

XÂY DỰNG MÔ HÌNH DỰ BÁO DỊCH TẢ NGẮN HẠN VÀ ĐÁNH GIÁ ẢNH HƯỞNG CỦA CÁC YẾU TỐ KHÍ HẬU VÀ ĐỊA LÝ

BUILDING SHORT-TERM CHOLERA FORECAST MODELS AND EFFECT EVALUATION OF CLIMATE AND GEOGRAPHICAL FACTORS

Lê Thị Ngọc Anh¹, Hoàng Xuân Dậu²

¹Trường Đại học Y Hà Nội; lengocanh@hmu.edu.vn

²Học viện Công nghệ Bưu chính Viễn thông; dauhx@ptit.edu.vn

Tóm tắt - Sự bùng phát của các bệnh truyền nhiễm nói chung và bệnh tả nói riêng có liên hệ chặt chẽ với các yếu tố như nguồn nước, thực phẩm và khí hậu. Bài báo này đề xuất xây dựng mô hình dự báo bệnh tả trong ngắn hạn dựa trên phương pháp rừng ngẫu nhiên, có xem xét toàn diện ảnh hưởng của các yếu tố khí hậu (nhiệt độ, độ ẩm...) và địa lý (sự lân cận về địa lý, hệ thống sông...) đến số ca mắc tả ở Hà Nội trong giai đoạn 2001-2012. Phân tích thực nghiệm cho thấy dạng "mô hình đầy đủ" có xem xét cả yếu tố khí hậu và địa lý cho kết quả dự báo tốt nhất cho từng quận/huyện của Hà Nội. Các kết quả cũng khẳng định sự lân cận về địa lý và số ca nhiễm bệnh ở các quận/huyện có liên kết mật thiết. Các yếu tố khí hậu có ảnh hưởng theo các mức khác nhau đến số ca nhiễm bệnh, trong đó nhiệt độ và độ ẩm có mức ảnh hưởng lớn nhất và chỉ số dao động Nam có mức ảnh hưởng thấp nhất.

Từ khóa - mô hình dự báo bệnh tả; dự báo bệnh tả ngắn hạn; các nhân tố khí hậu và thời tiết; các trường ngẫu nhiên; các chuỗi thời gian.

1. Giới thiệu

Bệnh tả vẫn là một vấn đề y tế công cộng toàn cầu mặc dù tỷ lệ mắc bệnh và tỷ lệ tử vong đã giảm mạnh trong những năm gần đây [1]. Bệnh tả là một tiêu chảy cấp tính gây ra bởi vi khuẩn *Vibrio Cholera*. Bệnh tả thường được xem xét trong mối quan hệ với nguồn nước bị ô nhiễm và cơ sở hạ tầng vệ sinh kém, đặc biệt là ở các nước có mức thu nhập thấp và trung bình [2], [3]. Cùng với nguồn nước và tình trạng vệ sinh, một số nghiên cứu trước đây đã chứng minh rằng sự biến đổi khí hậu cũng góp phần vào sự phổ biến của vi khuẩn tả [4]. Chẳng hạn, các nghiên cứu ở châu Phi cho thấy sự gia tăng của nhiệt độ và lượng mưa đã làm tăng số ca mắc tả [5], [6]. Hơn nữa, các nghiên cứu ở Bangladesh cũng cho thấy nhiệt độ và số giờ nắng có thể liên quan với sự xuất hiện dịch tả [7]. Trong một báo cáo gần đây, Tổ chức Y tế Thế giới nhấn mạnh các yếu tố khí hậu có vai trò quan trọng trong sự phân bố về không gian và thời gian của các bệnh truyền nhiễm [8], [9]. Vì vậy, việc thiết lập mô hình dự báo dịch tả dựa trên các yếu tố khí hậu là rất cần thiết để có các biện pháp phòng ngừa và can thiệp trong ngắn hạn cũng như dài hạn.

Việt Nam đã trải qua nhiều đợt dịch tả vào thế kỷ XX, đặc biệt là trong những năm 1960 và 1990, trong đó hầu hết các ca mắc bệnh được báo cáo ở các khu vực phía Nam [2], [10]. Tuy nhiên, trong các năm 2007 và 2008, các đợt dịch tả xảy ra ở các tỉnh chủ yếu ở khu vực phía Bắc, trong đó có Hà Nội [2], [10], [11], [12]. Tính đến tháng Tư năm 2008, đã có 3.271 ca mắc bệnh tả theo số liệu báo cáo từ 18 tỉnh [10], [11]. Nhiều nguyên nhân của dịch tả đã được đưa ra thảo luận, ngoài nguồn nước và thực phẩm bị ô nhiễm [2]. Do đó, việc nghiên cứu các mô hình dự báo trong đó có xem xét mối quan hệ giữa các ca mắc tả và các

Abstract - The outbreaks of infectious diseases in general and cholera in particular have a close relationship with factors such as water source, food and climate. This paper proposes building Random Forests-based models for short-term cholera forecast, which evaluate the effect of climate factors (temperature, humidity,...) and geographical factors (locality, river system,...) on the cholera cases in Hanoi city for the period of 2001-2012. Experimental analyses show that "complete model" has the best forecast accuracy for each district in Hanoi. The analysis results also confirm that the geographical locality and the number of cholera cases in Hanoi's districts have close relationships. Climate factors have different effect levels on the number of cholera cases. Particularly, the daily mean temperature and humidity have strongest effect, while southern oscillation index (SOI) has least effect.

Key words - cholera forecast model; short-term cholera forecast; climate and geographical factors; random forests; time series.

yếu tố khác như biến đổi khí hậu là cần thiết để xây dựng chiến lược phù hợp cho việc kiểm soát, giám sát và ngăn chặn sự bùng phát bệnh tả.

Phần tiếp theo của bài báo này được bố cục như sau: Mục 2 phân tích một số nghiên cứu có liên quan, Mục 3 trình bày quá trình xây dựng các mô hình dự báo ngắn hạn, Mục 4 trình bày các thực nghiệm mô hình và đánh giá ảnh hưởng của các yếu tố khí hậu và địa lý, và Mục 5 là phần kết luận.

2. Các nghiên cứu liên quan

Ali và cộng sự [1] nghiên cứu dữ liệu dịch tả ở Matlab, Bangladesh từ 1988 đến 2001 và rút ra kết luận rằng, số lượng các ca bệnh tả trong vùng có liên hệ mật thiết với nhiệt độ trên đất liền và nhiệt độ mặt biển (sea surface temperature - SST) trong khu vực nghiên cứu. Reiner và cộng sự [13] đã xây dựng thành công một mô hình cho phép dự báo số lượng các ca bệnh tả ở Matlab, Bangladesh trước 11 tháng. Các tập dữ liệu được sử dụng bao gồm thời tiết, chỉ số dao động phía Nam (southern oscillation index - SOI) và tình trạng ngập lụt từ năm 1995 đến 2008. Kết quả nghiên cứu này chỉ ra rằng SOI và tình trạng ngập lụt là các yếu tố ảnh hưởng chính đến lượng các ca bệnh tả ở Matlab.

Mở rộng theo hướng này, Xu và cộng sự [14] phân tích ảnh hưởng của khí hậu đến bệnh tả ở Trung Quốc từ năm 2001 đến 2008 và đưa ra kết luận rằng lượng mưa, nhiệt độ và độ cao so với mặt biển (sea surface height - SSH) có ảnh hưởng lớn nhất tới số ca bệnh tả. Khoảng cách tới bờ biển, độ ẩm tương đối và khí áp cũng có ảnh hưởng. Tuy nhiên số giờ nắng và quá trình giảm mức nước sông hầu như không có ảnh hưởng đến số ca bệnh.

Kelly-Hope và cộng sự [2] trong một nghiên cứu về

dịch tả ở Việt Nam đã kết luận có sự liên hệ rõ ràng giữa lượng mưa và sự bùng phát dịch tả với độ trễ 0 tháng trong giai đoạn 1991-2001. Nghiên cứu của Emch và cộng sự [4] chỉ ra các yếu tố ảnh hưởng khả năng nhiễm tả, bao gồm sự tăng nhiệt độ mặt biển và mực nước sông ở Việt Nam.

Kết quả của các nghiên cứu kể trên đều khẳng định các tham số thời tiết như nhiệt độ, độ ẩm, SOI, SST, SSH có liên hệ ở các mức khác nhau đến số lượng các ca bệnh tả ở các vùng nghiên cứu khác nhau. Tuy nhiên, các đợt bùng phát dịch tả ở Hà Nội từ năm 2007 đến 2009 đặt ra sự cần thiết xem xét toàn diện đến các tham số khí hậu và địa lý. Bài báo này đề xuất xây dựng các mô hình dự báo dịch tả trong ngắn hạn có xem xét toàn diện mức độ ảnh hưởng của các yếu tố khí hậu và địa lý đến số ca bệnh tả ở Hà Nội trong giai đoạn 2001-2012.

3. Các mô hình dự báo dịch tả ngắn hạn

3.1. Các tập dữ liệu sử dụng và tiền xử lý

3.1.1. Các tập dữ liệu sử dụng

Để xây dựng các mô hình dự báo bệnh tả trong ngắn hạn cho Hà Nội, chúng tôi sử dụng các tập dữ liệu sau: tập dữ liệu các ca bệnh tả, thời tiết, địa lý, hệ thống sông, hệ thống giao thông và SOI. Bảng 1 cung cấp thông tin chi tiết các tập dữ liệu này.

Bảng 1. Các tập dữ liệu sử dụng trong nghiên cứu

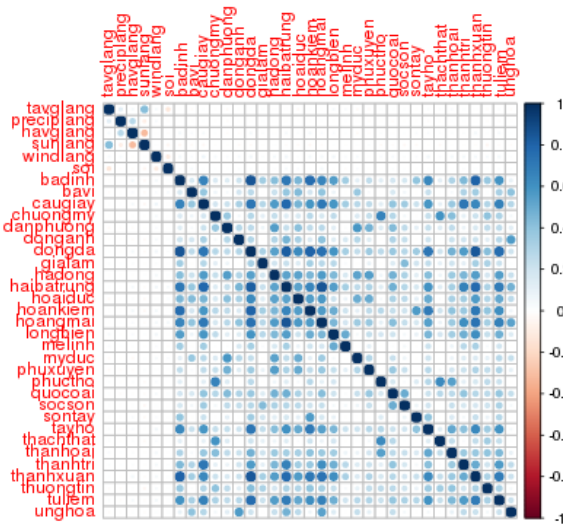
Các tập dữ liệu	Mô tả
Tập dữ liệu địa lý	Chứa bản đồ về các quận/huyện, phường/xã, hệ thống đường giao thông, hệ thống sông và mặt nước theo tỷ lệ 1:50.000. Hà Nội có 29 quận/huyện và 2 quận/huyện được xem là lân cận nếu có chung đường biên giới hành chính.
Tập dữ liệu thời tiết	Chứa dữ liệu theo ngày về: độ ẩm (thấp nhất, cao nhất và trung bình), nhiệt độ không khí (thấp nhất, cao nhất và trung bình), số giờ nắng, tốc độ gió và lượng mưa, cung cấp bởi trạm khí tượng Láng, Hà Nội từ 2001-2012.
Tập dữ liệu SOI	Dữ liệu SOI được thu thập từ một trang web của bang Queensland, Australia [15].
Tập dữ liệu các ca bệnh tả	Chứa dữ liệu về tất cả các ca tả tại Hà Nội từ 1/1/2001 đến 31/12/2012. Thông tin mỗi ca tả gồm họ tên, tuổi, giới tính, ngày nhiễm và địa chỉ (tối thiểu đến cấp phường/xã) của bệnh nhân. Dữ liệu ca tả được tổng hợp theo ngày, tháng và theo từng quận/huyện. Theo đó, các đợt dịch chỉ xảy ra trong các năm 2004 (25 ca), 2007 (1.179 ca), 2008 (2.057 ca), 2009 (890 ca) và 2010 (350 ca).

3.1.2. Tiền xử lý dữ liệu

Do dữ liệu các ca tả phân bố không đồng đều và chỉ phân bố tập trung trong 5 năm, chúng tôi quyết định sử dụng tổng hợp dữ liệu theo ngày cho việc xây dựng mô hình dự báo, trừ dữ liệu địa lý. Điều này giúp tăng số điểm dữ liệu trong giai đoạn nghiên cứu và thuận lợi hơn trong xây dựng mô hình dự báo ngắn hạn.

Các tập dữ liệu thời tiết, SOI và các ca bệnh được tổng hợp theo ngày và trộn thành một tập dữ liệu duy nhất, gọi

tên là FS. Tập FS có 35 biến và 4.383 quan sát, như minh họa trên hình 1. Trong số 35 biến, có 6 biến thời tiết, gồm nhiệt độ không khí, độ ẩm, lượng mưa, số giờ nắng, tốc độ gió và SOI. Các biến còn lại là số ca mắc tả cho 29 quận/huyện của Hà Nội.



Hình 1. Ma trận tương quan của tập dữ liệu FS

3.2. Xây dựng các mô hình dự báo dịch tả ngắn hạn

Để đánh giá ảnh hưởng của các yếu tố địa lý và khí hậu lên số ca mắc tả, chúng tôi coi mỗi quận/huyện của Hà Nội là một đơn vị địa lý và xây dựng 3 mô hình dự báo cho mỗi quận/huyện. Các mô hình dự báo bao gồm mô hình đầy đủ (ký hiệu là DD), mô hình độc lập khí hậu (ký hiệu là DLKH) và mô hình độc lập lân cận không gian địa lý (ký hiệu là DLDL). Bảng 2 cung cấp thông tin chi tiết về các nhóm biến sử dụng trong 3 dạng mô hình kể trên. Mục đích của việc xây dựng 3 mô hình cho mỗi quận/huyện là để lựa chọn được mô hình dự báo tốt nhất cho cho mỗi quận/huyện và đánh giá được mức độ ảnh hưởng của các tham số lân cận không gian địa lý và khí hậu đến độ chính xác của mô hình dự báo. Tất cả các mô hình đều có đầu ra là số ca bệnh tả.

Bảng 2. Mô tả mô hình dự báo với các nhóm biến đầy đủ, độc lập với khí hậu, độc lập lân cận địa lý

Nhóm dự báo	Mô hình		
	DD	DLKH	DLDL
Dữ liệu về khí hậu	Nhiệt độ trung bình		Nhiệt độ trung bình
	Độ ẩm trung bình		Độ ẩm trung bình
	Lượng mưa		Lượng mưa
	Chỉ số SOI		Chỉ số SOI
	Số giờ nắng		Số giờ nắng
Dữ liệu lân cận không gian địa lý	Tốc độ gió		Tốc độ gió
	Số lượng ca bệnh tả trong một quận D	Số lượng ca bệnh tả trong một quận D.	
	Số lượng ca bệnh tả của các quận lân cận quận D	Số lượng ca bệnh tả của các quận lân cận quận D	

Mỗi mô hình có một tham số độ trễ l tính theo ngày. Tham số này có nghĩa là sẽ sử dụng số ca bệnh tại thời điểm hiện tại và l ngày trước đó trong quận/huyện đang xem xét như là một biến dự báo cho mô hình. Điều này cũng có nghĩa là sẽ dự báo số ca bệnh của quận/huyện hiện tại trong l ngày tiếp theo. Ngoài ra, mỗi mô hình cũng sử dụng số ca bệnh trong quá khứ của tất cả các quận/huyện lân cận và dữ liệu khí hậu trong quá khứ như là các biến đầu vào bổ sung trong mô hình.

Để xây dựng các mô hình, chúng tôi sử dụng phương pháp học máy hồi quy Random Forests (RF) để xử lý tập dữ liệu FS là chuỗi thời gian theo phương pháp cửa sổ trượt song hành giữa tập huấn luyện và tập kiểm thử. Random Forests được đánh giá là phù hợp cho xử lý các bài toán chuỗi thời gian [16]. Trong phương pháp cửa sổ trượt, đầu tiên khởi tạo một cửa sổ s_1 tương ứng với tập dữ liệu huấn luyện ban đầu. Với tập dữ liệu kiểm thử lựa chọn cửa sổ s_2 . Chú ý rằng ở mỗi điểm dữ liệu trong tập huấn luyện bao gồm tất cả các biến đầu vào và đầu ra, còn mỗi tập dữ liệu kiểm thử sẽ chỉ bao gồm các biến dự báo. Cửa sổ trượt sẽ trượt dọc theo trục thời gian cho đến khi không còn dữ liệu. Mô hình được xây dựng trong sự chuyển dịch và cải thiện dọc theo trục thời gian. Chọn kích thước các cửa sổ trượt $s_1=s_2=l$ trong tất cả các mô hình. Độ trễ thời gian của mô hình được lựa chọn là $d=3, 7, 14$ hoặc 30 ngày, trong đó cửa sổ trượt có cỡ cố định với cỡ ban đầu là $d=3, 7, 14, 30$. Chuỗi thời gian được sử dụng để kiểm thử tương ứng là $n=3, 7, 14, 30$.

m1	m2	m3	m4	m5	m6	m7	m8
c4	c5	c6	c7	c8	c9	c10	c11
n1	n2	n3	n4	n5	n6	n7	n8
Dữ liệu huấn luyện 1				Dữ liệu kiểm thử 1			
Dữ liệu huấn luyện 2				Dữ liệu kiểm thử 2			
Dữ liệu huấn luyện 3				Dữ liệu kiểm thử 3			

Hình 2. Minh họa việc huấn luyện mô hình hồi quy RF theo phương pháp cửa sổ trượt có độ trễ thời gian

Hình 2 minh họa việc huấn luyện mô hình hồi quy RF theo phương pháp cửa sổ trượt với độ trễ thời gian là 3 ngày, kích cỡ cửa sổ trượt là 3 ngày, và số ngày dự báo trước là 3 ngày. Giá trị các tham số: $n=3, d=3$. Các ô $m1, m2, \dots, m8$ là các biến khí hậu từ ngày 1 đến ngày 8; các ô $c4, c5, \dots, c11$ là các biến ghi nhận các ca mắc bệnh của quận C ở các ngày 4, 5, ..., 11; và $n1, n2, \dots, n8$ là số các ca mắc tại ở các quận lân cận của các ngày 1 đến 8. Thời điểm bắt đầu huấn luyện mô hình là ngày 6. Dữ liệu huấn luyện là tập $\{m1, m2, m3, n1, n2, n3, c4, c5, c6\}$. Dữ liệu kiểm thử là tập $\{m4, m5, m6, n4, n5, n6\}$. Kết quả kiểm thử (dự báo) là tập $\{c7, c8, c9\}$. Quá trình này lặp lại cho các ngày 7, 8, ... Với dữ liệu 4.383 ngày trong giai đoạn nghiên cứu, số lần lặp trong quá trình huấn luyện và kiểm thử là 4.377.

4. Thực nghiệm và đánh giá

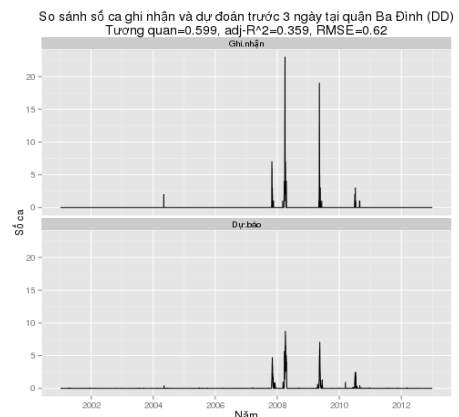
Chúng tôi đã xây dựng 29×3 mô hình hồi quy RF cho 29 quận/huyện sử dụng tập dữ liệu FS mô tả trong mục 3.1. Để đánh giá các mô hình hồi quy, các độ đo thường được sử dụng gồm sai số trung bình quân phương (Root mean squared error - RMSE) và hệ số xác định điều chỉnh

(Adjusted determination coefficient - R^2). Các giá trị RMSE và R^2 được tính cho 29×3 mô hình. Trên cơ sở các thực nghiệm, các phần việc sau được thực hiện: (1) so sánh ảnh hưởng của các yếu tố khí hậu và địa lý đến độ chính xác dự báo của các mô hình, (2) phân tích thống kê để tìm mối quan hệ giữa độ chính xác và khoảng thời gian dự báo, và (3) đánh giá tầm quan trọng của các biến khí hậu trong các mô hình hồi quy RF cho các quận/huyện. Phần tiếp theo trình bày chi tiết các phần việc trên.

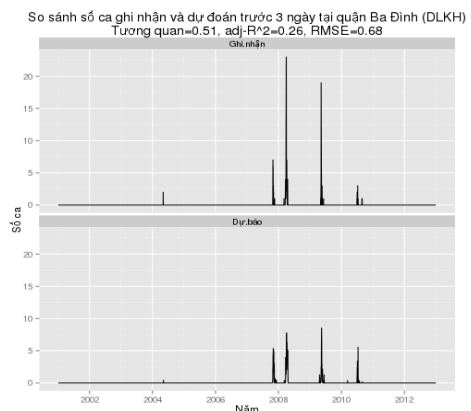
4.1. Ảnh hưởng của các yếu tố khí hậu và địa lý đến độ chính xác dự báo

Để so sánh ảnh hưởng của các yếu tố khí hậu và địa lý đến độ chính xác dự báo, cụ thể là các độ đo RMSE và R^2 , chúng tôi sử dụng phương pháp Tukey [17] với 4 khoảng dự báo trước là 3, 7, 14 và 30 ngày. Các kết quả được biểu diễn trên các hình 3-6. Xét khoảng cách của độ tin cậy và giá trị trung bình của các cặp mô hình DLDL-DD và DLKH-DD, có thể thấy các mô hình đầy đủ (DD) có độ đo R^2 cao nhất cũng là tốt nhất. Các mô hình độc lập địa lý (DLDL) có độ đo R^2 thấp nhất. Như vậy, có thể kết luận số ca mắc tại ở một quận/huyện có liên hệ chặt chẽ với số ca mắc tại ở các quận/huyện lân cận.

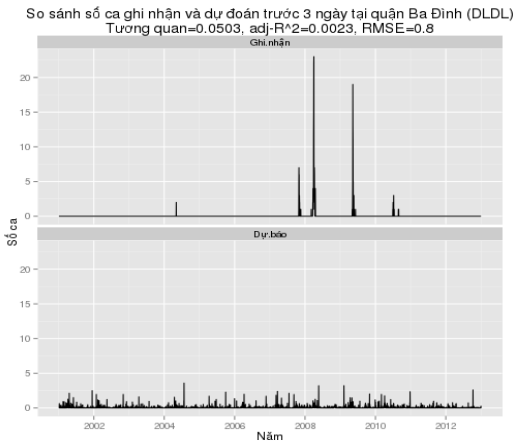
Tuy nhiên, các kết quả so sánh độ đo RMSE không cho thấy bất kỳ sự khác biệt nào trong độ chính xác của các mô hình. Hơn nữa, việc so sánh độ đo RMSE không cho phép chỉ ra mô hình nào tốt hơn. Do vậy, chúng tôi chỉ sử dụng độ đo R^2 để so sánh các mô hình.



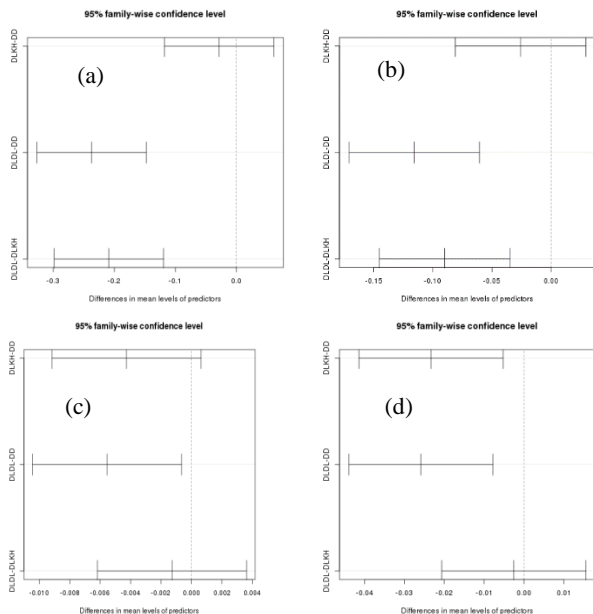
Hình 3. So sánh kết quả dự báo và thực tế với mô hình đầy đủ (DD) cho quận Ba Đình



Hình 4. So sánh kết quả dự báo và thực tế với mô hình độc lập khí hậu (DLKH) cho quận Ba Đình



Hình 5. So sánh kết quả dự báo và thực tế với mô hình đầy đủ (DLDL) cho quận Ba Đình



Hình 6. So sánh ảnh hưởng của nhóm biến khí hậu và nhóm biến địa lý đến độ chính xác của mô hình với độ đo R^2 : (a), (b), (c), (d) lần lượt ứng với khoảng dự báo trước là 3, 7, 14 và 30 ngày

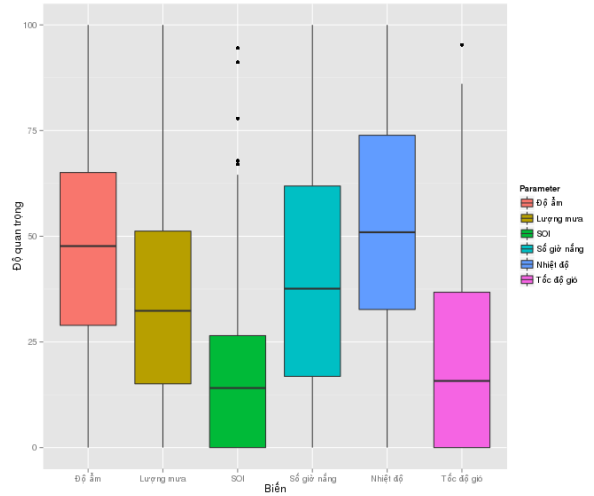
4.2. Mối quan hệ giữa độ chính xác và khoảng thời gian dự báo

Như đã phân tích trong mục 4.1, các mô hình đầy đủ là tốt nhất. Do vậy, chúng tôi sử dụng mô hình đầy đủ để dự báo số ca mắc tả cho 29 quận/huyện của Hà Nội với các khoảng dự báo là 3, 7, 14 và 30 ngày. Sau quá trình dự báo, các kết quả được so sánh với số liệu quan sát và độ đo R^2 được tính toán. Để quan sát sự thay đổi độ chính xác theo khoảng dự báo, các mô hình hồi quy tuyến tính với các tham số vào là số ngày dự báo trước và quận/huyện, còn đầu ra là độ đo R^2 . Kết quả cho thấy, nếu tất cả các tham số khác giữ nguyên và tăng độ dài dự báo lên 1 ngày thì độ đo R^2 giảm 0,0076 với khoảng tin cậy 95% là [-0,0095, -0,0057].

4.3. Tầm quan trọng của các biến khí hậu

Ảnh hưởng của các biến khí hậu được trích xuất từ các mô hình RF đã xây dựng cho các quận/huyện và được biểu

diễn ở hình 7. Theo đó, có thể thấy rằng các tham số nhiệt độ và độ ẩm trung bình ngày là các yếu tố quan trọng nhất, với khoảng 50% độ quan trọng so sánh với các biến khí hậu khác. Số giờ nắng chiếm khoảng 35% độ quan trọng. Các biến có độ quan trọng thấp nhất là tốc độ gió và SOI với độ quan trọng thấp hơn là 20%.



Hình 7. So sánh ảnh hưởng của các biến khí hậu lên mô hình đầy đủ (DD)

5. Kết luận

Chúng tôi đã xây dựng 29×3 mô hình hồi quy RF cho dự báo dịch tả cho từng quận/huyện của thành phố Hà Nội trong giai đoạn 2001 đến 2012. Kết quả so sánh, phân tích cho thấy mô hình đầy đủ cho kết quả dự báo chính xác nhất trong ngắn hạn do có xem xét đến tất cả các yếu tố khí hậu và địa lý.

Các kết quả so sánh, phân tích mức độ ảnh hưởng của các yếu tố địa lý và khí hậu khẳng định rằng sự lân cận về địa lý và số ca bệnh ở các quận/huyện lân cận có mối liên hệ chặt chẽ. Các yếu tố khí hậu cũng có ảnh hưởng theo các mức khác nhau đến số ca bệnh, trong đó nhiệt độ và độ ẩm trung bình ngày có mức ảnh hưởng lớn nhất, trong khi đó tốc độ gió và SOI có mức ảnh hưởng thấp nhất. Nghiên cứu cũng tìm ra rằng, độ chính xác của mô hình dự báo giảm nếu tăng khoảng dự báo, với hệ số R^2 giảm trung bình 0,0076 nếu khoảng dự báo tăng 1 ngày.

Trong tương lai, chúng tôi tiếp tục nghiên cứu, phân tích sâu chi tiết ảnh hưởng của các yếu tố lân cận địa lý, bao gồm hệ thống sông ngòi, mặt nước đến số ca bệnh.

TÀI LIỆU THAM KHẢO

- [1] Ali M, Lopez AL, You YA, et al, The global burden of cholera. Bulletin of the World Health Organization, Mar 1 2012, 90(3):209-218A.
- [2] Kelly-Hope LA, Alonso WJ, Thiem VD, et al, “Temporal trends and climatic factors associated with bacterial enteric diseases in Vietnam, 1991-2001”, Environmental health perspectives, Jan 2008, 116(1):7-12.
- [3] Organization WH. Cholera, Geneva, Switzerland: World Health Organization, 2003.
- [4] Emch M, Feldacker C, Yunus M, et al, “Local Environmental Predictors of Cholera in Bangladesh and Vietnam”, The American journal of tropical medicine and hygiene, May 1, 2008, 78(5):823-832.

- [5] Mendelsohn J, Dawson T, "Climate and cholera in KwaZulu-Natal, South Africa: the role of environmental factors and implications for epidemic preparedness", *International journal of hygiene and environmental health*, Mar 2008, 211(1-2):156-162.
- [6] Reyburn R, Kim DR, Emch M, Khatib A, von Seidlein L, Ali M, "Climate variability and the outbreaks of cholera in Zanzibar, East Africa: a time series analysis", *The American journal of tropical medicine and hygiene*, Jun 2011, 84(6):862-869.
- [7] Islam MS, Sharker MA, Rheman S, et al, "Effects of local climate variability on transmission dynamics of cholera in Matlab, Bangladesh", *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Nov 2009, 103(11):1165-1170.
- [8] Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A, El Nino and health, *Lancet*, Nov 1 2003, 362(9394):1481-1489.
- [9] Organization WH, *Using Climate to Predict Infectious Disease Outbreaks: A Review*, Geneva, Switzerland, 2004.
- [10] Control GTFoC, *Cholera country profile: Vietnam*, Geneva, Switzerland: World Health Organization, 2008.
- [11] Nguyen BM, Lee JH, Cuong NT, et al, "Cholera outbreaks caused by an altered *Vibrio cholerae* O1 El Tor biotype strain producing classical cholera toxin B in Vietnam in 2007 to 2008", *Journal of clinical microbiology*, May 2009, 47(5):1568-1571.
- [12] Organization WH, Outbreak news, *Severe acute watery diarrhoea with cases positive for *Vibrio cholerae*, Viet Nam*, Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. May 2 2008, 83(18):157-158.
- [13] Robert C. Reiner, A. A. King, M. Emch, M. Yunus, A. S. G. Faruque, and M. Pascual, *Highly localized sensitivity to climate forcing drives endemic cholera in a megacity*, *Proc. Natl. Acad. Sci. U. S. A.*, 109, 2033–2036 (2012).
- [14] Min Xu, Chunxiang Cao, Duochun Wang, and Biao Kan, Identifying Environmental Risk Factors of Cholera in a Coastal Area with Geospatial Technologies, *Int. J. Environ. Res. Public Health* 2015, 12, 354-370.
- [15] Daily SOI data set of the Queensland, Australia, available online at <https://www.longpaddock.qld.gov.au/seasonalclimateoutlook/southernoscillationindex/soidatafiles/DailySOI1887-1989Base.txt>
- [16] R. Hyndman, G. Athanasopoulos, *Forecasting: principles and practice*, Otexts, 2013.
- [17] Nguyễn Văn Tuấn (2015), Phân tích phương sai, [ykhoa.net/r/R/Chuong 11. Phan tich phuong sai.pdf](http://ykhoa.net/r/R/Chuong%2011.%20Phan%20tich%20phuong%20sai.pdf), Truy cập 5/2016.

(BBT nhận bài: 18/01/2017, hoàn tất thủ tục phản biện: 22/02/2017)