and Chung (2017)은 대설로 인한 피해를 예측하기 위한 모형으로 신적설심, 평균습도, 최저기온 등 수문기상인자와 농촌인구밀도, GRDP 등을 이용하여 다중선형회귀분석을 실시하였고, Choo et al. (2018)은 파랑과 관련된 풍속, 파고와 국립해양조사원에서 개발한 연안재해 관련 지수(KHOA, 2011)를 이용하여 연안지역에 위치한 시군구별 군집화를 시행하고, 이들 종속변수들과 피해액(독립변수)을 대상으로 다중선형회귀분석을 통하여 피해액산정모형을 개발하였다.

J.H. Kim et al. (2017), Y. Kim et al. (2018)은 태풍피해예측을 위하여 1-24시간 강우지속기간별 최대강우량, 총강우량, 1-5일 선행강우량, 최대풍속, 중심기압을 종속변수로 활용하였는데, 피해자료 부족을 고려하여 Leave One Out of Cross Validation (LOOCV) 방법을 적용하여 다중선형회귀분석을 시행하였다. 이 연구에서는 종속변수들간의 다중공선성 문제(지속기간별 최대강우량, 선행강우량)를 무시하고, 오직 예측력만을 고려하여 모형을 고려하였다.

한편 다중공선 문제를 해결하기 위하여 J.S. Kim et al. (2017), Kim et al. (2016)과 Choi et al. (2017)은 지속기간별 최대강우량과 총강우량, 선행강우량의 주성분분석을 시행하였고, 주성분분석 결과로 산정된 주성분과 호우피해를 관계를 도출하는 주성분회귀분석을 시행하였다. 또 이들 연구에서는 호우피해 예측함수는 피해액을 로지스틱회귀분석을 통해 호우피해액이 큰 범위와 호우피해액이 작은 범위로 분류한 혼합분포를 이용하여 개발하였다. 즉 주성분과 피해액을 로지스틱 회귀분석을 시행하여 피해가 큰 범위와 작은 범위로 구분하고, 크고 작은 2개 범위에 대하여 각각의

주성분분석을 새로 시행하고, 새로 산정된 각각의 주성분과 피해액을 회귀분석하여 호우피해함수를 개발하였다.

본 연구에서는 연안에 위치한 시군구를 대상으로 태풍피해함수를 회귀분석을 이용하여 개발하고자 한다. Jun et al. (2008) 연구에서 밝힌 바와 같이 연안지역에 위치한 시군구의 태풍피해 회수와 피해액이 클 뿐 아니라, 내륙지역에 비하여 풍랑으로 인한 피해가 추가적으로 발생하기 때문이다. 또 전술한 바와 같이, 태풍피해 중 가장 큰 피해액을 야기하는 호우의 경우에는 지속기간별 최대강우량, 선행강우량 등 호우특성치들의 다중공선성 문제를 해결하기 위하여 주성분회귀분석을 실시하였으며, 로지스틱분포를 이용하여 피해액이 큰 집단과 작은 집단으로 구분하여 피해예측함수를 개발하였다.

2. 통계적 방법론

본 연구에서는 태풍으로 인한 피해함수를 피해액을 종속변수로 하고, 각종 수문기상학적 인자를 독립변수로 하는 회귀분석모형을 이용하여 나타내고자 하며, 사용한 통계학적 방법론은 주성분 분석, 선형회귀분석, 로지스틱 회귀분석이다.

2.1 주성분 분석

주성분 분석(Principal Component Analysis)은 변수들의 공분산 행렬이나 상관행렬을 이용하여 변수들의 선형결합으로 적은 차원의 새로운 변수, 즉 주성분(Principal Component)이라고 부르는 서로 상관되어 있지 않은 새로운 변수를 찾아내는 방법을 말한다. 변수들의 전체분산에서 대부분을 포함하는 소수의 주성분을 결정하고, 이를 새로운 독립변수로 활용한다(J.S. Kim et al., 2017). 일반적으로 회귀모형의 독립변수 개수가 많으면 변수 간의 다중공선성(multi-collinearity)과 일부 중요하지 않은 변수로 인하여 모형의 예측 성능이 떨어진다. 이때 주성분 분석을 통해 도출된 소수의 주성분을 사용함으로써, 다중공선성을 해소하고 동시에 변수선택을 하는 효과를 얻을 수 있다. 특히 변수 사이의 강한 선형관계 때문에 독립변수를 선택하기 어려운 경우, 주성분 분석을 통해 소수의 주성분을 독립변수 대신에 사용하는 것이 가능하다(Choi et al., 2017).

2.2 선형회귀분석과 주성분회귀모형

선형회귀분석(Linear Regression)은 종속변수 y 에 유의한 관계를 가지는 독립변수 $x_1, x_2, \dots, x_k$ 를 이용하여, 종속변수를 설명하는 방법으로 Eq. (1)과 같은 형태로 나타낼 수 있다.

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k} + \epsilon_i \\ &= \beta_0 + \sum_{j=1}^k \beta_j x_{i,j} + \epsilon_i, \quad i = 1, 2, \dots, n \end{aligned} \quad (1)$$

Eq. (1)에서의 아래 첨자 i 는 i 번째 자료이며, 자료의 총 수는 n 개다. β_0 는 절편이며, β_j 는 j 번째 독립변수의 종속변수에 대한 회귀계수(regression coefficient)이다. ϵ_i 는 독립변수들로 설명할 수 없는 종속변수 값과의 오차(error)이다. 회귀 모형에서 오차항은 독립이며, 정규분포(normal distribution)를 따른다고 가정한다.

주성분회귀모형은 독립변수의 수가 대단히 많아 차원을 축소할 필요가 있거나, 독립변수들 간의 상관관계가 높아, 변수결정시 다중공선성이 문제될 경우에 적용할 수 있는 회귀분석방법으로, 2.1절에서 언급한 주성분분석에 의하여 산정된 주성분과 종속변수의 관계를 유도하는 것이다.

2.3 로지스틱 회귀분석

태풍 피해액의 경우 피해액의 범위가 매우 넓어서 단일 예측함수로는 태풍피해를 예측하는데 한계가 있다(J.H. Kim et al., 2017; Y. Kim et al., 2018). 본 연구에서는 로지스틱 회귀분석을 통해 태풍으로 인한 피해를 피해액이 큰 집단과 작은 집단으로 구분하고, 구분된 집단별로 적절한 회귀모형을 구축하여 예측성능을 향상시키고자 하였다. 로지스틱 회귀분석을 이용하여 피해액이 큰 집단과 작은 집단으로 분류하는 과정은 다음과 같다.

- (1) 피해액의 백분위수(10, 15, ..., 90, 95)를 기준으로 집단을 나누고, 피해액이 크면 1, 피해액이 작으면 0을 범주로 하는 범주형 변수를 생성하여 집단별 회귀모형을 생성한다.
- (2) 특정 피해액(백분위수)을 기준으로 분류한 이항(binary) 범주형 변수를 종속변수로 이용하여 로지스틱 회귀모형을 구축하였다. 구축된 로지스틱 회귀모형은 설명변수가 주어졌을 때, “피해액이 큰 집단”으로 분류될 확률 값을 제시한다.
- (3) 로지스틱 회귀모형의 학습과정에서 계산되는 검증용 데이터에 대한 Receiver Operating Characteristic (ROC) 곡선을 작성하고, 이 곡선 아래의 면적인 Area Under ROC (AUROC)를 최대화하는 확률 경계를 최적의 분류 경계값으로 선정하였다.

2.4 예측력 검증 방법

본 연구에서는 태풍피해예측함수의 예측력을 검증하기 위한 지표로 RMSE를 표준화한 NRMSE를 이용하였으며, 식은 다음과 같다.

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{y_{\max} - y_{\min}} \quad (2)$$

여기서 y_i 는 관측치, $\hat{y}_i$ 는 추정치(예측치), $y_{\max}$ 는 최대관측치, $y_{\min}$ 는 최소관측치, n 은 관측치의 총수이다.

3. 대상지역 및 피해자료 수집

태풍 피해의 특성인 호우, 강풍, 풍랑에 의한 피해를 모두 입을 수 있는 연안지역에 위치한 시군구에 대하여 태풍피해 예측함수를 개발하였다.

3.1 대상지역

본 연구에서는 부산시 기장군을 기준으로 북쪽에 위치한 연안지역 시군구를 대상으로 태풍피해자료를 수집하였으며, 강원도, 경상북도, 울산광역시, 부산광역시에 속한 총 17개 시군구가 해당되며, 시도별 분포가 Table 1에 나타나 있다.

본 연구의 대상지역의 특성은 태풍 내습시 백두대간과 태풍의 반시계방향 회전으로 인한 산악효과가 발생하는 지역이다.

3.2 태풍피해자료 수집

태풍피해자료의 수집은 행정안전부에서 발행하는 재해 연보를 이용하였으며, 기간은 1994년부터 2016년까지이다. 재해연보에서 자연재해의 원인 중 태풍과 관련된 것으로는 태풍, 태풍·호우로 총 2가지가 있으며, 본 연구에서는 선행 호우 또는 태풍 이후에 연속적으로 발생한 호우로 인한 피해를 제외하였다. 즉 태풍만으로 발생한 피해를 대상으로 하였으며, 시군구별 피해횟수와 피해액이 Fig. 1에 나타나 있다.

Table 1. Coastal Districts Used in This Study

Province	District	Province	District
Gangwondo	Gangneungsi	Gyeongsangbukdo	Ulleunggun
	Goseonggun		Pohangsi
	Donghaesi	Ulsan	Namgu
	Samcheoksi		Donggu
	Sokchosi		Bukgu
	Yangyanggun		Uljugun
Gyeongsangbukdo	Gyeongjusi	Busan	Gijanggun
	Yeongdeokgun

(a) No. of Damage (b) Amount of Damage

Fig. 1. Damage of Coastal Districts

2005년부터 2016년까지 국내에 피해를 야기한 태풍은 총 35개이며, 피해횟수는 193회이며, 최소값은 60만원에서 최대값은 태풍 ‘매미’로 인한 강릉시에 발생한 피해로 1조 805억원이었다. 과거에 발생한 피해액을 현재 가치로 변환하기 위하여 재해연보에서 사용하고 있는 생산자물가지수(Producer Price Index, PPI)와 그에 따라 금액환산계수(Conversion Index, CI)를 이용하였으며, Table 2에 제시한 바와 같다.

3.3 수문기상자료 수집

35개 태풍에 의한 193개 피해자료를 대상으로 수문기상자료를 수집하였다. 앞에서 언급한 바와 같이, 태풍은 호우, 강풍, 풍랑에 의한 피해가 중첩되어 발생하므로, 이들 요소를 모두 고려할 수 있는 변수들을 결정하고 수집하여야 한다.

호우피해와 관련하여 Choi et al. (2017)는 호우피해예측함수 개발을 위하여 1-24시간 지속기간별 최대강우량, 총강우량, 및 1-7일 선행강우량을 독립변수로 선정하였다. 본 연구

에서도 이와 동일한 독립변수를 선정하였으나, 선행강우량의 경우, NRCS 유출곡선지수법에서 5일 선행강우량을 기준으로 하므로, 1-5일 선행강우량을 채택하였다. 1994년부터 2005년까지는 ASOS 관측자료를 기준으로 Thiessen 다각형을 작성하여 시군구별 면적강우량을 산정하였으며, 이후 자료는 AWS자료를 이용하였다.

Song et al. (2016)은 경북지역에 대하여 강풍과 그로 인한 피해액의 관계를 회귀분석하였는데, 8개 시군구의 강풍피해함수에서 최대풍속과 최대순간풍속을 인자로 하는 식이 각각 4개로 비슷한 분포를 보이고 있다. 본 연구에서는 최대풍속을 강풍으로 인한 피해를 설명하기 위한 독립변수로 채택하였다.

태풍시 발생하는 풍랑은 강풍과 저기압에 의한 해수면 상승이 모두 원인이 되며, 이중 강풍은 최대풍속을 이미 선정하였으므로 저기압으로 인한 해수면 상승은 태풍이 위도 33°(마라도 인근 지점)를 지날 때의 중심기압으로 하였다.

태풍피해예측함수를 개발하는데 이용되는 독립변수는 호우관련 30개, 최대풍속, 위도 33° 위치에서의 중심기압로 총 32개 변수를 이용하였다.

4. 태풍피해예측함수의 개발

기존 자연재해로 인한 피해를 산정하기 위하여 개발된 회귀모형은 대부분 1개의 회귀분석식으로 피해액을 추정하는 경우가 많으나, 본 연구에서는 피해의 범위가 몇 십만 원에서 1조원 이상되는 넓은 범위를 가지므로 피해가 큰 집단과 작은 집단으로 구분하여 피해함수를 개발하고자 한다.

4.1 개발절차

태풍피해예측함수 개발은, 현재 가치로 환산된 피해액을 종속변수로 하고 32개의 수문기상자료를 독립변수로 하여

Table 2. Conversion Index

Year	PPI	CI	Year	PPI	CI
1994	62.87	1.6310	2006	87.67	1.1696
1995	65.81	1.5581	2007	88.93	1.1530
1996	67.94	1.5093	2008	96.53	1.0623
1997	70.52	1.4541	2009	96.33	1.0645
1998	79.13	1.2958	2010	100.00	1.0254
1999	77.49	1.3233	2011	106.71	0.9609
2000	79.09	1.2965	2012	107.45	0.9543
2001	78.72	1.3026	2013	105.73	0.9698
2002	78.49	1.3064	2014	105.17	0.9750
2003	80.20	1.2786	2015	100.95	1.0158
2004	85.07	1.2054	2016	99.11	1.0346
2005	86.88	1.1802			

개발하였다. J. Kim et al. (2018)은 회귀분석을 수행하는데 있어 설명변수들 사이에 상관관계가 강할 경우 다중공선성의 문제가 나타나며, 심각한 수준의 다중공선성은 회귀계수의 추정치와 표준오차를 왜곡시킴으로써 회귀계수의 해석과 가설검정의 결과를 왜곡할 수 있다고 하였고, 이를 방지하기 위한 통계학적 방법으로 주성분분석에 의한 주성분회귀분석과 능형회귀분석(Ridge Regression)을 제안하였다. 본 연구에서 주성분회귀분석을 이용하였으며, 태풍피해예측함수를 개발하는 절차는 다음과 같다.

- (1) 190개의 태풍피해자료의 독립변수중 상관관계가 높은 1-24시간 지속기간별 최대강우량, 총강우량(이하 강우 주성분)과 1-5일 선행강우량(이하 선행강우 주성분)을 대상으로 주성분분석을 실시하여 주성분을 도출한다.
- (2) 피해액(종속변수)과 도출된 강우 주성분과 선행강우 주성분, 최대풍속 및 중심기압을 이용하여 피해가 큰 집단과 작은 집단으로 구분하기 위한 로지스틱회귀분석을 실시하고, 두 집단의 경계값을 결정하고, 이를 기준으로 피해액 크기에 따라 두 집단으로 구분한다.
- (3) 피해액이 큰 집단과 작은 집단 각각에 대하여 강우와 선행강우에 대한 주성분분석을 실시하고, 산정된 주성분과 최대풍속, 중심기압을 이용하여 집단별 태풍피해예측함수를 개발한다.

개발된 태풍피해예측함수의 정확도를 평가하기 위하여 자료를 58개의 자료를 무작위로 선정하여 검증용 자료로 이용하고, 나머지 135개의 자료를 이용하여 태풍피해예측함수를 개발하였다.

4.2 로지스틱회귀분석과 집단 분류

전술한 바와 같이, 연안지역의 태풍피해는 수 십 만원에서

부터 1조원 이상의 피해가 발생하는 것처럼 크고, 작은 피해가 혼재해 있으며, 이를 1개의 식으로 나타내는 것은 부적절하다(Choi et al., 2017). 본 연구에서는 피해액의 크기에 따라 다른 2개의 태풍피해예측함수를 개발하였으며, 피해액의 작고, 큼을 구분하기 위한 경계를 결정할 수 있는 로지스틱회귀분석을 시행하기 위하여 강우와 선행강우에 대한 주성분 분석을 실시하였으며, 그 결과가 Table 3에 나타나 있다.

Table 3에서 Items는 ACV (accumulate covariance, 누가공분산)의 경우는 주성분의 순서를 의미하며, Coef. of PCA는 채택된 주성분 계수이다. Table 3에서 알 수 있는 바와 같이, 전체 자료로부터 추출된 주성분은 강우 주성분은 1개이며, 선행강우 주성분은 2개이고, Table 3의 주성분 계수를 이용하여 주성분을 도출하여 로지스틱회귀분석을 시행하였으며, 그 결과가 Tables 4와 5에 나타나 있다.

Table 4에서 Quantile은 피해액 분위, BVoD (boundary value of damage)는 피해액 분위에 해당하는 피해액에 대수를 취한 값, BVoP (boundary value of probability)는 피해액분위의 확률경계값, NRMSE는 피해액 분위에서의 RMSE를 표준화한 값으로, 해당지점의 NRMSE가 작으면 작을수록 해당 피해액분위 값을 채택할 때 오차가 작음을 의미한다. 피해액 분위 90%에서 가장 큰 AUROC를 가지나 관측치와 예측치의 오차정도를 나타내는 NRMSE가 80%일때가 가장 작으므로, 본 연구에서는 피해액분위 80%에 해당하는 피해액을 채택하기로 하며, 경계값은 피해액에 $\ln$ 를 취한 23.347 ($= \ln(13.787 \times 10^9)$), 즉 137.8억 정도를 기준으로 구분됨을 알 수 있다.

4.3 태풍피해예측모형 개발

태풍피해예측모형은 4.1절에서 언급한 137.8억의 피해를 기준으로 피해액이 큰 집단과 작은 집단으로 나누어 개발하였다. 태풍피해예측함수의 개발과정은 4.1절에서 언급한

Table 3. PCA Analysis

Class	Items	Logistic			Small Loss			Large Loss		
		ACV	Coef. of PCA		ACV	Coef. of PCA		ACV	Coef. of PCA	
Rain	1	94.8532	0.0327	-	91.2336	0.0394	-	97.7950	0.0237	-
	2	3.6950	0.0570	-	6.6960	0.0677	-	1.6015	0.0425	-
	3	0.9870	0.0798	-	1.5144	0.0906	-	0.3343	0.0643	-
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	23	0.0001	0.2676	-	0.0001	0.2562	-	0.0000	0.2750	-
	24	0.0001	0.2788	-	0.0001	0.2989	-	0.0000	0.2782	-
Anticipated Rain	1	89.4879	0.5981	-0.5087	89.6143	0.5983	-0.5473	88.6671	0.5947	-0.2583
	2	6.5644	0.5510	-0.1519	6.7505	0.5416	-0.1410	6.5812	0.5982	-0.1644
	3	2.2498	0.4651	0.2743	2.2062	0.4700	0.3057	2.8174	0.4368	0.0379
	4	1.2436	0.3242	0.5775	1.0784	0.3294	0.5772	1.7086	0.2966	0.5297
	5	0.4543	0.1314	0.5563	0.3507	0.1389	0.5040	0.2258	0.0991	0.7901

Table 4. Calculation of Maximum AUROC

Quantile	BVoD	AUROC	BVoP	NRMSE
50%	21.189	0.741	0.468	3.287
55%	21.345	0.734	0.366	3.655
60%	21.485	0.728	0.368	3.440
65%	21.941	0.783	0.253	3.572
70%	22.126	0.816	0.167	3.980
75%	22.858	0.875	0.117	3.870
80%	23.347	0.879	0.375	3.171
85%	23.810	0.878	0.386	3.210
90%	24.651	0.909	0.328	3.265

Fig. 2. AUROC by Quantiles**Table 5.** Typhoon Damage Prediction Model

Class	Function	NRMSE (R^2)
Logistic	$\logis(\pi) = 59.52322 + 0.00514PC_{r,1}$ $+ 0.00668P_{a,1} + 0.05572W_{max}$ $+ 0.06768Q_{press}$	
Small Loss	$\ln(y) = 7.523 + 0.00355PC_{r,1}$ $- 0.02536PC_{a,1} + 0.02202PC_{1,2}$ $+ 0.12233W_{max} + 0.00945Q_{press}$	12.4% (0.624)
Large Loss	$\ln(y) = 15.417 + 0.00118PC_{r,1}$ $- 0.01191PC_{a,1} + 0.01593PC_{1,2}$ $+ 0.10234W_{max} + 0.00649Q_{press}$	18.5% (0.431)

바와 같이, 강우 주성분과 선행강우 주성분을 분석하기 위하여 주성분분석을 실시하고, 산정된 주성분과 최대풍속, 중심기압을 독립변수로 하고 피해액을 종속변수로 하는 주성분회귀분석을 시행하였으며, 그 결과가 Table 5에 나타나 있다.

관측자료 중 검증용으로 자료를 이용하여, 큰 집단과 작은 집단의 태풍피해예측함수를 검증하였으며, Fig. 3에 나타나 있다.

Table 5에서 큰 집단과 작은 집단의 NRMSE (결정계수)가 0.624 (12.4%)와 0.421 (18.5%)로 높지 않은 것을 알 수 있는데, Fig. 3을 살펴보면 각각의 집단에서 피해액이 커질수록 추정치와 관측치 차가 큼을 알 수 있다. 피해액이 큰 집단의 경우는 강릉시의 ‘매미’와 같은 1조 이상 또는 수천억원에 달하는 피해는 일종의 이상치로 볼 수 있어, 선형회귀모형으로 설명하기에는 어려움이 있다. 또 피해액이 작은 집단은 Fig. 3(b)에서 많은 자료가 30억 미만의 피해를 가지는데 반하여 30억 이상의 피해액은 태풍피해예측함수가 잘 나타나지 못함을 알 수 있고, 이는 큰 집단과

동일한 이유로 보인다. Table 6은 검증용 자료에서 실제값(OBS)와 본 연구에서 개발한 모형을 이용한 추정치(EST)를 비교한 것이다.

5. 결 론

우리나라에서 발생하는 자연재해 중 큰 비중을 차지하는 태풍피해에 대한 대비를 위하여 연안에 위치한 시군구에 대하여 태풍피해예측함수를 개발하였다. 태풍피해는 호우, 강풍, 파랑에 의한 자연재해가 복합적으로 중첩되어 일어나는 특성이 있으므로, 이들 원인이 되는 인자들을 이용하여 태풍피해예측함수를 개발하였으며, 연구의 결과를 요약하면 다음과 같다.

- (1) 태풍피해예측함수를 개발하기 위하여 1-24시간 지속 기간별 최대홍수량, 총강우량 및 1-5일 선행강우량, 최대풍속, 태풍중심기압 등 총 32개의 인자를 이용하여 피해액을 설명하는 독립변수로 선정하였다.

(a) Large Loss

(b) Small Loss

Fig. 3. Comparison between Estimated & Observed Damage

Table 6. Verification of Estimated and Observed Data

Large Damage		Small Damage			
EST	OBS	EST	OBS	EST	OBS
59.8	14.5	1068.8	104.5	933.9	1841.8
34.1	16.3	194.2	164.2	2068.6	2034.7
19.2	21.9	433.7	190.1	1096.9	2176.1
37.2	25.8	174.6	294.1	78.3	2745.3
37.4	34.1	2203.5	404.6	4826.4	3670.3
48.1	41.3	453.2	483.7	3677.7	5122.7
22.4	50.6	61.0	557.4	5907.7	8064.3
26.5	67.4	573.6	668.9	6238.9	13011.9
247.3	91.0	298.1	971.2		
152.2	295.2	310.9	1138.7		
217.1	572.6	624.1	1601.7		

- (2) 이들 변수 중 강우관련 인자와 선행강우 관련 인자는 상관관계가 높아 다중공선성 문제를 일으킬 수 있고, 총 32개의 독립변수도 차원의 문제를 야기한다. 이를 해소하기 위하여 주성분분석을 강우와 선행강우에 대하여 각각 실시하여 차원축소와 다중공선성 문제를 해소하였다.
- (3) 최소 수 십 만원에서 1조원 이상의 피해를 기록하는 연안지역의 태풍피해를 설명하기 위하여 로지스틱회귀분석을 이용하여 피해액이 큰 집단과 작은 집단으로 구분하였으며, 두 집단 구분의 경계값은 137.8억으로 피해액분위 80%정도 해당하는 값이다.
- (4) 피해액이 큰 집단의 태풍피해함수에 대한 검증에서 $NRMSE = 12.2\%$ ($R^2 = 0.624$)로 나타났으며, 피해액이 작은 집단은 $NRMSE = 18.5\%$ ($R^2 = 0.421$)로 나타나, 피해액이 큰 집단의 더 좋은 결과를 나타내었다.

태풍피해예측함수 개발과 관련된 연구의 한계는 자료의 일관성부족과 자료부족을 들 수 있다. 방재기상관측소(AWS)의 본격적인 설치는 2005년부터이며, 본 연구에서 이용한 35개 태풍 중 2005년 이후 발생하여 영향을 미친 태풍은 17개에 불과하다. 피해자료를 확보하기 위하여 2005년 이전에 발생한 태풍에 대하여서는 중관기상관측소(ASOS)의 자료를 이용하였으므로, 자료의 일관성이 부족한 것이 사실이다. 또 최대 피해를 입힌 태풍으로 알려져있는 ‘루사(2002)’, ‘매미(2003)’의 경우는 중관기상관측소의 자료를 이용하였으므로, 정확도가 떨어질 것으로 판단되며, 이는 회귀모형에 큰 영향을 미칠 수 있다. 다음으로 방재관련 시스템이 발전함에 따른 방재관련 예방, 복구 사업의 효과를 반영하기 어렵다. 자연재해대책법 등에 따라 다양한 예방사업과 재해후 복구사업이 이루어져 재해위험은 감소하고 있다. 그러나 태풍피해예측함수에서 방재관련 사업을 반영하지 않았으므로, 이를 고려할 수 있는 방안을 모색하여야 하나, 관련 자료를 수집할 수 있는 기간이 짧아 과거의 태풍에 적용하기 어려운 점이 있다.

감사의 글

본 연구는 행정안전부 재난예측 및 저감연구개발사업의 지원을 받아 수행된 연구임(MOIS-재난-2015-05).

References

Choi, C., Kim, J., Lee, M., Kim, J., Lee, W., and Kim, H.S. (2017). Development of heavy rain damage prediction function using principal component analysis and logistic regression model. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 6, pp. 159-166.

Choi, C.H., Kim, J.H., Kim, J.S., Kim, D.H., Bae, Y.H., and Kim, H.S. (2018). Development of heavy rain damage prediction model using machine learning based on big data. *Advances in Meteorology*, Vol. 2018, Article ID 5024930. doi:10.1155/2018/5024930

Choo, T.H., Kim, Y.S., Sim, S.B., and Son, J.K. (2018). Development of the wind wave damage predicting functions in southern sea based on annual disaster reports. *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 19, No. 2, pp. 668-675.

Jun, H.D., Park, M.J., Kim, G.Y., and Kim, J.W. (2008). Damage analysis of meteorological disasters for each district considering the characteristics of a district. *J. Korean Soc. Hazard Mitig.*, Vol. 8, No. 2, pp. 75-82.

Kim, J., Choi, C., Oh, S., Han, D., and Kim, H.S. (2016). Development of heavy rain damage prediction functions in Gyeonggi Province. *Proceedings of 2016 Conference*, The Korean Society of Disaster Information, pp. 203-204.

Kim, J., Park, J., Choi, C., and Kim, H.S. (2018). Development of regression models resolving high-dimensional data and multicollinearity problem for heavy rain damage data. *Journal of the Korean Society of Civil Engineers*, Vol. 38, No. 6, pp. 801-808.

Kim, J.H., Kim, T., and Lee, B.R. (2017). An analysis of typhoon damage pattern type and development of typhoon damage forecasting function. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, pp. 339-347.

Kim, J.S., Choi, C., Lee, J., and Kim, H.S. (2017). Damage prediction using heavy rain risk assessment: (2) Development of heavy rain damage prediction function. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, pp. 371-379.

Kim, S.H., Kim, B.S., Lee, C.H., and Chung, J.H. (2014). Development of depth-damage function by investigating flooded area with focusing on building damage. *Journal of Korea Water Resources Association*, Vol. 47, No. 8, pp. 717-728.

Kim, Y., Kim, T., and Lee, B. (2018). Development of typhoon damage prediction function using Tukey's ladder of power transformation. *J. Korean Soc. Hazard Mitig.*, Vol. 18, No. 6, pp. 259-267.

Korea Hydrographic and Oceanographic Agency (KHOA). (2011). *Coastal disaster assessment system*.

Kwon, S.H., and Chung, G. (2017). Estimation of snow

- damages using multiple regression model: The case of Gangwon Province. *Journal of the Korean Society of Civil Engineers*, Vol. 37, No. 1, pp. 61-72.
- Lee, C.H., Kim, S.H., Hwang, S.B., and Kim, G.H. (2017). A study on development of flood depth-damage functions focused on school buildings. *J. Korea Water Resour. Assoc.*, Vol. 50, No. 8, pp. 513-520.
- Lee, S.W., Ahn, S.H., Lim, B., and Choi, G. (2016), Relationships between intensity of extreme climate events and magnitude of damages for different typhoon tracks in the Republic of Korea. *Journal of The Korean Association of Regional Geographers*, Vol. 22, No. 2, pp. 450-465.
- Ministry of the Interior and Safety (MOIS). (2018). *Statistical yearbook of natural disaster 2017*.
- Oh, Y.R., and Chung, G. (2017). Estimation of snow damage and proposal of snow damage threshold based on historical disaster data. *Journal of the Korean Society of Civil Engineers*, Vol. 37, No. 2, pp. 325-331.
- Song, C.Y., Lee, H.J., and Lee, C.J. (2016). A simple regression model for predicting the wind damage according to correlation analysis between wind speed and damage: Gyeongsangbuk-do. *Proceedings of 2016 Conference*, The Korean Society of Disaster Information, pp. 207-211.

Received	October 11, 2019
Revised	October 15, 2019
Accepted	November 11, 2019