

Development of an Automated Algorithm for Analyzing Rainfall Thresholds Triggering Landslide Based on AWS and AMOS

Donghyeon Kim*, Song Eu**, Kwangyoun Lee*, Sukhee Yoon*, Jongseo Lee*, Donggeun Kim***

*Researcher, Info. Mgmt. and Data Analytics Team, Divi. of Research Tech. Research Inst., Cheongju, Korea

**Research Official, Landslide Div., Dept. of Forest Disaster and Environ., National Inst. of Forest Science, Seoul, Korea

***Professor, Dept. of Forest Ecology and Prot., Kyungpook National University, Sangju, Korea

[Abstract]

This study presents an automated Python algorithm for analyzing rainfall characteristics to establish critical rainfall thresholds as part of a landslide early warning system. Rainfall data were sourced from the Korea Meteorological Administration's Automatic Weather System (AWS) and the Korea Forest Service's Automatic Mountain Observation System (AMOS), while landslide data from 2020 to 2023 were gathered via the Life Safety Map. The algorithm involves three main steps: 1) processing rainfall data to correct inconsistencies and fill data gaps, 2) identifying the nearest observation station to each landslide location, and 3) conducting statistical analysis of rainfall characteristics. The analysis utilized power law and nonlinear regression, yielding an average R^2 of 0.45 for the relationships between rainfall intensity-duration, effective rainfall-duration, antecedent rainfall-duration, and maximum hourly rainfall-duration. The critical thresholds identified were 0.9-1.4 mm/hr for rainfall intensity, 68.5-132.5 mm for effective rainfall, 81.6-151.1 mm for antecedent rainfall, and 17.5-26.5 mm for maximum hourly rainfall. Validation using AUC-ROC analysis showed a low AUC value of 0.5, highlighting the limitations of using rainfall data alone to predict landslides. Additionally, the algorithm's speed performance evaluation revealed a total processing time of 30 minutes, further emphasizing the limitations of relying solely on rainfall data for disaster prediction. However, to mitigate loss of life and property damage due to disasters, it is crucial to establish criteria using quantitative and easily interpretable methods. Thus, the algorithm developed in this study is expected to contribute to reducing damage by providing a quantitative evaluation of critical rainfall thresholds that trigger landslides.

▶ **Key words:** Algorithm, Rainfall characteristics, Landslide, AWS, AMOS

-
- First Author: Donghyeon Kim, Corresponding Author: Donggeun Kim
 - *Donghyeon Kim (d2h@kafet.or.kr), Info. Mgmt. and Data Analytics Team, Divi. of Research Tech. Research Inst.
 - **Song Eu (eusong89@korea.kr), Landslide Div., Dept. of Forest Disaster and Environ., National Inst. of Forest Science
 - *Kwangyoun Lee (lky@kafet.or.kr), Info. Mgmt. and Data Analytics Team, Divi. of Research Tech. Research Inst.
 - *Sukhee Yoon (shyoon@kafet.or.kr), Info. Mgmt. and Data Analytics Team, Divi. of Research Tech. Research Inst.
 - *Jongseo Lee (jongseo@kafet.or.kr), Info. Mgmt. and Data Analytics Team, Divi. of Research Tech. Research Inst.
 - ***Donggeun Kim (dgkim96@knu.ac.kr), Dept. of Forest Ecology and Prot., Kyungpook National University
 - Received: 2024. 07. 15, Revised: 2024. 09. 10, Accepted: 2024. 09. 10.

[요 약]

산사태 예방을 위한 비구조물 대책 중 하나로, 산사태 예·경보 시스템의 임계 강우 기준을 마련하기 위한 강우 특성 분석 자동화 알고리즘을 제안한다. 알고리즘은 파이썬으로 작성했다. 강우 자료는 기상청과 산림청의 관측소 정보를 활용하였고, 산사태 정보는 2020~2023년까지의 데이터를 기반으로 행정안전부의 생활안전지도를 통해 수집했다. 알고리즘은 3단계이다. 첫째, 강우 자료를 입력받아 지점 정보 불일치와 결측값을 정제한다. 둘째, 산사태 위치와 가장 가까운 관측소를 찾아내고, 무강우 기간 및 반감기를 고려하여 강우 사상을 분류하고 분석한다. 셋째, 강우 특성에 대해 기초 통계를 수행한 후, 적절한 통계 모델을 선정·분석한다. 분석 결과, 강우강도-지속 시간, 유효 강우량-지속 시간, 선행강우량-지속 시간, 최대강우량-지속 시간의 관계에 대해 멱법칙과 비선형 회귀분석을 실시하였으며, 평균 R^2 값이 0.45로 나타났다. 분석된 임계 강우 기준은 강우강도 0.9~1.4mm/hr, 유효강우량 68.5~132.5mm, 선행강우량 81.6~151.1mm, 최대강우량 17.5~26.5mm로 분석되었다. 기준 검증을 위해 AUC-ROC 분석을 수행한 결과, AUC 값이 0.5로 낮게 나타났다. 마지막으로 알고리즘의 속도 성능을 평가한 결과 총 30분의 시간이 소요되었다. 즉, 강우 자료만으로 재해를 예측하는 데에는 한계가 있음을 시사한다. 단, 재해로 인한 인명과 재산 피해를 막기 위해서는 정량적이나 해석이 편리한 방법으로 기준 선정이 필요하다. 따라서 본 연구에서 개발한 알고리즘을 활용하여 산사태 발생 임계 강우 기준의 정량적 평가를 통해 인명 및 재산 피해를 저감하는 데 기여할 수 있을 것으로 기대한다.

▶ **주제어:** 알고리즘, 강우특성, 산사태, 방재기상관측소, 산악기상관측소

I. Introduction

산사태 및 토석류와 같은 산지토사재해는 지표의 안정화 과정 또는 산지개발과 같은 인위적 행위로 인한 파괴과정이라 정의할 수 있다. 일반적으로 산지토사재해는 사면의 안정성이 감소하여 발생하는데 이와 관련된 주요 요인은 크게 자연적 요인과 인위적 요인으로 구분할 수 있다.

자연적 요인의 경우 강우, 해빙, 지질 구조, 임상 구조의 변화, 지진 등이며, 인위적 요인은 인간 활동으로 수반되는 지반의 절·성토작업 등이다. 특히 강우는 기반암과 상부 퇴적물 사이의 마찰력과 전단강도를 떨어뜨려 산사태를 유발하는 주요 요인으로 손꼽힌다[1-3].

우리나라의 연평균 강수량은 지역에 따라 다양하나, 일반적으로 1,000mm ~ 1,700mm 사이로 보고되고 있다. 2023년에는 전국 평균 강수량이 1,746mm를 기록하여[4] 2022년 대비 37%p 증가한 강수량을 보였다. 세계은행 자료에 의하면 전 세계 연평균 강수량인 약 900mm보다 41.5%p 많은 강수량을 나타내는 것으로 분석되었다[5].

이와 더불어 극한강우 및 폭설의 발생과 이상 고온 현상에 의해 기계적 및 화학적 풍화가 활발히 이루어져 암석 등이 붕괴, 풍화되어 모래질 등으로 변모할 가능성이 높아질 수 있다. 특히 모래질로 구성된 경사면은 산지토사재해

에 취약한 구조를 나타낸다. 이는 앞서 언급된 여러 가지 상황을 종합하였을 때, 산사태와 같은 재해 발생의 위험성이 증가한다는 것을 의미한다. 산림청에서 집계한 최근 10년간(2014 ~ 2023년) 산사태 발생 추이를 살펴보면 총 2,586ha의 피해 면적이 발생하였으며, 2023년에는 459ha로 2022년 대비 약 40%p 증가했다[25].

연도별 산사태 발생 추이는 당시의 기상 상황에 따라 차이가 존재하므로 일정한 패턴이 존재하지 않는다. 그러나 IPCC(Intergovernmental Panel on Climate Change) 6차 보고서에 따르면 온도 상승으로 인해 대기 중의 수증기 함량이 증가하여 극한 강우 증가율(매년 4 ~ 8%p)이 Clausius-Clapeyron 방정식과 유사하다고 보고되었다[6]. 또한 기상청에서 발표한 SSP(Shared Socioeconomic Pathways)의 SSP1-2.6(저탄소 시나리오)과 SSP5-8.5(고탄소 시나리오)의 글로벌 육상 강수량 시나리오 결과, SSP1-2.6에서는 2021~2040년간 1,366.6mm의 강수량을 시작으로 2100년간 10년당 0.47mm의 강수량 증가를 나타내어 1,373.8mm 강수량이 예측되었다. 그에 반해 SSP5-8.5는 2021 ~ 2040년간 1,318.7mm로 SSP1-2.6에 비해 47.9mm 적은 양이 예측

되었으나, 중국에는 1,571.4mm로 10년당 41.65mm의 강수량 증가가 예측되었다.

이 밖에도 2017년에 6,271ha였던 준보전산지의 산지전용 면적이 2018년에는 약 20%p 증가한 7,548ha였고, 산사태 취약지역은 2023년 6월 기준 27,948개소로 2020년 대비 5.5%p 증가하였다. 이어서 산사태 취약지역 거주민 또한 72,231명으로 4.2%p 증가하였다[7]. 이처럼 범지구적 기상이변에 따른 호우일수 및 국지성 호우의 증가 같은 강수 패턴의 변화와 동시에 산지 개발 증가로 산지토사재해(산사태 및 토석류 등)에 의한 인명 및 재산 피해 우려가 증대되고 있다.

산림청은 산지토사재해를 예방·방지하고자 사방구조물의 설치와 같은 구조물적 대책과 더불어 산사태 위험도 평가 및 산사태 예측정보의 제공 등과 같은 비 구조물적 대책을 시행 중이다[8]. 또한 산사태 위험지도와 산지토사재해를 실시간으로 제공하는 예·경보 시스템은 전 국민에게 제공되는 대표적인 비 구조물적 대책으로 최근 실시간 강우의 영향을 반영하는 동적 산사태 위험지도를 구현하고 있다. 일반적으로 국내에서 산사태 유발 강우 기준은 주의보에서 최대시우량: 20~30mm, 선행강우량: 80~150mm, 유발강우량: 100~200mm로 알려져 있다. 올바른 예·경보 강우 기준치 마련은 지역주민의 대피 등 대책 마련에 중요한 요소임에 따라 다양한 선행연구들이 수행된 바 있다 [9-11].

정리하자면 예측정보의 고도화를 위해서는 산지토사재해 발생지에 대한 토질 특성, 지형 및 임상 특성 등 공간적 자료와 재해 발생 당시의 강우 특성 등과 같은 시·공간적 자료가 요구된다. 특히, 강우 특성 자료의 경우 예·경보 강우 임계치 기준 마련에 활용된다. 임계치 마련을 위해 많은 선행연구에서는 통계적 방법을 이용하고 있다[12-14]. 통계적 방법은 많은 양의 정보를 필요하다. 그러나 자료의 용량 증가는 분석자가 수동 분석하는 과정에 과부하를 야기할 수 있다. 적은 양의 강우 정보를 이용하여 강우 임계치 기준을 마련할 경우 오경보 비율이 높아져 신뢰성 하락의 위험이 존재한다. 따라서 올바른 예·경보 강우 임계기준 마련을 위해선 대규모 정보를 자동으로 처리할 수 있는 알고리즘이 필요하다.

다음은 본 논문의 성과를 요약한 것이다.

- 기상자료개방포털(data.kma.go.kr)에서 취득한 방재 기상관측소(Automatic Weather System, AWS)와 산악기상관측소(Automatic Mountain Observation System, AMOS)의 강수량 정보를 정제하고 분석한다.
- 예·경보 강우 임계치 기준 마련을 위해 다양한 통계적

방법을 반영한 알고리즘을 구현한다.

- 사용자 사용 편의성 및 자료 시각화를 위해 GUI(Graphical User Interface)를 설계·구현한다.

본 논문의 구성은 다음과 같다. 2장은 산사태와 강우 특성 분석 및 프로그램 개발과 관련된 연구를 검토한다. 3장은 개발환경 및 활용된 자료 현황과 강우 특성 자동분석 과정을 기술한다. 4장은 알고리즘 수행 단계별로 산출된 연구 결과에 대해 기술 및 서술한다. 5장은 본 논문의 결론을 제시한다.

II. Preliminaries

1. Definition of the rainfall characteristic

강우와 산지토사재해와의 관계를 알아보기 위해서는 사면붕괴가 발생할 당시의 여러 가지 강우 자료를 수집하여 시간과의 관계 규명에 어떠한 역할로 작용하는 것인지 파악할 필요가 있다. 강우 자료 분석을 위해 필요한 용어와 기능은 아래의 설명 및 Fig. 1과 같다.

- 유효강우량(Effective Rainfall, E): 선행강우 이후 산사태가 발생한 강우사상 중 재해 발생 시점까지의 강우량
- 선행강우량(Antecedent Rainfall, RA): 유효강우 시작 전 발생한 n 기간 동안의 강우량
- 선행작용강우량(Antecedent Working Rainfall, RWA): 반감기별 감소계수를 선행강우량에 곱하여 얻어진 강우량
- 지속시간(Duration, D): 유효강우량이 발생한 총 시간
- 강우강도(Intensity, I): 단위시간 동안 내리는 강우량 (mm/hr)
- 최대시우량(Maximum Rainfall, RM): 하루 동안 발생한 강수 중 최대 강수량

파란색 막대그래프 구간은 선행강우량을 표현한 부분이며 해당 구간에 발생한 강우량을 대상으로 식 (1)을 이용하여 산출된 것이 RWA이다.

$$RWA = \sum_{t=1}^{24} (a_t \cdot d_t) \dots\dots\dots (1)$$

여기서 a_t 은 t일 전의 감소계수이며, d_t 은 t일 전의 강우량이다. 여기서 감소계수는 식 (2)로 계산한다.

$$a_t = 0.5^{t/T} \dots\dots\dots (2)$$

T는 반감기(강우량이 반으로 감소하는 일수)이다. 본 연구에서 적용한 반감기는 [26]에서 적용하고 있는 2일의 반

감기를 이용하였고, 총 24일간의 감소계수를 적용했다.

빨간색 막대그래프 구간은 산사태가 발생하였을 것으로 추정되는 구간 중, 산사태가 발생한 시점까지의 강우량인 E이며, 이때 산사태 발생 시점은 RM으로 특징한다.

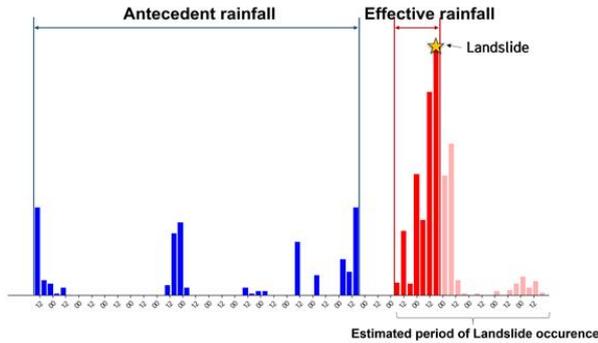


Fig. 1. Concept of a series of rainfall characteristic

2. Related works

국내에서는 산사태 유발 강우 임계치를 설정하기 위해 다양한 경험적 및 통계적 방법을 사용했다. 한국지질자원연구원(Korea Institute of Geoscience and Mineral Resources, KIGAM)에서는 과거 산사태 발생 자료를 분석하여 특정 강우량 이상일 때 산사태가 발생하는 경향을 파악하고, 이를 바탕으로 임계치를 설정했다. 이 과정에서 시간별 강우량과 누적 강우량, 강우강도 등의 요소가 고려되었다 [15]. 경험적 방법의 주요 장점은 실제 사례를 기반으로 하므로 실효성이 높다는 장점이 있다. [16]은 로지스틱 회귀분석(Logistic regression)을 통해 I와 D의 관계를 분석하여 강우 임계치를 정의하는 연구를 수행한 바 있다. 이처럼 국내 대부분의 연구에서는 I-D 곡선의 관계식을 대상으로 로지스틱 회귀 또는 분위수 회귀분석(Quantile regression)을 이용하여 강우 한계선을 정의하여 임계치를 분석하는 연구가 수행되었다 [17-18]. 다양한 강우량 변수를 종합적으로 검토한 [19]에서는 대규모 자료와 일관된 방법에 의한 자료 정제와 정렬을 통해 자료를 생산하고, 여러 가지의 강우 특성을 조합하고 분석하는 것이 산사태 발생 예측을 위한 강우량 임계치 개발 연구의 발전을 도모하는 것이라 언급했다. [20-21]은 OMIRL(Osservatorio Meteo Idrologico della Regione Liguria)에서 운영하는 172개의 우량계를 활용하여 산사태 발생 강우 임계값 자동 계산 도구인 CTRL-T(Calculation of Thresholds for Rainfall-induced Landslides Tool)를 개발한 바 있다. [22]는 CTRL-T를 보조하는 산사태 발생 위치기반의 대표 강우량계를 정량적으로 선정하는 QGIS 기반의 계산 도구를

개발한 바 있다.

종합하여 국내 연구는 AWS의 강우정보만을 기반으로 IETD를 조절하여 산사태를 유발할 수 있는 I-D 곡선도출 연구가 주를 이루고 있으며, 국외 연구는 산사태 발생 강우 임계값을 자동으로 계산하는 알고리즘 개발과 E-D와 같이 강우특성별 설명력이 높은 최적의 관계식을 산정하는 연구가 다수 존재한다.

따라서 본 연구는 AWS와 더불어 산악기상 상황을 반영할 수 있는 AMOS의 강수 자료를 자동으로 처리하고, 임계 강우 기준 분석을 수행할 수 있는 알고리즘을 개발했다.

III. The Proposed Scheme

1. Data sets and Environment

1.1 Data sets

연구에 활용된 자료는 다음과 같다(Table 1).

1) 기상청에서 제공하는 554개의 AWS 강수 정보와 산림청의 464개의 AMOS을 기상자료개방포털에서 2020 ~ 2023년 강수량 정보 전체를 수집했다. 2) AWS 관측소 위치 또한 기상자료개방포털에서 수집했다. 3) 2020 ~ 2023년 산사태 정보는 생활안전지도를 통해 9,594개(2020년: 5,852개, 2021년: 182개, 2022년: 1,235개, 2023년: 2,325개) 자료를 수집하고, 4) AMOS 관측소 위치자료는 국립산림과학원의 협조를 통해 수집했다.

Table 1. Characteristics of input data for rainfall analysis

Contents	Format	Source
• AWS rainfall data	CSV	KMA*
• AMOS rainfall data	.zip	
• AWS location data	CSV	MOIS**
• Landslide data	CSV	
• AMOS location data	CSV	NIFoS***

*: Korea Meteorological Administration

** : Ministry of the Interior and Safety

***: National Institute of Forest Science

1.2 Development environment

알고리즘 개발에 활용된 언어는 Python 3.9.13이며, numpy, pandas 등 다양한 API(Application Programming Interface)를 활용했다. GUI는 Qt Designer를 사용했다. 통합개발환경(Integrated Development Environment, IDE)은 Visual Studio Code(VSCode)이며, 운영체제는 Window 10에서 개발했다(Table 2).

Table 2. Algorithm development environment

OS	IDE	P.L.*	API
Window 10	VSCode	Python 3.9.13.	os, glob, csv, pandas, numpy, statsmodels, matplotlib

*: Programming Language

2. Description of algorithm

2.1 Algorithm

강우 자료 플랫폼별 강우 특성을 자동으로 분석하는 알고리즘의 흐름도는 Fig. 2와 같다.

1) BLOCK 1: 입력자료(강수량 정보, 관측소 위치 정보)에 대한 무결성을 확인하기 위한 자료 정제 단계이다. 정제 내용은 관측소별 결측된 강수량의 개수 및 결측치 보완 작업과 운영 종료된 관측소를 제거하는 작업으로 구분된다. 이때 결측치는 산사태가 발생 되었을 것으로 추정되는 기간(5 ~ 10월) 내에 한정하여 식 (3)에 따라 단순 비례법으로 보간했다.

$$P_x = \frac{P_y}{N_y} N_x \dots\dots\dots (3)$$

N_y 는 y 관측소의 정상 연평균 강수량, P_y y 관측소의 강우량, N_x 는 결측점의 정상 연평균 강수량이다.

2) BLOCK 2: ① 입력자료 중 산사태 위치 정보를 기반으로 최근접한 강우 관측소를 선별하는 작업을 우선적으로 수행한다. 최근접한 강우 관측소 선별 작업은 대원거리를 이용했다(식 4).

$$d = 2r \arcsin(\sqrt{(\phi_2 - \phi_1)^2 + \cos(\phi_1)\cos(\phi_2)(\lambda_2 - \lambda_1)^2}) \dots (4)$$

d 는 두 점 사이의 대원거리이며 r 은 지구의 반지름인 6,371km이다. ϕ_1 , λ_1 는 산사태 발생 위치의 위도와 경도

이며, ϕ_2 , λ_2 는 관측소의 위도와 경도이다.

② 산사태 발생에 RA의 영향을 고려하기 위하여 IETD에 따라 강우사상을 구분한다. IETD가 24시간일 경우, 0mm 강우가 연속적으로 24시간 동안 일어난 구간을 선별하고, 이에 해당되지 않는 구간을 하나의 강우사상으로 구분한다. [11]에서는 IETD를 6, 12, 24, 48, 72시간으로 조절하여 유발 강우 기준을 분석한 결과 48시간이 적절한 것으로 보고된 바 있다. 따라서 본 연구에서도 48시간으로 IETD를 설정하여 강우 특성을 계산하고, RA 기간은 2주로 한다.

③ 증발산, 침투 등 수문학적인 요인에 따른 감소효과를 적용하고자 국립방재연구소에서 개발한 감소계수를 선행 강우량에 적용하여 RWA를 계산한다. 감소계수는 반감기(1 ~ 3일)에 따라 감소계수가 상이하다. 일반적으로 산사태는 호우가 집중되는 7, 8월에 다수 발생되므로 3일의 반감기를 적용하는 것이 타당하나, 증발산 등의 조건은 기상 상황에 따라 상이하므로 본 연구에서는 2일로 설정하여 RWA를 계산한다. 이어서 산사태 위치별로 선정된 관측소의 강우특성(E, RA, RWA, D, I, RM)을 계산하여 CSV 파일로 출력한다. 이때 I는 E/D로 계산한다.

3) BLOCK 3: 산사태 임계 강우 기준을 분석하기 위해 기술통계를 시작으로 Q-Q(Quantile) 플롯, 상관성 및 정규성 분석을 실시한다. 이후, 정규성 만족 여부에 따라 선형 및 비선형 회귀분석으로 구분하여 통계분석을 실시한다.

여기서 IETD, RA 기간, 반감기, 통계분석은 사용자가 직접 설정 및 선택할 수 있다.

2.2 Design the GUI

사용자 친화적 환경을 조성하기 위해 스크립트 형식에서 벗어나, GUI 형태의 프로그램을 구현하기 위해 클래스 다이어그램을 설계한다(Fig. 3.).

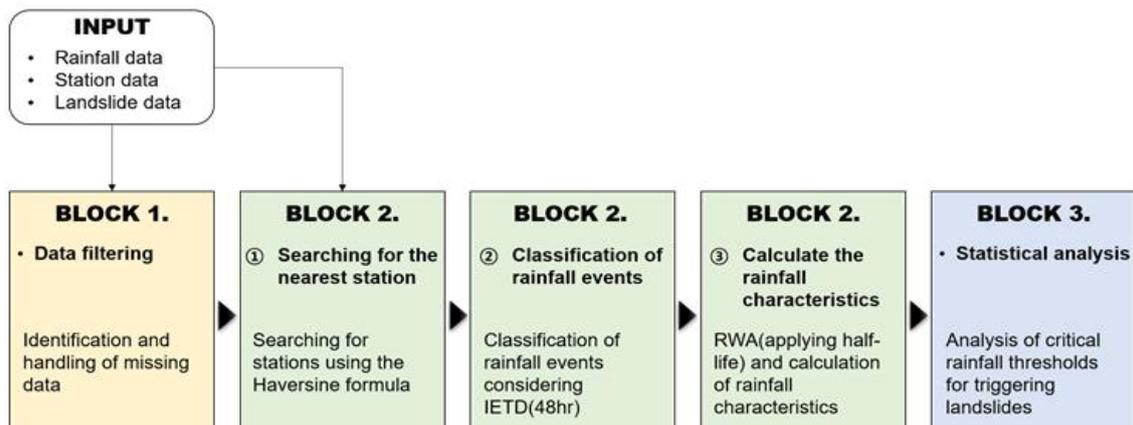


Fig. 2. Logical framework of the algorithm

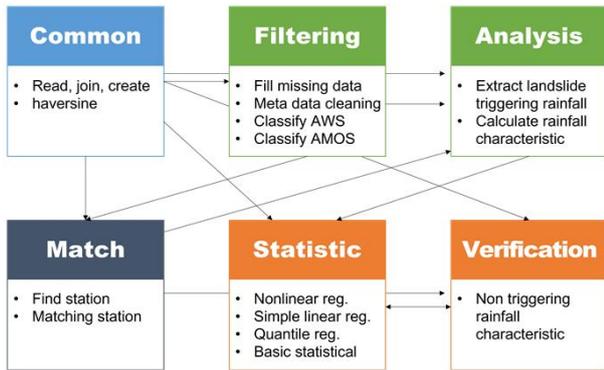


Fig. 3. Class diagram for design the GUI

Common 모듈에는 자료를 읽고, 폴더를 생성하고, 대원 거리를 계산하는 등 모든 모듈에 공통적으로 사용되는 함수로 구성했다. Filtering 모듈은 결측치를 보완하고, 관측소를 분류하는 역할을 수행한다. Analysis는 산사태가 발생한 시점을 특정하고 강우특성을 분석하는 함수로 구성했다.

Match 모듈은 산사태 위치별 최근접 관측소를 탐색하고 선정하는 함수로 구성했다. 마지막으로 Statistic과 Verification은 산사태 유발 임계강우 기준을 마련하기 위해 통계분석과 유의성 여부를 검증하는 함수로 구성했다.

IV. Performance Study

1. Results

1.1 Searching for the nearest station

식 (4)에 의해 선별된 산사태 발생 연도에 따른 관측소의 총개수 및 거리에 대한 정보는 Table 3과 같다.

Table 3. Number and distance of the first closest station by landslide location

Type		2020	2021	2022	2023	
AWS	N (ea)	238	42	89	138	
	D (km)	Min.	0.2	1.2	0.2	0.1
		Max.	23.7	13.4	17.1	19.6
		Avg.	7.9	6.4	7.7	8.5
AMOS	N (ea)	246	34	77	152	
	D (km)	Min.	0.03	0.8	0.09	0.04
		Max.	72.5	27.6	41.4	44.0
		Avg.	7.6	9.0	7.7	7.7
Combine	N (ea)	405	182	724	226	
	D (km)	Min.	0.03	0.8	0.09	0.04
		Max.	17.7	13.4	15.0	15.1
		Avg.	4.8	5.1	5.3	4.7

산사태 발생 위치별 AWS와 AMOS의 거리를 분석한 결과, AWS는 최소: 0.1km, 최대: 23.7km, 평균: 7.6km이

며, AMOS는 최소: 0.03km, 최대: 72.5km, 평균: 8.0km이다. 강우는 지형적인 특성에 따라 지역별로 상이한 분포 특성을 지닌다. 즉, 산사태를 유발한 강우 정보의 올바른 취득을 위해서는 최대한 근접한 강우 관측소를 선별하는 것이 올바른 임계 강우 기준 마련에 중요하다. 이에 따라 국지적 강우특성을 최대한 반영하기 위해 AWS와 AMOS를 병합하여 최근접 관측소를 탐색했다. 병합하여 탐색한 결과, 최소: 0.03km, 최대: 17.7km, 평균: 4.9km이다.

1.2 Data filtering

병합된 파일에서 선별된 관측소들의 결측률 분석 결과는 Table 4와 같다. 결측이 발생한 관측소는 식 (3)에 의하여 보간했다.

Table 4. Results of rainfall data missing rate by combined station data

Type	2020	2021	2022	2023
Total	1.4	2.1	2.7	0.4
May.	0.3	0	0	0
Jun.	0.3	0.2	0	0
Jul.	0	0.1	0.5	0
Aug.	0.1	0.3	0.2	0.1
Sep.	0	0	2.0	0.1
Oct.	0.7	1.5	0	0.2

2022년의 결측률이 2.7%로 분석연도 중 가장 높았으며, 5 ~ 10월 중 10월의 결측률이 2.4%로 나타났으나, 모든 구간에서 결측률이 5% 미만으로 미미하였다.

1.3 Calculate the rainfall characteristics

연도별로 분석된 강우특성 결과는 Table 5와 같다. 이 때 강우특성은 산사태 주의보에 대한 임계기준을 재산정하기 위하여 RWA, E, D, I, RM에 대한 기술통계에 대해 기술한다.

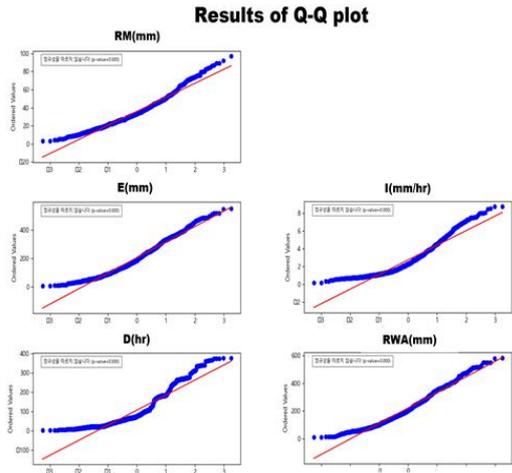
2020 ~ 2023년 중 산사태가 가장 많이 발생한 2020년에서 모든 강우특성의 값이 높게 나타났다.

1.4 Statistical analysis of landslide triggering rainfall thresholds

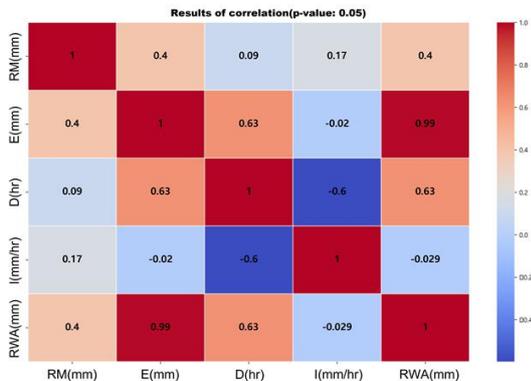
매개변수별 정규성 분포 검증, 상관관계 분석 결과는 Fig. 4와 같다. 매개변수 모두 정규성을 따르지 않아, 비선형 통계분석 방법이 적합한 것으로 분석되었다. 이어서 E-D, RWA-D의 상관관계가 0.63으로 높았으며, I-D는 -0.6으로 상관관계가 낮았다.

Table 5. Results of rainfall characteristics each by year

Type	RWA (mm)			E (mm)			D (hr)			I (mm/hr)			RM (mm)		
	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.
2020	15.9	931.9	326.8	8.8	915	304.7	2	558	137.5	0.4	36.9	4.4	5.8	434.8	62.8
2021	30.4	680.4	282.3	9.4	680.4	277.8	3	388	105.6	0.6	8	3.2	4.7	268.6	59.4
2022	15.3	940.4	304.8	5.1	934.6	288	2	268	79.7	0.1	28.8	4.3	3	203.5	62.9
2023	10	581.3	256.3	10	549	228.3	2	277	115.7	0.6	32.7	3.2	5.5	147.3	35.9



(a)



(b)

Fig. 4. Results of basic statistical. (a): Normality check results for each parameter, (b): Correlation analysis results for each parameter.

1.5 Results of non-linear and power law regression analysis

산사태 유발 강우 기준 마련을 위해 수행된 선행연구에서는 멱법칙(power law)을 이용한 통계분석을 주로 수행하고 있다. 멱법칙은 대부분의 자연현상을 설명하는 함수식으로 비선형적인 관계를 설명하는 데 유용하다. 그러나 매개변수별 관계는 기초통계에 기반하여 적절한 통계 방법을 선정하는 것이 타당할 수 있다. 이에 따라 본 연구에서는 멱법칙과 비선형 회귀분석을 수행한 뒤, 결정계수가 높은 함수식을 선정하여 임계 강우 기준을 산출했다.

멱법칙 함수식은 식 (5)와 같으며, 비선형 회귀분석의 곡선식은 식 (6)과 같다.

$$y = a \times x^{-\beta} \dots\dots\dots (5)$$

$$y = a \times x^{\beta} + c \dots\dots\dots (6)$$

a, β, c 는 보정 계수, x 는 독립변수이다.

멱법칙과 비선형 회귀분석의 결과는 Fig. 5. 및 Table 6.과 같다. Fig. 5.에서의 파란색 점선은 회귀식에서의 10%, 30%에 해당되는 지점을 알려주며, 빨간색 곡선은 식 (5), (6)에 의해 산정된 회귀식이다. 이에 대한 결과는 Table 6과 같다.

Table 6. Results of non-linear and power-law regression analysis

Type	Equation	R ²
E-D	non-linear $y = 0.753x^{0.353} - 0.080$	0.44
	power-law $E = 18.359 \times D^{-0.514}$	0.45
I-D	non-linear $y = -0.905x^{0.248} + 0.919$	0.46
	power-law $I = 18.363 \times D^{0.486}$	0.43
RWA-D	non-linear $y = 0.737x^{0.347} - 0.069$	0.44
	power-law $RWA = 24.331 \times D^{-0.473}$	0.46
RM-D	non-linear $y = 0.356x^{0.088} - 0.040$	0.03
	power-law $RM = 18.276 \times D^{-0.131}$	0.06

멱함수의 결정계수가 평균 0.35이며, 비선형 회귀분석은 0.34로 분석되어 큰 차이가 없었다. 두가지 통계분석 방법 모두 RM-D에서 낮은 결정계수를 나타내었으나, 나머지 관계에서는 평균 0.45로 나타났다.

임계 강우 기준은 10, 30%에 위치한 값을 멱함수 또는 비선형 회귀식에 대입하여 산출했다. E: 68.5~132.5mm, I: 0.9~1.4mm/hr, RWA: 81.6~151.1mm, RM: 17.5~26.5mm로 분석되었다. 기존에 설정된 유발 강우 기준에 비해 RM: 2.5~3.5mm, E: 31.5~67.5mm 낮았으나, RWA는 1.1~1.6mm 높게 나타났다.

2. Verification

임계 강우 기준의 적절성 여부를 판단하고자

AUC(Area Under the Curve)-ROC(Receiver Operating Characteristic) 분석을 실시했다.

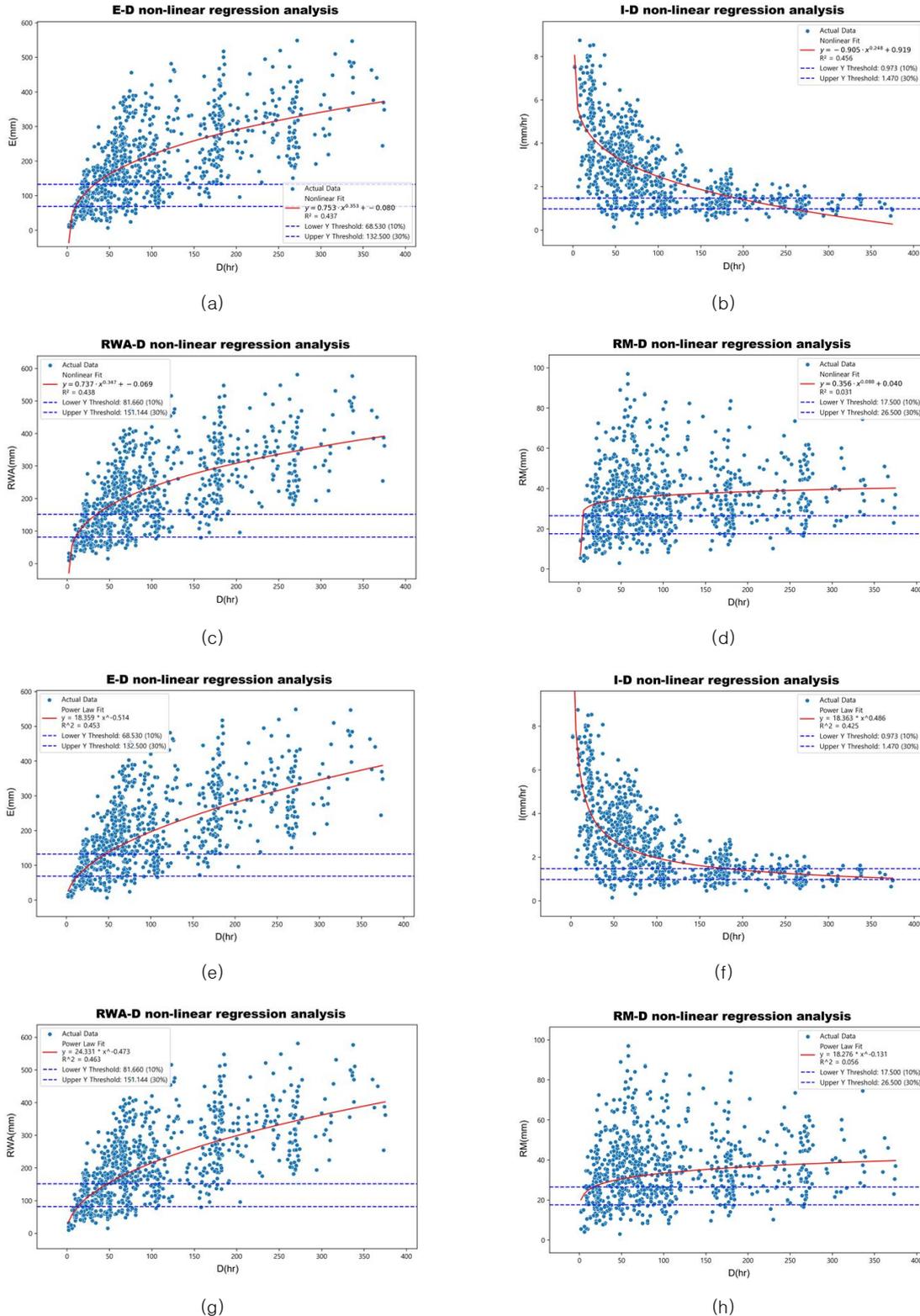


Fig. 5. Results of non-linear regression and power law regression analysis. (a): non-linear E-D curve, (b): non-linear I-D curve, (c): non-linear RWA-D, (d): non-linear RM-D, (e): power law E-D curve, (f): power law I-D curve, (g): power law RWA-D curve, (h): power law RM-D curve

본 연구에서는 관측소에서 산사태 발생 위치별 두 번째로 가까운 관측소를 선정했으며, 결과는 Table 7.과 같다.

Table 7. Number and distance of the second closest station by landslide location

Type		2020	2021	2022	2023	
Combine	N (ea)	437	52	136	260	
	D (km)	Min.	0.93	2.7	0.9	1.3
		Max.	19.8	17.5	19.0	18.3
		Avg.	7.9	8.8	8.7	8.0

선별된 관측소의 산사태 비유발 강우특성은 앞선 방법과 동일하게 추출하였다. 검증 결과, 모든 임계강우 기준에서 AUC=0.5으로 임계 강우 기준이 적절히 선정되지 않아 분류 성능이 떨어지는 것으로 분석되었다(Fig. 6.).

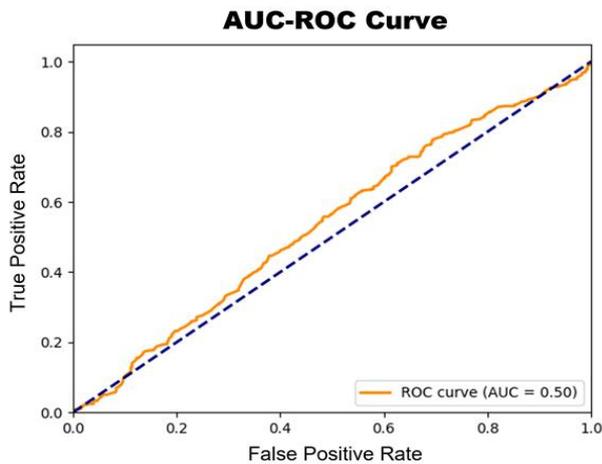


Fig. 6. Verification result of rainfall threshold value using AUC-ROC curve

3. Performance of the GUI

알고리즘의 사용 편리성을 확보하기 위하여 Fig. 3.에 따라 구현된 GUI는 Fig. 7.과 같다.

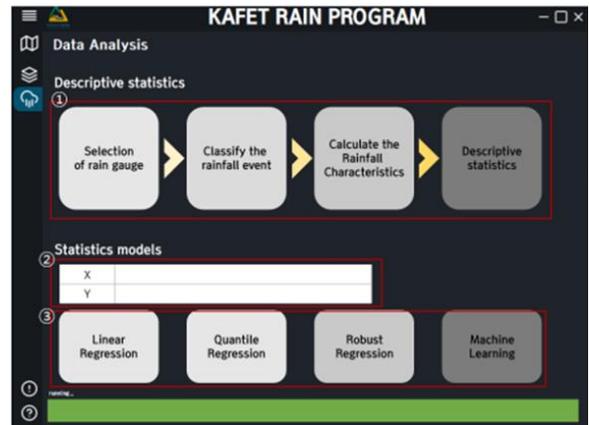
Fig. 7(a)는 실행화면 및 자료 정제 기능을 수행하는 화면이며, ① 자료를 준비하는 과정에서 필요한 기본경로, ② 사용자 정의 변수 입력창, ③ 자료 정제를 위한 단계를 수행할 수 있다. ④는 메뉴바이고, ⑤는 진행 표시줄이다. ③에서는 선택한 네모 상자에 삽입된 기능을 호출하여 수행되도록 하였다.

Fig. 7(b)는 기술통계 기능과 강우 임계치 분석을 위한 통계모델 선택 및 수행 기능을 수행하는 화면이다. ① 기술통계 기능에서는 최근접 관측소를 선택 후, IETD에 의한 강우사상을 구분한다. 다음으로 강우특성을 계산하고 기술통계를 수행한다.

기술통계가 마무리되면 matplotlib과 *.xlsx 파일로 제공되는 정보를 이용하여 ② 분석하고자 하는 매개변수를 입력하고, ③ 적절한 통계분석 모델을 선택할 수 있도록 구성하였다.



(a)



(b)

Fig. 7. GUI developed based on class diagram. (a): This page is for preparing and cleaning data. (b): This page is for performing descriptive statistics and analyzing landslide triggering thresholds using various statistical models.

본 연구에서는 IETD를 48시간, 반감기는 2일로 설정하여 분석하였을 때의 알고리즘 성능은 Table 8과 같다.

Table 8. Processing time of the Algorithm

Type	Processing Time(s)
BLOCK 1	300
BLOCK 2(①)	240
BLOCK 2(②)	900
BLOCK 2(③)	300
BLOCK 3	60
Total	1,800

79,201KB의 크기를 지닌 2020~2023년 동안의 AWS와 AMOS의 강수량 자료를 분석하여 임계 강우 기준을 분석하는데 소요된 총 시간은 30분으로 집계되었다. 가장 많은 시간이 소요되는 구간은 IETD 설정에 따른 강우사상을 추출하는 BLOCK 2(②)로 약 15분이 소요되었다.

본 알고리즘에서는 IETD를 48시간, 반감기를 2일로 설정하여 분석하였으나, IETD와 반감기의 조건을 변경할 경우, 소요시간은 증감할 수 있다.

4. Discussion

본 연구에서 개발한 알고리즘을 이용하여 임계 강우 기준을 분석한 결과, 기 설정되었던 임계 강우 기준에 비해 RM과 E는 낮게 나타났으나, RWA는 높게 산정됐다. 이는 기 설정된 기준치의 RA 기간이 공개되어 있지 않아 본 연구에서 설정한 2주의 기간보다 짧았을 경우 연구 결과와 같이 RWA가 높게 산정되었을 수 있다. 또는 RA 기간이 일치한다고 가정하였을 때, 기후변화로 인해 선행강우량이 증가하였다는 것을 확인할 수 있다. 즉, 본 연구에서 산출된 임계 강우 기준에 대한 해석은 많은 양의 RWA가 발생하고 있으며, 낮은 RM과 E 발생에도 산지토사재해가 발생하므로 재해 취약성이 높아진 것으로 해석할 수 있다.

이어서 도출한 임계 강우 기준의 적절성을 검증하고자 AUC-ROC 분석을 실시한 결과, 모든 임계 강우 기준에서 AUC=0.5로 나타났다.

선행연구에서는 동일한 관측소를 대상으로 산사태 발생 시점 기준 6, 12, 24시간 전의 강우량 자료를 산사태 비유발 강우로 확정하였다.

이런 경우, TP(True Positive)의 확률이 올라가므로 높은 AUC 값을 도출할 수 있으나, 객관적인 임계 강우 기준 평가가 어렵다. 이에 따라 본 연구에서는 유사한 강우특성 환경을 수집하기 위해 산사태 위치별로 2번째로 가까운 관측소를 선정하여 산사태 비유발 강우특성을 수집하여 AUC 값이 낮게 분석된 것으로 판단한다.

즉, 강우자료만을 이용하여 산지토사재해를 예측하는 것은 한계가 존재한다. 그럼에도 불구하고 산촌지역에서 다수 발생하는 재해는 인명과 재산에 큰 위협이 되므로 보수적인 방법에 의하여 선정되어야 한다. 다만, 선정되는 일련의 과정이 정량적이며 체계적이어야 한다. 본 알고리즘은 대용량 강우 자료를 손쉽게 처리하고 통계분석을 수행할 수 있다. 다만, 분석된 임계 강우 기준치에 대한 성능은 낮았다.

이를 타개하기 위해서는 관측소 선택 과정에서 개선이 필요할 것으로 예상된다.

[22]는 산사태 위치로부터 5km 반경을 기준으로 최근접한 관측소를 탐색 후, 유역, 고도, 사면방향 등을 고려하여 최근접 관측소를 선택하는 방식을 채택한 바 있다. 본 연구에서도 거리에 의한 강우의 공간적 분포 특성을 최대한 반영하고자 AWS와 AMOS를 병합하여 분석한 결과 평균 거리가 4.9km이었으나, 선행연구들에 비해 결정계수가 낮게 나타났다.

이러한 이유로는 산사태 피해 위치와 최근접한 관측소가 1:1 매칭되는 것이 아닌 N:1로 매칭되어 산사태 위치별로 동일한 관측소가 선택되어 다양한 강우특성을 반영하지 못하여 발생한 결과로 예상된다.

즉, 산사태 위치로부터 일정한 거리를 가진 관측소에서 의 강수량 자료를 예·경보 임계 강우 기준 마련에 활용하는 것은 한계점이 있을 수 있다.

2020년 기준 산림면적이 전국의 63%(6,298천ha)를 차지하는 우리나라에는 미기후의 영향이 극대화 될 수 있다.

이에 따라 강우의 공간적 해석을 위해 선행연구들에서는 역거리가중법(Inverse Distance Weighting, IDW), PRISM(Parameter-elevation Regression on Independent Slopes Model), 크리깅과 같은 보간법을 이용하여 강우를 예측하고있다[25, 26].

공간해상도를 높인 강우 자료를 획득하여 산사태 위치별 지형 특성을 반영한 강우특성이 확보된다면 본 연구와는 차별화된 결과를 도출할 수 있을 것으로 판단한다.

V. Conclusions

본 연구는 산지토사재해를 예방하는 비구조물적 대책 중 하나인 산사태 예·경보 시스템의 임계 강우 기준을 마련하기 위해 2020~2023년 동안의 AWS 및 AMOS의 강수량 자료를 자동으로 전처리하고 통계 분석을 수행하는 알고리즘과 GUI를 개발했다.

개발된 알고리즘의 기능을 종합하면 기상자료개방포털에서 다운로드 받은 강우 자료를 정제하고, 사용자가 설정한 매개변수에 따라 강우 사상이 자동으로 분류된다. 이어서 산사태 위치로부터 최근접한 관측소를 선택한 후, 강우특성을 자동으로 계산한 뒤, 기초통계를 수행하여 임계기준 분석에 적합한 통계모델을 직접 설정한 뒤, 분석할 수 있다.

본 연구에서 개발한 알고리즘은 정량적이고 체계적인 방법으로 임계 강우 기준 마련에 도움을 줄 수 있다. 또한 기후변화로 인해 강우패턴이 변화하는 현 시점에서 국내 연구 동향은 AWS 강수량자료를 기반으로 I-D 관계식만을

이용하거나 수동으로 분석을 수행했다. 본 연구에서는 AMOS의 강수량자료와 AWS를 동시에 병합하여 강우특성을 분석할 수 있고, 여러 가지 매개변수간의 관계성 등을 이용한 통계분석을 수행할 수 있다. 따라서 본 연구의 결과물이 산사태를 유발하는 강우특성에 대한 다양한 관계식 및 정의 등을 도출할 수 있는 도구로 활용될 수 있을 것으로 생각한다.

ACKNOWLEDGEMENT

This work was supported by “Advanced Landslide Risk Assessment Using Machine Learning (FE0400-2023-01-2024)” grant funded by National Institute of Forest Science the Republic of Korea.

This study was carried out with the support of “R&D Program for Forest Science Technology (Project No. RS-2024-00404816)” provided by Korea Forest Service (Korea Forestry Promotion Institute).

REFERENCES

- [1] R. M. Iverson, “Landslide Triggering by Rain Infiltration.” *Water Resources Research*, Vol. 36, No. 7, pp.1897-1910, Jul. 2000. DOI: 10.1029/2000WR900090
- [2] R. M. Iverson, and J. J. Major, “Rainfall, Ground-water Flow, and Seasonal Movement at Minor Creek Landslide, Northwestern California: Physical Interpretation of Empirical Relations,” *Geological Society of America Bulletin*, Vol. 99, No. 4, pp.579-594, Oct. 1987. DOI: 10.1130/0016-7606(1987)99<579:RGFASM>2.0.CO;2
- [3] J. A. Coe, W. L. Ellis, J. W. Godt, W. Z. Savage, J. E. Savage, J. A. Michael, J. D. Kibler, P. S. Powers, D. J. Lidke, and S. Debray, “Seasonal Movement of the Slumgullion Landslide Determined from Global Positioning System Surveys and Field Instrumentation, July 1998-March 2002,” *Engineering Geology*, Vol. 68, No. 1-2, pp. 67-101, Feb. 2003. DOI: 10.1016/S0013-7952(02)00199-0
- [4] Trading Economics, “South Korea Average Precipitation,” <https://tradingeconomics.com/south-korea/precipitation>
- [5] Korea Meteorological Administration, “Seoul Annual Weather Data,” <http://www.weather.go.kr/weather/climate/annual.jsp?>
- [6] V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, and S. Berger, “Climate Change 2021: The Physical Science Basis,” IPCC, pp. 1-219, 2021.
- [7] M. K. Park, “Landslide Vulnerable Areas in Korea, Increasing Every Year!,” <https://www.hidonmin.com/news/articleView.html?idxno=519220>
- [8] D. H. Ham, and S. H. Hwang, “Review of Landslide Forecast Standard Suitability by Analysing Landslide-inducing Rainfall,” *Journal of the Korean Society of Hazard Mitigation*, Vol. 14, No. 3, Jun. 2014, pp. 299-310, DOI: 10.9798/KOSHAM.2014.14.3.299
- [9] W. Y. Lee, and H. H. Sung, “Analysis of Rainfall Thresholds for Landslide Occurrence in Chuncheon Gangwon Province, Korea,” *Journal of the Korean Geographical Society*, Vol. 53, No. 5, Oct. 2018, pp. 669-689.
- [10] J. Oh, and H. J. Park, “Analysis of Landslide Triggering Rainfall Threshold for Prediction of Landslide Occurrence,” *Journal of the Korean Society of Hazard Mitigation*, Vol. 14, No. 2, Apr. 2014, pp. 115-129, DOI: 10.9798/KOSHAM.2014.14.2.115
- [11] M. H. Hong, J. H. Kim, G. J. Jung, and S. S. Jeong, “Rainfall Threshold (ID curve) for Landslide Initiation and Prediction Considering Antecedent Rainfall,” *Journal of the Korean Geotechnical Society*, Vol. 32, No. 4, Apr. 2016, pp. 15-27, DOI: 10.7843/kgs.2016.32.4.15
- [12] J. S. Lee, and Y. T. Kim, “Development of Optimum Rainfall Threshold to Predict of Rainfall-Induced Landslides Occurrence,” *Journal of the Korean Society of Hazard Mitigation*, Vol. 17, No. 6, Dec. 2017, pp. 333-340, DOI: 10.9798/KOSHAM.2017.17.6.333
- [13] J. Lee, and G. Chung, “Estimation of InterEvent Time Definition Using in Urban Areas,” *Journal of the Korean Society of Hazard Mitigation*, No. 17., No. 4, Aug. 2017, pp. 287-294, DOI: 10.9798/KOSHAM.2017.17.4.287
- [14] J. Y. Park, S. R. Lee, D. H. Lee, Y. T. Kim, S. Oh, and H. J. Park, “Development of Continuous Rainfall-Based Citywide Landslide Early Warning Model,” *Journal of the Korean Society of Hazard Mitigation*, Vol. 18, Np. 1, Jan. 2018, pp. 99-111, DOI: 10.9798/KOSHAM.2018.18.1.99
- [15] J. S. Lee, H. S. Kang, J. W. Suk, and Y. T. Kim, “Development of Hazard Level-based Rainfall Threshold for Prediction of Rainfall-induced Landslide Occurrence in Korea,” *Journal of the Korean Society of Hazard Mitigation*, Vol. 19, No. 5, Oct. 2019, pp. 225-236. DOI: 10.9798/KOSHAM.2019.19.5.225
- [16] B. G. Chae, W. Y. Kim, Y. C. Cho, K. S. Kim, C. O. Lee, and Y. S. Choi, “Development of a Logistic Regression Model for Probabilistic Prediction of Debris Flow,” *The Journal of Engineering Geology*, Vol. 14, No. 2, Jun. 2004, pp. 211-222.
- [17] D. G. Park, T. H. Kim, J. R. Oh, and J. H. Park, “A Study on the Monitoring & Detection of Slope Failure (II) - Focusing on the Application of Rainfall Data -,” *NDMRI*, pp. 163-207,

- 2005.
- [18] S. W. Kim, G. W. Chun, J. H. Kim, M. S. Kim, and M. S. Kim, "Characteristics of Heavy Rainfall for Landslide-triggering in 2011," *Journal of Korean Society fo Forest Science*, Vol. 101, No. 1, Jan. 2012, pp. 28-35.
- [19] A. M. Maturidi, N. Kasim, K. A. Taib, and W. N. Azahar, "Rainfall-Induced Landslide Thresholds Development by Considering Different Rainfall Parameters: A Review," *Journal of Ecological Engineering*, Vol. 22, No. 10, pp. 85-97, DOI: 10.12911/22998993/142183
- [20] M. Melillo, M. T. Brunetti, S. Peruccacci, and S. L. Gariano, "An Algorithm for the Objective Reconstruction of Rainfall Events Responsible for Landslides," *Landslides*, Vol. 12, Apr. 2014, pp. 311-320, DOI: 10.1007/s10346-014-0471-3
- [21] M. Melillo, M. T. Brunetti, S. Peruccacci, S. L. Gariano, A. Rocaati, and F. Guzzetti, "A Tool for the Automatic Calculation of Rainfall Thresholds for Landslide Occurrence," *Environmental Modelling & Software*, Vol. 105, Jul. 2018, pp. 230-243. DOI: 10.1016/j.envsoft.2018.03.024
- [22] O. F. Al-Thwaynee, M. Melillo, S. L. Gariano, H. J. Park, S. W. Kim, L. Lombardo, P. Hader, M. Mohajane, R. P. Quevedo, F. Catani, and A. Aydda, "DEWS: A QGIS Tool Pack for the Automatic Selection of Reference Rainfall Gauges for Landslide-Triggering Rainfall Thresholds," *Environmental Modelling & Software*, Vol 162. Apr. 2023, 105657, DOI: 10.1016/j.envsoft.2023.105657
- [23] M. T. Abraham, N. Satyam, A. Rosi, B. Pradhan, and S. Segoni, "The Selection of Rain Gauges and Rainfall Parameters in Estimating Intensity-Duration Thresholds for Landslide Occurrence: Case Study from Wayanad (India)," *Water*, Vol. 12, No. 4, Apr. 2020, pp. 1000, DOI: 10.3390/w12041000
- [24] S. L. Gariano, M. T. Brunetti, G. Iovine, M. Melillo, S. Peruccacci, O. Terranova, C. Vennari, and F. Guzzetti, "Calibration and validation of rinfall thresholds for shallow landslide forecasting in Sicily, southern Italy," *Geomorphology*, Vol. 228, Jan. 2015, pp. 653-665, DOI: 10.1016/j.geomorph.2014.10.019
- [25] J. C. Park, and M. K. Kim, "Comparison of Precipitation Distributions in Precipitation Data Sets representing 1km Spatial Resolution over South Korea Produced by PRISM, IDW, and Cokriging," *Journal of the Korean Association of Geographic Information Studies*, Vol. 16, Sep. 2013, pp. 147-163, DOI: 10.11108/kagis.2013.16.3.147
- [26] M. J. Um, and C. S. Jeong, "Spatial Analysis of Precipitation with PRISM in Gangwondo", *Journal of Korea Water Resources Association*, Vol. 44, Mar. 2011, pp. 179-188.
- [27] Korea Forest Service, Forestry Statistics Platform, <https://kfss.forest.go.kr/>
- [28] National Disaster Management Institute, "A Study on the

Monitoring & Detection of Slope Failure (II) - Focusing on the Application of Rainfall Data -", pp.165-167, 2005

Authors



Donghyeon Kim is a Ph.D. in the division of research technical research institute. His current research interests include forestry, landslide, rainfall, and algorithm.



Song Eu is a research official at the Landslide Division, Department of Forest Disaster and Environment, National Institute of Forest Science. His current research interests include landslide, debris-flow,

early-warning, and numerical simulation.



Kwangyoun Lee is a Ph.D. in the division of research technical research institute. His interesting subject and area of research is GIS/RS etc.



Sukhee Yoon holds a Ph.D. in Disaster Engineering and is currently works as a Manager at Korea Association of Forest Enviro-conservation Technology. His expertise includes the development and operation of

information systems, maintenance of landslide monitoring system and research in disaster meteorology.



Jongseo Lee is a Master in the division of research technical research institute. His current research interests include forestry, GIS simulation, landslide, and algorithm



Donggeun Kim is an full professor at the department of ecology and protection at Kyungpook National University. His current research interest include forestry, forest measurement, forest management, and forest

engineering.