

# Ứng dụng phương pháp hồi quy logistic xác định tổ hợp tối ưu các yếu tố ảnh hưởng và xây dựng bản đồ tai biến trượt lở đất huyện Khánh Vĩnh, tỉnh Khánh Hòa

Nguyễn Thanh Danh, Đậu Văn Ngọ, Tạ Quốc Dũng

**Tóm tắt**—Nghiên cứu này áp dụng phương pháp hồi quy logistic (LR) tích hợp với GIS thành lập bản đồ tai biến trượt đất huyện Khánh Vĩnh, tỉnh Khánh Hòa dựa vào mối liên hệ không gian giữa các yếu tố ảnh hưởng liên quan đến địa hình; thạch học; thực vật; lượng mưa lớn nhất năm; khoảng cách đến đường giao thông, sông suối, đứt gãy và sự phân bố các điểm trượt lở. Sử dụng đường cong tỷ lệ thành công (success rate) và tỷ lệ dự báo (prediction rate) để đánh giá mức độ phù hợp và độ chính xác của phương pháp hồi quy logistic. Kết quả cho thấy phương pháp này có mức độ phù hợp và độ chính xác cao (phần diện tích bên dưới đường cong: Areas Under Curves - AUC = 0,8-0,9). Thuật toán mô hình trung bình Bayesian (BMA) của phần mềm thống kê R được áp dụng để xác định các yếu tố ảnh hưởng nhất và các mô hình tối ưu tổ hợp yếu tố ảnh hưởng. Có bốn yếu tố ảnh hưởng quan trọng nhất và năm mô hình tối ưu tổ hợp yếu tố ảnh hưởng. Mô hình 3 (góc dốc, hướng dốc, cao độ, khoảng cách đến đường giao thông và lượng mưa lớn nhất năm) là mô hình tối ưu tốt nhất.

**Từ khóa**—Trượt đất, hồi quy logistic, tỷ lệ thành công, tỷ lệ dự báo.

## 1 MỞ ĐẦU

Vùng nghiên cứu nằm trọn trong ranh giới hành chính huyện Khánh Vĩnh, là một vùng núi phía tây tỉnh Khánh Hòa. Khánh Vĩnh là huyện miền núi, tiếp giáp với đồng bằng, bắc giáp huyện Ninh Hòa và tỉnh Đắk Lắk, đông giáp huyện Diên

Khánh, nam giáp huyện Khánh Sơn, tây giáp tỉnh Đắk Lắk và tỉnh Lâm Đồng (hình 1). Trượt lở đất ở các sườn dốc, mái dốc tự nhiên và mái dốc công trình thường xảy ra khi có mưa to kéo dài, đặc biệt là vào mùa mưa bão. Một trong những biện pháp cấp thiết hiện nay để các cấp chính quyền phòng, tránh và giảm thiểu những thiệt hại do trượt lở đất gây ra là thành lập bản đồ phân vùng tai biến trượt lở. Mục đích chính của việc thành lập bản đồ này là nhằm cảnh báo trước vùng có tai biến trượt lở xảy ra trong tương lai, làm cơ sở khoa học cho quy hoạch phát triển bền vững kinh tế - xã hội và bảo vệ môi trường trên địa bàn huyện Khánh Vĩnh.

Hầu hết các bản đồ nguy cơ trượt lở được thành lập theo một trong các phương sau: Phương pháp trực tiếp; phương pháp tính toán địa kỹ thuật và phương pháp xác suất thống kê [1, 2]. Các phương pháp xác suất thống kê thường sử dụng gồm: Phương pháp tỷ số tần suất - Frequency Ratio; phương pháp chỉ số thống kê - Statistical Index; phương pháp trọng số các chứng cứ - Weights of Evidence; phương pháp hồi quy logistic - Logistic Regression; phương pháp mạng nơron nhân tạo - Artificial Neural Network. Trong nghiên cứu này sẽ sử dụng phương pháp hồi quy logistic.

Do đó, bản đồ chỉ số nguy cơ trượt lở (Landslide Susceptibility Index - *LSI*) xem như là một hàm của các điểm trượt lở và các yếu tố ảnh hưởng (1); bản đồ chỉ số tai biến trượt lở (Landslide Hazard Index - *LHI*) xem như là một hàm của các điểm trượt lở, các yếu tố ảnh hưởng và yếu tố kích thích (2) [3].

*Bản thảo nhận được vào ngày 7 tháng 8 năm 2017. Bản sửa đổi bản thảo ngày 25 tháng 12 năm 2017.*

Nguyễn Thanh Danh - Trường Đại học Xây dựng Miền Trung, Bộ Xây dựng (e-mail: nguyenthandanh@cuc.edu.vn).

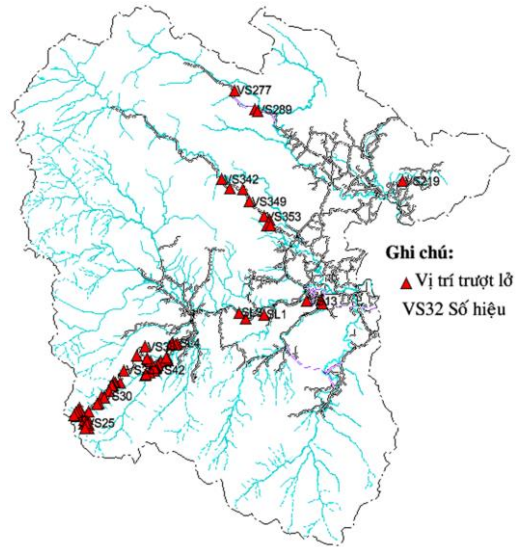
Đậu Văn Ngọ - Khoa Kỹ thuật Địa chất và Dầu khí, Trường Đại học Bách khoa, ĐHQG-HCM.

Tạ Quốc Dũng - Khoa Kỹ thuật Địa chất và Dầu khí, Trường Đại học Bách Khoa - ĐHQG-HCM (e-mail: [tqdung@hcmut.edu.vn](mailto:tqdung@hcmut.edu.vn)).

\* Tác giả chính: Email: nguyenthandanh@cuc.edu.vn

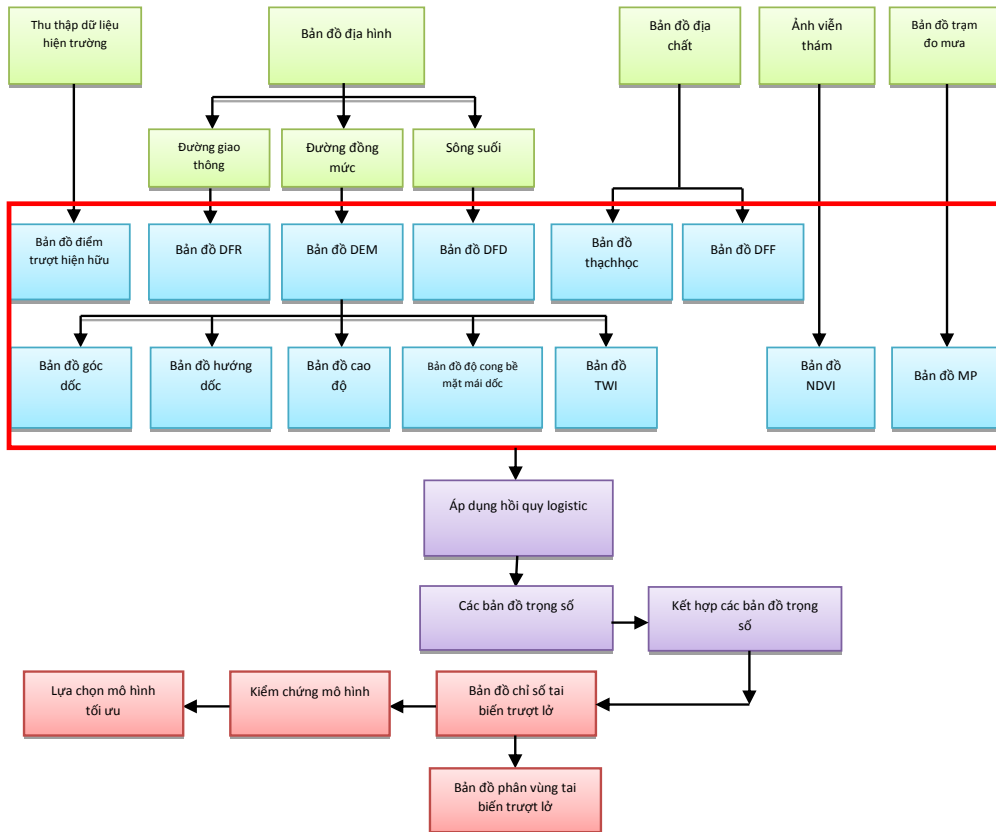


Hình 1. Vị trí vùng nghiên cứu



Ghi chú:  
▲ Vị trí trượt lở  
VS32 Số hiệu

Hình 2. Vị trí điểm trượt lở trên các tuyến giao thông huyện Khánh Vinh



Hình 3. Quy trình xây dựng bản đồ tại biến trượt lở

Trong nghiên cứu này, lượng mưa lớn nhất năm được xem là yếu tố kích thích bên ngoài khi thành lập bản đồ tai biến trượt lở và điểm trượt lở trên các tuyến giao thông (hình 2).

$$LSI = f(\text{điểm trượt, yếu tố ảnh hưởng}) \quad (1)$$

$$LHI = f(\text{điểm trượt, yếu tố ảnh hưởng, yếu tố kích thích}) \quad (2)$$

## 2 PHƯƠNG PHÁP

Phương pháp hồi quy logistic là một hình thức của phương pháp hồi qui đa biến thể hiện mối quan hệ giữa một biến phụ thuộc và một vài biến độc lập [3, 5-7]. Phương pháp hồi quy logistic thường được sử dụng để dự báo sự có mặt hay vắng mặt của biến phụ thuộc dựa vào giá trị của các biến độc lập. Biến phụ thuộc có dạng nhị phân còn các biến độc lập có thể ở dạng nhị phân, liên tục, rời rạc và phân nhóm. Trong nghiên cứu này, biến phụ thuộc nhận giá trị là 1 nếu có xảy ra trượt lở và nhận giá trị là 0 nếu không xảy ra trượt lở; các biến độc lập là các yếu tố ảnh hưởng đến trượt lở.

Theo phương pháp hồi quy logistic, quan hệ giữa xác suất xảy ra trượt lở và các yếu tố ảnh hưởng được thể hiện qua phương trình:

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3)$$

Trong đó:  $P$ : Xác suất xảy ra trượt lở;  $\alpha_0$ : Hằng số;  $\beta_i$  ( $i = 1, 2, \dots, n$ ): Hệ số hồi quy;  $x_i$  ( $i = 1, 2, \dots, n$ ): Biến độc lập (yếu tố ảnh hưởng).

## 3 DỮ LIỆU VÀ QUY TRÌNH LẬP BẢN ĐỒ

### 3.1 Phân bố điểm trượt lở

Để xây dựng bản đồ phân bố điểm trượt lở vùng nghiên cứu, có ba phương pháp được sử dụng: Phương pháp thu thập dữ liệu lịch sử trượt lở đã có; phương pháp đi thực địa, xác định tọa độ vị trí trượt lở bằng máy GPS cầm tay (GARMIN eTrex30) có độ chính xác  $\pm 5$  m; phương pháp phân tích ảnh viễn thám và ảnh Google Earth kết hợp với kiểm tra ở hiện trường. Trong vùng nghiên cứu có 231 điểm trượt lở được ghi nhận và được chia ngẫu nhiên thành 2 nhóm dữ liệu: Nhóm dữ liệu huấn luyện chiếm 70% (161 điểm trượt lở); nhóm dữ liệu kiểm tra chiếm 30% (70 điểm trượt lở). Nhóm dữ liệu huấn luyện dùng để tính toán; nhóm dữ liệu kiểm tra dùng để đánh giá độ chính xác các phương pháp.

### 3.2 Các yếu tố ảnh hưởng

Lựa chọn các yếu tố ảnh hưởng đến trượt lở là rất cần thiết đối với việc thành lập bản đồ tai biến

trượt lở. Tuy nhiên, cho tới nay tất cả các nghiên cứu vẫn chưa thống nhất trong lựa chọn các yếu tố ảnh hưởng. Do đó, các yếu tố ảnh hưởng được lựa chọn chủ yếu dựa vào dữ liệu sẵn có và kinh nghiệm chuyên gia đối với vùng nghiên cứu [1, 4, 8]. Nghiên cứu này lựa chọn 11 yếu tố ảnh hưởng đến trượt lở để đưa vào phân tích trong các mô hình gồm: Góc dốc, hướng dốc, cao độ, chỉ số ẩm ướt, độ uốn cong bề mặt mái dốc, thạch học, khoảng cách đến đường giao thông, khoảng cách đến sông suối, khoảng cách đến đứt gãy, chỉ số thực vật và lượng mưa lớn nhất năm. Bản đồ tai biến trượt lở đất được thành lập từ 11 yếu tố ảnh hưởng theo phương pháp hồi quy logistic dựa trên quy trình như hình 3.

## 4 KẾT QUẢ

### 4.1 Xác định các hệ số hồi quy

Tất cả các bản đồ yếu tố ảnh hưởng đã được phân nhóm đều chuyển sang dạng raster (lưới) với kích thước mỗi pixel là  $20 \times 20$  m<sup>2</sup> và được phủ lên bởi bản đồ phân bố của 161 điểm trượt lở huấn luyện (dùng để tính toán). Trong toàn vùng nghiên cứu có tổng số 2915267 pixel với 1435 pixel trượt lở. Sử dụng phần mềm thống kê R có mã nguồn mở để tính toán xác định các hệ số hồi quy. Kết quả tính toán cho trong bảng 1.

BẢNG 1. CÁC HỆ SỐ HỒI QUY

Yếu tố	Hệ số hồi quy ( $\beta$ )	Sai số chuẩn (SE)	Wald Z	Pr(> Z )
Cao độ (DEM)	1,4144	0,2193	6,45	<0,0001
Góc dốc (SLOPE)	0,372	0,1043	3,57	0,0004
Hướng dốc (ASPECTD)	0,3432	0,1408	2,44	0,0148
Độ uốn cong bề mặt mái dốc (CURVATURE)	0,999	0,6286	1,59	0,112
Chỉ số ẩm ướt địa hình (TWI)	-0,011	0,3196	-0,03	0,9726
Thạch học (LITHOLOGY)	0,259	0,0869	2,98	0,0029
Lượng mưa lớn nhất năm (MP)	0,6136	0,1114	5,51	<0,0001
Khoảng cách đến sông suối (DFD)	0,3264	0,187	1,75	0,0809
Khoảng cách đến đường giao thông (DFR)	0,7938	0,0631	12,57	<0,0001
Khoảng cách đến đứt gãy (DFF)	0,5047	0,1895	2,66	0,0077
Chỉ số thực vật (NDVI)	-0,034	0,2036	-0,17	0,8659

Ghi chú: Hằng số  $\alpha_0 = -13,1181$ ; Wald Z: Số thống kê Wald Z; Pr(>|Z|): Giá trị P.

4.2 Bản đồ phân vùng tai biến trượt lở

Bản đồ tai biến trượt lở được thành lập dựa vào sự kết hợp của 11 bản đồ trọng số của các yếu tố ảnh hưởng theo các phương trình (3). Để thuận tiện trong việc trình bày và đánh giá kết quả, bản đồ chỉ số tai biến trượt lở được phân vùng theo các mức độ tai biến khác nhau. Trong nghiên cứu này, có 5 mức độ phân vùng tai biến trượt lở được áp dụng: Rất thấp, thấp, trung bình, cao và rất cao.

Dựa vào biểu đồ histograms của giá trị chỉ số tai biến trượt lở, phương pháp độ lệch chuẩn (Standard Deviation) được sử dụng để phân vùng tai biến trượt lở. Trong thực tế, hầu hết các biểu đồ histograms này thường không có phân bố chuẩn. Do đó, các bản đồ chỉ số tai biến trượt lở trước khi phân vùng mức độ tai biến trượt lở cần được chuẩn hóa để biểu đồ histograms có phân bố chuẩn hoặc gần phân bố chuẩn. Phương trình chuẩn hóa chỉ số tai biến trượt lở:

$$LHI_{sd} = \frac{LHI - LHI_{min}}{LHI_{max} - LHI_{min}} \quad (4)$$

Trong đó:

$LHI_{sd}$ : Giá trị chỉ số tai biến trượt lở đã chuẩn hóa;  $LHI$ : Giá trị chỉ số tai biến trượt lở chưa chuẩn hóa;  $LHI_{min}$ : Giá trị chỉ số tai biến trượt lở nhỏ nhất;  $LHI_{max}$ : Giá trị chỉ số tai biến trượt lở lớn nhất.

Từ phương trình (3), chỉ số tai biến trượt lở ( $LHI$ ) và xác suất trượt lở đã được chuẩn hóa ( $P_{sd}$ ) xác định như sau:

$$LHI = -13,1181 + 1,4144*DEM + 0,372*SLOPE + 0,3432*ASPECT + 0,999*CURVATURE - 0,011*TWI + 0,259*LITHOLOGY + 0,6136*MP + 0,3264*DFD + 0,7938*DFR + 0,5047*DFP - 0,034*NDVI \quad (5)$$

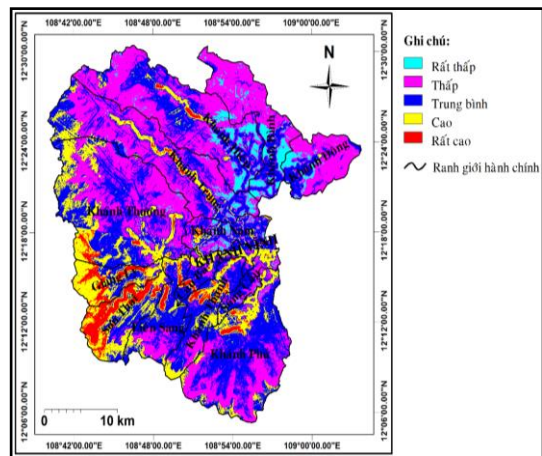
$$P_{sd} = EXP(LHI_{sd}) / (1 + EXP(LHI_{sd})) \quad (6)$$

Các khoảng giá trị phân vùng các mức độ tai biến trượt lở dựa vào giá trị trung bình và độ lệch chuẩn (Bảng 2). Bản đồ phân vùng các mức độ tai biến trượt lở được thành lập từ bản đồ chỉ số tai biến trượt lở và xác suất trượt lở đã được chuẩn hóa (Hình 4).

BẢNG 2  
KHOẢNG GIÁ TRỊ PHÂN VÙNG MỨC ĐỘ TAI BIẾN TRƯỢT LỖ

Phân vùng tai biến trượt lở	Khoảng giá trị
Rất thấp	$Min \sim Mean - 1,5Std.Dev$
Thấp	$Mean - 1,5Std.Dev \sim Mean - 0,5Std.Dev$
Trung bình	$Mean - 0,5Std.Dev \sim Mean + 0,5Std.Dev$
Cao	$Mean + 0,5Std.Dev \sim Mean + 1,5Std.Dev$
Rất cao	$Mean + 1,5Std.Dev \sim Max$

Ghi chú:  $Std.Dev$  (0,033): Độ lệch chuẩn;  $Mean$  (0,59): Trị trung bình;  $Median$  (0,58): Số trung vị;  $Min$  (0,5),  $Max$  (0,73): Giá trị nhỏ nhất và lớn nhất.



Hình 4. Bản đồ phân vùng tai biến trượt lở

4.3 Kiểm chứng phương pháp

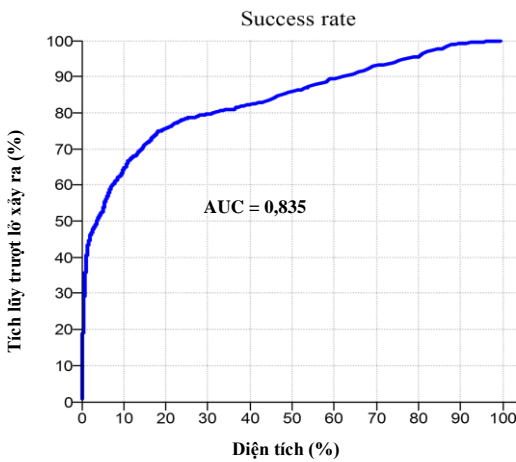
Kiểm chứng phương pháp là một nhiệm vụ rất quan trọng và không thể thiếu. Nếu phương pháp không được kiểm chứng thì phương pháp không thể sử dụng và không có ý nghĩa khoa học. Một phương pháp sử dụng phổ biến hiện nay là dựa vào các đường cong tỷ lệ, có 2 loại: Đường cong tỷ lệ thành công (Success rate) và đường cong tỷ lệ dự báo (Prediction rate). Đường cong tỷ lệ thành công dùng để đánh giá mức độ phù hợp của phương pháp so với dữ liệu quan sát, đường cong tỷ lệ dự báo dùng để đánh giá mức độ chính xác của phương pháp.

Đường cong tỷ lệ thành công (Hình 5) được thành lập từ bản đồ xác suất tai biến trượt lở với các điểm trượt dùng để huấn luyện (161 điểm trượt). Đường cong tỷ lệ dự báo (Hình 6) được thành lập từ bản đồ xác suất tai biến trượt lở với các điểm trượt dùng để kiểm tra (70 điểm trượt).

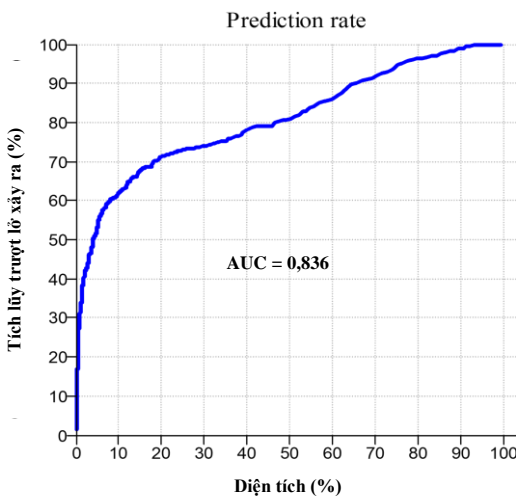
Giá trị các pixel của bản đồ xác suất tai biến trượt lở được sắp xếp theo thứ tự từ lớn đến bé để

tính phần trăm diện tích. Các điểm trượt lở được chồng lên các bản đồ này để tính phần trăm tích lũy trượt lở đã xảy ra.

Giá trị của phần diện tích bên dưới các đường cong tỷ lệ (Areas Under Curves - AUC) được sử dụng như một thông số định lượng để kiểm chứng phương pháp. Giá trị AUC thay đổi từ 0,5~1. Phương pháp tốt nhất là phương pháp có giá trị AUC gần bằng 1. Quan hệ giữa AUC và chất lượng mô hình có thể phân chia như sau [7]: AUC = 0,5~0,6: Kém; AUC = 0,6~0,7: Trung bình; AUC = 0,7~0,8: Tốt; AUC = 0,8~0,9: Rất tốt; AUC = 0,9~0,1: Xuất sắc. Kết quả tính toán cho giá trị AUC thuộc khoảng 0,8~0,9 nên có mức độ phù hợp và độ chính xác cao.



Hình 5. Đường cong tỷ lệ thành công



Hình 6. Đường cong tỷ lệ dự báo

4.4 Đánh giá các yếu tố ảnh hưởng và lựa chọn mô hình tối ưu

Tiêu chí chung để tìm mô hình tối ưu là mô hình phải đảm bảo các điều kiện sau: Đơn giản (Mô hình phải có ít yếu tố ảnh hưởng, vì nếu có quá nhiều yếu tố ảnh hưởng sẽ dẫn đến khó khăn trong giải thích kết quả và không thực tiễn); đầy đủ (Mô hình mô tả dữ liệu một cách tốt nhất, nghĩa là kết quả mô hình gần đúng nhất với dữ liệu quan sát và có độ chính xác cao); có ý nghĩa thực tế (Các yếu tố ảnh hưởng trong mô hình phải giải thích được bằng lý thuyết và bằng thực tiễn).

Hiện nay, có các thuật toán để tìm mô hình tối ưu như: Stepwise, Backward, Deviance, AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) và BMA (Bayesian Model Average). Trong các thuật toán này, thuật toán BMA không chỉ đưa ra một mô hình tối ưu nhất mà còn cho biết một vài mô hình tối ưu khác có thể để lựa chọn cho phù hợp với điều kiện thực tế. Phần mềm thống kê R có mã nguồn mở được sử dụng để giải thuật toán BMA.

Kết quả tính toán có 7 yếu tố ảnh hưởng liên quan đến trượt lở (DEM, DFR, MP, SLOPE, LITHOLOGY, DFF, ASPECT) vì xác suất xuất hiện khác không. Trong đó, có 4 yếu tố ảnh hưởng quan trọng nhất (DEM, DFR, MP và SLOPE) vì xác suất xuất hiện 100%. Có 5 mô hình tối ưu có thể để lựa chọn, nhưng mô hình tối ưu nhất là mô hình 1 vì có ít yếu tố ảnh hưởng nhất.

Các mô hình tối ưu đều có giá trị AUC từ 0,8~0,9 nên có độ chính xác cao (Bảng 3). Mặc dù mô hình 1 là mô hình tối ưu nhất nhưng mô hình 3 với 5 yếu tố ảnh hưởng là mô hình tối ưu tốt nhất vì có giá trị AUC cao hơn.

BẢNG 3  
GIÁ TRỊ AUC CỦA CÁC MÔ HÌNH TỐI ƯU

Mô hình	Yếu tố ảnh hưởng	AUC
1	DEM, DFR, MP và SLOPE	0,801
2	DEM, DFR, LITHOLOGY, MP và SLOPE	0,807
3	ASPECT, DEM, DFR, MP và SLOPE	0,812
4	DEM, DFF, DFR, MP và SLOPE	0,801
5	DEM, DFF, DFR, LITHOLOGY, MP và SLOPE	0,806

5 KẾT LUẬN

Tai biến trượt lở đất thường xuyên xảy ra vào mùa mưa ở huyện miền núi Khánh Vĩnh, đặc biệt là dọc theo các tuyến đường giao thông. Trượt lở chủ yếu xảy ra ở cao độ từ 1200~1500 m; góc dốc

từ 40~60°; hướng dốc Đông - Nam; trong nhóm sét kết, bột kết, cát kết, sạn kết, cuội kết và trong vỏ phong hóa dày của nhóm andesit, dacsit, andesittodacit; vùng phân bố lượng mưa lớn nhất nằm từ 650~750 mm; trong phạm vi khoảng cách đến sông suối, đường giao thông từ 0~100 m và khoảng cách đến đứt gãy từ 400~500 m.

Phương pháp hồi quy logistic được sử dụng để thành lập bản đồ tai biến trượt lở dựa vào 11 yếu tố ảnh hưởng (DEM, SLOPE, ASPECT, CURVATURE, TWI, LITHOLOGY, DFR, DFD, DFF, NDVI và MP) có mức độ phù hợp, độ chính xác cao (AUC của đường cong tỷ lệ thành công là 0,835; AUC của đường cong tỷ lệ dự báo là 0,836).

Các bản đồ chỉ số tai biến trượt lở cần được chuẩn hóa để có phân bố chuẩn hoặc gần phân bố chuẩn. Áp dụng phương pháp độ lệch chuẩn (Standard Deviation) để phân vùng mức độ tai biến trượt lở. Bản đồ phân vùng tai biến trượt lở gồm 5 vùng theo mức độ từ thấp đến cao: Rất thấp, thấp, trung bình, cao và rất cao.

Trong nghiên cứu này, lần đầu tiên thuật toán BMA (Bayesian Model Average) được sử dụng để tìm các yếu tố ảnh hưởng quan trọng nhất và các tổ hợp yếu tố ảnh hưởng tối ưu. Kết quả có 7 yếu tố ảnh hưởng liên quan đến trượt lở (DEM, DFR, MP, SLOPE, DFF, LITHOLOGY, ASPECT), 4 yếu tố ảnh hưởng quan trọng nhất (DEM, DFR, MP và SLOPE) và 5 mô hình tối ưu tổ hợp các yếu tố ảnh hưởng.

Phương pháp hồi quy logistic lần lượt được áp dụng cho các mô hình tổ hợp để dự báo tai biến trượt lở cho thấy có độ chính xác cao (AUC = 0,8-0,9). Tuy nhiên, mô hình 3 là mô hình tối ưu tốt nhất, tiếp đến là các mô hình 2, 5, 4 và 1.

## REFERENCES

[1] A. M. Youssef, H. R. Pourghasemi, B. A. El-Haddad, and B. K. Dhahry, "Landslide susceptibility map using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia," *Bull Eng Geol Environ*, 2015.

[2] N. R. Regmi, J. R. Giardino, and J. D. Vitek, "Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA," *Geomorphology*, vol. 115, pp. 172-187, 2010.

[3] B. Pradhan and S. Lee, "Delineation of landslide hazard reason Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models," *Environ Earth Sci*, vol. 60, pp. 1037-1054, 2010.

[4] Sở Khoa học và Công nghệ Khánh Hòa, "Báo cáo tổng hợp kết quả khoa học và công nghệ - Nghiên cứu điều kiện địa chất, thủy văn hai huyện miền núi Khánh Sơn và Khánh Vĩnh, đề xuất giải pháp xây dựng các công trình giao thông, thủy lợi bền vững" Khánh Hòa, 2011.

[5] A. Yalcin, S. Reis, A. C. Aydinoglu, and T. Yomralioglu, "GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey," *Catena*, vol. 85, pp. 274-287, 2011.

[6] D. T. Bui, O. Lofman, I. Revhaug, and O. Dick, "Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression," *Nat Hazards*, 2011.

[7] I. Yilmaz, "Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey)," *Computer & Geoscience* vol. 35, pp. 1125-1138, 2009.

**Nguyễn Thanh Danh** sinh ngày 16/ 05/ 1978, năm 2001 tốt nghiệp đại học chuyên ngành Địa kỹ thuật tại Trường Đại học Bách Khoa – ĐHQG-HCM. Năm 2012 tốt nghiệp thạc sĩ chuyên ngành Địa kỹ thuật tại Trường Đại học Bách Khoa – ĐHQG-HCM. Năm 2014 là nghiên cứu sinh chuyên ngành Kỹ thuật Địa chất tại Trường Đại học Bách Khoa – ĐHQG- HCM.

Từ 2001 đến nay là giảng viên Trường Đại học Xây dựng Miền Trung. Năm 2013 đến nay là Trưởng bộ môn Địa kỹ thuật, Khoa Cầu đường, Trường Đại học Xây dựng Miền Trung. Chủ nhiệm 4 đề tài cấp Trường, chủ biên và tham gia 3 sách Đại học. Có hơn 7 bài báo đăng trên tạp chí, kỷ yếu hội nghị trong nước và quốc tế. Các lĩnh vực nghiên cứu chính gồm: Địa chất công trình; địa chất thủy văn; xử lý nền đất yếu; ứng dụng địa thống kê trong nghiên cứu cấu trúc địa chất và tính chất cơ lý đất; ứng dụng các mô hình xác suất thống kê và GIS trong nghiên cứu trượt lở đất đá.

**Đậu Văn Ngộ** sinh ngày 20/ 05/ 1955, tốt nghiệp đại học 1979 và bảo vệ thành công luận án tiến sĩ năm 2001 chuyên ngành Địa chất công trình tại Trường Đại học Mở Địa chất.

Từ 1979 đến 1988 là giảng viên Trường Đại học Hàng Hải. Từ 1988 đến nay là giảng viên Trường Đại học Bách Khoa – ĐHQG-HCM. Năm 2010 được phong học hàm Phó Giáo sư. Chủ nhiệm hơn 6 đề tài cấp Trường, cấp Đại học Quốc Gia và cấp Nhà nước. Hướng dẫn hơn 10 học viên cao học và 4 nghiên cứu sinh, 1 nghiên cứu sinh đã bảo vệ thành công luận án tiến sĩ. Chủ biên và tham gia 4 sách Đại học, có hơn 60 bài báo đăng trên tạp chí,

kỹ yếu hội nghị trong nước và quốc tế. Các lĩnh vực nghiên cứu chính gồm: Địa chất công trình; địa chất thủy văn; xử lý nền đất yếu; địa chất động lực công trình; biến đổi môi trường địa chất dưới ảnh hưởng của biến đổi khí hậu và nước biển dâng; các vấn đề về sử dụng hợp lý và bảo vệ môi trường địa chất.

**Tạ Quốc Dũng** sinh ngày 02/ 05/ 1976, năm 1998 tốt nghiệp đại học chuyên ngành Dầu khí tại Trường Đại học Bách Khoa – ĐHQG-HCM. Năm 2003 tốt nghiệp thạc sĩ chuyên ngành Cơ học ứng dụng thuộc chương trình liên kết giữa Trường Đại học Liege (Bi) và Trường Đại học Bách Khoa – ĐHQG-HCM. Năm 2007 bảo vệ thành công luận án tiến sĩ chuyên ngành Kỹ thuật Dầu khí tại Trường Đại học Adelaide (Úc).

Từ 1998 đến 2008 là giảng viên Trường Đại học Bách Khoa – ĐHQG-HCM. Năm 2008 đến 2015 là Phó trưởng khoa Khoa Kỹ thuật Địa chất và Dầu

khí. Năm 2015 đến nay là Trưởng khoa Khoa Kỹ thuật Địa chất và Dầu khí. Chủ nhiệm hơn 4 đề tài cấp Trường, cấp Đại học Quốc Gia và cấp Bộ. Hướng dẫn hơn 7 học viên cao học và 2 nghiên cứu sinh. Có hơn 29 bài báo đăng trên tạp chí, kỹ yếu hội nghị trong nước và quốc tế. Các lĩnh vực nghiên cứu chính gồm: Ứng dụng các mô hình địa cơ học và địa thống kê trong các lĩnh vực tìm kiếm, thăm dò, khoan và khai thác dầu khí; dự báo khả năng sinh cát theo độ sâu của các giếng khai thác dầu; ứng dụng mô hình địa cơ học trong phân tích ổn định thành giếng khoan.

# Applying logistic regression method to determine combinatorial optimization of landslide-related factors and construct landslide hazard map in Khanh Vinh district, Khanh Hoa Province

Nguyen Thanh Danh, Dau Van Ngo, Ta Quoc Dung

**Abstract**—The purpose of this study is to produce landslide hazard map in Khanh Vinh district, Khanh Hoa province using logistic regression method integrated with GIS analytical tools. The spatial relationship between landslide-related factors such as topography; lithology; vegetation; maximum precipitation in year; distance from roads; distance from drainages; distance from faults and the distribution of landslides were used in the landslide hazard analyses. Using success rate and prediction rate curve assess the fit and accuracy of logistic regression method. The results show that this method have the goodness of fit and the high accuracy (Areas Under Curves - AUC = 0.8 ~ 0.9). Bayesian Model Average (BMA) of the R statistical software was applied to identify the most influential factors and the combinatorial optimization models of landslide-related factors. There are four the most important landslide-related factors and five combinatorial optimization models of landslide-related factors. Model 3 (slope angle, slope aspect, altitude, distance from roads and maximum precipitation in year) is the best optimization.

**Index Terms**—Landslide, logistic regression, success rate, prediction rate.