빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 지표 개발 및 평가: (2) 복원탄력성 평가

Indicator Development and Evaluation of Storm and Flood Resilience Using Big Data Analysis: (2) Evaluation of Resilience

최창현* · 김연수** · 김종성*** · 김동현**** · 김정욱***** · 김형수*****

Choi, Changhyun*, Kim, Yonsoo**, Kim, Jongsung***, Kim, Donghyun****, Kim, Jungwook*****,
and Kim, Hung Soo*****

Abstract

In this study, we proposed a method to evaluate the resilience of storm and flood using big data analysis. Standardization method and factor analysis were applied to storm and flood resilience indicators with big data analysis technique for indicator selection. And the weights for each indicator were calculated using the TF-IDF (Term Frequency-Inverse Document Frequency). The storm and flood resilience was evaluated municipality of city, town, and county in Anyang river basin using the storm and flood resilience evaluation method developed in this study and compared with the result of the regional safety assessment. The results of this research suggested the application methodology of big data analysis techniques which were not applied to the existing disaster management field. And it is expected that it will be used as basic data for effective disaster management in the situation where the intensity and frequency of natural disasters are increasing due to climate change.

Key words : Big Data, Disaster Management, Resilience Evaluation, Storm and Flood, TF-IDF

요 지

본 연구에서는 빅 데이터 분석 기법을 이용하여 풍수해 복원탄력성을 평가할 수 있는 방안을 제시하였다. 적정 지표 선정을 위해 빅 데이터 분석 기법이 적용된 풍수해 복원탄력성 지표에 표준화 방법 및 요인분석을 적용하였고, TF-IDF (Term Frequency-Inverse Document Frequency)를 이용하여 각 지표별 가중치를 산정하였다. 본 연구에서 개발된 풍수해 복원탄력성 평가 방안을 이용하여 안양천 유역의 시군구별 풍수해 복원탄력성을 평가하였고, 이를 지역안전도 평가 결과와 비교 및 검토하였다. 개발된 연구 성과는 기존의 재난관리 분야에 적용이 미비하였던 빅 데이터 분석 기법의 활용 방안을 제시하였고, 기후변화로 인해 자연재난의 강도 및 빈도가 증가하고 있는 상황에서 효율적인 재난관리를 실시하기 위한 기초자료로 활용할 수 있을 것으로 기대된다.

핵심용어 : 빅 데이터, 재난관리, 복원탄력성 평가, 풍수해, TF-IDF

*정회원, 인하대학교 토목공학과 박사수료(E-mail: karesma0cch@naver.com)

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

**정회원, ㈜LIG 시스템

Member, LIG system

***정회원, 인하대학교 토목공학과 박사과정

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

****정회원, 인하대학교 토목공학과 석사과정

Member, Master's Course, 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

국민들의 소득 및 생활수준 향상으로 인해 안전에 대한 관심이 매우 높아지면서 자연재난에 소극적인 대응이나 순응하는 삶에서 벗어나, 발생 가능한 자연재난의 피해를 적극적으로 예방하고 대비하기 위한 노력들이 이루어지고 있다. 그러나, 기존의 재난관리 전략에서 주로 사용된 자연재난을 견딜 수 있는 능력을 증대시키는 저항력(Resistance) 강화 전략은 일정 수준 이하의 재난은 견딜 수 있지만 기후변화로 인한 자연재난의 발생 빈도와 강도가 증대될 것으로 예상되고 있기 때문에 미래의 강력한 재난에는 취약한 한계점을 가지고 있다(Etkin, 1999; De Bruijn, 2004; Choi et al., 2016). 따라서, 기후변화로 인한 자연재난 발생의 충격으로부터 신속하게 원래 상태로 회복할 수 있는 능력인 복원탄력성(Resilience)을 평가하여 적절한 재난관리를 실시하는 것이 중요한 전략으로 제시되고 있다(Adger et al., 2005; Norris et al., 2008; Cutter, 2013; World bank, 2013). 효율적인 재난관리를 위해 지역별 복원탄력성을 평가하는 연구에 대한 국외 연구사례로 Bruneau et al. (2003)은 지진 발생으로 인한 지역사회의 복원탄력성을 평가하기 위한 중요 기능으로 4R (Robustness, Redundancy, Resourcefulness, Rapidity)을 제시하였고, TOSE (Technical, Organizational, Societal, Economic) 차원을 고려하여 복원탄력성을 평가하였다. Rose (2004)는 복원탄력성을 평가하기 위한 단계별(conceptual level, operational level, empirical level) 고려 요소를 제시하였고, CGE (Computable General Equilibrium) 모델을 이용하여 지역경제의 복원탄력성을 평가하였다. Cutter et al. (2008)은 취약성(Vulnerability)과 복원탄력성을 고려한 DROP (Disaster Resilience Of Place) 모델을 이용하여 지역사회 복원탄력성을 평가하였고, Boon et al. (2012)은 Bronfenbrenner's bioecological theory를 이용하여 태풍, 홍수, 산불 등의 재난으로 인한 지역사회의 복원탄력성을 평가하였다. World Economic Forum (2013)은 복원탄력성을 5가지 차원(Economic, Environmental, Governance, Infrastructure, Social)으로 구성하였고, 설문조사 자료를 이용하여 국가 수준에서 복원탄력성을 평가하는 방법을 제시하였다. Da Silva and Morera (2014)는 도시지역의 복원탄력성 구성요소로 4가지 차원(Leadership & strategy, Health & wellbeing, Economy & society, Infrastructure & ecosystems)으로 구성하였고, 최근 재해가 발생한 도시(Surat, Concepción, New Orleans, Semarang, Cali, Cape Town)를 대상으로 복원탄력성을 평가하였다. Cai et al. (2016)은 3가지(exposure, damage, recovery) 차원의 복원탄력성 지표를 z-score를 이용하여 표준화하였고, k-means clustering 기법을 이용하여 지표를 취합하였다. 그리고 RIM (Resilience Inference Measurement) 모델을 이용하여 Mississippi River의 하류 지역에 대한 복원탄력성을 정량적으로 평가하였다.

효율적인 재난관리를 위해 지역별 복원탄력성을 평가하는 국내 연구 사례를 살펴보면, Kim et al. (2011)은 복원탄력성 평가를 위해 5개의 기능(내구성, 대체성, 신속성, 자원동원력, 지역경쟁력)과 15개의 요소를 제시하였다. 이를 2007년 충남 태안군 기름유출사고 피해, 2010년 서해 연평도 포격 사고 피해, 2005년 허리케인 '카트리나' 피해, 2009년 대만 '모라꽃' 태풍 피해, 2010년 아이티 지진 피해, 2010년 칠레 지진 피해 등의 국내·외 주요 재난 사례에 적용하여, 해당 지역의 복원탄력성을 평가하였다. Lee et al. (2013)은 복원탄력성을 평가하기 위해 4개의 차원(대응·복구성, 물리적 차원, 사회적 차원, 경제적 차원)과 12개의 지표를 제시하였다. 각 지표별 가중치를 산정하기 위해 엔트로피, 수정된 엔트로피, AHP 방법을 사용하였고, 복원탄력성 평가의 적정성을 수도권 지역에서의 총 피해액 자료와 비교하여 검토하였다. Shin et al. (2014)은 조선 시대의 재난 대응체계를 복원탄력성 개념과 비교 및 평가하였다. 조선 시대에는 기후변화의 영향으로 지속적인 기근이 발생하였는데, 천혜의 자연환경(내구성, 대체성), 특유의 상부상조 정신(자원동원력, 지역경쟁력), 진흥청이란 전담 조직 강화(대체성, 신속성, 자원동원력, 지역경쟁력), 기근 구휼을 위한 비축곡 조성(대체성, 신속성, 자원동원력) 및 이동수단 체계화(신속성, 자원동원력), 중앙집권제 강화를 위하여 구축된 지방조직을 활용한 보고체계(신속성), 필요 시 신속하게 이루어진 민생정책(자원동원력) 그리고 조선 후기 실학사상의 발전 등과 맞물린 농업생산성 강화(내구성, 대체성) 등의 복원탄력성 증대 방안으로 인해 조선시대에 빈번하게 발생한 기근을 지속적으로 적응하고 극복할 수 있었던 것으로 제시하였다. Bak et al. (2016)은 복원탄력성의 구성요소를 3가지 차원(Social, Economic, Living), 32개의 지표로 구성하였고, 인천지역을 대상으로 기후변화에 따른 복원탄력성의 영향을 분석하였다. 각각의 복원탄력성 지표들은 Re-scaling 방법을 이용하여 표준화 과정을 진행하였고, 엔트로피, 수정된 엔트로피, AHP 방법을 이용하여 지표별 가중치를 산정하였다. 기후변화 시나리오로 RCP 8.5 시나리오를 이용하였고, 2014년부터 2040년까지 시뮬레이션 방법으로 인천지역의 기후변화로 인한 복원탄력성 변화를 검토하였다. Jeon and Byun (2017)은 선행 연구들을 검토하여 복원탄력성을 6개 차원(사회, 경제, 제도, 인프라, 커뮤니티 역량, 환경) 및 20개의 변수로 재정립하였다. 각각의 복원탄력성 지표들은 Re-scaling 방법을 이용하여 표준화 과정을 진행하고, 수도권 66개 시군구에 복원탄력성을 평가하였다.

국내·외 선행연구들을 살펴보면 주로 복원탄력성을 3~7개의 차원으로 구분하고, 각 차원별 지표를 설정하여 복원탄력성을 평가하였다. 그러나 대부분의 연구들이 선행 연구를 참고하여 연구자의 주관에 따라 차원 및 지표를 설정하였기 때문에 새로운 복원탄력성 지표를 개발하는데

미흡한 한계점이 있었고, 다양한 선행연구에서 제시한 유사한 지표들을 무분별하게 사용하여 중복되는 지표들이 있었다. 또한 지표별 가중치 산정의 경우 특정 전문가들의 주관에 반영되는 설문방법을 기반으로 하거나 특정 정보량의 영향을 많이 받는 엔트로피 방법들을 사용하는 한계가 있는 것으로 파악되었다. 특히 최근 다양한 분야에서 빅 데이터를 활용하거나 빅 데이터 분석 기법을 적용한 연구들이 진행되고 있는데, 빅 데이터 분석기법을 이용하여 복원탄력성을 평가하는 연구들은 미비한 실정이다. 따라서 본 연구에서는 비교적 객관적인 방법으로 기후재난으로 인한 복원탄력성 지표와 가중치를 산정할 수 있는 빅 데이터 분석 기법을 사용하였다.

이를 위해 Kim et al. (2018)의 연구결과인 83개의 상세지표별 대리변수를 빅 데이터 분석기법을 통해 산출된 풍수해 복원탄력성 지표로 사용하였고, 이를 표준화 및 요인분석을 실시하여 풍수해 복원탄력성 지표로 제시하였다. 제시된 지표의 가중치를 산정하기 위해 빅 데이터 분석기법인 TF-IDF (Term Frequency - Inverse Document Frequency) 방법을 사용하였고, 이를 안양천 유역에 해당하는 시군구에 적용하여 복원탄력성을 산정하고 지역안전도 평가 결과와 비교하여 적합성을 평가하였다.

2. 풍수해 복원탄력성 평가를 위한 방법론

2.1 표준화 방법

수집된 자료들의 통계단위와 범위가 매우 다르기 때문에, 자료를 적절한 형태로 표준화하는 작업이 선행되어야 한다. 변수들의 단위와 범위를 표준화하는 대표적인 방법은 Re-scaling, Z-score, T-score 방법 등이 있다. Re-scaling 방법은 변수들의 최대값과 최소값을 이용하는 방법으로 Eq. (1)과 같이 해당 변수값에서 변수들의 최소값을 빼고, 이를 변수들의 최대값과 최소값의 차이로 나누어 표준화하는 방법이다. 즉, Re-scaling은 변수의 최소값과 최대값을 이용하여 0~1의 범위를 가지는 값으로 표준화 된다(Kim, 2016). 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-score는 Eq. (2)와 같이, 해당 변수값을 평균으로 뺀 차이를 표준편차로 나누는 방법을 사용한다. 그러나 Z-score 방법은 음수 값이 발생하고, 해당 값의 수준을 직관적으로 파악하기 어렵다는 단점이 있다(McCall, 1923). 따라서 Eq. (3)과 같이 Z-score에 10을 곱하고 50을 더해준 평균 50, 표준편차 10인 분포로 전환하는 T-Score를 사용하기도 한다. T-score는 주로 20~80 사이의 값을 갖게 되어 전통적인 100점 단위와 유사한 결과를 얻게 된다(Song et al., 2013).

$$Re-scaling = \frac{X_i - \min}{\max - \min} \quad (1)$$

$$Z-score = \frac{X_i - \mu}{\sigma} \quad (2)$$

$$T-score = 10 \times \left(\frac{X_i - \mu}{\sigma} \right) + 50 \quad (3)$$

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

2.2 요인분석

요인분석(factor analysis)은 다양하고, 복잡한 변수들을 유사한 특성을 지닌 변수들끼리 묶어 적은 수의 요인으로 축소시키는 방법이다(Lee and Nho, 2012). 즉, 요인분석은 상관(혹은 공분산)행렬의 구조에 관한 통계적 모형을 구축하고, 그와 같은 구조를 생성시키는 소수 몇 개의 요인을 유도하여 측정변수들 간의 구조적 관계를 해석하는 자료 분석 기법이라고 할 수 있다(Kang et al., 2005). 요인분석 과정을 간단하게 정리하면 다음과 같다.

- ① 먼저, 요인분석을 수행하기 전에 자료가 최소한의 조건을 가지고 있는지 검토하는 과정이 선행되어야 한다. 일반적으로 KMO (Kaiser-Meyer-Olkin)의 표본 적합도 검정과 Bartlett의 구형성 검정을 실시하여 요인분석의 적용 가능성을 판단한다(Baek et al., 2016; Park and Park, 2016). KMO의 표본 적합도 검정은 관측된 상관관계수들의 값과 편상관계수들의 값을 비교하는 지수로서, 이 값이 클수록 측정변수들 저변에 공통적인 잠재요인이 존재함을 나타낸다. Bartlett의 구형성 검정은 “상관계수 행렬이 단위행렬이다.”라는 귀무가설을 기각할 수 있는지를 검정하는 것으로, “귀무가설: 공통요인이 존재하지 않는다.”와 “대립가설: 공통요인이 존재한다.”를 대상으로 검정을 실시하는 것이다. 따라서, Bartlett 검정의 p-값이 유의수준보다 작아서 귀무가설이 기각되어야 그 자료에 요인분석을 실시할 가치가 있음을 나타낸다(Kang, 2013). 일반적으로 KMO 검정 결과값이 0.5 이상이고, Bartlett 검정의 경우 p-value가 0.1 이하이면 요인분석을 수행하기에 적절하다고 판단하고 있다(Kim et al., 2017).
- ② KMO의 표본 적합도 검정과 Bartlett의 구형성 검정을 통해 요인분석을 실행하기에 적합하다고 판단되면, 연구 목적에 맞는 요인 추출방법을 선택해야 한다. 대표적인 요인 추출방법으로 주성분 분석법(principal component method)과 공통요인 분석법(common factor analysis) 등이 있다. 주성분 분석법은 주로 서로 상관관계가 높은 변수들을 조합해서 그 변수들의 정보를

가능한 많이 함축하고 있는 새로운 인위적인 변수를 만들어 내기 위한 기법이고, 공통요인 분석법은 최초 변수들을 통해 쉽게 파악되지 않는 잠재적인 공통요인이나 차원을 알고자 할 때 주로 사용되는 방법이다 (Kim, 2016).

- ③ 요인 추출방법을 통해 변수들의 정보를 요약한 요인을 추출하게 되는데, 이 때 변수들의 정보 손실을 최소화 하면서 동시에 요인의 개수도 최소화 할 수 있는 적절한 요인의 개수를 결정해야 한다. 적절한 요인의 개수를 결정하기 위해서는 고유값(eigen value), 설명분산(explained variance) 비율 및 누적 설명분산 비율, 공통성(communality) 등을 고려한다. 일반적으로 고유값이 1.0보다 큰 요인만을 추출하고, 추출된 요인들이 분산의 60% 이상을 설명할 수 있으면 적절한 것으로 판단되며, 추출된 요인에 속해 있는 변수의 공통성은 적어도 0.4 이상이 되어야 한다(Hair et al., 1995; Kang, 2013; Kim, 2016).
- ④ 변수들이 여러 요인에 비슷하게 요인 적재량을 나타낼 경우, 변수들을 각 요인에 분류하기 힘들기 때문에 변수들을 어느 한 요인에 쏠리도록(관찰변수와 요인 간의 관계가 명확히 드러나도록) 요인구조의 축을 변화시키는 요인회전 방법을 적용한다. 요인회전의 방법에는 회전축이 직각을 유지하는 직각회전(orthogonal rotation)과 사각회전(oblique rotation) 등이 있다. 직교회전은 요인의 직교성(독립성)을 유지하면서 요인구조를 회전하는 방법이며, 사각회전은 요인의 직교성을 포기하고 단순성이라는 측면에서 요인구조의 해석 그 자체에 더 큰 비중을 둔 개념이라고 할 수 있다.
- ⑤ 앞에서 진행된 요인분석을 통해 변수들이 새로운 지표군인 요인으로 묶이게 되면, 추출된 요인들이 어느 정도 일관성을 갖는지 신뢰도 검정을 통해 확인해야 한다. 일반적으로 요인분석시에 신뢰도 검정을 위해 크론바 알파(Cronbach α) 계수를 이용하는데, Eq. (4)는 크론바 알파 계수를 산정하는 식이다(Nam and Kim, 2013; Baek et al., 2016). 크론바 알파 계수는 절대적인 임계값이 있지 않으나 일반적으로 0.6 이상이면 요인 분석 결과를 신뢰할 수 있다(Nunnally, 1978).

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum s_i^2}{s_x^2} \right) \quad (4)$$

여기서, n : 항목의 수, s_i^2 : 각 항목의 분산, s_x^2 : 총 분산

2.3 가중치 산정 방법

다양한 지표를 이용하여 특정 요소를 평가하는 경우, 선정된 지표별로 해당 요소에 어느 정도 영향을 미치는지 분석하

는 것은 평가 결과에 매우 중요한 부분이라 할 수 있다. 각각의 지표가 해당 요소를 설명하는 정도가 다르고, 영향의 정도가 상이하기 때문에 지표들 사이의 상대적 중요도를 정량화하기 위한 방법이 필요하다. 이와 같이 지표별로 영향의 정도를 검토하는 방법을 지표의 가중치를 산정 방법이라고 하며, 가중치 산정방법에는 델파이법(Delphi method), AHP (Analytic Hierarchy Process), PROMETHEE (Preference Ranking Organization METHOD for Enrichment Evaluation), 엔트로피(Entropy) 방법, TF-IDF (Term Frequency - Inverse Document Frequency) 방법 등이 있다(Khang et al., 1998; Choi et al., 2013; Lee et al., 2013; Oh et al., 2015; Kim, 2016; Kim et al., 2017).

델파이법은 전문가들의 경험적 지식과 직관적인 판단을 이용하는 탐구적 기법으로, 특정한 전문가 그룹에 대한 설문 조사를 통해 해당 요소에 대한 가중치를 산정하는 방법이다(Khang et al., 1998). AHP는 의사결정의 전 과정을 다수의 계층(level)으로 설계한 후, 계층별 기준 및 대안들의 중요도를 서로 비교하는 쌍대응비교법(paired comparisons method)에 의해 단계별로 평가함으로써 가중치를 산정하는 방법이다(Lee et al., 2013). PROMETHEE는 세부평가기준 항목별 선호함수(preference function)와 선호유출량(leaving flow) 및 선호유입량(entering flow), 그리고 순흐름량(net flow)을 이용하여 대안들의 순위선호를 도출하여 가중치를 산정하는 방법이다(Choi et al., 2013). 엔트로피 방법은 대안과 속성을 많이 포함하고 있는 문제에 대해 의사결정권자가 비교적 이해하기 쉬운 정보이론(information theory) 방법을 적용하는 것으로, 정보이론이란 각 속성의 빈도수를 기반으로 응집도가 높은 속성을 발견하여 가중치를 높게 부여하는 방법이다(Song et al., 2013).

TF-IDF는 “어떤 단어의 중요도(Weight)는 그 단어가 문서에 나온 횟수(Term Frequency)에 비례하고, 그 단어가 있는 모든 문서의 총 수에 반비례한다(Inverse Document Frequency)”라는 명제에 기초한 가중치 산정 방법이다(Lee and Bae, 2002). 주로 정보 검색분야에서 사용하는 가중치 산정방법으로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 통계적 수치로 제시해 준다(Kim, 2016). 여기서 TF (Term Frequency)는 문서에서 단어의 출현 빈도수를 의미하며, IDF (Inverse Document Frequency)는 단어의 일반적인 중요도를 나타내는 수치로 전체 문서의 수를 해당단어가 포함된 문서들의 수로 나눈 값에 로그를 취하는 형태를 가진다. 예를 들어 문서에 등장하는 모든 단어 빈도가 1,000번이고 ‘재난관리’라는 단어의 빈도가 110번이라면 TF값은 $110/1,000=0.110$ 이 된다. 또한, 분석에 필요한 50,000개의 문서 중 ‘재난관리’라는 단어가 포함된 문서가 2,500건이라고 하면 IDF는 $\log(50,000/2,500)=1.301$ 과 같이 계산한다. 따라서 최종적으로 TF-IDF 가중치는 $0.110 \times 1.301=0.143$ 이 된다. TF값은

문서 내부의 단어 출현 빈도를 모든 단어의 총 출현 회수로 나누어 정규화한 형태를 가지며, 이는 문서의 크기에 따른 TF값의 편중 현상을 방지하기 위해서이다. 또한, IDF의 경우 상대적으로 많은 문서에 출현한 단어의 값은 작아지고 반대로 한쪽으로 편중하여 나타난 단어의 IDF값이 커지는 특징을 가지고 있다. 이는 IDF값이 작은 단어는 보편적인 단어일 가능성이 크며, 반대로 IDF값이 큰 단어는 문서 내에서 주요 의미를 가지는 단어로 분별하기 위한 의도를 반영한 것이다. TF-IDF를 식으로 나타내면 Eqs. (5)~(7)와 같이 표현할 수 있다.

$$\text{TF: } TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (5)$$

$$\text{IDF: } IDF = \log \frac{|D|}{|\{d_j | t_j \in d_j\}|} \quad (6)$$

$$\text{TF-IDF 가중치: } w_{i,j} = TF_{i,j} \times IDF \quad (7)$$

여기서, $n_{i,j}$ 는 문서 d_j 에서 단어 t_i 의 빈도수, $\sum_k n_{k,j}$ 는 문서 d_j 에 등장한 모든 단어의 빈도수, $|D|$ 는 문서집합에 포함되어 있는 문서의 수, $|\{d_j | t_j \in d_j\}|$ 는 단어 t_j 가 등장하는 문서의 개수이다.

3. 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 지표 결정

3.1 복원탄력성 지표의 표준화

빅 데이터 분석기법을 이용한 풍수해 복원탄력성 지표를 결정하기 위해 Kim et al. (2018)의 연구결과인 빅 데이터 분석기법을 통해 산출된 83개의 상세지표별 대리변수를 사용하였다. 수집된 상세지표의 대리변수들은 통계단위와 범위가 각자 다르기 때문에, 각각의 대리변수들에 대한 데이터를 동일한 단위와 범위를 가지도록 표준화를 실시하였다. 표준화 방법으로 Re-scaling, Z-score 방법이 많이 적용되고 있는데, Z-score 방법은 지표가 평균보다 낮은 경우 음수의 값이 나오기 때문에 복원탄력성을 비교 및 검토하는데 적합하지 않다고 판단하였다. 따라서 Re-scaling 방법을 이용하여 상세지표별 대리변수를 0~1 사이의 값으로 표준화하였다.

3.2 범주별 요인분석을 통한 최적 풍수해 복원탄력성 지표의 결정

빅 데이터 분석기법을 통해 산출된 83개의 다양한 상세지표별 대리변수들을 유사한 특성을 지닌 변수들끼리 묶어 적은 수의 요인으로 축소시키는 요인분석을 수행하기 전에, 차원별로 KMO (Kaiser-Meyer-Olkin)의 표본 적합도 검정과

Bartlett의 구형성 검정을 통해 요인분석의 적용 가능성을 판단하였다. Table 1에서 나타난 바와 같이 모든 차원의 KMO 검정 결과값은 모두 0.5 이상이고, Bartlett의 검정 결과는 모두 유의확률 $p = 0.1$ 이하로 나타났다. 따라서 모든 차원의 상세지표별 대리변수들에 요인분석이 가능하다고 판단하였다.

상세지표별 대리변수들이 요인분석을 실행하기에 적합하다고 판단되었기 때문에, 본격적인 요인분석을 진행하기 위해 연구 목적에 맞는 요인 추출방법을 선택해야한다. 본 연구에서는 서로 상관관계가 높은 대리변수들을 조합해서 변수들의 정보를 가능한 많이 함축하는 새로운 요인을 구축하기 위해, 주성분 분석을 요인 추출방법으로 사용하였다. 주성분 분석을 통해 변수들의 정보를 요약한 요인을 추출하게 되는데, 이 때 변수들의 정보를 가능한 많이 함축하면서 동시에 요인의 개수도 최소화 할 수 있는 적절한 요인의 개수를 결정해야 한다. 본 연구에서는 선행 연구사례들을 참고하여 고유값이 1.0보다 큰 요인만을 추출하고, 추출된 요인들의 누적 설명분산이 60% 이상이며, 추출된 요인에 속해 있는 변수의 공통성은 적어도 0.4 이상이 되도록 요인의 개수를 결정하였다.

변수와 요인 간의 관계가 명확히 드러나도록 요인구조의 축을 변화시키기 위해 직교회전(orthogonal rotation) 중 많은 연구에서 적용성을 인정받은 베리맥스 회전(varimax rotation)을 요인회전 방법으로 이용하였다. 베리맥스 회전은 요인행렬의 열을 단순화 해주는 방법으로 요인행렬의 요인 적재량을 0 또는 1에 가깝게 하여 분산을 최대화 한 후, 이에 해당하는 새로운 요인 적재량을 얻는 방법으로 요인과 변수간의 관계를 명확하게 해석하는데 용이하다. 일반적으로 해당 요인 적재량이 0.1 이하인 경우 같은 요인으로 묶일 수 없음을 의미하고, 수용기준은 0.5-0.7 이상이다(Kim, 2016). 따라서 본 연구에서는 요인 적재량이 0.6 이상인 변수만을 사용하였다. 요인분석을 통해 추출된 요인들이 어느 정도 일관성을 갖는지 크론바 알파(Cronbach α)를 이용하여 신뢰도 검정을 진행하였고, 모든 차원의 최종 추출된 요인들의 내적일관도를 측정하는 크론바 알파값은 모두 0.6 이상으로 충분한 신뢰도를 확보하고 있는 것으로 나타났다.

3.3 요인 명명

요인분석 과정을 통해 5개 차원(물리적, 사회적, 경제적, 제도적, 자연적)에 대하여 11개 요인과 36가지 상세지표별 대리변수를 최종적으로 선정하였다. 요인분석을 통해 추출된 요인들을 구성하는 변수들의 특성을 반영하여 명칭을 부여하였다. 먼저 물리적 차원의 요인 1을 구성하는 상세지표별 대리변수는 하수도시설 연장, 상수도시설 연장, 포장도로 연장, 자가소유 주택수, 침수되지 않은 주택 비율, 가스공급시설 면적 순서로 나타났고, 각각의 변수는 공공 시설물과 사유 시설물에 대한 변수들이기 때문에 ‘시설물’이라고 명명

하였다. 요인 2는 상수도 급수량, 도시가스 보급률, 전력 사용량 순서로 나타나고 있으며, 이는 각각 상수도, 가스, 전력 등의 사회기반시스템에 대한 공급량과 사용량을 나타내기 때문에 ‘공급 및 사용’이라고 명명하였다.

사회적 차원의 요인 1에서는 교육기관 수, 사회복지시설 수, 재난취약자 수, 공동체 활동 순서로 나타나고 있으며, 각 변수들은 지역사회를 구성하는 집단의 변수들이기 때문에 ‘사회 집단’이라고 명명하였다. 요인 2는 건강보험 가입자, 인구증가율 순서로 나타나고 있으며, 이는 지역사회의 활동으로 인한 사회보장과 구성원으로서의 참여에 대한 변수이기 때문에 ‘사회 참여’라고 명명하였다.

경제적 차원의 요인 1은 금융기관 수, 사업체 수, 지방세 부담액 순서로 나타나고 있으며, 각각의 변수는 지역사회의 경제적인 수준을 나타내고 있기 때문에 ‘경제적 수준’이라고 명명하였다. 요인 2는 타 지역 통근 취업자 수, 여성취업자 수 순서로 나타나고 있으며, 개인 소득과 관련된 경제활동을 설명하고 있으므로 ‘경제적 활동’이라고 명명하였다.

제도적 차원의 요인 1은 공무원 수, 관내 관공서 및 주요기관 개수, 이송 및 구조 건수, 소방공무원 수, 소방 및 건설장비 개수의 순서로 나타나고 있으며, 이는 재난에 대응하기 위한 공공기관의 역할 및 활동을 나타내고 있기 때문에 ‘공공기관 활동’이라고 명명하였다. 요인 2는 언론매체 개수, 비영리단체 개수 순서로 나타나고 있으며, 각각의 변수는 지역사회 내 비공공기관의 역할 및 활동을 설명하고 있기 때문에 ‘사설기관 활동’이라고 명명하였다.

자연적 차원의 요인 1은 $R10$, $R25$, $R20$ 의 순서로 도출되

었고, 각각의 변수는 10 mm 이상, 25 mm 이상, 20 mm 이상 강우의 발생일수(days)로 ‘자연재난 빈도’라고 명명하였다. 요인 2는 $R95p$, $RX5day$, $RX1day$ 순서로 나타났고, 각각의 변수는 강우 발생일 중 95%에 해당하는 일강수량(mm), 지속기간 5일과 1일 최대 일강수량(mm)으로 강우의 강도에 해당하기에 ‘자연재난 강도’라고 명명하였다. 요인 3은 생태계 및 수변지구 개수, 환경분야 검사실적 건수, 공원 개수의 순서로 나타났으며, 각각의 변수는 생태 및 녹지 관련 개수와 환경검사 실적을 나타내기 때문에 ‘생태계 및 환경’이라고 명명하였다.

Table 1은 KMO의 표본 적합도 검정과 Bartlett의 구형성 검정 결과를 나타낸 것이고, Table 2는 요인분석 결과를 정리한 것이다.

3.4 가중치 산정

요인분석을 통하여 풍수해 복원탄력성 지표를 선정 및 제시하였다. 도출된 지표들은 복원탄력성에 미치는 영향이 서로 상이하기 때문에, 복원탄력성 평가 전에 지표들 사이의 상대적 중요도를 정량화하기 위한 가중치 산정이 필요하다. 가중치 산정방법에는 설문조사를 이용한 델파이법(Delphi method), AHP (Analytic Hierarchy Process), PROMETHEE (Preference Ranking Organization METHOD for Enrichment Evaluation) 등과 정보이론을 이용한 엔트로피(Entropy) 방법, 빅 데이터 분석 기법에서 주로 사용되는 TF-IDF (Term Frequency - Inverse Document Frequency) 방법 등이 있다.

설문조사 기반의 가중치 산정방법인 델파이법, AHP,

Table 1. Results of KMO's and Bartlett's Test

Dimension	Method		Result
Physical	Kaiser-Meyer-Olkin's test for sampling adequacy		0.743
	Bartlett's test of sphericity	Chi-Square	1198.589
		Degree of freedom	276
		p-value	0.000
Social	Kaiser-Meyer-Olkin's test for sampling adequacy		0.620
	Bartlett's test of sphericity	Chi-Square	543.443
		Degree of freedom	105
		p-value	0.000
Economic	Kaiser-Meyer-Olkin's test for sampling adequacy		0.522
	Bartlett's test of sphericity	Chi-Square	602.007
		Degree of freedom	66
		p-value	0.000
Institutional	Kaiser-Meyer-Olkin's test for sampling adequacy		0.676
	Bartlett's test of sphericity	Chi-Square	689.670
		Degree of freedom	105
		p-value	0.000
Natural	Kaiser-Meyer-Olkin's test for sampling adequacy		0.614
	Bartlett's test of sphericity	Chi-Square	1280.568
		Degree of freedom	136
		p-value	0.000

Table 2. Result of Factor Analysis

Dimension	Factor	Detailed indicator (Proxy variable)	Factor Analysis							Cronbach α	
			Eigen value	Variance (%)	Cumulative (%)	Communality	Factor 1	Factor 2	Factor 3		
Physical	Facilities	Waste pipe (Extension of waste pipe)	4.88	54.18	54.18	0.91	0.95	-	-	0.94	0.74
		Water pipe (Extension of water pipe)				0.89	0.94	-	-		
		Network of road (Extension of paved road)				0.84	0.91	-	-		
		Ownership of house (Number of self-owned houses)				0.72	0.84	-	-		
		Lowland, Underground, Inundation (Percentage of houses without flood damage)				0.63	0.77	-	-		
		Gas facilities (Area of gas facilities)				0.62	0.77	-	-		
	Supply and use	Water use (Amount of water supply)	1.73	19.24	73.42	0.69	-	0.81	-	0.63	
		Gas supply (City gas distribution rate)				0.63	-	0.76	-		
		Electricity, Use (Amount of electricity used)				0.67	-	0.64	-		
Social	Social group	School, Education facilities (Number of education facilities)	3.00	50.07	50.07	0.90	0.95	-	-	0.89	0.71
		Social welfare facilities (Number of social welfare facilities)				0.84	0.87	-	-		
		Flooded area, Resident evacuation (Population with physical or language disabilities)				0.79	0.86	-	-		
		Community, Community involvement (Welfare facilities and public physical facilities)				0.62	0.77	-	-		
	Social participation	Health Insurance (Population of health insurance)	1.82	30.39	80.46	0.86	-	0.92	-	0.82	
		Life expectancy (Population growth rate)				0.82	-	0.91	-		
Economic	Economic level	Saving, Deposit (Number of financial institution)	2.61	52.17	52.16	0.95	0.94	-	-	0.93	0.81
		Businesses (Number of businesses)				0.89	0.94	-	-		
		Income level, National income (Amount of local tax)				0.84	0.91	-	-		
	Economic activity	Commuters in other regions (Number of commuters in other regions)	1.90	37.91	90.07	0.91	-	0.95	-	0.90	
		Female employment (Number of female employee)				0.91	-	0.94	-		
Institutional	Public institution activity	Public officer (Number of public officer)	3.69	52.71	52.71	0.92	0.90	-	-	0.92	0.89
		Related organization (Number of government office and major organizations in jurisdiction)				0.93	0.89	-	-		
		Emergency aid (Number of transportation and rescue)				0.83	0.89	-	-		
		Fire officer (Number of fire officer)				0.85	0.83	-	-		
		Fire equipment, Recovery equipment (Number of fire equipment and construction equipment)				0.54	0.72	-	-		
	Private institution activity	Press (Number of press)	2.27	32.49	85.20	0.97	-	0.98	-	0.96	
		NGOs (Number of NGOs)				0.92	-	0.92	-		

Table 2. (Continued)

Dimension	Factor	Detailed indicator (Proxy variable)	Factor Analysis							Cronbach α	
			Eigen value	Variance (%)	Cumulative (%)	Communality	Factor 1	Factor 2	Factor 3		
Natural	Frequency of natural disaster	Rainfall (R10)	2.93	32.57	32.57	0.96	0.97	-	-	0.98	0.84
		Rainfall (R25)				0.99	0.93	-	-		
		Rainfall (R20)				0.97	0.92	-	-		
	Intensity of natural disaster	Rainfall (R95p)	2.82	31.36	63.94	0.94	-	0.94	-	0.96	
		Rainfall (RX5day)				0.94	-	0.89	-		
		Rainfall (RX1day)				0.89	-	0.87	-		
	Ecosystem and environment	Biodiversity, Protection of habitat (Number of ecosystem and watershed)	2.00	22.26	86.20	0.71	-	-	0.83	0.74	
		Environmental management, Environmental inspection (Number of environmental inspection)				0.74	-	-	0.83		
		Greens of park (Number of park)				0.62	-	-	0.73		

PROMETHEE 등은 전문가의 지식, 경험 및 직관을 포착하는 장점이 있지만, 설문에 응답하는 대상에 따라 의사결정의 주관성이 발생할 수 있다. 또한, 정보이론을 이용한 엔트로피 방법의 경우 통계자료만을 이용하기 때문에 비교적 객관적이라는 장점이 있지만, 정보량이 충분하지 않을 경우에는 특정 값의 영향을 받아 가중치의 편차가 크게 나타나는 단점이 있다. 따라서 본 연구에서는 이러한 주관적 응답에 의해 발생할 수 있는 왜곡과 정보량 부족으로 인한 오류를 최소화하기 위해, 다양한 자료의 정보를 수렴하여 가중치를 비교적 객관적으로 표현하는 방법인 TF-IDF 방법을 적용하였다. Kim et al. (2018)에서 수집한 비정형 데이터인 전문가자료와 3.2절에서 최종 선정된 상세 지표를 기초자료로 이용하였고, 가중치 산정을 위해 상세 지표별 TF-IDF를 산정하였다. 가중치 산정 결과 물리적 차원은 0.360, 사회적 차원은 0.132, 경제적 차원은 0.031, 제도적 차원은 0.206, 자연적 차원은 0.272로 물리적 차원과 자연적 차원의 가중치가 비교적 높게 산정되었다. 요인별로는 물리적 차원의 ‘시설물’, 제도적 차원의 ‘공공기관의 활동’, 자연적 차원의 ‘생태계 및 환경’, 사회적 차원의 ‘사회 집단’ 순으로 높은 가중치가 산정되었다. Table 3과 Fig. 1은 TF-IDF를 이용한 가중치 산정 결과를 나타낸다.

4. 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 산정 및 평가

4.1 대상지역 및 기후재난관련 현황

3장에서 개발된 빅 데이터 분석 기법을 적용한 지표와 가중치를 이용하여, 풍수해 복원탄력성을 평가하기 위해 안양천 유역에 해당하는 지자체를 대상지역으로 선정하였

다. 안양천 유역은 하천연장 32.5 km에 유역면적 286 km²으로 경기도 의왕시 지지대 고개에서 발원하여 군포시를 경유, 안양시 도심을 중앙으로 관류하여 광명, 서울시를 거쳐 한강에 유입되는 도시형 하천으로 서울시 양천구, 강서구, 구로구, 금천구, 영등포구, 동작구, 관악구와 경기도 안양시, 부천시, 광명시, 과천시, 시흥시, 군포시, 의왕시 등 총 14개의 지자체를 포함하고 있다.

4.2 빅 데이터 분석 기법을 이용한 풍수해

복원탄력성 산정 결과의 비교군 구축

빅 데이터 분석기법을 이용하여 개발된 풍수해 복원탄력성 지표의 적용 결과를 검토 및 비교하기 위해 매년 행정안전부에서 전국 지자체를 대상으로 실시하는 “자연재해에 대한 지역안전도 진단” 결과를 이용하였다. 지역안전도 진단은 ‘자연재해대책법’ 제75조의2(지역안전도 진단)에 따라 시군구별로 매년 실시하고 있으며, 재해위험요인 조사, 예방대책 추진 실태, 예방시설정비 현황 등의 3개 분야로 나누어 서면 및 현지진단 등을 통해 자연재해에 대한 지역의 안전도를 평가하는 제도이다. 지역안전도 진단 결과가 직접적으로 해당 지자체의 풍수해 복원탄력성을 나타내는 것은 아니지만, 지역의 위험요인과 방재성능 등을 분석하여 지자체의 방재역량 제고를 위한 방안으로 활발하게 사용되고 있기 때문에 본 연구에서 개발한 풍수해 복원탄력성 평가 결과와 비교하기에 가장 적절할 것으로 판단하였다.

1회의 지역안전도 진단 결과만을 이용할 경우 해당 지자체의 지역안전도가 왜곡될 수 있기 때문에, 2011~2017년(최근 7개년) 지역안전도 진단 결과의 평균을 이용하였다. 또한 풍수해 복원탄력성과 지역안전도 결과를 가시적인 점수로 확인하기 위해 T-Score 방법을 이용하여 0~100을 가지는

표준화 점수(Value)로 변환하고, 1~5등급(1등급: 81~100, 2등급: 60~80, 3등급: 41~60, 4등급: 21~40, 5등급: 0~20)까지 5단계(1등급에 가까울수록 복원탄력성과 지역안전도가 높음)로 적정성을 평가하였다. 지역안전도의 경우 복원탄력성과 반대로 값이 클수록 안전도가 낮기 때문에 동일한 등급을 주기 위해 역산하는 작업을 선행하였다.

4.3 빅 데이터 분석 기법을 이용한 풍수해

복원탄력성 산정 및 검토

본 연구에서 개발한 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성(Storm and Flood Resilience, SFR)을 산정한 결

과 양천구, 동작구, 영등포구의 복원탄력성이 높은 것으로 나타났고, 광명시, 의왕시, 강서구의 복원탄력성은 상대적으로 낮은 것으로 나타났다(Table 4). 또한 Fig. 2에서 복원탄력성이 높은 지역과 낮은 지역의 예시를 나타냈는데, 대체로 복원탄력성이 높은 지역은 물리적 차원이 50% 이상을 차지하는 것으로 나타났지만, 복원탄력성이 낮은 지역은 물리적 차원의 영향이 50% 이하로 낮은 경향을 보였다. Table 5와 Fig. 3은 본 연구에서 개발한 복원탄력성 평가와 지역안전도 평가 결과를 비교하였는데, 최근 7년간(2011~2017)의 지역안전도 평가 결과 관악구와 동작구의 지역안전도가 높은 것으로 나타났고, 의왕시와 시흥시의 지역안전도가 상대적

Table 3. Result of TF-IDF

Detailed indicator	TF-IDF	Weight	Factor	Weight	Dimension	Weight
Waste pipe	0.00611	0.06246	Facilities	0.24792	Physical	0.35965
Water pipe	0.00367	0.03748				
Network of road	0.00631	0.06450				
Ownership of house	0.00345	0.03522				
Lowland, Underground, Inundation	0.00459	0.04692				
Gas facilities	0.00013	0.00134	Supply and use	0.11173		
Water use	0.00358	0.03663				
Gas supply	0.00108	0.01107				
Electricity, Use	0.00626	0.06403	Social group	0.12138	Social	0.13156
School, Education facilities	0.00290	0.02962				
Social welfare facilities	0.00042	0.00431				
Flooded area, Resident evacuation	0.00582	0.05945				
Community, Community involvement	0.00274	0.02800	Social participation	0.01018		
Health Insurance	0.00076	0.00776				
Life expectancy	0.00024	0.00242	Economic level	0.02620	Economic	0.03094
Saving, Deposit	0.00033	0.00340				
Businesses	0.00108	0.01107				
Income level, National income	0.00115	0.01173	Economic activity	0.00474		
Commuters in other regions	0.00033	0.00340				
Female employment	0.00013	0.00134	Public institution activity	0.15448	Institutional	0.20603
Public officer	0.00344	0.03518				
Related organization	0.00476	0.04863				
Emergency aid	0.00241	0.02464	Private institution activity	0.05155		
Fire officer	0.00183	0.01869				
Fire equipment, Recovery equipment	0.00267	0.02734	Frequency of natural disaster	0.07146	Natural	0.27182
Press	0.00309	0.03155				
NGOs	0.00196	0.02000				
Rainfall	0.00233	0.02382	Intensity of natural disaster	0.07146		
Rainfall	0.00233	0.02382				
Rainfall	0.00233	0.02382				
Rainfall	0.00233	0.02382	Ecosystem and environment	0.12890		
Rainfall	0.00233	0.02382				
Rainfall	0.00233	0.02382				
Biodiversity, Protection of habitat	0.00781	0.07985				
Environmental management, Environmental inspection	0.00157	0.01600				
Greens of park	0.00323	0.03305

(a) Weight of factor

(b) Weight of detailed indicator (Physical)

(c) Weight of detailed indicator (Social)

(d) Weight of detailed indicator (Economic)

(e) Weight of detailed indicator (Institutional)

(f) Weight of detailed indicator (Natural)

Fig. 1. Result of TF-IDF

으로 낮은 것으로 나타났다. 복원탄력성 산정 결과와 지역안전도를 검토한 결과 각 지자체별로 복원탄력성과 지역안전도 사이의 점수 차이가 있지만, 전체적으로 모든 지역이 2~4등급에 해당하는 것으로 분석되었다.

5. 결 론

국내의 경우 예방, 대비, 대응, 복구의 4단계 재난관리가 시행되고 있지만, 최근 10년간의 자연재난 피해액을 살펴보면 연간 약 5천억 원의 피해액이 발생하는 등 아직까지 효과적인 재난관리가 시행되지 않는 것으로 보인다(Choi et al., 2017). 특히 최근에는 다양한 분야에서 빅 데이터를

구축 및 활용하고 있지만, 재난관리에 빅 데이터를 활용한 연구들은 다소 미비한 것으로 보인다. 따라서 본 연구에서는 효과적인 재난관리를 위해 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성(Storm and Flood Resilience, SFR) 평가 방안을 제시하고자 하였다. 연구 결과를 요약한 내용은 다음과 같다.

- (1) 본 연구에서는 Kim et al. (2018)에서 제시한 상세지표별 대리변수를 사용하였다. 빅 데이터 분석기법을 통해 선정된 5개 차원, 27개 요소, 83개의 상세지표별 대리변수들은 통계단위와 범위가 각자 다르고, 유사한 특성을 지닌 변수들이 존재한다. 따라서, Re-scaling

방법을 이용하여 표준화를 선행하였고, 요인분석을 통해 총 5개 차원, 11개 요인의 36가지 상세지표를 빅 데이터 분석기법을 이용한 풍수해 복원탄력성 지표로 결정하였다.

- (2) 결정된 풍수해 복원탄력성 지표들 사이의 가중치를 산정하기 위해, TF-IDF 방법을 이용하였다. 가중치 산정 결과 물리적, 자연적, 제도적 차원의 가중치가 가장 높은 것으로 나타났고, 요인별로는 물리적 차원

의 ‘시설물’, 제도적 차원의 ‘공공기관의 활동’, 자연적 차원의 ‘생태계 및 환경’, 사회적 차원의 ‘사회 집단’ 순으로 높은 가중치가 산정되었다.

- (3) 개발된 지표와 가중치를 이용하여 풍수해 복원탄력성을 평가하기 위해 안양천 유역에 해당하는 지자체를 대상지역으로 선정하였고, 풍수해 복원탄력성 산정 결과의 적정성을 검토하기 위해 “자연재해에 대한 지역안전도 진단” 결과를 비교군으로 설정하였다.

Table 4. Result of SFR by Dimensions

District	Dimensions	SFR	District	Dimensions	SFR
Yangcheon-gu	Physical	0.185	Anyang-si	Physical	0.116
	Social	0.071		Social	0.069
	Economic	0.012		Economic	0.012
	Institutional	0.010		Institutional	0.013
	Natural	0.055		Natural	0.067
	Sum	0.334		Sum	0.277
Gangseo-gu	Physical	0.116	Bucheon-si	Physical	0.154
	Social	0.065		Social	0.070
	Economic	0.012		Economic	0.012
	Institutional	0.010		Institutional	0.010
	Natural	0.049		Natural	0.059
	Sum	0.252		Sum	0.305
Guro-gu	Physical	0.163	Gwangmyeong-si	Physical	0.114
	Social	0.043		Social	0.062
	Economic	0.011		Economic	0.013
	Institutional	0.014		Institutional	0.012
	Natural	0.053		Natural	0.047
	Sum	0.283		Sum	0.248
Geumcheon-gu	Physical	0.173	Gwacheon-si	Physical	0.077
	Social	0.043		Social	0.081
	Economic	0.011		Economic	0.022
	Institutional	0.028		Institutional	0.050
	Natural	0.054		Natural	0.058
	Sum	0.310		Sum	0.289
Yeongdeungpo-gu	Physical	0.176	Siheung-si	Physical	0.122
	Social	0.035		Social	0.058
	Economic	0.013		Economic	0.011
	Institutional	0.041		Institutional	0.019
	Natural	0.050		Natural	0.062
	Sum	0.315		Sum	0.273
Dongjak-gu	Physical	0.179	Gunpo-si	Physical	0.115
	Social	0.052		Social	0.065
	Economic	0.012		Economic	0.012
	Institutional	0.014		Institutional	0.014
	Natural	0.067		Natural	0.084
	Sum	0.325		Sum	0.290
Gwanak-gu	Physical	0.143	Uiwang-si	Physical	0.104
	Social	0.057		Social	0.069
	Economic	0.012		Economic	0.013
	Institutional	0.011		Institutional	0.023
	Natural	0.062		Natural	0.039
	Sum	0.285		Sum	0.249

(a) Result of SFR by dimensions (Yangcheon-gu)

(d) Result of SFR by dimensions (Gwangmyeong-si)

(b) Result of SFR by dimensions (Dongjak-gu)

(e) Result of SFR by dimensions (Uiwang-si)

(c) Result of SFR by dimensions (Yeongdeungpo-gu)

(f) Result of SFR by dimensions (Gangseo-gu)

[High SFR]

[Low SFR]

Fig. 2. Result of SFR (example)

Table 5. Comparison of SFR with Regional Safety

District	SFR		Regional safety	
	Value	Grade	Value	Grade
Yangcheon-gu	66.5	2	62.3	2
Dongjak-gu	63.2	2	65.1	2
Yeongdeungpo-gu	59.9	3	55.2	3
Geumcheon-gu	57.7	3	53.1	3
Bucheon-si	56.3	3	58.9	3
Gunpo-si	50.8	3	52.6	3
Gwacheon-si	50.1	3	44.1	3
Gwanak-gu	49.0	3	71.3	2
Guro-gu	48.6	3	49.0	3
Anyang-si	46.0	3	48.3	3
Siheung-si	44.2	3	38.1	4
Gangseo-gu	36.9	4	53.2	3
Uiwang-si	35.4	4	38.8	4
Gwangmyeong-si	35.4	4	49.5	3

(a) SFR

(b) Regional safety

Fig. 3. Comparison of SFR with Regional Safety

(4) 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 평가 결과 양천구, 동작구, 영등포구의 복원탄력성이 높은 것으로 나타났고, 광명시, 의왕시, 강서구의 복원탄력성은 상대적으로 낮은 것으로 나타났다. 대체로 복원탄력성이 높은 지역은 물리적 차원의 복원탄력성이 높은 경향을 나타냈지만, 복원탄력성이 낮은 지역은 물리적 차원의 복원탄력성이 낮은 것으로 나타났다. 이는 복원탄력성이 높은 지역의 경우 물리적 차원에 해당하는 ‘시설물’ 요인이 잘 갖추어져 있지만, 복원탄력성이 낮은 지역의 경우 ‘시설물’ 요인이 비교적 미비한 것으로 판단된다.

(5) 각 지자체별로 복원탄력성과 지역안전도는 대부분 비슷한 수준의 등급을 나타내는 것으로 분석되었다. 특히 복원탄력성이 높은 양천구와 동작구는 지역안전도에서도 비슷한 수준의 높은 값이 나왔고, 비교적 지역안전도가 낮다고 평가된 의왕시도 낮은 복원탄력성 점수가 나타났다. 그러나 관악구의 경우 복원탄력성은 낮고 지역안전도가 높은 차이가 발생하였는데, 해당 지자체의 경우 과거에는 지역안전도가 낮았지만 최근에 개선사업을 통해 지역안전도가 높아진 지역인 것으로 파악되었다.

본 연구에는 몇 가지 한계점과 보완점 있다. 우선 가중치를 산정하기 위해 사용한 TF-IDF 방법은 빅 데이터의 일종인 비정형 데이터의 영향을 크게 받지만, 재난관리 분야에서는 실질적으로 활용 가능한 비정형 데이터가 매우 부족한 실정이다. 특히 비정형 데이터를 수집하고 분석하기 위한 방법론과 시스템 또한 아직은 초기 단계로 적절한 자료의 구축과 분석 방법의 개선이 필요할 것으로 판단된다. 또한, 개발한 빅 데이터 분석 기법을 이용한 복원탄력성의 경우 참 값이 없기 때문에, 옳게 평가되었는지 판단하기 어려운 부분이 있었다. 따라서 비교군으로 선정한 지역안전도의 산정 방법과 평가 항목이 본 연구의 복원탄력성 평가와 상이하기에 정성적인 수준에서 비교 및 검토만 가능하였다. 그리고, 대리변수들을 수집할 때에 안양천 유역에 해당되는 시군구의 자료들을 2013년 기준으로 수집하였기 때문에, 전국적인 적용은 어려울 것으로 판단된다. 따라서 향후에는 최근 통계 자료를 업데이트 하고, 전국의 시군구를 대상으로 자료를 수집하여 풍수해 복원탄력성을 평가하는 연구가 필요할 것으로 판단된다.

그러나 본 연구에서는 기존에 재난관리 분야에서 고려되지 못했던, 빅 데이터(Big Data) 분석 기법을 이용하여 풍수해 복원탄력성 평가 방안을 개발 및 적용하는 것에 의의를 두었다. 특히 재난관리의 기본 단위로 볼 수 있는 지자체 단위의 풍수해 복원탄력성을 평가하는 데 있어서, 단순히 기후변화와 자연재난 등과 같은 물리적, 자연적 차원만을 고려하는 것이 아니라 사회적, 경제적, 제도적 차원을 종합적으로 검토하였다. 또한, 본 연구에서 제시한 빅 데이터를 활용한 TF-IDF 기법은 새로운 지표를 제시하기 위한 방법으로 이용할 수 있고, 특정 전문가의 의견에 따라 특정 지표의 가중치가 크게 나타날 수 있는 설문조사에 의한 기법과 데이터 정보량에만 의존하는 엔트로피 기법을 보완하는 대안으로 활용할 수 있다. 따라서, 본 연구의 결과는 재난관리 분야에 빅 데이터 및 빅 데이터 분석기법의 활용성을 제시하였다는 점에서 의미가 있다고 판단된다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 18AWMP-B083066-05).

References

Adger, W.N., Hughes, T.P., Folke, C., Carpenter, S.R., and Rockström, J. (2005) Social-ecological Resilience to Coastal Disasters. *Science*, Vol. 309, No. 5737, pp. 1036-1039.

Baek, S.G., Yoo, S.P., and Hwang, J.W. (2016) A Study

on the Decision-making Factors of Living-in Idea into Elderly Welfare Housing of Middle-aged Classes. *Journal of The Korean Regional Development Association*, Vol. 28, No. 1, pp. 151-169.

Bak J.U., Kim D.H., and Kim H.S. (2016) Assessment of Resilience in Incheon City: (1) Estimation Model Development and Assessment of Resilience. *J. Korean Soc. Hazard Mitig.*, Vol. 16, No. 6, pp. 109-118.

Boon, H.J., Cottrell, A., King, D., Stevenson, R.B., and Millar, J. (2012) Bronfenbrenner's Bioecological Theory for Modelling Community Resilience to Natural Disasters. *Natural Hazards*, Vol. 60, No. 2, pp. 381-408.

Bruneau, M., Chang, S.E., Eguchi, R.T., Lee, G.C., O'Rourke, T.D., Reinhorn, A.M., Shinozuka, M., Tierney, K., et al. (2003) A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra*, Vol. 19, No. 4, pp. 733-752.

Cai, H., Lam, N.S.N., Zou, L., Qiang, Y., and Li, K. (2016) Assessing Community Resilience to Coastal Hazards in the Lower Mississippi River Basin. *Water*, Vol. 8, No. 2, Article No. 46, pp. 1-18.

Choi, C.H., Han, D.G., Kim, J.W., Jung, J.W., Kim, D.H., and Kim, H.S. (2016) Mega Flood Simulation Assuming Successive Extreme Rainfall Events. *Journal of Wetlands Research*, Vol. 18, No. 1, pp. 76-83.

Choi, C.H., Park, K.H., Park, H.K., Lee, M.J., Kim, J.S., and Kim, H.S. (2017) 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, H.I., Park, S.Y., Song, J.H., and Park, M.J. (2013) Identification of Flood Risk Areas using a Multi-criteria Decision Making Method. *J. Korean Soc. Hazard Mitig.*, Vol. 13, No. 2, pp. 237-243.

Cutter, S.L. (2013) Building Disaster Resilience: Steps Toward Sustainability. *Challenges in Sustainability*, Vol. 1, No. 2, pp. 72-79.

Cutter, S.L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., and Webb, J. (2008) A Place-based Model for Understanding Community Resilience to Natural Disasters. *Global Environmental Change*, Vol. 18, No. 4, pp. 598-606.

Da Silva, J., and Morera, B. (2014) *City Resilience Framework*. Arup & Rockefeller Foundation.

De Bruijn, K.M. (2004) Resilience and Flood Risk

- Management. *Water Policy*, Vol. 6, No. 1, pp. 53-66.
- Etkin, D. (1999) Risk Transference and Related Trends: Driving Forces Towards More Mega-disasters. *Global Environmental Change Part B: Environmental Hazards*, Vol. 1, No. 2, pp. 69-75.
- Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. (1995) *Multivariate Data Analysis* (4th ed.). Prentice Hall, Upper Saddle River, NJ.
- Jeon, E.Y., and Byun, B.S. (2017) A Study on the Development and Application of Community Resilience Evaluation Indicators for Responding to Climate Change. *The Geographical Journal of Korea*, Vol. 51, No. 1, pp. 47-58.
- Kang, H., Han, S.T., Kim, K., and Jhun, M. (2005) *Multivariate data Analysis using SAS by Examples*. Freedom Academy, Paju.
- Kang, H.C. (2013) A Guide on the Use of Factor Analysis in the Assessment of Construct Validity. *Journal of Korean Academy of Nursing*, Vol. 43, No. 5, pp. 587-594.
- Khang, Y.H., Yoon, S.J., Kang, G.W., Kim, C.Y., Yoo, K.Y., and Shin, Y.S. (1998) An Application of Delphi Method to the Assessment of Current Status of Cancer Research. *Korean Journal of Preventive Medicine*, Vol. 31, No. 4, pp. 844-856.
- Kim, J.S., Choi, C.H., Lee, J.S., and Kim, H.S. (2017) Damage Prediction Using Heavy Rain Risk Assessment: (1) Estimation of Heavy Rain Damage Risk Index. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, 361-370.
- Kim, T.H., Kim H.J., and Lee K.J. (2011) The Concept and Functional Objectives of the Urban Resilience for Disaster Management. *Journal of the Korean Society of Safety*, Vol. 26, No. 1, pp. 65-70.
- Kim, Y.S. (2016) *Development of Resilience Indicator Based on Big Data Analysis Under Climate Change*. Ph.D. dissertation, Inha University.
- Kim, Y.S., Choi, C.H., Bae, Y.H., Kim, D.H., Kim, D.H., and Kim, H.S. (2018) Indicator Development and Evaluation of Storm and Flood Resilience Using Big Data Analysis: (1) Development of Resilience Indicators. *J. Korean Soc. Hazard Mitig.*, Vol. 18, No. 4, pp. 97-107.
- Lee, H.Y., and Nho, S.C. (2012) *Advanced Statistical Analysis: Theory and Practice*. Bubmunsu.
- Lee, L.Y., Kim S.J., Kim Y.S., and Kim H.S. (2013) Evaluation of Resilience in Anyang River Basin for Flood Damage Mitigation. *J. Korean Soc. Hazard Mitig.*, Vol. 13, No.2, pp. 291-298.
- Lee, M.R., and Bae, H.K. (2002) Design of Keyword Extraction System Using TFIDF. *Korean Journal of Cognitive Science*, Vol. 13, No. 1, pp. 1-11.
- McCall, W.A. (1923) Science of Education. *The Journal of Educational Research*, Vol. 7, No. 5, pp. 384-396.
- 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.
- Norris, F.H., Stevens, S.P., Pfefferbaum, B., Wyche, K.F., and Pfefferbaum, R.L. (2008) Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness. *American Journal of Community Psychology*, Vol. 41, No. 1-2, pp. 127-150.
- Nunnally, J.C. (1978) *Psychometric Theory* (2nd ed.). McGraw-Hill, NY.
- Oh, S.M., Kim, G.H., Jung, Y.H., Kim, D.H., and Kim, H.S. (2015) Risk Assessment of Heavy Snowfall Using PROMETHEE: The Case of Gangwon Province. *J. Korean Soc. Hazard Mitig.*, Vol. 15, No. 1, pp. 87-98.
- Park, M.M., and Park, J.W. (2016) Development of Resilience Scale for Nurse. *Journal of Korean Academy of Fundamentals of Nursing*, Vol. 23, No. 1, pp. 32-41.
- Rose, A. (2004) Defining and Measuring Economic Resilience to Disasters. *Disaster Prevention and Management: An International Journal*, Vol. 13, No. 4, pp. 307-314.
- Shin, J.D., Kim, M.S., Kim, T.H., and Kim, H.J. (2014) A Conceptual Review of Resilience from a Disaster Perspective. *Crisisonomy*, Vol. 10, No. 12, pp. 93-107.
- Song, J.H., Kim, S.D., Park, M.J., and Choi, H.I. (2013) Estimation of Flood Risk Index for the Nakdong River Watershed. *Journal of Korean Water Resources Association*, Vol. 46, No. 1, pp. 35-45.
- World Bank (2013) *Building Resilience: Integrating Climate and Disaster Risk into Development*.
- World Economic Forum (2013) *Global Risks 2013* (8th ed.).

Received	■	April 4, 2018
Revised	■	April 9, 2018
Accepted	■	April 17, 2018