관측강우와 랜덤포레스트 기반 주요 침수영향 강우 인자 분석 Analysis of Major Rainfall Factors Affecting Inundation Based on Observed Rainfall and Random Forest

김현일* · 김병현**

Kim, Hyunil*, and Kim, Byunghyun**

Abstract

It is important to analyze urban floods according to uncertain rainfall patterns. This study was carried out to determine key statistical characteristics of rainfall affecting the occurrences of flooding by using observed rainfall, one- and two-dimensional hydraulic analysis models, and random forests. The target area selected was the Centum City area of Busan, where flooding occurs frequently with heavy rainfall. For analysis, the rainfall data from the Automatic Weather System (AWS) and radar observatory in the entire city of Busan were collected. Using the results from the one- and two-dimensional numerical models, the occurrence of flooding was determined according to the observed rainfall events. The random forest was used to classify the presence or absence of flooding according to the statistical characteristics of rainfall, and the importance evaluation function was applied to determine the impact of each statistical factor. The three main statistical factors of rainfall with the greatest impact on the occurrence of flooding were analyzed as the 1-hr maximum rainfall, the average rainfall, and the skewness. On the other hand, the three main factors affecting the maximum flood depth were the 3-hrs maximum rainfall, the 1-hr maximum rainfall, and the total rainfall. It is judged that the results of this study will be useful in suggesting quantitative urban flood forecasting standards when rainfall is examined in real-time by using various observation equipment.

Key words : Urban Flood, Observed Rainfall, Criterion of Flooding, Random Forest

요 지

불확실한 강우발생 패턴에 따른 도시홍수의 사전 분석을 위하여 본 연구에서는 관측강우, 1, 2차원 수리학적 해석 모형, 랜덤포레스트 기법을 이용하여 침수 발생에 영향을 주는 주요 강우 통계학적 특성을 파악하고자 한다. 대상지역은 집중호우에 따라 침수가 빈번히 발생한 부산광역시 센텀지구로 선정하였으며, 자동기상관측장비(Automatic Weather System, AWS)와 강우레이더를 통하여 관측된 강우 자료를 부산광역시 전역에서 조사하였다. 1, 2차원 모형 결과를 이용하여 관측 강우 사상에 따라 침수발생 유무를 파악하였으며, 다양한 강우의 통계적 특성에 따른 관계를 랜덤포레스트를 통하여 정의하고자 하였다. 랜덤포레스트를 이용하여 강우 통계적 특성에 따른 침수유무를 분류하였으며, 중요도 산정 기능을 적용하여 각 특성의 영향도를 파악하였다. 지속시간 6시간 강우 사상에서 침수 발생에 가장 큰 영향을 주는 강우의 통계학적 주요 세 가지 인자는 1시간 단위 최대강우량, 평균 강우량, 왜곡도로 분석되었다. 최대 침수심의 크기에 영향을 주는 주요 세 가지 인자는 3시간 최대강우량, 1시간 최대강우량, 총강우량으로 파악되었다. 다양한 관측장비를 이용하여 실시간으로 강우를 조사하는 현시점에서 정량적 도시 홍수예보 기준을 마련하는데 본 연구결과가 유용하게 사용될 수 있을 것으로 판단된다.

핵심용어 : 도시홍수, 관측강우, 침수위험기준, 랜덤포레스트

*정회원, 낙동강홍수통제소 예보통제과 시설연구사(E-mail: hyunn228@korea.kr)

Member, Senior Officer, Forecast and Control Division of Nak-Dong River Flood Control Office

**평생회원, 교신저자, 경북대학교 토목공학과 교수(Tel: +82-53-950-7819, Fax: +82-53-950-6564, E-mail: bhkime@knu.ac.kr)

Corresponding Author, Member, Assistant Professor, Department of Civil Engineering, Kyungpook National University

1. 서 론

도시화가 상당히 이뤄진 지역에 대하여 최근 기후변화의 영향과 함께 국지성 집중호우가 발생하고 있다. 이는 복잡한 시스템을 가지고 자연배수능력이 상대적으로 낮은 도시유역에 대해 침수를 일으킬 수 있는 주요 원인이 된다. 도시홍수는 재산 및 인명피해로 이어질 수 있기에 침수의 양상과 주요 위험인자를 사전에 파악하는 것은 중요하다. 배수를 위하여 복잡하게 관망체계가 구축되어 있는 도시지역에 발생하는 홍수에 대한 주요 요인으로서 설계홍수량을 초과하는 강우의 발생, 하천 여유고 부족, 배수시설능력의 미달, 빗물받이 시설의 부족, 우수유입시설 관리의 미흡, 상대적인 저지대의 특성 등이 있다(Lee and Yeon, 2008). 실제 홍수 발생 이력을 살펴보면 대부분 집중호우와 하수관거 통수능의 부족으로 인하여 홍수가 발생한 것으로 확인되었다.

도시유역에 위험이 되는 강우 특성에 대한 연구가 다양하게 수행되어왔다. 가장 기본적인 방법으로 1, 2차원 수치해석 모형에 관측 또는 설계강수량을 입력함으로써, 강우에 대한 유출 양상을 파악할 수 있다. 도시지역에 대한 홍수분석은 1차원 도시유출해석뿐만 아니라, 침수심의 공간적분포를 파악하기 위한 2차원 해석 또한 수행되고 있다(Yoon et al., 2014). 수치해석뿐만 아니라 통계학적 분석에 근거한 위험 강우량을 산정하는 연구도 수행되었다. Montesarchio et al. (2011)은 심각한 홍수를 일으킬 수 있는 하천 홍수량을 발생시키는 강우 특성에 대하여 분석하였으며, 이를 위하여 베이지안, 엔트로피, 레이더 강우를 이용하여 지속시간별 임계강우를 추정하였다. 연구결과로 홍수 경보를 위한 정량적 강우 기준을 마련하였으며, 더욱 풍부한 자료와 함께 발전 가능성을 나타내었다. Golian et al. (2011)은 무차원 임의의 강우 생성, 다양한 통계학적 기법을 통한 강우의 공간적 분포, 1차원 강우-유출 해석을 수행하는 HEC-HMS를 통하여 위험 유출량을 발생시키는 강우량에 대하여 분석하고 Rainfall Threshold (RT)곡선을 제시하였다. Candela and Aronica (2016)는 홍수 경보 기준을 마련하기 위하여 홍수범람 범위와 발생 수위를 고려하였으며, 다양한 강우 시나리오와 FLURB-2D 모형을 이용하여 발생 침수심에 따른 임계강우량을 제시하였다. Cho et al. (2018)은 피해이력이 없거나 한계강우량 산정이 어려운 지역에 대하여, 유출특성과 유역

특성에 기반하여 한계강우량을 추정할 수 있는 기법을 제시하였다. 뉴로퍼지 모형을 이용하였으며, 실제 침수사상에 적용한 결과 실제 한계강우량과 매우 근사하게 예측하였다. Kim and Han (2019)은 로지스틱 회귀와 1, 2차원 침수해석 모형을 연계하여 도시유역에 대한 강우 지속시간별 침수발생 기준이 될 수 있는 한계강우량을 산정하였다. Automated Weather System (AWS)자료와 Automated Synoptic Observing System (ASOS)자료를 이용하였으며, 방재성능목표 강우량과 한계강우량 산정결과를 비교하였다.

침수를 발생시키고 이를 심화시키는 중요한 강우특성을 파악하기 위한 다양한 연구가 수행될 수 있다. 대표적으로 전문가 설문조사(Delphi-AHP) 방법이 적용될 수 있지만, 본 연구에서는 자료에 기반하여 빠르게 도시유역 침수발생에 영향을 주는 강우 통계학적 특성을 파악하는 방법을 제시하고자 한다. 이를 위하여 1차원 도시유출해석을 수행하는 EPA-SWMM, 2차원 침수해석을 실시하는 FLO-2D 그리고 랜덤포레스트 기법을 적용하였다. 대상유역에 대한 침수 발생유무를 수치해석 프로그램을 통하여 파악하고, 이때의 입력 강우사상의 통계적 특성을 랜덤포레스트와 함께 분석하고자 한다. 실시간으로 관측 강우 자료가 생성되는 단계에서 본 연구의 결과는 도시유역에 대한 홍수예보 기준을 마련하는데 도움이 될 것이다. 본 연구의 흐름도는 Fig. 1과 같다.

2. 연구 방법

2.1 관측 강우자료의 조사

부산시에 대하여 침수를 일으킨 2009년 7월 7일, 2009년 7월 16일, 2012년 7월 15일, 2017년 9월 11일에 대한 Automatic Weather System (AWS) 관측 강우와 부산(160) 레이더 관측 자료를 조사하였다. 강우 지속시간은 6시간으로 선정하였으며, 기상청(Korea Meteorological Administration, KMA) 웹사이트를 통하여 조사하였다. 강우 자료는 1분 단위 관측자료를 활용하였으며, 1차원 도시유출해석을 위하여 10분 단위의 강우량자료로 변환하여 사용하였다. 침수를 일으킬 만한 실제 강우사상은 부족하므로, 부산광역시 전역에서 관측된 자료를 사용함으로써 다양한 강우 통계학적 특성에 따른 도시홍수의 양상을 검토하고자 하였다. Table 1에 관측

Fig. 1. Flowchart for Study

Table 1. Observed Rainfall Data

Rainfall Stations	Observed Date	Total Rainfall (mm)
Busan (160) Weather Radar	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	141.0, 194.0, 48.0, 194.0
Sasang (904) AWS	2009/07/07, 2009/07/16, 2017/09/11	92.0, 137.0, 167.5
Yeongdo (910) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	218.5, 110.5, 9.5, 355.5
Gadeokdo (921) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	101.5, 90.5, 91.5, 265.5
Gijang (923) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	165.5, 178.5, 66.5, 21.5
Haeundae (937) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	278.0, 239.5, 79.0, 225.5
Busanjin (938) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	155.5, 194.0, 9.5, 181.0
Geumjeong (939) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	116.0, 167.5, 50.0, 96.5
Dongnae (940) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	116.5, 176.5, 78.5, 154.0
Bukgu (941) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	118.5, 166.0, 37.0, 89.0
Busannamgu (942) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	293.0, 258.5, 101.5, 263.0
Saha (950) AWS	2009/07/07, 2009/07/16, 2012/07/15, 2017/09/11	259.0, 221.5, 125.5, 249.0
Namhang (968) AWS	2012/07/15, 2017/09/11	64.5, 257.5
Bukhang (969) AWS	2012/07/15, 2017/09/11	8.5, 240.0

소별 관측날짜와 지속시간 6시간의 총 강우량을 나타내었다. 동일 날짜에 관측된 자료는 같은 시간대에서 관측된 강우 정보를 수집하였다. 총 51개의 관측강우 사상을 수집하였다.

2.2 1-2차원 홍수분석

도시유역에 대한 홍수량 데이터베이스를 만들기 위하여 사용한 SWMM 모형은 도시 및 비도시지역에 대하여 널리 사용되는 1차원 도시유출해석 모형이다(Rai et al., 2017). 앞서 조사한 관측 강우에 따른 유출량을 EPA-SWMM (EPA's Storm Water Management Model, 5.1 version)을 통하여 산출하였다. SWMM 모형의 실행블록 중 Runoff와 Extran을 활용하였으며, Runoff 블록은 강우사상 기반의 초기연산이 수행되는 블록으로써 배수유역 내의 유출현상, 맨홀에 대한 유입수문곡선을 계산한다. Extran 블록의 경우 Runoff 블록의 출력자료를 활용하여 배수관망 내에 흐르는 유량과 관내 수심을 산정하기 위하여 사용되며, 유량 및 수위 계산결과를 반영하여 관내의 역류현상, 월류량 해석이 가능하다(Park et al., 2019). 연구 대상지역에서 다양한 강우 사상에 따른 월류량 해석을 위하여 Eqs. (1), (2)에 나타난 Saint-Venant의 연속방정식과 운동량 방정식을 활용하였다. Q는 유출량(m^3), W는 소유역의 폭(m), n은 Manning의 조도계수, d는 수심(m), d_p 는 지면저류 손실수심(m), S는 소유역 경사를 의미하며, A는 소유역의 지표호를 단면적(m^2), V는 지표호를 속도(m/sec)를 나타낸다(Huber et al., 1988).

$$Q = W \times \frac{1}{n} (d - d_p)^{\frac{5}{3}} S^{\frac{1}{2}} \quad (1)$$

$$\frac{\partial Q}{\partial t} + gAS_f - 2V \frac{\partial A}{\partial t} - V^2 \frac{\partial A}{\partial x} + gA \frac{\partial H}{\partial x} = 0 \quad (2)$$

침수 범위에 대한 분석을 위한 2차원 홍수해석은 유한차분 기법을 기반으로 하는 수리해석 프로그램을 이용하였다. 적용한 FLO-2D 모형은 Federal Emergency Management Agency (FEMA, 미국 연방재난관리청)가 승인한 하천흐름과 홍수해석을 위한 수리모형으로 유한차분법을 이용한다(Risi et al., 2015). 2차원 수리학적 침수해석의 결과를 이용하여 각 강우조건에 따른 최대 침수 지도를 생성하였으며, 결과는 로지스틱 회귀의 종속 변수로서 사용된다. 홍수와 해석을 위한 2차원 격자의 크기는 5 m 정방형으로 하였다.

2.3 랜덤포레스트

랜덤포레스트(Random forest) 모델은 특정 사건 발생에 대한 분류 및 회귀를 수행하는데 다수의 의사결정나무를 구성하는 앙상블 학습을 사용하는 기법이다(Fig. 2). Zhou et al. (2018)이 시도한 서로 다른 종류의 인공신경망 간의 앙상블 학습을 이용하여 원하는 수문 자료를 예측할 수도 있지만, 본 연구에서 적용하는 랜덤포레스트는 유사한 의사결정나무를 사용자가 원하는 개수만큼 사용하고 각각의 결과를 종합하는 모형이다. 랜덤포레스트는 단순하지만 자연현상에 대한 해석을 실시하는데 높은 예측력을 나타내고 있다(Tyralis et al., 2019). 훈련 과정에서 다양한 의사결정나무를 구성하고 클래스 또는 평균인 클래스를 출력하는 방식으로 작동된다. 랜덤포레스트의 중요 매개변수는 max_features, bootstrap의 사용여부, n_estimator로 볼 수 있다. max_features 매개변수는 각 노드에서 사용할 특성의 최대 개수를 결정한다. bootstrap은 각각의 분류 모델을 위한 자료 샘플링 시 중복을 허용하는가에 대한 옵션이다. n_estimator의 경우 랜덤포레스트에서 만들 의사결정나무 개수를 결정한다. 일반적으로 랜덤포레스트를 구성하는 의사결정나무의 개수는

입력되는 변수의 개수 m 에 대한 임의의 $m/3$ 개의 변수를 선택하여 결정하게 된다(Choi et al., 2017).

랜덤포레스트의 알고리즘은 다음 네 단계로 요약될 수 있다.

- 1) n 개의 임의의 bootstrap 샘플을 추출한다.
- 2) bootstrap 샘플에서 의사결정 나무를 학습하는데, 각 노드에서는 중복을 허용하지 않고 랜덤하게 d 개의 특성을 선택한다. 정보 이득과 같은 목적 함수를 기준으로 최선의 분할을 만드는 특성을 사용해서 노드를 분할한다.
- 3) 단계 1), 2)를 k 번 반복한다.
- 4) 각 의사결정 나무의 예측을 모아 다수결 투표(majority voting)로 클래스 레이블을 할당한다.

Fig. 2. Typical Structure of Random Forest

가장 정보가 풍부한 특성으로 노드를 나누기 위해 최적화할

목적 함수를 정의한다. 랜덤포레스트에서 사용될 수 있는 목적 함수는 각 분할에서 정보 이득을 최대화한다. 정보 이득은 Eq. (3)과 같이 정의할 수 있다.

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j) \quad (3)$$

여기서 f 는 분할에 사용할 특성이며, D_p 와 D_j 는 부모와 j 번째 자식 노드의 데이터 세트이다. I 는 불순도(impurity) 지표이다. N_p 는 부모 노드에 있는 전체 샘플 개수이며, N_j 는 j 번째 자식 노드에 있는 샘플 개수를 나타낸다. 정보 이득은 단순히 부모 노드의 불순도와 자식 노드의 불순도 합의 차이이다. 자식 노드의 불순도가 낮을수록 정보이득이 더 커지게 된다.

3. 대상지역과 침수흔적도와의 비교

강우 사상에 따른 침수 발생을 분석하기 위하여, 부산시에 서 빈번하게 침수가 발생한 이력이 있는 해운대구 우동지역의 센텀시티 지구를 선정하였다. 본 지역에서는 2009년 7월 7일 집중호우와 2009년 7월 19일 집중호우로 인해 벅스코 지역 및 센텀시티 주변 도로 침수 및 지하침수 발생하였다. 저지대에 대한 빈번한 침수를 막기 위하여 2011년 7월에 올림픽공원 지하에 대규모 빗물 저장공간을 건설하였지만, 비교적 최근인 2017년 9월 11일에 벅스코 앞 APEC로에 대한 침수가 다시 발생하였다. 자연재해위험지구로 지정된 이후 부산시가 대규모의 예산을 들여 침수를 방지하기 위한 시설을 설치하였지만, 기후의 아열대로 인한 국지성 호우의 증가로 지속적으로 침수가 발생하고 있다.

연구 대상지역의 건물 높이를 고려한 Triangular Irregular Networks (TIN) 지형자료는 Fig. 3(a)와 같으며, 정방형 5 m간

(a) Triangular Irregular Network

(b) Drainage System

Fig. 3. Geographical Data for Study Area

(a) Inundation Trace Map (2009.7.7.)

(b) Result of Two-dimensional Simulation

Fig. 4. Comparison of Floodmark and Simulation Result

격의 Digital Elevation Model (DEM)을 이용하여 구축하였다. 배수관망은 Fig. 3(b)에 나타나있다. EPA-SWMM 구축시 600 mm 이상의 관거를 사용하였으며, 1개의 유출구, 78개의 맨홀, 78개의 관거를 사용하였다. 본 연구의 1, 2차원 수리학적 해석 결과의 적합성을 판단하기 위하여, 2009년 7월 7일에 대한 침수흔적도와 2차원 침수해석 결과를 비교하고자 하였다. 최근 침수 사상에 대한 관측자료가 부재하여, 침수흔적도는 선행연구에서 제시된 침수흔적도를 이용하였다. 대상구역에 대한 침수피해 주요 원인으로는 복잡한 배수 관망과 수영강 해수면 만조위일때 강우강도가 높은 강우발생 시기와 겹쳐 우수배제가 원활하게 이뤄지지 못하였기 때문이다(Cho et al., 2011). Fig. 4에 보고된 침수흔적도와 FLO-2D를 이용한 2차원 침수해석 결과를 나타내고 있으며, 선행연구와는 다른 2차원 침수해석 프로그램을 사용했음에도 불구하고 유사한 침수양상이 계산되었다.

4. 강우에 따른 수치해석 결과 검토

부산광역시 전역에서 관측된 51개 관측강우에 대하여 1, 2차원 수리학적 해석을 실시하였다. 먼저 SWMM을 이용한 1차원 도시유출해석을 통하여 맨홀지점에 대한 월류량을 계산하였다. 계산된 지점별 월류량은 2차원 침수해석을 실시하는 FLO-2D 프로그램의 입력 자료로서 적용되었으며, 각 강우사상에 대한 최대침수심 지도를 계산하였다. 조사된 강우 사상을 이용하여 단일 연구대상지역에 대한 1, 2차원 해석을 실시하고 침수 발생유무를 판단하였다. 침수 발생에 대한 기준은, 2차원 침수해석 시 최대 침수심이 도로의 연석 높이가 20 cm 이상일 때를 기준으로 하였다. Table 2에 모의된 총 월류량과 최대 침수심이 나타나 있다. 총 51개의 강우사상 중에 33개의 강우사상에서 침수가 발생하였다.

모의 결과를 살펴보았을 때에, 총 강우량과 발생 최대 침수심이 완전한 선형관계가 아님을 확인할 수 있다. 48.0 mm의 강우량에서 침수가 발생하였지만, 50.0~66.5 mm 범위의 강우 시나리오에 대해서는 침수가 발생하지 않는 것으로 모의결과가 확인되었다. 또한 167.5 mm로 동일한 총 강우량 특성을 가진 두 개의 강우 시나리오가 있지만, 발생 최대 침수심은 상이하게 나타났으며 0.35 m의 차이를 나타내었다. 수리학적 해석에 기반하였을 때에 총 강우량이 아닌 다른 통계학적 특성이 침수 발생에 적지 않은 영향을 주는 것으로 판단된다. 총 강우량과 최대침수심간의 R-square은 0.7로 나타났다.

5. 중요도 산정

5.1 침수양상에 대한 중요도

대상 지역에서 침수를 일으키는 강우의 통계학적 인자에 대하여 면밀히 파악하기 위해, 랜덤포레스트 분류기법을 사용하였다. 분류의 대상은 침수 발생 유무(0: 침수 발생하지 않음, 1: 침수가 발생함)이며, 이에 대한 영향 인자는 6시간 지속시간 관측 강우의 통계학적 특성이다. 높은 침수심이 나타나는데 영향을 주는 강우 통계학적 인자를 파악하기 위하여 랜덤포레스트 회귀기법을 사용하였으며, 회귀의 대상은 대상구역 내에서 발생하는 최대침수심이다.

랜덤포레스트의 목표값은 앞서 제시한 1, 2차원 수치해석 결과를 참조하였다. 각 관측 강우에 대하여 총강우량(mm), 평균강우량(mm), 10분 단위 최대강우량(mm), 강우강도(mm/hr), 침도, 왜곡도, 1시간부터 5시간 단위 최대강우량(mm)을 고려하였다. 1시간 단위 최대강우량의 경우 지속시간 6시간 관측 강우 내에서 1시간(60분) 단위로 강우량의 합을 산정했을 때에, 최대로 나타나는 강우량을 나타낸다.

Table 2. One- and Two-dimensional Flood Simulation Results

Total Rainfall (mm)	Total Overflow (m ³ /s)	Max. Flood Depth (m)	Total Rainfall (mm)	Total Overflow (m ³ /s)	Max. Flood Depth (m)
8.50	0.00	0.00	155.50	1.95	0.36
9.50	0.00	0.00	165.50	1.35	0.27
9.50	0.00	0.00	166.00	0.41	0.00
21.50	0.00	0.00	167.50	1.30	0.23
37.00	0.00	0.00	167.50	8.38	0.58
48.00	8.13	0.50	176.50	1.85	0.35
50.00	0.00	0.00	178.50	2.53	0.46
64.50	0.00	0.00	181.00	17.90	0.66
66.50	0.00	0.00	194.00	15.10	0.70
78.50	0.61	0.20	194.00	7.33	0.67
79.00	0.00	0.00	194.00	9.76	0.52
89.00	0.81	0.22	218.50	5.42	0.54
90.50	3.47	0.51	221.50	19.25	0.75
91.50	6.10	0.60	225.50	21.03	0.66
92.00	0.00	0.00	239.50	26.36	0.79
96.50	0.00	0.00	240.00	27.53	0.88
101.50	0.00	0.00	249.00	23.89	0.69
101.50	2.77	0.47	257.50	19.12	0.72
110.50	0.00	0.00	258.50	40.04	1.20
116.00	0.00	0.00	259.00	17.23	0.71
116.50	0.00	0.00	263.00	30.79	0.87
118.50	0.01	0.00	265.50	18.77	0.73
125.50	17.78	0.73	278.00	24.43	0.87
137.00	14.91	0.61	293.00	31.11	1.06
141.00	1.01	0.25	355.50	88.99	1.44
154.00	10.81	0.67	-	-	-

Table 3. Statistical Features for 51 Observed Rainfall

Statistical Feature	For 51 Rainfall Scenario			
	Max.	Min.	Average	Std.
Total rainfall (<i>mm</i>)	355.50	8.50	151.33	83.09
Avg. rainfall (<i>mm</i>)	9.88	0.24	4.20	2.31
10 min. max. rainfall (<i>mm</i>)	26.00	2.00	14.53	5.48
Intensity (<i>mm/1hr</i>)	59.25	1.42	25.22	13.85
Kurtosis	31.35	-1.38	2.76	6.81
Skewness	5.49	0.14	1.40	1.14
1hr max. rainfall (<i>mm</i>)	114.00	5.00	57.52	23.42
2hr max. rainfall (<i>mm</i>)	219.00	8.50	89.15	43.71
3hr max. rainfall (<i>mm</i>)	290.00	8.50	111.38	60.28
4hr max. rainfall (<i>mm</i>)	329.50	8.50	131.63	73.58
5hr max. rainfall (<i>mm</i>)	352.00	8.50	143.30	80.52

이는 총강우량에 지속시간 6시간을 단순히 나눠 산정한 강우 강도와 다른 결과이다. 51개 관측강우 사상에 대하여 산정한 통계학적 강우 특성은 Table 3과 같다. 각 통계학적 특성에

따른 최대값, 최소값, 평균값, 표준편차를 산정하였다.

대상유역에 대하여 침수 발생에 영향을 주는 강우 통계학적 특성은 랜덤포레스트 분류를 통하여 분석되었으며, 침수

심에 영향을 주는 특성에 대해서 랜덤포레스트 회귀를 적용하였다. 랜덤포레스트 분류 시에 지니 불순도(Gini impurity, I_g)를 기준으로 분석을 수행하였으며, 각 의사결정나무의 자식노드에서 나타나는 지니 불순도가 낮게 나타나 정보이득이 최대로 되는 방향으로 최적화 된다. 지니 불순도는 Eq. (4)와 같이 나타낼 수 있다. 여기서 c 는 클래스 환경을 나타내며, $p(i|t)$ 는 특정노드 t 에서 클래스 i 에 속한 샘플 비율을 나타낸다.

$$I_G(t) = \sum_{i=1}^c p(i|t)(1-p(i|t)) = 1 - \sum_{i=1}^c p(i|t)^2 \quad (4)$$

랜덤포레스트 회귀를 수행할 때에는 불순도 측정을 평균

제곱 오차(Mean Square Error, MSE)를 기준으로 수행하게 되며(Eq. (5)), 예상 추정값과 실제 추정값 간의 평균 제곱 오차를 계산하게 된다. 랜덤포레스트 분류와 마찬가지로 자식노드들에서 평균 제곱 오차가 작게 나타나야, 정보이득이 증가한다. $\hat{y}$ 는 실제 예측값이며, y 는 노드 조건에 따른 예상값으로 볼 수 있다.

$$MAE = \frac{1}{n} \sum |\hat{y} - y| \quad (5)$$

강우 통계학적 특성자료와 랜덤포레스트를 연계한 분류 및 회귀 과정은 Fig. 5와 같이 나타낼 수 있다. 분류 수행을 위한 의사결정나무의 노드 클래스는 홍수 또는 비홍수로

(a) A decision tree in random forest (classification)

(b) A decision tree in random forest (regression)

Fig. 5. A Decision Tree in Random Forest

Table 4. Statistical Features for 51 Observed Rainfall

Statistical Feature	Feature Importance for	
	Flood Occurrence	Flood Depth
Total rainfall (mm)	0.040	0.126
Avg. rainfall (mm)	0.157	0.028
10 min. max. rainfall (mm)	0.118	0.064
Intensity (mm/1hr)	0.128	0.035
Kurtosis	0.014	0.104
Skewness	0.147	0.033
1hr max. rainfall (mm)	0.195	0.134
2hr max. rainfall (mm)	0.038	0.106
3hr max. rainfall (mm)	0.015	0.205
4hr max. rainfall (mm)	0.120	0.092
5hr max. rainfall (mm)	0.028	0.073
Sum.	1.000	1.000

(a) Feature importance for flood occurrence

(b) Feature importance for flood depth

Fig. 6. Estimation of Feature Importance with Random Forest Model

나타나며, 회귀 수행을 위한 의사결정나무의 경우 각 최대침수심을 추정하게 된다. 다만, 랜덤포레스트 회귀 시 침수가 발생하지 않은 사상에 대해서는 고려하지 않았다. 랜덤포레스트를 구성하는 의사결정나무의 개수는 10개로 하였으며, 그 중 첫 번째 의사결정나무를 나타내었다. 지식노드에서 지니 계수 또는 평균 제곱 오차가 적게 나타나는 방향으로 트리가 구성되는 것을 확인할 수 있다. 정보이득을 높여줄 수 있는 특정 기준(강우 통계학적 특성)에 대하여 중요도가 산정되게 된다.

침수 발생에 대하여 영향을 주는 강우 통계학적 특성은 1시간 최대강우량, 평균 강우량, 왜곡도, 강우강도 순으로 나타났으며, 침수심에 대한 영향도는 3시간 최대강우량,

1시간 최대강우량, 총강우량, 2시간 최대강우량 순으로 나타났다. 랜덤포레스트 기법을 이용하여 중요도를 산정한 결과는 Table 4에 나타나있으며, Fig. 6과 같이 중요도 내림차순으로 나타낼 수 있다.

6. 결 론

본 연구에서는 침수발생 이력이 있는 부산광역시 센텀지구에 대하여, 다양한 관측 강우사상과 함께 1, 2차원 수리학적 해석을 실시하였다. 수치해석 결과와 랜덤포레스트 기법을 이용하여 침수의 발생과 침수심의 증가에 대한 영향인자를 강우의 통계학적 특성에 초점을 맞추어 분석하였다.

총강우량, 평균 강우량, 10분 단위 최대강우량, 강우강도, 침도, 왜곡도, 1~5시간 최대 강우량을 고려하였으며, 주요 연구결과는 다음과 같다.

- (1) 다양한 강우 시나리오에 따른 침수양상을 파악하기 위하여, 부산광역시 전역에서 관측된 지속시간 6시간을 가지는 강우 사상을 적용하였다. 강우 자료를 바탕으로 1차원 도시유출해석과 2차원 침수해석을 실시하였다. 침수발생 유무와 최대침수심을 비교한 결과로, 동일한 총강우량에 불구하고 침수 발생의 차이가 나타나거나 상대적으로 낮은 총강우량에서 더 높은 최대침수심이 발생하기도 하였다. 이는 총강우량이 도시 배수분구의 침수발생 상황에 정비례적인 관계가 아님을 나타낸다. 이에 관측 강우별 통계학적 특성에 따른 침수 발생 영향을 파악하기 위하여, 1, 2차원 수치해석 결과 기반 랜덤포레스트 입력 자료를 생성하였다.
- (2) 침수의 발생 유무에 따른 분석을 랜덤포레스트 분류로 수행하였으며, 목표값은 침수가 발생하지 않았을 시에 0, 침수가 발생했을 경우 1로 선정하였다. 발생 최대 침수심에 대한 강우 특성의 영향을 파악하기 위해 랜덤포레스트 회귀를 이용하였으며, 침수가 발생한 관측 강우 사상만을 적용하였다. 랜덤포레스트의 입력 인자의 중요도 산정 기법을 사용하였으며, 침수발생에 대한 주요 강우 특성은 1시간 최대 강우량, 평균강우량, 왜곡도 순으로 나타났다. 상대적으로 높은 침수심을 발생시키는데 주요한 인자는 3시간 최대강우량, 1시간 최대강우량, 총강우량 순으로 나타났다. 침수발생 유무에 대해서는 정량적 강우 크기 외에 강우 분포의 형상과 강우 강도에 큰 영향을 받았다. 침수심의 크기에 대하여 강우의 정량적 크기에 더 영향을 받으며, 침수발생 유무에 따른 중요도와 달리 강우 분포의 침도에 상대적으로 더 영향을 받는 것으로 나타났다. 이는 유사한 강우량일 때 강우의 집중도에 따라 발생하는 침수심이 상이하게 계산된 것으로 보인다.
- (3) 본 연구에서는 단순한 총강우량, 강우강도 외에 시간 단위별 최대강우량, 강우의 왜곡도 및 침도를 고려하여 홍수발생에 대한 영향을 분석하였다. 종합한 결과 1시간 단위 최대 강우량이 침수 상황에 대한 주요 영향인자로 판단되었으며, 침수가 발생한 후 최대로 발생하는 침수심에 대해서는 강우의 정량적 크기에 영향을 많이 받는 것으로 확인되었다. 침단 장비와 함께 실시간으로 강우가 관측되는 현시점에서 침수 발생에 따른 주요 강우 인자를 파악하는 것은 홍수예보 기술 발전을 위한 기초자료가 될 것으로 판단된다. 또한, 비선형 관계를 가질 수 있는 침수 유무에 따른 강우 통계학적 인자의 영향도를 랜덤포레스트 모델을 통해 충분히 분석이 가능함을 나타내었다. 이에 대한

분석결과는 도시 배수분구 특징별 상이하게 나타날 수 있으며, 홍수예보 발령을 위한 기상학적 근거를 마련하기 위해 더욱 다양한 배수분구에서 연구될 필요가 있다.

감사의 글

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2020R1I1A3074459).

References

- Candela, A., and Aronica, G.T. (2016). Rainfall thresholds derivation for warning pluvial flooding risk in urbanised areas. *3rd European Conference on Flood Risk Management (FLOODrisk 2016), E3S Web of Conferences*, Vol. 7. doi:10.1051/e3sconf/20160718016.
- Cho, J.W., Bae, C.Y., and Kang, H.S. (2018). Development and application of urban flood alert criteria considering damage records and runoff characteristics. *J. Korea Water Resour. Assoc.*, Vol. 51, No. 1, pp. 1-10.
- Cho, W.H., Han, K.Y., Hwang, T.J., and Son, A.L. (2011). 2-D inundation analysis in urban area considering building and road. *J. Korean Soc. Hazard Mitig.*, Vol. 11, No. 5, pp. 159-168.
- 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.
- Golian, S., Saghaian, B., Elmi, M., and Maknoon, R. (2011). Probabilistic rainfall thresholds for flood forecasting: Evaluating different methodologies for modelling rainfall spatial correlation (or dependence). *Hydrological Processes*, Vol. 25, No. 13, pp. 2046-2055. doi:10.1002/hyp.7956
- Huber, W.C., Dickson, R.E., Roesner, L.A., and Aldrich, J.A. (1988). *Storm water management model. User's manual, Version 4*. EPA/600/S3-88/001, U.S. Environmental Protection Agency.
- Kim, H.I., and Han, K.Y. (2019). Computation of criterion rainfall for urban flood by logistic regression. *Journal of the Korean Society of Civil Engineers*, Vol. 39, No. 6, pp. 713-723.

- Lee, J.H., and Yeon, K.S. (2008). Flood inundation analysis using XP-SWMM model in urban area. *J. Korean Soc. Hazard Mitig.*, Vol. 8, No. 5, pp. 155-161.
- Montesarchio, V., Ridolfi, E., Russo, F., and Napolitano, F. (2011). Rainfall threshold definition using an entropy decision approach and radar data. *Nat. Hazards Earth Syst. Sci.*, Vol. 11, pp. 2061-2074. doi:10.5194/nhess-11-2061-2011
- Park, J.H., Kim, S.H., and Bae, D.H. (2019). Evaluating appropriateness of the design methodology for urban sewer system. *Journal of Korea Water Resource Association*, Vol. 52, No. 6, pp. 411-420.
- Rai, P.K., Chahar, B.R., and Dhanya, C.T. (2017). GIS-based SWMM model for simulating the catchment response to flood events. *Hydrology Research*, Vol. 48, No. 2, pp. 384-394.
- Risi, R.D., Jalayer, F., and De Paola, F. (2015). Meso-scale hazard zoning of potentially flood prone areas. *Journal of Hydrology*, Vol. 527, pp. 316-325.
- Tyralis, H., Papacharalampous, G., and Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, Vol. 11, No. 5, 910. doi:10.3390/w11050910
- Yoon, S.S., Bae, D.H., and Choi, Y.J. (2014). Urban inundation forecasting using predicted radar rainfall: Case study. *J. Korean Soc. Hazard Mitig.*, Vol. 14, No. 3, pp. 117-126.
- Zhou, J.Z., Peng, T., Zhang, C., and Sun, N. (2018). Data pre-analysis and ensemble of various artificial neural networks for monthly streamflow forecasting. *Water*, Vol. 10, No. 5, 628. doi:10.3390/w10050628

Received	October 21, 2020
Revised	October 21, 2020
Accepted	November 5, 2020