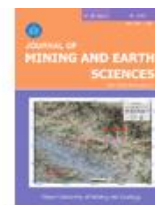




## Journal of Mining and Earth Sciences

Website: <http://jmes.humg.edu.vn>



# Methods of building database to establish flooding map for coastal areas using a combination of artificial intelligence and GIS technology



Trong Gia Nguyen <sup>1, 2\*</sup>, Nghia Viet Nguyen <sup>1</sup>, Quang Ngoc Pham <sup>1,2</sup>, Cuong Van Nguyen <sup>3</sup>, Quan Anh Duong <sup>1</sup>, Hai Dinh Nguyen <sup>4</sup>, Nhi Hoang Nguyen <sup>5</sup>

<sup>1</sup> Hanoi University of Mining and Geology, Hanoi, Vietnam

<sup>2</sup> Geodesy and Environment research group, Hanoi University of Mining and Geology, Hanoi, Vietnam,

<sup>3</sup> The Vietnam Agency of Seas and Islands, Hanoi, Vietnam

<sup>4</sup> Nautical chart surveying and marine research team, Vietnamese People Navy, Haiphong, Vietnam

<sup>5</sup> An Giang Construction and Traffic Consulting Joint Stock Company, Angiang, Vietnam

### ARTICLE INFO

#### Article history:

Received 20<sup>th</sup> Mar. 2023

Revised 23<sup>rd</sup> July 2023

Accepted 17<sup>th</sup> Aug. 2023

#### Keywords:

AI,  
Flood,  
Flood sensitive,  
GIS,  
Machine learning,  
Weka.

### ABSTRACT

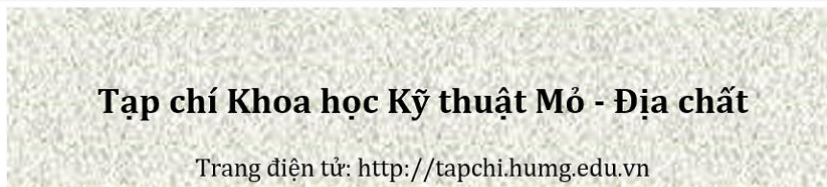
*As a country with a coastline stretching from North to South, in recent years natural disasters, especially floods and inundation, have severely affected people and properties in Vietnam. In order to prevent and control natural disasters and adapt to climate change, there have been many researches to establish the flood-related map in the country. Among the methods of creating flood maps, the application of AI (Artificial Intelligence) combined with GIS (Geography Information System) has outstanding advantages due to its ability to handle a mixture of many types of input data in a geographical space unification. This method is also used widely in the world in general and Vietnam in particular. When applying the aforementioned method, building the input database of machine learning and artificial intelligence models is an essential issue. Based on the Sentinel-1, Landsat 8/9 images, digital elevation model (DEM), and soil maps, the authors have built the input database for modeling by using AI models. This paper introduces the method of building the input database for making flood maps using machine learning, and artificial intelligence combined with GIS. The computation process is divided into two steps: (1) Editing the component data layers from input data and (2) Standardization of data to transfer the component data layers into the same unit with the standard data format of Weka software. The research's results are 11 data layers including the flood map in the past, elevation, slope, slope direction, curvature, terrain energy, geology, land use, soil, NDVI, NDWI for Quang Nam province.*

Copyright © 2023 Hanoi University of Mining and Geology. All rights reserved.

\*Corresponding author

E - mail: [nguyengiatrong@humg.edu.vn](mailto:nguyengiatrong@humg.edu.vn)

DOI: 10.46326/JMES.2023.64(4).02



## Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

Trang diện tử: <http://tapchi.humg.edu.vn>

# Phương pháp xây dựng cơ sở dữ liệu phục vụ thành lập bản đồ ngập lụt cho các vùng ven biển sử dụng kết hợp trí tuệ nhân tạo và công nghệ GIS

Nguyễn Gia Trọng<sup>1,2\*</sup>, Nguyễn Việt Nghĩa<sup>1</sup>, Phạm Ngọc Quang<sup>1,2</sup>, Nguyễn Văn Cương<sup>3</sup>, Dương Anh Quân<sup>1</sup>, Nguyễn Đình Hải<sup>4</sup>, Nguyễn Hoàng Nhi<sup>5</sup>

<sup>1</sup> Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam

<sup>2</sup> Nhóm nghiên cứu Trắc địa cao cấp - môi trường, Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam

<sup>3</sup> Cục Biển và Hải đảo Việt Nam, Hà Nội, Việt Nam

<sup>4</sup> Đoàn Đo đạc, biên vẽ hải đồ và nghiên cứu biển, Hải Phòng, Việt Nam

<sup>5</sup> Công ty Cổ phần tư vấn xây dựng giao thông An Giang, An Giang, Việt Nam

### THÔNG TIN BÀI BÁO

### TÓM TẮT

Quá trình:

Nhận bài 20/3/2023

Sửa xong 23/7/2023

Chấp nhận đăng 17/8/2023

Từ khóa:

GIS,

Học máy,

Ngập lụt,

Nhạy cảm lũ lụt,

Trí tuệ nhân tạo,

Weka.

Là một quốc gia có đường bờ biển trải dài từ Bắc xuống Nam, trong những năm gần đây, thiên tai, đặc biệt là lũ lụt, ngập lụt đã gây ảnh hưởng nặng nề về người và tài sản tại Việt Nam. Để phục vụ cho công tác phòng, chống thiên tai, thích ứng với biến đổi khí hậu, đã có nhiều công trình nghiên cứu thành lập các loại bản đồ liên quan đến ngập lụt ở nước ta. Trong số các phương pháp thành lập bản đồ ngập lụt, phương pháp ứng dụng trí tuệ nhân tạo kết hợp với công nghệ GIS có ưu điểm nổi trội do có khả năng xử lý hỗn hợp nhiều loại dữ liệu đầu vào trong một không gian địa lý thống nhất. Đây là phương pháp đang được sử dụng một cách phổ biến trên thế giới và tại Việt Nam. Khi ứng dụng phương pháp nêu trên, một vấn đề hết sức quan trọng là phải xây dựng được cơ sở dữ liệu đầu vào của các mô hình học máy, trí tuệ nhân tạo. Bài báo giới thiệu phương pháp xây dựng cơ sở dữ liệu là dữ liệu đầu vào khi thành lập bản đồ ngập lụt sử dụng học máy, trí tuệ nhân tạo kết hợp với công nghệ GIS. Trên cơ sở dữ liệu ảnh vệ tinh Sentinel-1, ảnh vệ tinh Landsat 8/9, mô hình số độ cao (DEM), bản đồ thổ nhưỡng, các tác giả đã xây dựng được bộ cơ sở dữ liệu đầu vào để mô hình hóa bằng các mô hình trí tuệ nhân tạo. Quy trình tính toán được chia làm 2 bước: (1) biên tập các lớp dữ liệu thành phần từ dữ liệu đầu vào; (2) chuẩn hóa để đưa các lớp dữ liệu thành phần đã xây dựng về cùng một đơn vị với định dạng dữ liệu theo chuẩn của phần mềm Weka. Kết quả nghiên cứu đã xây dựng được 11 lớp dữ liệu bao gồm bản đồ lũ trong quá khứ, độ cao, độ dốc, hướng dốc, độ cong, năng lượng địa hình, địa chất, sử dụng đất, thổ nhưỡng, NDVI, NDWI cho tỉnh Quảng Nam.

© 2023 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

\*Tác giả liên hệ

E - mail: [nguyengiatrong@humg.edu.vn](mailto:nguyengiatrong@humg.edu.vn)

DOI: 10.46326/JMES.2023.64(4).02

## 1. Mở đầu

Sự tăng lên của nhiệt độ toàn cầu đã làm trầm trọng thêm ảnh hưởng của các tai biến thiên tai như bão, áp thấp nhiệt đới, ngập lụt,... Là một quốc gia trải dài theo đường bờ biển, trong những năm qua Việt Nam đã chịu ảnh hưởng nặng nề do lũ lụt gây ra. Do đó, thành lập các bản đồ cảnh báo, dự báo, nhạy cảm với lũ lụt là việc làm hết sức cần thiết phục vụ ra quyết định nhằm làm giảm thiểu thiệt hại do ảnh hưởng của thiên tai cũng như công tác tìm kiếm cứu hộ cứu nạn.

Theo thống kê, hiện nay có khoảng 40% dân số thế giới sinh sống trong vùng 100 km tính từ bờ biển do đó các nghiên cứu về mô hình ngập lụt đã diễn ra từ lâu với nhiều chủ đề khác nhau. Các mô hình nghiên cứu ngập lụt trên thế giới có thể chia ra làm các nhóm cơ bản như: các mô hình ngập lụt ven biển, các mô hình xâm lấn nước biển dâng, các mô hình mô phỏng không gian cảnh quan sinh thái, các mô hình đánh giá tổn thương tác động, nhóm sử dụng phần mềm mô hình hóa hệ thống (phần mềm SimCLIM), các mô hình thủy văn và thủy lực và nhóm phân tích quyết định đa tiêu chí (McLeod và nnk., 2010). Trong những năm gần đây, học máy và trí tuệ nhân tạo được ứng dụng rộng rãi trong nghiên cứu về tai biến thiên tai nói chung và ngập lụt ven biển nói riêng. Ưu điểm của việc tích hợp học máy, trí tuệ nhân tạo với công nghệ GIS cho phép xử lý hỗn hợp nhiều loại dữ liệu đầu vào khác nhau để thành lập bản đồ ngập lụt do đó giúp nâng cao độ chính xác và độ tin cậy của bản đồ được thành lập.

Thành lập bản đồ ngập lụt trên cơ sở ứng dụng học máy, trí tuệ nhân tạo kết hợp công nghệ GIS sử dụng nhiều loại dữ liệu đầu vào khác nhau cần thiết phải có một phương pháp xây dựng cơ sở dữ liệu đầu vào một cách thống nhất.

Khi nghiên cứu đánh giá mức độ nhạy cảm với lũ lụt tại Bangladesh, Rahman và nnk. (2019) đã sử dụng kết hợp học máy, phân tích quyết định đa tiêu chí và GIS. Dữ liệu thực nghiệm mà các tác giả sử dụng là lượng mưa, mô hình số độ cao (DEM), bản đồ sử dụng đất, bản đồ địa chất, bản đồ thổ nhưỡng, khu vực thoát nước,... Từ các dữ liệu đầu vào nêu trên, sử dụng công cụ GIS đã thành lập được các lớp dữ liệu đầu vào của mô hình trí tuệ nhân tạo bao gồm lượng mưa, độ cao, độ dốc, bản đồ sử dụng đất, bản đồ địa chất, bản đồ đường đất, vùng thoát nước, độ sâu ngập nước

và thời gian ngập nước. Costache và nnk. (2020a) đã xây dựng các lớp cơ sở dữ liệu bao gồm: vị trí lũ trong quá khứ, độ dốc, chỉ số năng lượng địa hình (TPI), sử dụng đất, độ cong, chỉ số độ ẩm địa hình (TWI), hướng dốc (aspect), chỉ số hội tụ, nhóm đất thủy văn, khoảng cách tới sông, độ cao và thạch học.

Khi thành lập bản đồ về tính dễ bị tổn thương do lũ lụt với khu vực thực nghiệm tại Ấn Độ, Akshayasimha và Chandra (2023) đã sử dụng lớp mật độ thoát nước và mật độ dân số cùng với các lớp dữ liệu khác.

Dodangeh và nnk. (2020) đã sử dụng lớp cơ sở dữ liệu chỉ số thực vật khác biệt (NDVI), thạch học cùng với các lớp cơ sở dữ liệu khác để lấy mẫu lại phục vụ dự đoán tính nhạy cảm của lũ lụt. Trong các công bố của Costache và nnk. (2020b), Diaconu và nnk. (2021), các lớp cơ sở dữ liệu về cơ bản tương tự như các công bố đã được liệt kê trước đó.

Tại Việt Nam, có 28 tỉnh, thành phố trực thuộc trung ương giáp biển và có tới hơn 50% dân số sinh sống ven bờ biển. Vì vậy, nghiên cứu về ngập lụt cho các vùng ven biển là một chủ đề được quan tâm nghiên cứu.

Pham và nnk. (2021) đã đánh giá rủi ro lũ lụt bằng mô hình trí tuệ nhân tạo lai tích hợp với phân tích quyết định đa tiêu chí tại tỉnh Quảng Nam. Các lớp dữ liệu đầu vào của mô hình trí tuệ nhân tạo mà các tác giả sử dụng ở đây bao gồm: độ dốc, độ cong, độ cong dự kiến, biên độ cong, tích tụ dòng chảy, độ cao, chỉ số độ ẩm địa hình, chỉ số vận chuyển trầm tích (STI), chỉ số năng lượng dòng chảy (SPI), mật độ sông và khoảng cách tới sông. Trong các nghiên cứu của mình, Luu và nnk. (2017), Luu và Von Meding (2018) đã dựa vào mô hình số độ cao, bản đồ sử dụng đất, mật độ dân số kết hợp điều tra thực địa để đánh giá lũ lụt, nguy cơ lũ lụt cho tỉnh Quảng Nam. Chau và nnk. (2015) đã sử dụng bản đồ lũ lịch sử để đánh giá tác động của lũ lụt đối với kinh tế nông nghiệp cho tỉnh Quảng Nam.

Khi ứng dụng mạng nơ ron nhân tạo đa lớp trong thành lập phân vùng lũ quét cho tỉnh Yên Bái, Nghĩa và Cường (2020) đã xây dựng cơ sở dữ liệu đầu vào tương tự như các nghiên cứu khác trên thế giới và tại Việt Nam.

Mặc dù đã có các công bố về kết quả thành lập bản đồ ngập lụt, bản đồ nhạy cảm với lũ lụt trên cơ sở ứng dụng trí tuệ nhân tạo, máy học kết hợp



công nghệ GIS tại Việt Nam. Các nghiên cứu chưa trình bày rõ về phương pháp xây dựng cơ sở dữ liệu đầu vào phục vụ huấn luyện, kiểm tra các mô hình trí tuệ nhân tạo được sử dụng. Bài báo này giới thiệu về phương pháp xây dựng cơ sở dữ liệu đầu vào phục vụ huấn luyện, kiểm tra mô hình trí tuệ nhân tạo trong thành lập bản đồ ngập lụt ven biển.

## 2. Dữ liệu và phương pháp nghiên cứu

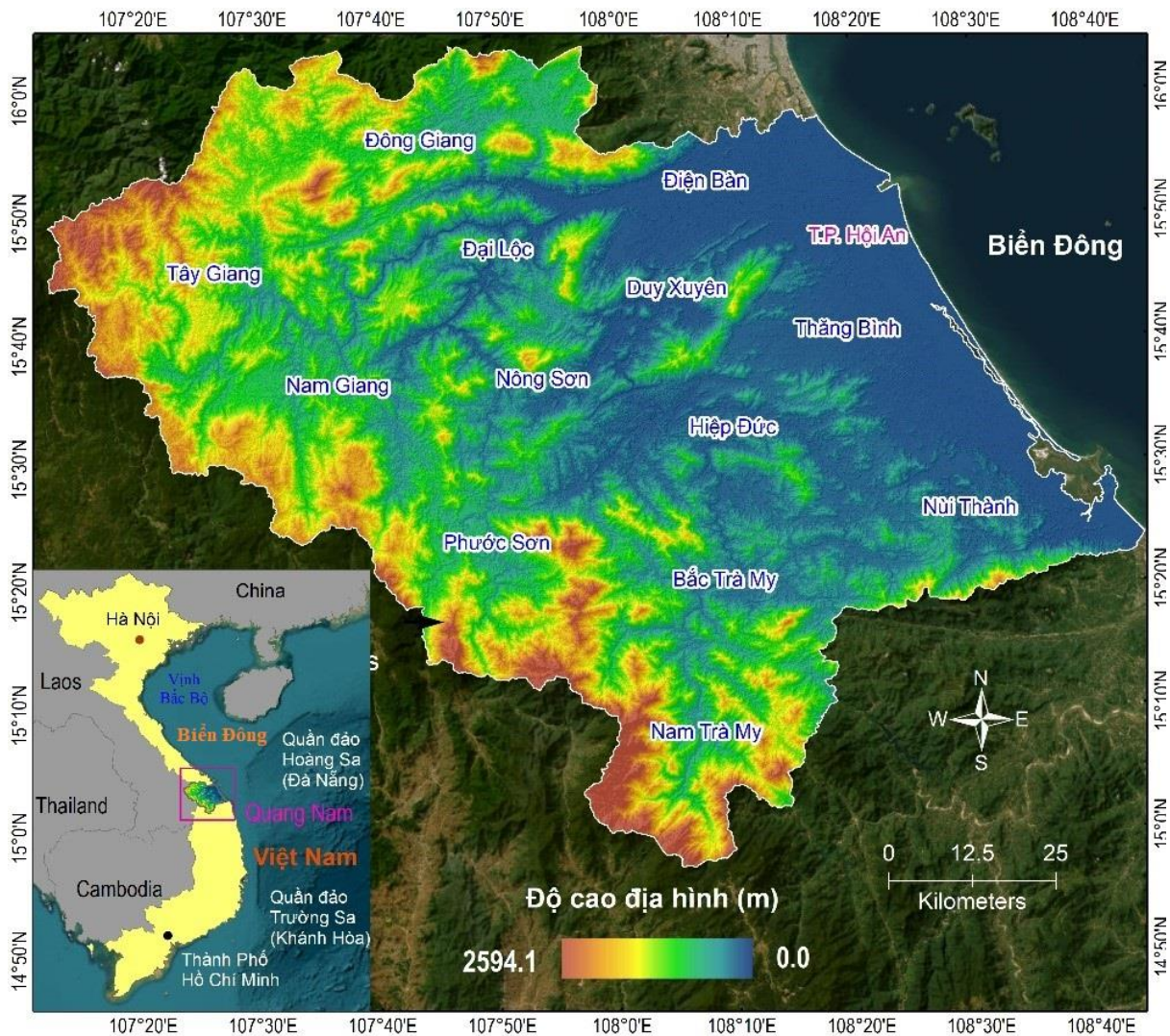
### 2.1. Giới thiệu về khu vực và dữ liệu thực nghiệm

Khu vực thực nghiệm của nghiên cứu này là tỉnh Quảng Nam, nơi có hai lưu vực sông thường xuyên xảy ra hiện tượng lũ lụt hàng năm là sông Vu Gia và sông Thu Bồn (Hình 1).

Các dữ liệu đầu vào sử dụng trong nghiên cứu này bao gồm:

- Mô hình số độ cao: trong nghiên cứu mô hình số độ cao có độ phân giải 30 x 30 m được cung cấp bởi Cơ quan thám hiểm hàng không vũ trụ Nhật Bản (JAXA).

- Lượng mưa: Bên cạnh các yếu tố khác, lượng mưa là yếu tố quyết định có xảy ra lũ lụt hay không. Khi nghiên cứu về lũ lụt, cần xác định được lượng mưa lớn nhất trong thời điểm xảy ra lũ lụt. Dữ liệu về lượng mưa có thể thu nhận thông qua Trung tâm thông tin của Tổng cục Khí tượng thủy văn hoặc thu nhận từ các cơ quan khí tượng trên thế giới. Trong nghiên cứu này, lượng mưa được thu thập từ Cơ quan Hàng không và Vũ trụ quốc gia Hoa Kỳ chi tiết ở Bảng 1.



Hình 1. Vị trí và địa hình của tỉnh Quảng Nam.

- Bản đồ sử dụng đất: bản đồ sử dụng đất có thể thường xuyên cập nhật thông qua xử lý dữ liệu ảnh vệ tinh và có thể được cung cấp bởi một số cơ quan nghiên cứu trên thế giới. Bài báo sử dụng bản đồ sử dụng đất được cung cấp bởi JAXA.

- Bản đồ thổ nhưỡng và bản đồ địa chất được thành lập bởi Cục Địa chất và Khoáng sản Việt Nam.

- Dữ liệu lũ trong quá khứ: Để thành lập dữ liệu lũ trong quá khứ, sử dụng dữ liệu ảnh vệ tinh Sentinel-1 cung cấp bởi Cơ quan Vũ trụ Châu Âu (thông tin chi tiết trong Bảng 2) và sử dụng phần mềm SNAP để xác định các khu vực bị ngập lụt.

- Chỉ số khác biệt thực vật (NDVI): Được xác định bằng ảnh vệ tinh Landsat 8/9 sử dụng phần mềm ArcGIS Pro. Thông tin về các ảnh vệ tinh Sentinel-1 và Landsat 8/9 đã sử dụng được nêu trong Bảng 2.

*Bảng 1. Định dạng dữ liệu lượng mưa cung cấp bởi NASA.*

Độ kinh(°)	Năm	Tháng	Ngày	Lượng mưa (mm)
106.75	2019	3	1	0,00
107.25	2019	3	1	0,00
107.75	2019	3	1	0,00
108.25	2019	3	1	0,02
108.75	2019	3	1	0,04
109.25	2019	3	1	0,08
106.75	2019	3	1	0,00
107.25	2019	3	1	0,00
107.75	2019	3	1	0,01
108.25	2019	3	1	0,02
108.75	2019	3	1	0,06
109.25	2019	3	1	0,11
106.75	2019	3	1	0,00

*Bảng 2. Thông tin về ảnh vệ tinh được sử dụng để tính thực nghiệm.*

TT	Loại ảnh vệ tinh
Ảnh vệ tinh Sentinel-1	
1	S1A_IW_GRDH_1SDV_20181208T224356_20181208T224421_024940_02BF81_E5F8
2	S1A_IW_GRDH_1SDV_20181211T105627_20181211T105652_024977_02CODE_F3DA
3	S1A_IW_GRDH_1SDV_20201001T105641_20201001T105706_034602_04074E_F12E
4	S1A_IW_GRDH_1SDV_20201013T105641_20201013T105706_034777_040D6B_2EDD
Ảnh vệ tinh Landsat 8/9	
1	LC08_L2SP_124050_20230429_20230509_02_T1
2	LC08_L2SP_125049_20230114_20230130_02_T1
3	LC09_L2SP_124049_20230421_20230423_02_T1

## 2.2. Phương pháp nghiên cứu

Để thành lập được bản đồ ngập lụt trên cơ sở ứng dụng trí tuệ nhân tạo kết hợp với công nghệ GIS, cần phải xây dựng cơ sở dữ liệu đầu vào của các mô hình trí tuệ nhân tạo trong một không gian thống nhất, đồng nhất về đơn vị tính.

Phương pháp thành lập cơ sở dữ liệu trong nghiên cứu này bao gồm các bước như sau: (1) từ các dữ liệu đầu vào bao gồm bản đồ địa chất, DEM, ảnh vệ tinh biên tập thành các lớp dữ liệu thành phần và (2) chuẩn hóa để đồng nhất đơn vị của các lớp dữ liệu thành phần với đầu ra là cơ sở dữ liệu theo định dạng của phần mềm Weka để có thể mô hình hóa bằng các mô hình học máy, trí tuệ nhân tạo.

Để có thể thống nhất về mặt không gian của các lớp dữ liệu, trong nghiên cứu này sử dụng công cụ ArcGIS Pro. ArcGIS Pro là phần mềm phân tích dữ liệu địa không gian mạnh, được tích hợp nhiều công cụ học máy và trí tuệ nhân tạo. Trình tự xây dựng cơ sở dữ liệu đầu vào của các mô hình trí tuệ nhân tạo phục vụ thành lập bản đồ ngập lụt như sau:

### 2.2.1. Thành lập các lớp dữ liệu từ ảnh vệ tinh

- Xác định khu vực ngập lụt trong quá khứ: Sử dụng dữ liệu ảnh vệ tinh Sentinel-1, xác định được khu vực ngập trong quá khứ sử dụng phần mềm SNAP theo chu trình như trong Hình 2.

Kết quả xử lý xác định được vùng ngập lụt cho tỉnh Quảng Nam như trong Hình 3.

Sau khi đã xác định được khu vực ngập lụt qua các năm, tiến hành số hóa bằng cách vẽ các polygon cho các vùng bị ngập lụt sử dụng phần mềm ArcGIS Pro (Hình 4).

- Xác định chỉ số thực vật: Để xác định chỉ số thực vật có thể sử dụng phần mềm ArcGIS Pro. Tuy nhiên cần lưu ý rằng trong trường hợp khu vực thực nghiệm lớn thì 1 ảnh vệ tinh có thể không bao trùm được toàn bộ cả khu vực thực nghiệm. Trong trường hợp này, cần phải tìm kiếm các ảnh liên tiếp nhau về mặt không gian bao phủ hết toàn bộ khu vực đã chọn. Để ghép dữ liệu đầu vào trong ArcGIS Pro, sử dụng lệnh mosaic trong công cụ Geoprocessing. Trong trường hợp cần cắt bỏ dữ liệu ở dạng raster sử dụng lệnh Extract by mask.

2.2.2. Thành lập lớp dữ liệu lượng mưa

Để thành lập được lượng mưa, trước tiên cần phải xác định được lượng mưa lớn nhất tại thời điểm có xảy ra lũ lụt tại các điểm quan trắc mưa. Trong trường hợp dữ liệu được lấy từ các cơ quan khí tượng thế giới cung cấp theo vị trí mắt lưới thì cần phải thống kê lượng mưa lớn nhất tại các điểm

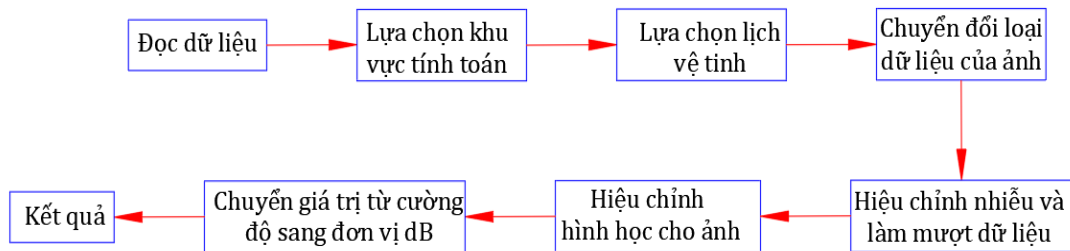
mắt lưới. Sau khi đã có lượng mưa lớn nhất, tiến hành biên tập cơ sở dữ liệu lượng mưa.

2.2.3. Thành lập các lớp dữ liệu liên quan đến mô hình số địa hình

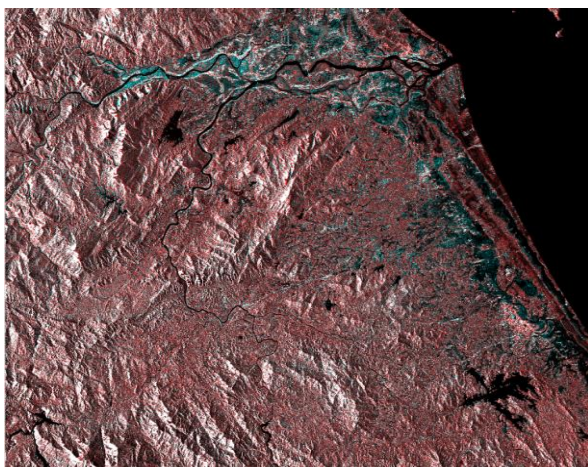
Các lớp dữ liệu liên quan đến mô hình số địa hình như độ cao, độ dốc, độ cong, năng lượng địa hình, chỉ số độ ẩm địa hình có thể áp dụng phương pháp phân tích không gian sử dụng công cụ Geoprocessing trong ArcGIS Pro với các tùy biến khác nhau.

2.2.4. Thành lập các lớp dữ liệu khác

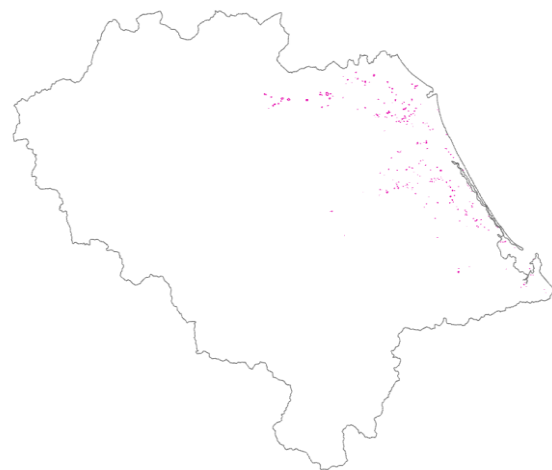
Lớp dữ liệu địa chất, thổ nhưỡng đầu vào thường có số lượng các thành phần đất đá, loại đất rất nhiều trong đó có những thành phần chiếm số lượng và phân bố ít và ít ảnh hưởng đến lũ lụt. Trong trường hợp như vậy, cần tiến hành biên tập lại dữ liệu đầu vào bằng cách gộp các thành phần có phân bố ít vào các thành phần tương tự có phân bố nhiều hơn.



Hình 2. Quy trình xác định khu vực ngập lụt sử dụng ảnh vệ tinh Sentinel-2 bằng phần mềm SNAP.



Hình 3. Khu vực ngập lụt trong quá khứ tại Quảng Nam xác định được sử dụng ảnh Sentinel (2018).



Hình 4. Kết quả số hóa các vùng ngập trong lịch sử từ năm 2018 - 2020.



2.2.5. Chuẩn hóa các lớp dữ liệu

Các lớp dữ liệu thành phần được xây dựng như đã trình bày ở trên có giá trị biểu thị rất khác nhau. Để có thể đưa các thành phần vào trong cùng một bài toán cần phải quy đổi đơn vị tương đương với nhau. Công thức xác định đơn vị tương đương trong trường hợp này được cho như sau:

$$\mu(C_{ij}) = \frac{FR_{ij} - \text{Min}(FR_{ij})}{\text{Max}(FR_{ij}) - \text{Min}(FR_{ij})} [\text{Max}(\mu(C_{ij})) - \text{Min}(\mu(C_{ij}))] + \text{Min}(\mu(C_{ij})) \quad (1)$$

Trong đó:  $\mu(C_{ij})$  - giá trị thành viên mờ;  $\text{Max}(\mu(C_{ij}))$  và  $\text{Min}(\mu(C_{ij}))$  - giới hạn chuẩn hóa trên và dưới tương ứng.

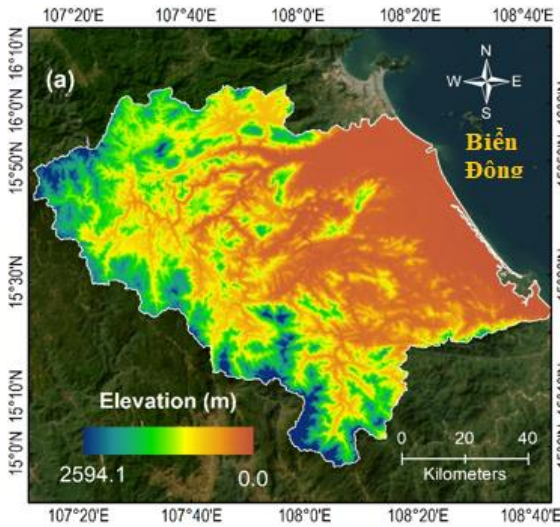
Kết quả của bước này sẽ là giá trị của các pixel của các lớp dữ liệu đã được thành lập. Giá trị của các pixel sẽ nằm trong khoảng từ 0,001 đến 0,999.

3. Kết quả và thảo luận

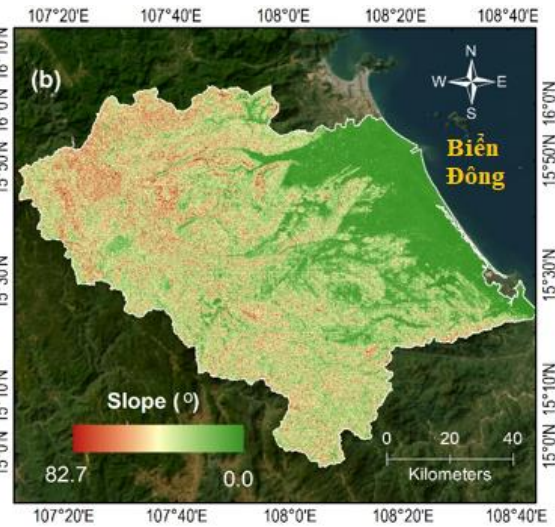
3.1. Kết quả thành lập các lớp dữ liệu

Từ phương pháp đã lựa chọn ở trên, tiến hành biên tập các lớp dữ liệu thành phần như sau:

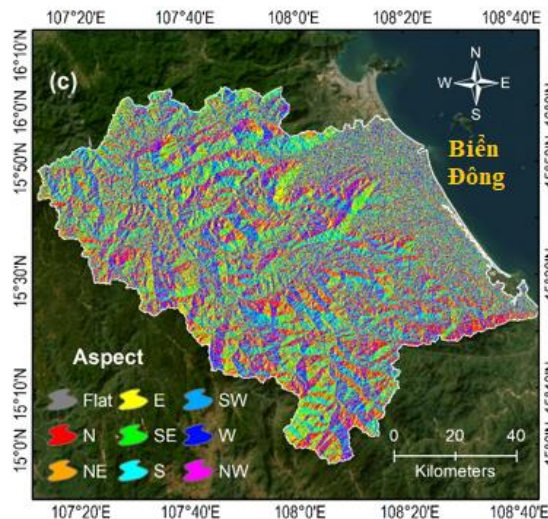
- Kết quả số hóa vùng ngập lụt trong lịch sử tại Quảng Nam.
- Kết quả xây dựng các lớp dữ liệu thành phần như Hình 5.



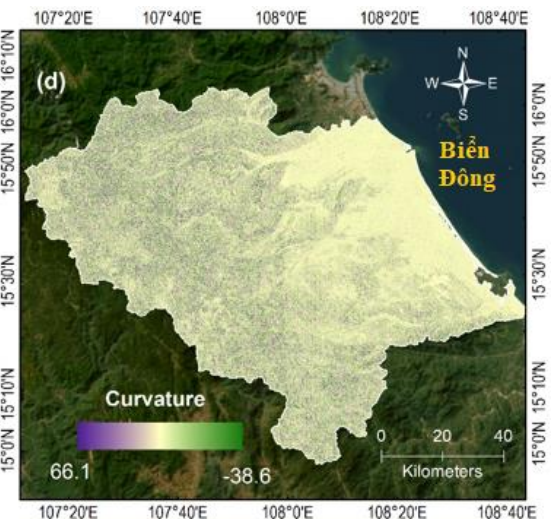
Hình 5a. Lớp dữ liệu thành phần độ cao.



Hình 5b. Lớp dữ liệu thành phần độ dốc.

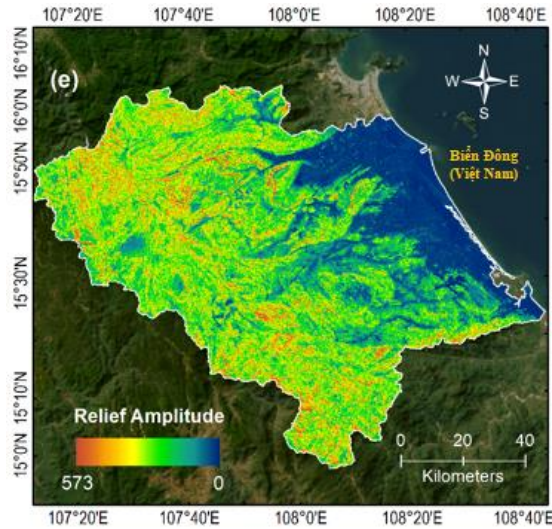


Hình 5c. Lớp dữ liệu thành phần hướng dốc.

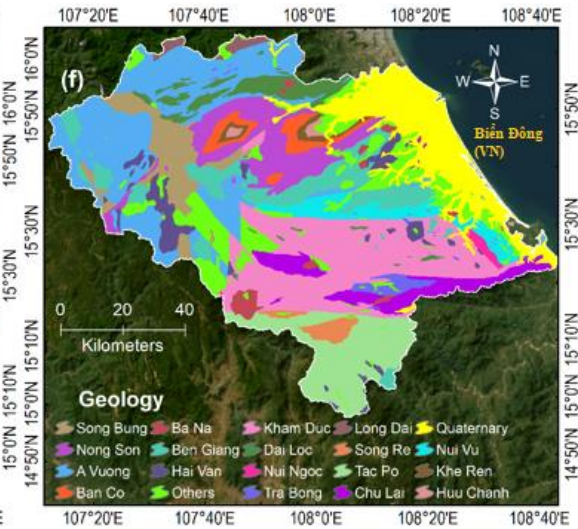


Hình 5d. Lớp dữ liệu thành phần độ cong.





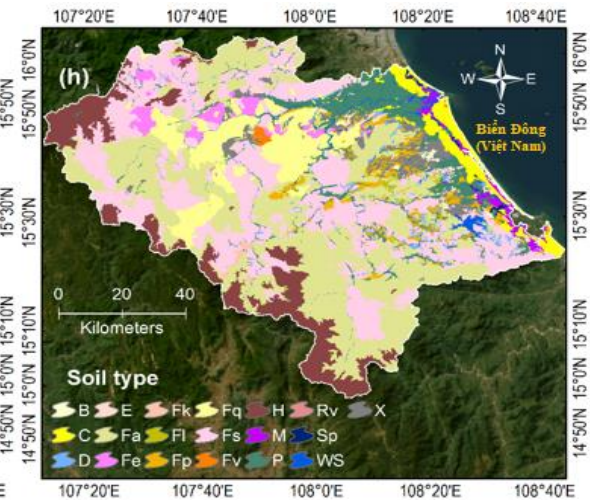
Hình 5e. Lớp dữ liệu thành phần năng lượng địa hình.



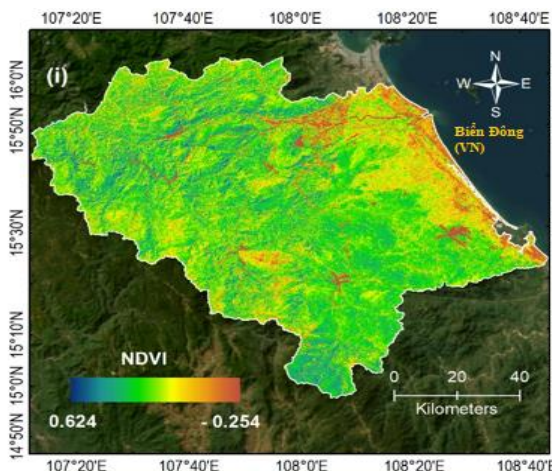
Hình 5f. Lớp dữ liệu thành phần địa chất.



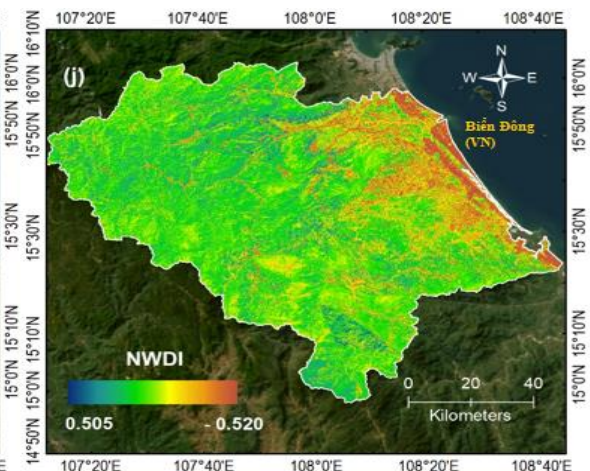
Hình 5g. Lớp dữ liệu thành phần sử dụng đất.



Hình 5h. Lớp dữ liệu thành phần thổ nhưỡng.



Hình 5i. Lớp dữ liệu thành phần chỉ số thực vật khác biệt chuẩn hóa (NDVI).



Hình 5j. Lớp dữ liệu thành phần chỉ số khác biệt nước được chuẩn hóa (NWDI).



### 3.2. Kết quả chuẩn hóa các lớp dữ liệu

Trên cơ sở các bản đồ thành phần đã thành lập, sử dụng công thức (1) để chuẩn hóa các lớp dữ liệu làm dữ liệu đầu vào khi sử dụng các mô hình học máy, trí tuệ nhân tạo. Kết quả chuẩn hóa như sau:

```
relation WekaExcel
@attribute Geol numeric
@attribute Soil numeric
@attribute Lulc numeric
@attribute NDVI numeric
@attribute NDWI numeric
@attribute Rainfall numeric
@attribute Ele numeric
@attribute Slop numeric
@attribute Aspect numeric
@attribute ReliefA numeric
@attribute FL {1,0}
@data
0.841 0.056 0.112 0.234 0.717 0.717 0.902 0.024 0.015 0.325 1
0.841 0.056 0.112 0.246 0.700 0.700 0.903 0.024 0.011 0.378 1
0.841 0.056 0.112 0.248 0.699 0.699 0.903 0.024 0.021 0.449 1
0.841 0.056 0.112 0.262 0.715 0.715 0.902 0.024 0.015 0.457 1
0.841 0.056 0.112 0.269 0.717 0.717 0.902 0.024 0.015 0.431 1
0.841 0.056 0.112 0.276 0.718 0.718 0.903 0.024 0.019 0.351 1
0.841 0.056 0.112 0.260 0.631 0.631 0.901 0.024 0.019 0.432 1
0.841 0.056 0.112 0.277 0.719 0.719 0.902 0.024 0.012 0.399 1
0.841 0.056 0.112 0.284 0.626 0.626 0.901 0.024 0.016 0.346 1
0.841 0.056 0.112 0.289 0.620 0.620 0.901 0.024 0.012 0.364 1
0.841 0.056 0.112 0.292 0.632 0.632 0.902 0.023 0.016 0.547 1
0.841 0.056 0.112 0.294 0.605 0.605 0.902 0.024 0.025 0.422 1
0.841 0.056 0.112 0.182 0.563 0.563 0.926 0.024 0.019 0.325 1
0.841 0.056 0.112 0.183 0.636 0.636 0.926 0.024 0.012 0.408 1
0.841 0.056 0.112 0.183 0.636 0.636 0.926 0.023 0.019 0.432 1
0.841 0.056 0.112 0.183 0.561 0.561 0.926 0.023 0.027 0.474 1
```

Kết quả chuẩn hóa như trên đã được chuẩn hóa về định dạng dữ liệu của phần mềm Weka để có thể sử dụng các hàm sẵn có trong phần mềm này. Trong cấu trúc dữ liệu trên, mỗi dòng dữ liệu bao gồm 11 cột tương ứng với 11 lớp dữ liệu được giải thích bằng 11 hàng bắt đầu bằng kí tự @ (không phải dòng @data).

### 4. Kết luận

Thiên tai, lũ lụt đặc biệt là ngập lụt các vùng ven biển tại Việt Nam diễn ra thường xuyên gây ảnh hưởng nghiêm trọng đến người và tài sản do có đường bờ biển dài và nằm trong vùng nhiệt đới gió mùa ẩm. Do đó, công tác thành lập bản đồ ngập lụt cho các vùng ven biển có tính thời sự và ý nghĩa khoa học.

Với ưu điểm cho phép xử lý hỗn hợp nhiều loại dữ liệu đầu vào trong một không gian địa lý thống nhất, phương pháp kết hợp trí tuệ nhân tạo với công nghệ GIS đang được sử dụng một cách phổ biến trong nghiên cứu về thiên tai, lũ lụt nói chung và ngập lụt bờ biển nói riêng. Muốn vậy, cần phải xây dựng được cơ sở dữ liệu đầu vào một cách thống nhất để xử lý bằng các mô hình học máy, trí tuệ nhân tạo.

Trong quy trình xây dựng bộ cơ sở dữ liệu đó, trước tiên phải xây dựng các bản đồ thành phần là các yếu tố có ảnh hưởng đến ngập lụt. Sau đó, tiến hành chuẩn hóa các lớp bản đồ để đưa các bản đồ về cùng một đơn vị mới có thể xử lý bằng các mô hình học máy, trí tuệ nhân tạo. Bài báo này đã đề xuất quy trình và xây dựng được bộ cơ sở dữ liệu gồm 11 lớp phục vụ thành lập bản đồ ngập lụt cho tỉnh Quảng Nam ứng dụng mô hình học máy, trí tuệ nhân tạo kết hợp công nghệ GIS với kết quả cuối cùng là tệp dữ liệu ở định dạng của phần mềm Weka.

### Lời cảm ơn

Tập thể tác giả xin gửi lời cảm ơn tới đề tài cấp Bộ Tài nguyên và Môi trường “Nghiên cứu ứng dụng trí tuệ nhân tạo xây dựng hệ thống cảnh báo thiên tai, lũ lụt cho các vùng ven biển Việt Nam có sự tham gia tương tác của cộng đồng, thử nghiệm tại 1 tỉnh thuộc khu vực ven biển miền Trung” mã số TNMT.2022.04.09 đã cung cấp tài liệu, dữ liệu để thực hiện bài báo này.

### Đóng góp của các tác giả

Nguyễn Gia Trọng - xây dựng ý tưởng bài báo, viết bản thảo; Nguyễn Việt Nghĩa - xây dựng ý tưởng bài báo; Nguyễn Văn Cương - xây dựng ý tưởng bài báo; Dương Anh Quân - xử lý dữ liệu, viết bản thảo; Phạm Ngọc Quang - xử lý dữ liệu, viết bản thảo, kiểm tra và chỉnh sửa bài báo; Nguyễn Đình Hải - xử lý dữ liệu, kiểm tra và chỉnh sửa bài báo; Nguyễn Hoàng Nhi - xử lý dữ liệu, kiểm tra và chỉnh sửa bài báo.

### Tài liệu tham khảo

Akshayasimha Channarayapatna Harshasimha, Chandra Mohan Bhatt, (2023), Flood Vulnerability Mapping Using MaxEnt Machine Learning and Analytical Hierarchy Process (AHP) of Kamrup Metropolitan District, Assam,

- Environ. Sci. Proc. 2023, 25, 73.  
<https://doi.org/10.3390/ECWS-7-14301>.
- Chau, V. N., Cassells, S., & Holland, J., (2015). Economic impact upon agricultural production from extreme flood events in Quang Nam, central Vietnam. *Natural Hazards*, 75, 1747-1765, DOI 10.1007/s11069-014-1395-x.
- Costache, R., Țincu, R., Elkhachy, I., Pham, Q. B., Popa, M. C., Diaconu, D. C., ... & Bui, D. T., (2020a). New neural fuzzy-based machine learning ensemble for enhancing the prediction accuracy of flood susceptibility mapping. *Hydrological Sciences Journal*, 65(16), 2816-2837, DOI: 10.1080/02626667.2020.1842412.
- Costache, R., Popa, M. C., Bui, D. T., Diaconu, D. C., Ciubotaru, N., Minea, G., & Pham, Q. B., (2020b). Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. *Journal of Hydrology*, 585, 124808, <https://doi.org/10.1016/j.jhydrol.2020.124808>.
- Diaconu, D. C., Costache, R., & Popa, M. C., (2021). An Overview of Flood Risk Analysis Methods. *Water* 2021, 13, 474. <https://doi.org/10.3390/w13040474>.
- Dodangeh, E., Choubin, B., Eigdir, A. N., Nabipour, N., Panahi, M., Shamshirband, S., & Mosavi, A., (2020). Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Science of the Total Environment*, 705, 135983, <https://doi.org/10.1016/j.scitotenv.2019.135983>.
- Mcleod, E., Poulter, B., Hinkel, J., Reyes, E., & Salm, R., (2010). Sea-level rise impact models and environmental conservation: A review of models and their applications. *Ocean & Coastal Management*, 53(9), 507-517.
- Nghĩa, N. V., & Cường, N. C., (2020). Ứng dụng mạng Nơ-ron nhân tạo đa lớp trong thành lập mô hình phân vùng lũ quét khu vực miền núi Tây Bắc, thực nghiệm tại tỉnh Yên Bái. *Tạp chí Khoa học Đo đạc và Bản đồ*, (44), 56-64.
- Pham, B. T., Luu, C., Van Phong, T., Nguyen, H. D., Van Le, H., Tran, T. Q., ... & Prakash, I., (2021). Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. *Journal of Hydrology*, 592, 125815, <https://doi.org/10.1016/j.jhydrol.2020.125815>.
- Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T., (2019). Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. *Earth Systems and Environment*, 3, 585-601, <https://doi.org/10.1007/s41748-019-00123-y>.
- Luu, C., Von Meding, J., & Kanjanabootra, S. (2018). Assessing flood hazard using flood marks and analytic hierarchy process approach: a case study for the 2013 flood event in Quang Nam, Vietnam. *Natural Hazards*, 90, 1031-1050; DOI 10.1007/s11069-017-3083-0.
- Luu, C., & Von Meding, J. (2018). A flood risk assessment of Quang Nam, Vietnam using spatial multicriteria decision analysis. *Water*, 10(4), 461; DOI:10.3390/w10040461.