위험 등급 분류를 통한 호우피해 예측함수 개발

Development of a Heavy Rain Damage Prediction Function by Risk Classification

김동현* · 최창현** · 김종성*** · 주홍준**** · 김정욱***** · 김형수*****

Kim, Donghyun*, Choi, Changhyun**, Kim, Jongsung***, Joo, Hongjun****, Kim, Jungwook*****,
and Kim, Hung Soo*****

Abstract

In this study, risks were classified into Red Zone, Orange Zone, Yellow Zone, and Green Zone using investment factors and socioeconomic factors. We developed a heavy rain damage prediction function considering the risk class. In order to develop the heavy rain damage prediction function, the total amount of damage due to heavy rain was used as the dependent variable, and the hydrological weather data, investment factors, and socioeconomic factors were used as the independent variables. A multiple regression model and artificial neural network model were used as the statistical prediction methods. In order to evaluate the predictive power, we set the learning period as 2005 to 2012 and the test period as 2013 to 2016. The NRMSE was estimated to be about 12%-13%, and the predictive power was the highest in the artificial neural network model considering the risk class. Based on the results of this study, if we identify the characteristics of different cities and districts and conduct disaster management in advance, then is possible to reduce the damage caused by natural disasters.

Key words : Investment Factor, Risk Class Classification, Statistical Prediction Techniques, Heavy Rain Damage

요 지

본 연구에서는 한강 권역을 대상으로 기 투자 요소 및 사회·경제적 요소 등을 활용하여 Red Zone, Orange Zone, Yellow Zone, Green Zone으로 위험 등급을 분류하였고, 위험 등급을 고려한 호우피해 예측함수를 개발하였다. 호우피해 예측함수를 개발하기 위하여 종속변수로는 호우로 인한 총 피해액, 독립변수로는 수문기상 자료, 기 투자 요소 그리고 사회·경제적 요소 등을 활용하여 구축하였다. 통계적 예측 기법으로는 다중회귀 모형, 인공신경망 모형을 활용하였다. 예측력을 평가하기 위하여 2005년부터 2012년까지는 학습구간, 2013년부터 2016년까지는 평가구간으로 구분하였다. 예측력 평가 결과 NRMSE는 약 12%~13%로 나타났으며, 위험등급을 고려한 인공신경망 모형에서 가장 높은 예측력을 보였다. 본 연구의 결과를 바탕으로 각기 다른 시군구별 특성을 파악하고 사전 대비차원에서 재난관리를 실시한다면, 자연재해의 피해를 경감시킬 수 있을 것으로 판단된다.

핵심용어 : 기 투자 요소, 위험 등급 분류, 통계적 예측 기법, 호우피해

*정회원, 인하대학교 토목공학과 석사과정(E-mail: yesdktpdi@naver.com)

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

**정회원, 인하대학교 토목공학과 박사수료

Member, Ph.D. Candidate, Department of Civil Engineering, Inha University

***정회원, 인하대학교 토목공학과 박사과정

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

****정회원, 인하대학교 토목공학과 박사수료

Member, Ph.D. Candidate, Department of Civil Engineering, Inha University

*****정회원, 인하대학교 토목공학과 박사수료

Member, Ph.D. Candidate, Department of Civil Engineering, Inha University

*****교신저자, 정회원, 인하대학교 사회인프라공학과 교수(Tel: +82-32-874-0069, Fax: +82-32-876-9787, E-mail: sookim@inha.ac.kr)

Corresponding Author, Member, Professor, Department of Civil Engineering, Inha University

1. 서 론

최근 자연재난으로 인한 피해는 급증하고 있으며, 국내의 자연재난의 원인 중 호우로 인한 피해는 절반가량을 차지하고 있다(Ministry of the Interior and Safety, 2018). 자연재난으로 인한 피해를 경감시키기 위해서는 방재시설 및 재해 강도, 재해 빈도 등의 재해 특성을 파악하는 것이 선행되어야 한다(Kim et al., 2018). 또한, 지역마다 방재 대응 능력이나 재해가 발생한 이후의 대응 능력 등을 반영한 사전 대비차원의 재난 관리가 필요한 실정이다.

효율적인 재난 관리를 위하여 재해 예측에 대한 국외 연구를 살펴보면 Defu Liu et al. (2009)은 1950년부터 2004년까지의 중국 전역의 태풍 피해액과 허리케인의 풍속 등을 고려하여 Multivariate Compound Extreme Value Distribution (MCEVD)라는 새로운 모형을 제안하였다. 제안된 모형을 활용하여 해안 지역, 해양 구조물 및 하구 도시에 대한 예방 기준을 제시하였다. Mendelsohn and Saher (2011)은 기상 관측 자료 및 사회·경제적 요소 등을 설명변수로 사용하여 전 세계 재난피해에 대하여 재난피해를 예측하였다. Kung et al. (2012)은 대만 전역에서 발생하는 debris-flow에 대하여 예측 모델을 설계하였다. 실제 피해가 일어난 지역의 기상 자료를 이용하였고, 설계된 Realtime Mobile Debris Flow Disaster Forecast System ($RMDF^2$)을 활용하여 재난 피해를 예측하였다. Mumane and Elsner (2012)은 1900년대부터 2000년대까지의 미국 전역에서 발생한 허리케인 피해액과 풍속 자료를 활용하여 분위회귀분석(quantile regression analysis)을 실시하였다. 이를 바탕으로 2011년에 발생되어진 Hurricane Irene 허리케인의 피해액을 예측할 수 있었다.

효율적인 재난 관리를 위하여 재해 예측에 대한 국내 연구를 살펴보면 Chung et al. (2005)은 1990년부터 2000년까지 침수피해 정보를 바탕으로 유역특성과 강우량을 변수로 활용하였다. 이러한 변수를 활용하여 유역 내 침수피해에 대한 예측 모형을 개발하였다. Jang and Kim (2009)은 홍수피해와 강우량을 활용하여 비선형회귀분석을 실시하였고, 회귀식의 계수를 사회·경제적 인자들로 나타냈다. Kim, Choi, Lee, et al. (2017b)은 지역별 호우 위험등급에 따른 피해 예측 함수를 개발하였고, 사용된 변수로는 재해기간별 총강우량, 재해일수, 강우자료 등을 활용하였다. Kim, Choi, Kim, et al. (2017)은 호우로 인한 피해를 예측하기 위하여 수문기상 자료들을 분석하고, 통계적 기법으로는 다중회귀 모형, 주성분 회귀 모형, 인공신경망을 이용하여 호우피해 예측함수를 개발하였다. Choi, Kim, et al.(2017)은 선행강우량, 총 강우량, 지속시간별 최대강우량 등을 이용하여 로스틱 회귀모형을 통하여 피해액이 큰 집단과 피해액이 작은 집단으로 분류하였다. 또한, 주성분 분석을 통하여 새로운 설명변수를 도출하였고, 호우피해 예측함수를 개발하였다. Choi, Park, et al. (2017)은 경기도 지역의 호우로 인한 공공시

설물에 대하여 피해를 예측하는 함수를 개발하였다. 호우피해 예측함수를 개발하기 위하여 기상요소 및 사회·경제적 요소를 고려하여 3가지의 머신러닝 기법을 활용하였다. Kwon et al. (2017), Oh and Chung (2017), Choo, Cho, et al. (2018), Choo, Kim, et al. (2018)은 기상 요소 및 사회·경제적인 요소를 고려하여 국내 자연재해 중 대설피해 및 풍랑 피해에 대한 예측함수를 개발하였다.

국내·외 선행 연구를 살펴보면 주로 기상 요소와 사회·경제적인 요소를 활용하여 지역적 특성을 분석한 연구가 수행되어 왔다. 효율적 재난관리를 위해 수문·기상 요소 및 기투자 요소 등을 고려하고, 세분화된 지역적 특성을 반영한 위험 등급을 고려한 예측 기법에 대한 연구는 미흡한 실정이다. 따라서 본 연구에서는 국내 자연재난으로 인한 피해 중 약 50% 이상을 차지하고 있는 호우 피해를 대상으로 세분화된 지역적 특성을 분석하여 위험 등급을 제시하고 위험 등급을 반영한 호우피해 예측함수를 개발하고자 하였다.

이를 위해 호우 피해에 영향을 미치는 요소들을 선정하였고, 호우피해 위험 지표로 정의하였다. 각 지표들의 상대적인 크기를 비교할 수 있도록 표준화하였고, 이렇게 정의된 지표를 활용하여 위험 평가 기법중 하나인 Pressure-State-Response (PSR) 모형을 활용하여 4가지 Red Zone, Orange Zone, Yellow Zone, Green Zone으로 위험 등급을 분류하였다. 통계적 예측 기법인 다중회귀 모형 및 인공신경망 모형을 활용하였다. 따라서 위험 등급 별 호우피해를 예측하기 위하여 위험 등급을 고려하지 않은 다중회귀 모형, 인공신경망 모형과 위험 등급을 고려한 다중회귀 모형, 인공신경망 모형의 예측력 평가 결과를 비교하였다.

2. 호우피해 위험 평가 방법 및 호우피해 예측함수 개발을 위한 모형

2.1 위험 평가 방법

PSR 모형은 평가 지표를 구분하는 방법으로 경제 협력 개발기구(Organization for Economic Co-operation and Development, OECD, 1993)에서 개발되어 지표설정에 주로 활용되고 있다. PSR 모형은 각각 압력지수(Pressure), 상태지수(State), 반응지수(Response) 구성에 의한 평가지표 선정을 보여주고 있다. PSR 모형은 다음 Eq. (1)과 같다.

$$PSR = \frac{Pressure \times State}{Response} \quad (1)$$

여기서, PSR 모형에서 사용되고 있는 정의를 살펴보면, 압력(Pressure)은 다양한 자연적 과정을 통해 전달되고 변형되어 환경 조건의 변화를 나타낸다. 상태(State)는 특정 영역의 물리 현상, 화학적 현상을 뜻한다. 반응(Response)은 사회의 반응과 환경 상태의 변화를 방지, 보완, 개선 또는 적응을 뜻한다.

2.2 자료의 표준화 방법

통상적으로 지표를 이용하여 지수를 계산하는 경우, 자료의 표준화(normalization) 과정이 필요하다. 각 지표들을 이용하여 지수를 계산하는 경우 수집된 자료들의 통계단위와 범위가 매우 다르기 때문이다. 각 지표 및 지수들을 상대적인 크기를 비교할 수 있도록 단위가 없는 무차원 값으로 변환하여 자료를 적절한 형태로 표준화하는 작업이 선행되어야 한다(Kim, Choi, Lee, et al, 2017a). 선행연구들을 살펴보면, 여러 가지 지표를 표준화 하는 방법 중 보편적으로 Z-score, Re-scaling, T-score 방법을 많이 활용하였다(Nardo et al, 2005; Kim, Choi, Lee, et al, 2017a; Choi et al, 2018; Kim et al, 2018). Z-score 방법은 모든 자료들의 평균을 0, 표준편차는 1이 되도록 표준화를 시켜서, 해당 자료의 수치가 평균으로부터 표준편차의 몇 배 정도 떨어져 있는지 나타내는 방법이다(Nam and Kim, 2013). Z-score 방법은 Eq. (3)과 같이 표현 된다. Re-scaling 방법은 변수들의 최대값과 최소값을 이용하는 방법으로 변수의 최소값과 최대값을 이용하여 0~1의 범위를 가지는 값으로 표준화 된다(Kim, 2016). Re-scaling 방법은 Eq. (4)와 같이 표현 된다. T-score 방법은 주로 20~80 사이의 값을 갖게 되어 100점 단위와 유사한 결과를 얻게 된다(Song et al., 2013). T-score 방법은 Eq. (5)와 같이 표현된다.

$$Z\text{-score} = \frac{X_i - \mu}{\sigma} \quad (3)$$

$$Re\text{-scaling} = \frac{X_i - \min}{\max - \min} \quad (4)$$

$$T\text{-score} = 10 \times \left(\frac{X_i - \mu}{\sigma} \right) + 50 \quad (5)$$

여기서, X_i 는 해당 변수의 값, μ 는 평균, σ 는 표준편차를 의미한다.

2.3 엔트로피를 이용한 가중치 산정 방법

지표 및 지수에 사용되는 관측자료 및 기초 통계자료를 종합하여 산정할 경우 상대적 중요도 즉 가중치(weighting)를 고려하게 된다(Kim, 2013). 엔트로피 방법은 일반적으로 불확실성의 척도로 알려져 있고, 정보이론에서는 신호가 가지고 있는 정보용량으로 정의되고 있다(Shannon and Weaver, 1949). 엔트로피 이론에서는 정보량을 정량화하기 위하여 최대 엔트로피, 조건 엔트로피, 결합 엔트로피 및 정보 전달량의 개념을 사용한다. 최대 엔트로피 $H(X)$ 는 X 가 가지고 있는 불확실성 또는 정보량을 의미하며, 무작위 이산변수 X 에 대해 최대엔트로피 $H(X)$ 는 다음 Eq. (6)과 같다(Lee, et al., 2016).

$$H(X) = - \sum_{\forall x}^N p(x) \cdot \ln p(x) \quad (6)$$

보편적으로 이산확률 변수 X 가 균일 분포 일 때 엔트로피는 최대가 된다. 이는 불확실성 감소의 최대화된 엔트로피의 최대화를 의미한다. 엔트로피 방법을 이용하여 지표별 가중치를 산정하기 위한 절차로는 첫째, 각 지표의 구축된 값을 행렬로 구성하고 구성된 지표들을 표준화한다. 둘째, 표준화된 지표를 이용하여 엔트로피를 산정한 후 최종적으로 지표들 간의 가중치를 결정한다. 의사결정 문제는 다음 Eq. (7)과 같이 나타낼 수 있으며, 모든 지표들에 대해서 표준화한 결과를 p_{ij} 라고 하면, p_{ij} 는 Eq. (8)과 같이 표현된다. 또한 각 지표의 엔트로피를 E_j 라고 하면, 엔트로피는 Eq. (9)로 나타내어진다. 지표의 가중치를 구하기 위하여 다양성 정도 d_j 를 사용하고, 다양성 정도는 $d_j = 1 - E_j$ 와 같이 계산된다. 각 지표에 대하여 표준화한 것이 Eq. (10)과 같으며 이에 대한 속성이 가중치(w_j)이다.

$$D = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{m1} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix} \quad (7)$$

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (8)$$

($i = 1, 2, \dots, m; j = 1, 2, \dots, n$)

$$E_j = -k \sum_{i=1}^m p_{ij} \log p_{ij} \quad (9)$$

($k = \frac{1}{\log m}; j = 1, 2, \dots, n$)

$$w_j = \frac{d_j}{\sum_{j=1}^n d_j} \quad (j = 1, 2, \dots, n) \quad (10)$$

2.4 다중회귀 모형

다중회귀 모형은 독립변수가 2개 이상으로 여러 개의 독립변수의 값을 이용하여 종속변수의 값을 예측하는 통계적 기법을 말한다(Kim, Choi, Lee, et al., 2017b). 다중회귀 모형은 선형함수라고 가정하여 종속변수 y 에 대하여 n 개의 독립변수 $x_1, x_2, x_3 \dots x_n$ 의 관계를 다음 Eq. (11)과 같이 나타낸다.

$$y_i' = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i + \epsilon_i \quad (11)$$

($i = 1, 2, \dots, n$), $\epsilon_i \sim n(0, \sigma^2)$)

여기서, β_0 은 상수, $\beta_1, \beta_2, \dots, \beta_i$ 은 독립변수의 회귀계수, ϵ_i 은 종속변수를 예측할 때 발생하는 오차를 뜻한다.

회귀모형에서는 표본의 수가 적거나 독립변수들 간의 상관관계가 높을 경우 다중공선성의 문제가 발생한다. 다중공선성을 판별하기 위하여 분산 분산팽창요인(Variance Inflation Factor, VIF)을 통해 평가하며, 통상적으로 VIF가 10 이상일 때 다중공선성이 높다고 판단한다. 다중공선성을 해결하기 위하여 통계적 유의성이 높은 변수를 추출할 수 있는 변수선택법을 고려할 수 있다(Kim, Choi, Kim, et al., 2017). 변수선택법으로는 단계적 선택법(Stepwise), 전진 선택법(Forward), 후방 소거법(Backward)이 있다.

2.5 인공지능망

인공신경망은 인간의 뇌가 수많은 신경들로부터 입력과 출력의 신호를 전달하는 과정을 착안하고 시냅스의 결합으로 네트워크를 형성한 인공 뉴런이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반을 가리킨다(Kim et al., 2017). 인공신경망의 중요한 특징은 학습(training)을 통하여 예측인자(predictor)와 예측변수(predictand) 사이의 관계를 학습하고, 예측인자와 예측변수 사이의 관계를 추정한다(Kang et al., 2008). 인공신경망의 구조는 예측하고자 하는 변수가 입력되어지는 최하위의 층인 입력층(input layer), 외부와 직접연결이 이루어지지 않는 은닉층(hidden layer), 신경망에 의해서 예측결과를 출력해 주는 출력층(output layer)으로 구분된다(Choi et al., 2009). 다음 Fig. 1에서 인공신경망의 구조를 도식화 하였다.

Fig. 1. The Structure of Artificial Neural Network

3. 대상지역 선정 및 입력 자료 구축

3.1 대상지역 선정

1994년부터 2016년까지의 권역별 호우피해 횟수를 분석하고 호우로 인한 총 피해금액을 살펴본 결과 한강 권역이 가장 많은 호우 피해가 발생하였고, 피해 금액도 가장 큰

것으로 나타났다(Table 1). 따라서 본 연구에서는 한강 권역을 대상지역으로 선정하였고, 한강 권역은 강원도(17개), 경기도(31개), 서울특별시(25개), 인천광역시(10개), 충청북도(3개) 시군구로 이루어져 있으며, 총 86개의 시군구를 포함하고 있다.

Table 1. Number of Damage and Total Damage for Each River Zone (1994~2016)

Each river zone	Number of Damage	Total Damage (1,000,000 won)
Han River	1,442	5,278,078
Nakdong River	824	1,922,116
Geum River	726	1,338,162
Seomjin and Yeongsan River	590	569,008

3.2 호우피해 위험 평가 관련 자료 구축

호우피해에 대한 위험 평가를 위해서는 PSR에 따른 세부 지수에 대한 정의가 필요하며, 각 세부 지수를 구성하기 위한 지표들을 구축하는 과정이 가장 먼저 이루어져야 한다. PSR 관련 위험 평가를 수행한 선행 연구사례(C. Giupponi, 2008; Chung et al., 2008; Kim et al., 2018)를 통해 호우에 대한 압력 지수, 상태 지수, 대책 지수를 다음과 같이 정의하였다. 압력 지수는 호우에 직접적으로 영향을 주는 요소로 정의하였다. 압력 지수에 구성된 지표들은 평균기온($^{\circ}\text{C}$), 연 최저기온($^{\circ}\text{C}$), 연 최고기온($^{\circ}\text{C}$), 연 최대 순간 풍속(m/s), 1일 강수량이 나타난 날 등으로 분류하여 자료를 구축하였다. 상태 지수는 호우에 의해 예상되는 상태 또는 상태의 변화가 예상되는 요소로 정의하였다. 상태 지수에 구성된 지표들은

Table 2. Index Constituting the PSR

Classification	No.	Index
Pressure Index	1	Average temperature
	2	Maximum temperature
	3	Minimum temperature
	4	Maximum instantaneous wind velocity
	5	The day when precipitation appeared
State Index	1	Fiscal self-reliance ratio
	2	Urbanization rate
	3	Gross Regional Domestic Product
	4	Property tax
Response Index	1	Total length of river levee
	2	Recovery cost
	3	Stream improvement rate
	4	Number of drainage pump stations
	5	Drainage pump capacity

제정자립도(%), 도시화율(%), 지역내 총생산(GRDP: Gross Regional Domestic Product), 재산세 등으로 분류하여 자료를 구축하였다. 반응 지수는 호우에 대한 방재 대응 요소로 정의하였다. 반응지수에 구성된 지표들은 하천 제방 총 길이, 복구비, 하천개수율, 배수펌프장 개수, 배수펌프장 처리능력 등으로 분류하여 자료를 구축하였다. PSR의 세부 지수를 구성하는 지표를 Table 2와 같이 나타냈다.

3.3 자료의 표준화

구축한 지표들은 각자의 단위가 모두 다르기 때문에 직접적인 연산과정이 불가능하다. 따라서 다양한 세부지표들의 직접적인 연산을 위해 표준화를 수행하여 단위를 무차원화하였다. 본 연구에서는 표준화에서 주로 활용되는 Z-Score, Re-scaling, T-Score방법을 고려하고자 하였다. 아래 Table 3에서 지표들을 3 가지 방법으로 기 투자 요소 중의 하나인 복구비를 활용하여 표준화한 결과를 예시로 나타냈다.

Table 3. Three Standardization Methods

City	Z-Score	Re-scaling	T-Score
Gangneung	0.1303	0.0701	51.31
Goseong	-0.1097	0.0362	48.91
Donghae	-0.2894	0.0109	47.11
⋮	⋮	⋮	⋮

3가지 방법을 통해 표준화를 모두 적용한 결과 T-Score 방법은 모두 양수의 값이 도출되었으나 Z-Score방법과 Re-scaling에서는 음수의 값과 0의 값이 함께 도출되었다. 그러나 가중치 산정을 위하여 표준화된 지표들의 연산과정이 필수적인데, 음수 값과 0의 값을 연산과정에 적용할 경우 가중치가 올바르게 도출되지 않을 것이라 판단하였다. 따라서 향후 엔트로피를 이용한 가중치 산정에서는 Z-Score방법과 Re-scaling 방법은 제외하고, 100점 단위와 유사한 결과를 도출 할 수 있는 T-Score방법을 적용하였다.

3.4 엔트로피를 이용한 가중치 산정

엔트로피 방법을 통한 가중치 산정 절차는 먼저 압력지수, 상태지수, 대책지수를 구성하는 지표들을 통해 각 세부 지수를 산정한다. 그리고 그 지수들을 활용하여 호우피해 지수를 구성하는 가중치를 산정한다. T-Score 방법으로 표준화된 지표들을 활용하였고, 압력지수, 상태지수, 대책지수를 산정하기 위한 가중치 산정 결과는 아래 Table 4와 같이 표현된다.

압력지수를 구성하는 지표의 경우 1일 강수량이 나타난 날과 상태지수를 구성하는 지표의 경우 GRDP, 재산세가 높았으며 반응지수를 구성하는 복구비, 배수펌프장의 처리능력의 가중치가 높았다. 세부 지수간의 가중치는 압력지수, 상태지수, 반응지수 순으로 가중치가 높았다.

Table 4. The Weight of the Index Constituting the PSR

Index	Weight	P-S-R Index	Weight
1_Pressure	0.128	Pressure Index	0.371
2_Pressure	0.089		
3_Pressure	0.207		
4_Pressure	0.153		
5_Pressure	0.423		
1_State	0.086	State Index	0.328
2_State	0.033		
3_State	0.424		
4_State	0.376		
1_Response	0.078	Response Index	0.301
2_Response	0.335		
3_Response	0.010		
4_Response	0.165		
5_Response	0.412		

3.5 PSR 모델을 활용한 지역별 위험 등급화

한강 권역은 총 86개의 시군구로 구성되며, 각 시군구의 피해 금액과 피해 빈도가 상이하게 분포되고 있다. 이는 86개 시군구 마다 특성이 다르기 때문이며, 지역적 특성을 고려한다면 호우피해 예측 모형의 정확도가 높아질 것이라 판단된다. 따라서 본 연구에서는 PSR 모형을 활용하여 한강 권역에 대하여 위험한 순서대로 Red Zone, Orange Zone, Yellow Zone, Green Zone으로 위험 등급을 부여하였다. 부여한 위험 등급을 활용하여 범주형 변수로 설정하였다. 다음 Table 5와 같이 각 시군구별 위험 등급을 나타내었다.

3.6 호우피해 예측 함수 개발을 위한 입력 자료 구축

호우피해 예측 함수를 개발하기 위하여 종속변수와 독립변수를 설정하고 해당하는 자료를 수집하여 입력 자료를 구축하였다. 종속변수로는 호우 총 피해액으로 설정하였다. 호우피해 횟수 및 호우피해금액을 분석하기 위하여 행정안전부에서 매년 발간되어 지고 있는 재해연보 자료를 활용하였다. 재해연보의 체계화된 자료는 1994년부터 제공되고 있으며, 본 연구에서는 2005년부터 2016년까지의 한강 권역에 해당하는 86개 시군구의 호우피해액 자료를 수집하였다. 또한, 2005년부터 2016년까지의 화폐가치가 상이하기 때문에 재해연보에서 제공되어지는 생산자 물가지수를 고려하여 모든 기간의 피해금액을 2016년 기준으로 환산하였다.

독립변수로는 Kim, Choi, Lee, et al. (2017b); Kim, Choi, Kim, et al.(2017); Choi, Kim, et al. (2017); Choi, Park, et al. (2017) 선행 연구 사례를 통하여 재해일수(date) 선행강수량(1day~7day), 지속시간별 최대강수량(1_x~24_x), 시군구 면적(area), 재해기간의 총 강수량(tot)의 변수를 활용하였으

Table 5. Results of PSR Analysis for Distribution of Zones in Han River Basin

City	PSR value	Rank	Zone	City	PSR value	Rank	Zone
Gangneung	18.31204	40	Orange Zone	Paju	12.95437	78	Green Zone
Goseong	15.97273	63	Yellow Zone	Pyeongtaek	17.94321	44	Orange Zone
Donghae	17.81069	47	Yellow Zone	Pocheon	17.93975	45	Yellow Zone
Samcheok	16.41891	61	Yellow Zone	Hanam	17.21328	55	Yellow Zone
Sokcho	18.0087	43	Orange Zone	Hwaseong	17.05521	56	Yellow Zone
Yanggu	13.83237	75	Green Zone	Gangnam	31.53558	4	Red Zone
Yangyang	14.80406	69	Green Zone	Gangdong	35.96159	1	Red Zone
Yeongwol	15.2809	66	Green Zone	Gangbuk	24.24275	12	Red Zone
Wonju	17.37415	52	Yellow Zone	Gangseo	19.19092	34	Orange Zone
Inje	10.96547	84	Green Zone	Kwanak	21.28392	21	Red Zone
Jeongseon	12.18377	80	Green Zone	Gwangjin	20.93074	25	Orange Zone
Cheorwon	14.27649	71	Green Zone	Guro	20.38371	29	Orange Zone
Chuncheon	15.212	67	Green Zone	Geumcheon	21.56798	19	Red Zone
Pyeongchang	14.10008	73	Green Zone	Nowon	21.54261	20	Red Zone
Hongcheon	10.11964	85	Green Zone	Dobong	20.99162	24	Orange Zone
Hwacheon	14.27834	70	Green Zone	Dongdaemun	19.72763	33	Orange Zone
Hoengseong	14.85056	68	Green Zone	Dongjak	21.96308	17	Red Zone
Gapyeong	14.02403	74	Green Zone	Mapo	19.72879	32	Orange Zone
Goyang	11.78894	81	Green Zone	Seodaemun	25.70791	10	Red Zone
Gwacheon	23.73832	13	Red Zone	Secho	19.08173	36	Orange Zone
Gwangmyeong	16.84733	57	Yellow Zone	Seongdong	32.93476	2	Red Zone
Gwangju	17.46804	51	Yellow Zone	Seongbuk	22.52874	15	Red Zone
Guri	17.77753	48	Yellow Zone	Songpa	20.3316	30	Orange Zone
Gunpo	17.54248	50	Yellow Zone	Yangcheon	25.09069	11	Red Zone
Gimpo	16.23747	62	Yellow Zone	Yeongdeungpo	21.07614	23	Orange Zone
Namyangju	19.07541	37	Orange Zone	Yongsan	29.28141	5	Red Zone
Dongducheon	17.29582	53	Yellow Zone	Eunpyeong	23.64052	14	Red Zone
Bucheon	15.73093	65	Yellow Zone	Jongno	20.84831	27	Orange Zone
Seongnam	22.46123	16	Red Zone	Junggu	28.46139	6	Red Zone
Suwon	26.40013	9	Red Zone	Jungnang	32.04878	3	Red Zone
Siheung	26.7092	8	Red Zone	Ganghwa	17.92332	46	Yellow Zone
Ansan	20.1775	31	Orange Zone	Gyeyang	16.80282	59	Yellow Zone
Anseong	20.92961	26	Orange Zone	Namgu	17.25281	54	Yellow Zone
Anyang	17.62244	49	Yellow Zone	Namdong	18.3657	39	Orange Zone
Yangju	21.21708	22	Red Zone	Donggu	21.80904	18	Red Zone
Yangpyeong	16.84272	58	Yellow Zone	Bupyeong	9.311433	86	Green Zone
Yeoju	13.44279	77	Green Zone	Seogu	18.67142	38	Orange Zone
Yeoncheon	13.79731	76	Green Zone	Yeonsu	20.65617	28	Orange Zone
Osan	14.15556	72	Green Zone	Ongjin	19.13225	35	Orange Zone
Yongin	16.74229	60	Yellow Zone	Bupyeong	18.29289	42	Orange Zone
Uiwang	27.59631	7	Red Zone	Danyang	11.296	83	Green Zone
Uijeongbu	12.6952	79	Green Zone	Jecheon	18.30221	41	Orange Zone
Icheon	15.7506	64	Yellow Zone	Chungju	11.50251	82	Green Zone

며, PSR 모형에 구성된 지표(14개)와 위험 등급(zone)으로 독립변수 총 48개를 구축하였다. 구축된 독립변수를 다음 Table 6과 같이 나타내었다.

Table 6. Dependent Variable and Independent Variable

Dependent variable			
Heavy rain damage			
Independent variable			
1_Pressure	4_Response	3_x	15_x
2_Pressure	5_Response	4_x	16_x
3_Pressure	date	5_x	17_x
4_Pressure	1_day	6_x	18_x
5_Pressure	2_day	7_x	19_x
1_State	3_day	8_x	20_x
2_State	4_day	9_x	21_x
3_State	5_day	10_x	22_x
4_State	6_day	11_x	23_x
1_Response	7_day	12_x	24_x
2_Response	1_x	13_x	area
3_Response	2_x	14_x	zone

4. 호우피해 예측함수 개발 및 예측력 평가

본 연구에서는 호우피해를 예측하기 위하여 다중회귀모형과 인공신경망을 구축하였다. 모형을 구축할 때 학습과 검증을 위하여 자료의 기간을 구분하였다. 2005년부터 2012년 까지를 학습기간, 2013년부터 2016년까지를 평가구간으로 설정하였다. 모형의 검증을 위하여 예측력 평가 방법으로는 Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE)를 활용하여 각 모형의 예측력을 평가하였다.

4.1 다중회귀 모형 구축

다중회귀 모형을 구축함에 있어서 독립변수들 간의 상관성으로 인하여 다중공선성의 문제가 발생할 수 있다. 이러한 문제점을 해결하기 위하여 분산 분산팽창요인(VIF)을 통해 이를 검토하였다. 통상적으로 VIF가 10 이상인 경우 다중공선성의 문제가 발생한다고 판단하고 있다(Kim, Choi, Kim, et al., 2017). 모형을 구축할 때 대부분의 독립변수들은 강우자료로 구성되어 있어 상관성이 강하기 때문에 다음 Table 7.에서 나타나듯이 대부분의 변수에서 다중공선성이 존재함을 알 수 있다. 따라서 본 연구에서는 다중공선성 문제를 해결하기 위하여 변수선택법 중 단계적 선택법(Stepwise)을 활용하여 다중회귀 모형을 구축하였다. 단계선택법을 활용하여 변수를 선택하였을 때 다음 Table 7과 같이 13개의 변수로 모형이 구성되었고, 다음 Eq. (12) 같이 구축되었다.

$$\hat{y}_i = 8.4913 + 0.0021x_1 + 0.0354x_2 + \dots + 0.0606x_{13} \quad (12)$$

Table 7. Dependent Variable and Independent Variable

Variable	VIF	Variable	VIF
2_State	30.96801	6_day	25.10726
4_State	2.411149	1_x	10.263318
4_Response	1.263318	16_x	62.77074
5_Response	1.886653	19_x	123.7541
date	2.465301	area	2.70645
1_day	10.14899	zone	2.687115
4_day	23.51315		

4.2 인공신경망 모형 구축

인공신경망 모형을 구축하기 위하여 종속변수 및 독립변수들에 대한 표준화가 선행되어야 한다. 표준화 방법으로는 Re-scaling방법을 활용하였다. 본 연구에서는 1개의 종속변수 및 48개의 독립변수에 대하여 표준화 하였다. 또한, 인공신경망 모형은 블랙박스 모형이기 때문에 변수들 간의 다중공선성을 고려하지 않아도 무방하며, 회귀모형과는 다르게 회귀식이 도출되지 않는다. 반면 다음 Fig. 2와 같이 도식화 할 수 있으며, 이를 통하여 각 노드들의 가중치를 파악할 수 있다.

Fig. 2. Visualization of Artificial Neural Network

인공신경망 모형에서는 매개변수들을 변경하여 예측하고자 하는 값에 대하여 정확도를 높일 수 있다는 장점이 있다. 본 연구에서는 매개변수 중 히든노드의 수를 1씩 증가시키며 매개변수를 보정하였다.

4.3 예측력 평가

본 연구에서는 한강 권역의 위험 등급 별 호우피해를 예측하기 위하여 위험 등급을 고려하지 않은 다중회귀 모형,

인공신경망 모형과 위험 등급을 고려한 다중회귀 모형, 인공신경망 모형 총 4가지 모형을 구축하였다. 구축된 총 4가지 모형에 대하여 예측력을 평가 하였다. 예측력을 평가하기 위하여 학습 구간은 2005년부터 2012년으로 설정하였으며, 평가 구간은 2013년부터 2016년으로 설정하였다. 예측력 평가 방법으로는 RMSE와 RMSE를 표준화한 NRMSE를 활용하였다. 두 가지 평가 방법 모두 예측값과 실제값에 대한 오차를 통하여 예측력을 평가하는 방법이며, 0에 가까울수록 예측력이 높음을 의미한다. Eqs. (11)과 (12)와 같이 RMSE와 NRMSE를 수식으로 표현하였다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (11)$$

$$NRMSE = \frac{RMSE}{Max(y_i) - Min(y_i)} \quad (12)$$

여기서 y_i 는 i 번째 실제 호우피해액을 의미하며 $\hat{y}_i$ 는 i 번째 구축된 호우피해액을 의미한다. 선행연구를 살펴보면 통상적으로 NRMSE 결과가 20% 이내 일 경우 예측력이 높고, 50% 이하일 경우 예측력이 보통, 50% 이상일 경우 예측력이 낮다고 의미한다(Kim et al., 2017?). 본 연구에서는 위험 등급을 고려하지 않은 다중회귀 모형, 위험 등급을 고려하지 않은 인공신경망 모형, 위험 등급을 고려한 다중회귀 모형, 위험 등급을 고려한 인공신경망 모형을 순선대로 Model 1 ~ Model 4로 정의 하였다. 다음 Table 8과 같이 각각의 Model에 대한 예측력 평가 결과를 나타내었다.

Table 8. Predictability Evaluation of Models

Model		RMSE	NRMSE (%)
Not Considering of Risk Assessment			
Model 1	Multiple regression model	3,259,068	13.33
Model 2	Artificial neural network	3,243,976	13.27
Considering of Risk Assessment			
Model 3	Multiple regression model	2,956,221	12.09
Model 4	Artificial neural network	2,938,546	12.02

예측력 평가 결과 Model 1 ~ Model 4까지 모두 12%에서 13%내외의 오차를 보이고 있으며, 한강 권역의 호우피해를 적절하게 예측 하는 것으로 판단된다. 또한, 위험 등급을 고려하지 않은 모형보다 위험 등급을 고려한 모형의 예측력 평가 결과가 더 효과적임을 알 수 있다. 최종적으로 Model

4인 위험 등급을 고려한 인공신경망 모형이 가장 높은 예측력을 나타내었다.

5. 결 론

본 연구에서는 각기 다른 시군구별 특성을 반영한 호우피해 예측 함수를 개발하고자 하였다. 각 시군구별 호우피해액과 기상자료 개방 포털을 활용한 수문기상자료를 구축하였다. 또한, 각기 다른 시군구별 특성을 고려하기 위하여 기투자 요소 및 사회·경제적 요소를 고려하였고, 위험 평가 방법인 PSR모형을 활용하여 위험 등급을 부여하였다. 통계적 예측기법으로는 다중회귀 모형 및 인공신경망 모형을 활용하였다. 최종적으로 위험 등급과 통계적 예측기법을 활용하여 총 4가지 호우피해 예측함수를 개발하였다.

- (1) 기 투자 요소 및 사회·경제적 요소를 활용하여 각기 다른 시군구별 특성을 고려하였고, 한강 권역 86개 시군구의 위험 등급은 Red Zone 22개, Orange Zone 22개, Yellow Zone 21개, Green Zone 21개의 지역으로 구분 되었다.
- (2) 위험 등급을 고려하지 않은 통계적 모형(Model 1, 2)과 위험 등급을 고려한 통계적 모형(Model 3, 4)을 비교한 결과 위험 등급을 고려한 통계적 모형이 높은 예측력 나타내었다. 또한, Model 4가 약 12.02%로 가장 높은 예측력 나타내었다.
- (3) 총 4가지로 구축된 통계적 모형의 NRMSE는 약 12%~13%로 나타났으며 예측력 평가 결과는 크게 상이하지 않았다. 그러나 구축된 다중회귀 모형과 인공신경망 모형에 대하여 인공신경망 모형이 호우피해를 예측하는데 효과적임을 알 수 있다.

본 연구에서의 한계점으로는 호우피해 예측함수를 개발함에 있어서 호우에 직접적인 관계가 없더라도 잠정적인 영향을 주는 요소와 호우로 인하여 야기되는 영향을 나타내는 요소 등을 반영하였다면 더 좋은 예측력 평가 결과가 도출 되었을 것이라고 판단된다. 이러한 한계점을 극복하기 위하여 향후 연구에서는 더욱 세분화된 위험 평가 기법을 활용할 방안을 고려할 필요가 있다.

기존 국내·외 선행 연구 사례들에서는 예측함수를 개발함에 있어서 기 투자 요소 및 사회·경제적 요소, 지역적인 특성을 복합적으로 반영하지 못하였다. 이러한 문제점을 해결하기 위하여 본 연구에서는 수문 요소, 기상 요소, 기투자 요소 및 사회·경제적 요소, 지역적인 특성을 복합적으로 고려하여 호우피해 예측함수를 개발하였다.

본 연구의 결과를 바탕으로 각기 다른 시군구별 특성을 파악하고 사전 대비차원에서의 재난관리를 실시한다면 자연재해의 피해를 경감 시킬 수 있을 것으로 판단된다.

감사의 글

본 연구는 정부(행정안전부)의 재원으로 재난안전기술 개발사업단의 지원을 받아 수행된 연구임[MOIS-재난-2015-05], 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2017R1A2B3005695).

References

- Choi, C.H., Kim, J.S., Kim, J.H., Kim, H.Y., Lee, W.J., and Kim, H.S. (2017a). Development of heavy rain damage prediction function using statistical methodology. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 3, pp. 331-338.
- Choi, C.H., Kim, Y.S., Kim, J.S., Kim, D.H., Kim, J.W., and Kim, H.S. (2018). Indicator development and evaluation of storm and flood resilience using big data analysis: (2) Evaluation of resilience. *J. Korean Soc. Hazard Mitig.*, Vol. 18, No. 4, pp. 109-123.
- Choi, C.H., Park, K.H., Park, H.K., Lee, M.J., Kim, J.S., and Kim, H.S. (2017b). Development of heavy rain damage prediction function for public facility using machine learning. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 6, pp. 443-450.
- Choi, K.S., Kang, K.R., Kim, D.W. and Kim, T.R. (2009). Prediction of tropical cyclone intensity and track over the western North Pacific using the artificial neural network method. *Journal of Korean Earth Science Society*, Vol. 30, No. 3, pp. 294-304.
- Choo, T.H., Cho, H.M., Shim, S.B., and Park, S.J. (2018). Development of the wind wave damage estimation functions based on annual disaster reports: Focused on the western coastal zone. *Journal of the Korea Contents Association*, Vol. 18, No. 1, pp. 154-163.
- Choo, T.H., Kim, Y.S., Shim, 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.
- Chung, E.S., Lee, K.S., and Park, K.S. (2008). Development of alternative evaluation index using multicriteria decision making techniques. *Journal of Korea Water Resources Association*, Vol. 41, No. 1, pp. 87-100.
- Chung, K.J., Chen, H., and Kim, A.S. (2005). A new model for forecasting inundation damage within watersheds: An artificial neural network approach. *J. Korean Soc. Hazard Mitig.*, Vol. 5, No. 2, pp. 9-16.
- Feás, J., Giupponi, C., and Rosato, P. (2004). Water management, public participation and decision support systems: The MULINO approach. *Proceedings of 2nd International Congress on Environmental Modelling and Software*, University of Osnabrück, Germany.
- Jang, O.J., and Kim, Y.O. (2009). Flood risk estimation using regional regression analysis. *J. Korean Soc. Hazard Mitig.*, Vol. 9, No. 4, pp. 71-80.
- Kim, D.H., Choi, C.H., Kim, J.S., Lee, J.H., Bae, Y.H., and Kim, H.S. (2018). Analysis of heavy rain damage considering regional characteristics. *J. Korean Soc. Hazard Mitig.*, Vol. 18, No. 4, pp. 311-320.
- Kim, G.H. (2013). *A study on estimation of flood damage and development of flood damage index on public facilities*. Ph.D. dissertation, Inha University.
- Kim, J.S., Choi, C.H., Kim, D.H., Lee, M.J., and Kim, H.S. (2017c). Development of heavy rain damage prediction function using artificial neural network and multiple regression model. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 6, pp. 73-80.
- Kim, J.S., Choi, C.H., Lee, J.S., and Kim, H.S. (2017a). Damage prediction using heavy rain risk assessment: (1) Estimation of heavy rain damage risk index. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, pp. 361-370.
- Kim, J.S., Choi, C.H., Lee, J.S. and Kim, H.S. (2017b). 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, M.J., and Kim, G.S. (2018). Analysis of the applicability of flood risk indices according to flood damage types. *Journal of the Korean Society of Civil Engineers*, Vol. 38, No. 1, pp. 29-39.
- Kim, Y.S. (2016). *Development of resilience indicator based on big data analysis under climate change*. Ph.D. dissertation, Inha University.
- Kung, H.Y., Chen, C.H., and Ku, H.H. (2012). Designing intelligent disaster prediction models and systems for debris-flow disasters in Taiwan. *Expert Systems with Applications*, Vol. 39, No. 5, pp. 5838-5856.
- Kwon, S.H., Lee, J.W., and Chung, G.H. (2017). Snow damages estimation using artificial neural network and multiple regression analysis. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, pp. 315-325.

- Lee, H.D. (2016). Comparison of condition assessment item weights in reinforced earth retaining walls using multi-criteria decision making and entropy technique. Master's thesis, Chungbuk National University.
- Mendelsohn, R., and Saher, G. (2011). *The global impact of climate change on extreme events*. Policy Research Working Paper 5566, World Bank.
- Ministry of the Interior and Safety. (2018). *The 2017 Annual Natural Disaster report*.
- Murnane, R.J., and Elsner, J.B. (2012). Maximum wind speeds and US hurricane losses. *Geophysical Research Letters*, Vol. 39, No. 16, L16707.
- Nam, K.P., and Kim, C.H. (2013). Study on sensitivity of different standardization methods to climate change vulnerability index. *Journal of Environmental Impact Assessment*, Vol. 22, No. 6, pp. 677-693.
- Nardo, M., Saisana, M., Saltelli, A., and Tarantola, S. (2005). *Tools for Composite Indicators Building*. EUR 21682 EN, Joint Research Centre, European Commission.
- OECD (1993). *OECD core set of indicators for environmental performance reviews*. OECD Environment Monographs No. 83.
- Oh, Y.R., and Chung, G.H. (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.
- Shannon, C.E., and Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Received	October 24, 2018
Revised	October 25, 2018
Accepted	November 6, 2018