

# ỨNG DỤNG THÔNG TIN KHÍ HẬU TRONG VIỆC CẢNH BÁO DỊCH BỆNH TIÊU CHẢY VÙNG TÂY BẮC

Nguyễn Hữu Quyền\*, Nguyễn Văn Thắng, Dương Văn Khảm, Lê Thị Phương Mai, Phạm Đức Phúc  
Viện Khoa học Khí tượng Thủy văn và Biến đổi khí hậu

Ngày nhận bài 10/10/2017; ngày chuyển phản biện 11/10/2017; ngày chấp nhận đăng 14/11/2017

**Tóm tắt:** Biến đổi khí hậu là một trong những mối đe dọa lớn nhất đến sức khỏe con người trên toàn cầu trong thế kỷ 21. Vì vậy, việc khai thác sử dụng các thông tin khí hậu để dự báo, cảnh báo sớm các dịch bệnh ở người và giảm thiểu tính dễ bị tổn thương do biến đổi khí hậu là việc làm rất cần thiết, nhưng cho tới nay còn chưa được nghiên cứu đầy đủ. Mục tiêu chính của nghiên cứu này nhằm xây dựng mô hình ứng dụng thông tin khí hậu phục vụ cảnh báo dịch bệnh trên người, đặc biệt là bệnh tiêu chảy ở vùng Tây Bắc. Nghiên cứu đề xuất sử dụng mô hình động thái ARIMAX với các bộ số liệu dịch bệnh và khí hậu theo bước thời gian tháng, thời kỳ 2004-2014 để cảnh báo dịch bệnh tiêu chảy, trong đó số liệu năm 2014 đã được sử dụng để kiểm chứng mô hình. Kết quả nghiên cứu cho thấy mô hình động thái ARIMAX có thể cảnh báo số ca mắc bệnh tiêu chảy khá phù hợp với thực tế.

**Từ khóa:** Bệnh tiêu chảy, thông tin khí hậu, biến đổi khí hậu, mô hình cảnh báo dịch bệnh

## 1. Đặt vấn đề

Theo báo cáo của Ủy ban Liên Chính phủ về Biến đổi Khí hậu (IPCC), sự thay đổi nhiệt độ và môi trường sống đã dẫn đến dịch bệnh gia tăng [4], Tổ chức khí tượng thế giới (WMO) và Tổ chức Y tế thế giới (WHO) đã có kết luận về mối quan hệ giữa khí hậu và một số loại bệnh truyền nhiễm như sốt rét, sốt xuất huyết, tiêu chảy... [12]. Vì vậy, đã có những khuyến cáo sử dụng các thông tin về thời tiết khí hậu trong các hệ thống cảnh báo dịch bệnh để bảo vệ sức khỏe cộng đồng thông qua các biện pháp để thích ứng, giảm nhẹ tác động và sẵn sàng đối phó với các nguy cơ dịch bệnh [4, 8, 12].

Tại Việt Nam, hướng nghiên cứu khí hậu ứng dụng liên quan đến dịch bệnh trên người đang rất được quan tâm trong những năm gần đây, một số kết quả bước đầu đã cho thấy có sự ảnh hưởng của các yếu tố khí hậu đến nguy cơ phát sinh dịch bệnh, đặc biệt là một số nhóm bệnh có mối tương quan khá rõ rệt với các yếu tố khí hậu và dị thường khí hậu ở các bước trễ thời gian khác nhau [1, 5, 6]. Đây là cơ sở khoa học

rất quan trọng trong việc thiết lập mô hình ứng dụng thông tin khí hậu phục vụ cảnh báo nguy cơ phát sinh dịch bệnh.

Bệnh tiêu chảy là một trong những bệnh chịu tác động trực tiếp bởi các yếu tố khí hậu, đặc biệt khi thời tiết nóng ẩm tạo điều kiện thuận lợi để các loại vi khuẩn phát triển và xâm nhập vào cơ thể, góp phần khiến cho dịch bệnh này bùng phát. Đây là bệnh đứng thứ hai trong số những nguyên nhân gây tử vong ở trẻ em dưới 5 tuổi [9, 10, 12, 13]. Bệnh này có tỷ lệ mắc rất cao, đặc biệt là ở các tỉnh vùng Tây Bắc, nơi có nhiều đồng bào dân tộc thiểu số sinh sống, trình độ dân trí thấp, nhiều phong tục tập quán lạc hậu, khả năng tiếp cận dịch vụ y tế của người dân còn hạn chế, do vậy tỷ lệ mắc dịch bệnh thường cao hơn so với các vùng khác [1].

## 2. Phương pháp và số liệu nghiên cứu

### 2.1. Phương pháp nghiên cứu

Nghiên cứu này sử dụng mô hình động thái ARIMAX để xử lý và phân tích các thông tin khí hậu nhằm cảnh báo dịch bệnh tiêu chảy. ARIMAX là mô hình phân tích chuỗi thời gian, nó không chỉ xem xét các chu kỳ tự vận động của chuỗi phụ thuộc, các mối tương tác trong quá trình tự vận động của các nhân tố ảnh hưởng

\*Liên hệ tác giả: Nguyễn Hữu Quyền  
Email: nhquyen13@gmail.com

khác mà nó còn đánh giá được các quy luật sai số trong quá trình mô phỏng để nâng cao độ chính xác của mô hình cảnh báo [2,11]. Cho đến nay cũng đã có một số tác giả sử dụng mô hình phân tích chuỗi thời gian trong đánh giá và mô phỏng tác động của các yếu tố khí hậu đến số ca bệnh [3,5].

### 2.1.1. Cấu trúc thuật toán của mô hình động thái ARIMAX

Giả sử ta có các chuỗi độc lập  $X_{it}$  ( $i = 1, 2, \dots, m$ ;  $t=1, 2, \dots, n$ ) và chuỗi phụ thuộc  $Y_t$  ( $t=1, 2, \dots, n$ ), khi đó mô hình động thái ARIMAX được trình bày dưới dạng tổng quát như sau [3]:

$$Y_t = \mu + \sum_{i=1}^m \frac{U_i^s(B)}{S_i^r(B)} B^{k_i} X_{i,t} + \frac{q(B)}{p(B)} a_t \quad (1)$$

Trong đó:

-  $Y_t$  là giá trị quan trắc ở các bước thời gian  $t$ ;  $\mu$  là hằng số;

-  $B$  là phép toán dịch chuyển lùi theo quy tắc:  $BX_t = X_{t-1}$ ,  $B^k X_t = X_{t-k}$ ;

-  $U_i^s(B) = U_{i0} + U_{i1}B + \dots + U_{is}B^s$ ;  $S_i^r(B) = S_{i0} + S_{i1}B + \dots + S_{is}B^s$  là những trọng số động thái của chuỗi độc lập thứ  $i$ ;

-  $k$  là thời điểm tác động của chuỗi độc lập thứ  $i$  tại thời điểm  $t = k$ ;

-  $p(B) = (1 - p_1B - p_2B^2 - \dots - p_pB^p)$ ;  $q(B) = (1 - q_1B - q_2B^2 - \dots - q_qB^q)$  là phép toán tự hồi quy và trung bình trượt của chuỗi phụ thuộc;

-  $a_t$  là sai số ngẫu nhiên (giữa giá trị thực và giá trị tính toán).

### 2.1.2. Các bước xây dựng mô hình cảnh báo

Theo công thức tổng quát của mô hình động thái ARIMAX, yếu tố cần cảnh báo phụ thuộc vào 3 thành phần sau: 1) Thành phần ảnh hưởng của các chuỗi độc lập đến chuỗi phụ thuộc theo các bước trễ thời gian khác nhau (transfer function); 2) Thành phần tự hồi quy của chuỗi phụ thuộc theo các bước trễ thời gian; 3) Thành phần trung bình trượt của chuỗi sai số theo các bước trễ thời gian. Bài toán cần giải quyết ở đây là chọn ra được thành phần nào, bước trễ thời gian nào có ý nghĩa về mặt thống kê để tham gia vào mô hình cảnh báo. Đây là bài toán khá phức tạp, độ chính xác của mô hình cảnh báo không chỉ phụ thuộc vào các chuỗi độc lập, chuỗi phụ thuộc mà còn phụ thuộc vào việc lựa chọn chính xác các bước thời gian tham gia vào mô hình cảnh báo có thể tóm lược các bước thực hiện chính

như sau:

1) Áp dụng phương pháp phân tích tự tương quan để kiểm định tính ổn định ngẫu nhiên của các chuỗi độc lập và chuỗi phụ thuộc, nếu chuỗi chưa đạt độ ổn định, sẽ thông qua bước sai phân để đưa chuỗi về dạng ổn định ngẫu nhiên [2].

2) Kế thừa phương pháp Box Tao đối với mô hình động thái ARIMAX trong việc nhận dạng các thành phần tự hồi quy của chuỗi yếu tố cảnh báo, chuỗi sai số cảnh báo và các thành phần ảnh hưởng của các chuỗi độc lập đến yếu tố cảnh báo thông qua việc xem xét sự biến đổi các hàm tự tương quan, tương quan từng phần và tương quan chéo [11];

3) Sử dụng phương pháp bình phương tối thiểu trong việc xác định các tham số trong mô hình ARIMAX;

4) Áp dụng các phương pháp kiểm nghiệm giả thiết thống kê trong khí hậu để chọn lựa các tham số có đủ độ tin cậy tham gia trong mô hình động thái ARIMAX;

5) Sử dụng công cụ phần mềm thống kê SAS để tính toán các đặc trưng của chuỗi thời gian và các tham số trong mô hình động thái ARIMAX.

## 2.2. Số liệu sử dụng

Bài báo đã sử dụng các bộ số liệu dịch bệnh và khí hậu theo bước thời gian tháng thời kỳ 2004-2014, số liệu năm 2014 được sử dụng để kiểm chứng mô hình. Các loại số liệu đã được sử dụng cụ thể như sau:

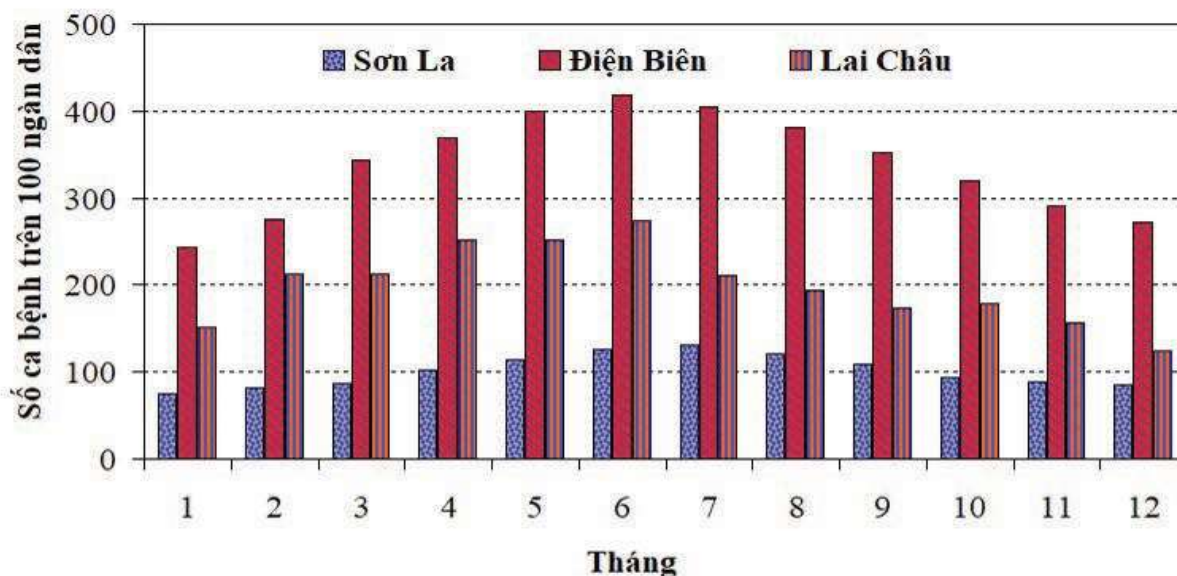
**Chuỗi phụ thuộc:** Số liệu cấp tỉnh về số trường hợp mắc bệnh tiêu chảy trên 100 nghìn dân của 3 tỉnh Sơn La, Điện Biên và Lai Châu, nguồn số liệu này được cập nhật từ Niên giám Thống kê của Bộ Y tế.

**Chuỗi độc lập tham gia tuyển chọn:** Số liệu cấp tỉnh về các yếu tố khí hậu được tính trung bình theo lát cắt thời gian từ các trạm đại diện trong mỗi tỉnh, bao gồm 11 yếu tố: Tổng lượng bốc hơi tháng; tổng lượng mưa tháng; lượng mưa ngày lớn nhất tháng; nhiệt độ trung bình tháng; nhiệt độ tối cao trung bình tháng; nhiệt độ tối thấp trung bình tháng; nhiệt độ tối cao tuyệt đối tháng; nhiệt độ tối thấp tuyệt đối tháng; độ ẩm tương đối trung bình tháng; độ ẩm tương đối tối thấp tháng; tổng số giờ nắng tháng.

Ngoài ra, các chỉ số ENSO như chỉ số dao

động Nam (SOI) và nhiệt độ mặt nước biển trên các vùng NINO1.2, NINO3, NINO4, NINO3.4

cũng được sử dụng làm yếu tố tuyển chọn trong quá trình xây dựng mô hình động thái ARIMAX.



Hình 1. Diễn biến dịch bệnh tiêu chảy trung bình theo tháng thời kỳ 2004 -2014 ở các tỉnh vùng Tây Bắc

### 3. Kết quả

#### 3.1. Diễn biến về số ca bệnh tiêu chảy ở vùng nghiên cứu

Từ Hình 1 có thể nhận thấy dịch bệnh tiêu chảy ở vùng nghiên cứu có thể xảy ra ở tất cả các tháng trong năm. Các tháng cao điểm xảy ra vào mùa hè (từ tháng 4 đến tháng 9) với số ca mắc bệnh tùy thuộc vào từng tỉnh, dao động trong khoảng từ 100 đến 400 ca bệnh trên 100 nghìn dân.

#### 3.2. Bộ tham số trong mô hình ARIMAX

Trên cơ sở các bước thực hiện được nêu trong mục 2.1.2, xác định được bộ tham số hồi quy phù hợp nhất đối với bài toán cảnh báo dịch bệnh tiêu chảy cho từng tỉnh. Với mục đích cảnh báo dịch bệnh trước một tháng, nên các tham số hồi quy chỉ được xem xét khi thỏa mãn tiêu chuẩn thống kê và có bước trễ thời gian  $\geq 1$ . Kết quả được trình bày trong Bảng 1.

Bảng 1 cho thấy các tham số hồi quy được chọn đều đạt mức ý nghĩa thống kê với  $p$  value  $< 0,05$ . Các thành phần AR và MA đều có xuất hiện trong các mô hình ARIMAX đối với cả 3 tỉnh, điều này cho thấy chuỗi thời gian về số ca bệnh tiêu chảy ở vùng nghiên cứu đều có tính mùa rõ rệt.

Mặc dù các biến khí hậu được chọn trong mô hình ARIMAX đối với 3 tỉnh là khác nhau (cụ

thể NINO3.4 được chọn đối với tỉnh Sơn La,  $U_{tb}$  được chọn đối với tỉnh Điện Biên và  $Tmax_{tb}$  được chọn đối với tỉnh Lai Châu), nhưng các biến được chọn này đều có liên quan với chế độ nhiệt ẩm ở vùng nghiên cứu; chỉ số NINO3.4 thường có quan hệ với nhiệt độ và lượng mưa ở Việt Nam với bước trễ thời gian khoảng từ 3-6 tháng [7]; các biến  $Tmax_{tb}$  và  $U_{tb}$  cũng đã được các nghiên cứu trước đây đánh giá là có ảnh hưởng đến dịch bệnh tiêu chảy [5, 6].

#### 3.3. Khả năng cảnh báo của mô hình ARIMAX

Hiện nay, có nhiều chỉ tiêu đánh giá khả năng mô phỏng của mô hình, mỗi chỉ tiêu đánh giá có những mặt mạnh riêng, do vậy tùy thuộc vào mục tiêu chính cần đánh giá và loại mô hình được đánh giá, sẽ chọn loại chỉ tiêu phù hợp. Trong bài báo, ngoài những nhận xét mang tính định tính, so sánh, ước lượng giữa sản phẩm của mô hình và số liệu thực tế thì các chỉ số thống kê như sai số quân phương RMSE (Root Mean Square Error) và điểm kỹ năng MSSS (Mean Square Skill Score) được áp dụng để đánh giá khả năng mô phỏng của các mô hình. Chi tiết về thuật toán và ý nghĩa khoa học của các chỉ số thống kê này được trình bày trong [7].

Trong quá trình xây dựng mô hình cảnh báo, cần phải qua các bước sai phân và chọn các bước

Bảng 1. Các đặc trưng thống kê trong mô hình động thái ARIMAX các tỉnh vùng Tây Bắc

Biến số	Bước trễ thời gian (tháng)	Hệ số hồi quy	Sai số chuẩn	Trị thống kê (t)	Giá trị p
Mô hình động thái ARIMAX đối với tỉnh Sơn La					
MA	1	-0,626	0,074	-8,45	< 0,0001
MA	4	-0,370	0,090	-4,13	<0,0001
AR	12	0,557	0,080	6,95	<0,0001
NINO3.4	5	12,865	3,214	4,00	0,0001
Mô hình động thái ARIMAX đối với tỉnh Điện Biên					
MA	12	0,545	0,084	6,50	<0,0001
AR	3	-0,228	0,091	-2,50	0,0139
U <sub>tb</sub>	2	3,242	1,201	2,70	0,0080
Mô hình động thái ARIMAX đối với tỉnh Lai Châu					
MA	1	0,544	0,080	6,79	<0,0001
AR	12	-0,241	0,090	-2,66	0,0088
Tmax <sub>tb</sub>	1	4,314	2,481	1,74	0,0848
<p><i>Trong đó:</i></p> <ul style="list-style-type: none"> <li>- AR và MA là các thành phần tự hồi quy và trung bình trượt của chuỗi dịch bệnh</li> <li>- NINO3.4 là chỉ số nhiệt độ bề mặt nước biển ở vùng NINO3.4</li> <li>- U<sub>tb</sub> là độ ẩm không khí trung bình tháng;</li> <li>- Tmax<sub>tb</sub> là nhiệt độ không khí cao nhất trung bình tháng</li> </ul>					

trễ về thời gian nên các chuỗi mô phỏng từ mô hình bị ngắn lại, vì vậy thời kỳ đánh giá ở đây được thực hiện từ tháng 1/2006. Hình 2 trình bày diễn biến về số ca bệnh thực tế và mô phỏng thời kỳ 2006-2013 đối với chuỗi phụ thuộc.

Từ Hình 2 có thể nhận thấy về mặt định tính, phần lớn các kết quả mô phỏng có sự trùng pha tốt so với số ca bệnh thực tế, đặc biệt là các thời kỳ cao điểm xảy ra dịch bệnh hầu hết các mô hình đều cho kết quả mô phỏng khá phù hợp.

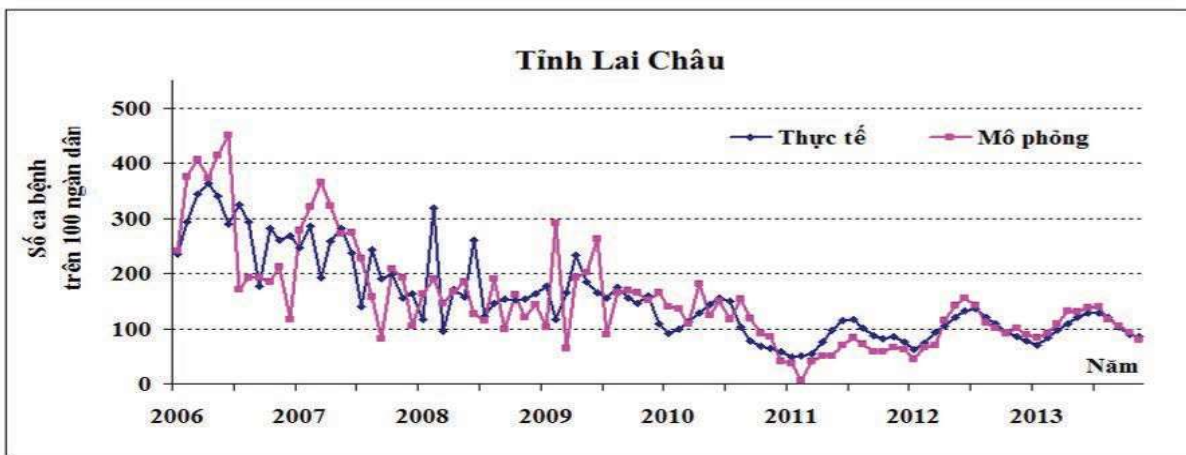
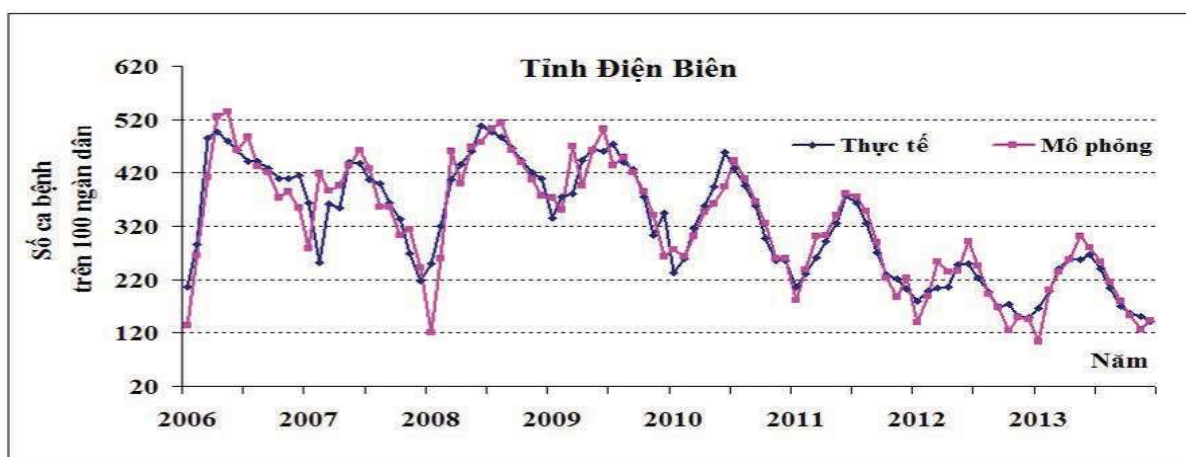
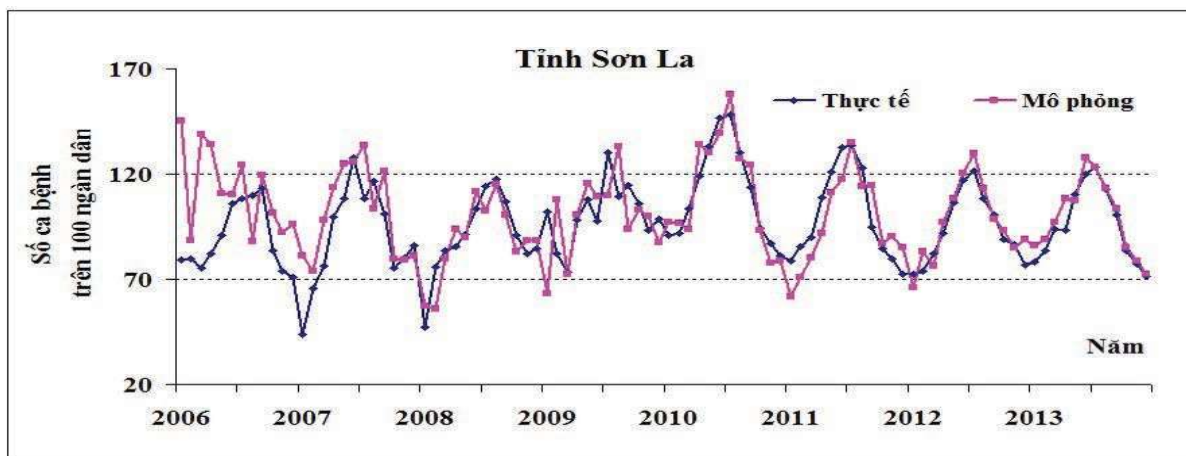
Về mặt định lượng, thông qua các chỉ số thống kê trong Bảng 2 nhận thấy điểm kỹ năng MSSS (Mean Square Skill Score) của tất cả các mô hình đều có giá trị >0, dao động trong khoảng từ 0,22 đến 0,69 đối với chuỗi phụ thuộc và từ 0,05 đến 0,57 đối với chuỗi số liệu độc lập (trường hợp lý tưởng chỉ số này bằng 1; và < 0 được xem là cảnh báo không có ý nghĩa). Vì vậy có thể sử dụng các mô hình này trong việc cảnh báo dịch bệnh tiêu chảy ở vùng nghiên cứu.

#### 4. Kết luận

Trên cơ sở ứng dụng mô hình động thái ARIMAX để phân tích các chuỗi số liệu về số ca bệnh tiêu chảy (trên 100 nghìn dân), các yếu tố khí hậu và các chỉ số ENSO, bài báo đã bước đầu thử nghiệm dự báo, cảnh báo một cách có hiệu quả để tiến tới xây dựng được mô hình cảnh báo dịch bệnh tiêu chảy phù hợp ở 3 tỉnh vùng Tây Bắc. Các kết quả mô phỏng có sự trùng pha khá tốt so với số ca bệnh thực tế. Chỉ số đánh giá sai số MSSS trên số liệu độc lập thể hiện khá tốt, MSSS dao động trong khoảng từ 0,05 đến 0,57, điều này chứng tỏ khả năng mô phỏng của các mô hình này là có thể chấp nhận được.

Với tính ưu việt trong quá trình lựa chọn nhân tố cảnh báo và khả năng xử lý sai số, chúng tôi hi vọng trong tương lai mô hình động thái ARIMAX sẽ được sử dụng phổ biến hơn trong nghiên cứu cảnh báo dịch bệnh ở Việt Nam.





Hình 2. Kết quả thực tế và mô phỏng số ca bệnh tiêu chảy trên 100 nghìn dân vùng Tây Bắc

Bảng 2. Chỉ tiêu đánh giá khả năng mô phỏng của mô hình ARIMAX

Tên tỉnh	Trung bình tháng (Số ca bệnh trên 100 nghìn dân)	Sai số quân phương (RMSE)	Điểm kỹ năng (MSSS)
Đối với chuỗi số liệu phụ thuộc (2006 -2013)			
Sơn La	96,6	16,4	0,22
Điện Biên	306,3	38,2	0,69
Lai Châu	149,9	39,6	0,30
Đối với chuỗi số liệu độc lập (năm 2014)			
Sơn La	92,8	25,6	0,05
Điện Biên	156,4	34,8	0,57
Lai Châu	125,3	43,5	0,28

#### Tài liệu tham khảo

1. Bộ Y Tế (2014), *ATLAS các bệnh truyền nhiễm tại Việt Nam giai đoạn 2000-2011*.
2. Box G.E.P., Jenkins, G.M., and Reissel, G.C., (1994), *Time Series Analysis Forecasting and Control. 3rd edition*. Prentice Hall.
3. Chun-Yu Chuang et al., (2010), *Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996-2007)*, *Sci Total Environ*, 409 (1), 43-51.
4. Intergovernmental Panel on Climate Change (IPCC) (2001), *Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge.
5. Lê Thị Phương Mai và nnk (2015), *Nghiên cứu ảnh hưởng của biến đổi khí hậu đến sức khỏe một số cộng đồng dễ bị tổn thương ở Việt Nam và giải pháp ứng phó*, Đề tài KHCN-BĐKH.47.
6. Nguyễn Hữu Quyền và nnk (2017), "Tác động của nhiệt độ và lượng mưa đến dịch bệnh tiêu chảy ở một số huyện vùng Tây Bắc", *Tạp chí Khí tượng Thủy văn*, số 677, tr.40-45
7. Nguyễn Văn Thắng và nnk (2007), *Nghiên cứu và xây dựng công nghệ dự báo và cảnh báo sớm hạn hán ở Việt Nam*, Báo cáo tổng kết đề tài NCKH cấp Bộ.
8. Nguyễn Văn Thắng và nnk (2010), *Nghiên cứu ảnh hưởng của biến đổi khí hậu đến các điều kiện tự nhiên, tài nguyên thiên nhiên và đề xuất các giải pháp chiến lược phòng tránh, giảm nhẹ và thích nghi, phục vụ phát triển bền vững kinh tế - xã hội ở Việt Nam*, Đề tài KC08.13/06-10.
9. Reena Singh et al., (2001), *The Influence of Climate Variation and Disease and Change on Diarrheal Disease in the Pacific Islands*, *Environmental Health Perspectives* 109, no. 2, 155-159.
10. Sushenjit Bandyopadhyaya et al., (2012), *The impact of rainfall and temperature variation on diarrheal prevalence in SubSaharan Africa*, *Applied Geography* 33.
11. Tim Arnold (2010), *SAS/ETS9.22User's Guide*. SAS Institute Inc, ISBN 978-1-60764-543-6.
12. WHO/WMO (2012), *Atlas of health and climate*, World Health Organization/World Meteorological Organization, Geneva, Switzerland.
13. Yoon Ling Cheong et al., (2012), *Assessing Weather Effects on Dengue Disease in Malaysia*, *Int. J. Environ. Res. Public Health* 2013, 10, 6319-6334.

## APPLYING CLIMATE INFORMATION FOR DIARRHEAL DISEASE WARNING IN THE NORTHWEST REGION OF VIET NAM

Nguyen Van Thang, Duong Van Kham, Nguyen Huu Quyen,  
Le Thi Phuong Mai, Pham Duc Phuc

*Viet Nam Institute of Meteorology, Hydrology and Climate Change*

**Abstract:** *Climate change is one of the most serious threats to global human health in the 21<sup>st</sup> century. Therefore, using climate information in epidemic forecasting and early warning systems and reducing vulnerability to climate change is very important but so far this was not well studied. The main objective of this study is to apply and further develop a model using climate information for epidemic warning, especially for diarrheal disease, in the Northwest region of Viet Nam. The ARIMAX dynamic model has been applied using monthly climate and morbidity data sets of period 2004-2014 with 2014 data used for model calibration. The results have showed that ARIMAX dynamic model can simulate diarrheal disease cases very closely to the observation data, therefore can provide a good basis for diarrheal disease warning system.*

**Keywords:** *Diarrheal disease, climate information, climate change, epidemic disease warning model.*