

BÀI BÁO NGHIÊN CỨU GỐC

**Đánh giá mối liên quan giữa ô nhiễm không khí và sức khỏe tại Việt Nam:
Kỹ thuật phân tích chuỗi thời gian**

Nguyễn Thùy Linh^{1*}, Lê Tự Hoàng¹, Nguyễn Thị Kim Ngân¹, Nguyễn Thị Bích Liên²,
Nguyễn Thị Trang Nhung¹, Trần Minh Điền²

TÓM TẮT

Mục tiêu: Phân tích chuỗi thời gian được sử dụng phổ biến trong các nghiên cứu dịch tễ học môi trường; đặc biệt trong việc đánh giá tác động tức thời (short-term effect) giữa ô nhiễm không khí bên ngoài và tình trạng sức khỏe.

Phương pháp nghiên cứu: Nghiên cứu sử dụng số liệu gồm biến phụ thuộc và độc lập được ghi nhận theo ngày (ví dụ như số ca nhập viện hằng ngày và nồng độ chất ô nhiễm hằng ngày), để đánh giá sự thay đổi của biến độc lập lên biến phụ thuộc trong một thời gian ngắn.

Kết quả: Trong bài báo này, chúng tôi mô tả các bước thực hiện phân tích chuỗi thời gian và những điểm cần lưu ý khi phân tích mối tác động ngắn hạn giữa ô nhiễm không khí bên ngoài và số ca nhập viện. Nghiên cứu sử dụng số liệu trong nghiên cứu đánh giá tác động của ô nhiễm không khí bên ngoài lên sức khỏe trẻ em để làm ví dụ minh họa. Chúng tôi cũng trình bày các câu lệnh dùng để xây dựng mô hình trên phần mềm R và cách phiên giải số liệu ở dạng cơ bản.

Kết luận và khuyến nghị: Cần mô hình hóa tính mùa và tác động dài hạn, mối quan hệ tính tự tương quan và xử lý được các tác động của biến nhiễu. Mô hình này có thể tận dụng được các số liệu ghi nhận hằng ngày để đánh giá những tác động theo thời gian hoặc can thiệp.

Từ khóa: phân tích chuỗi thời gian, ô nhiễm không khí, phân tích tác động ngắn hạn.

ĐẶT VẤN ĐỀ

Phân tích chuỗi thời gian (time series) được xem như là một trong những phương pháp hiệu quả và mạnh trong các nghiên cứu dịch tễ học môi trường. Phương pháp này đánh giá mối liên quan bằng cách xây dựng mô hình hồi quy mô tả sự thay đổi của biến đầu ra được quan tâm theo thời gian. Một trong những ứng dụng của phân tích chuỗi thời gian là dùng để mô tả mối liên quan giữa ô nhiễm không khí bên ngoài và sức khỏe. Thông thường, những nghiên cứu đánh giá tác động

ngắn hạn thường đánh giá mối liên quan giữa ô nhiễm không khí và số ca nhập viện hoặc tử vong trong thời gian tối đa 7 ngày sau phơi nhiễm. Ví dụ, năm 2019, trong một phân tích của Tapia, Steenland (1) tại thành phố Lima, Peru cho thấy, khi nồng độ $PM_{2.5}$ tăng thêm khoảng $6,1\mu g/m^3$ thì tỷ lệ nhập viện cấp cứu do bệnh đường hô hấp tăng thêm 4% (95%CI: 0 – 5%), đột quỵ tăng thêm 10% (95%CI: 3 – 18%) và thiếu máu cục bộ cơ tim tăng 11%. Tại Việt Nam, cũng có một số bài báo sử dụng phương pháp phân tích chuỗi thời gian để đánh giá mối liên quan giữa sức khỏe và



*Địa chỉ liên hệ: Nguyễn Thùy Linh

Email: ntl@huph.edu.vn

¹ Trường Đại học Y tế Công cộng

² Bệnh viện Nhi Trung ương

Ngày nhận bài: 18/6/2020

Ngày phản biện: 30/7/2020

Ngày đăng bài: 29/12/2020

ô nhiễm không khí. Ví dụ, nghiên cứu năm 2012 đánh giá tác động của ô nhiễm không khí lên tình trạng nhập viện do nhiễm khuẩn cấp đường hô hấp (NKHH) dưới ở trẻ em dưới 5 tuổi tại bệnh viện Nhi đồng 1 và 2, thành phố Hồ Chí Minh trong vòng 3 năm 2003, 2004 và 2005 (2). Kết quả cho thấy có mối liên quan giữa NKHH dưới và ô nhiễm không khí do NO_2 và PM_{10} trong mùa khô. Tương tự, nghiên cứu của Nhung, Schindler (3) sử dụng kỹ thuật phân tích chuỗi thời gian để đánh giá tác động ngắn hạn của ô nhiễm không khí lên tỷ lệ nhập viện ở trẻ em 0-17 tuổi tại Hà Nội trong giai đoạn 2007-2014; kết quả cho thấy nồng độ trung bình NO_2 trong 7 ngày tăng một khoảng tứ phân vị ($21,9 \mu\text{g}/\text{m}^3$) thì tỷ lệ nhập viện do viêm phổi tăng 6,1% (95%CI: 2,5%-9,8%).

Trong bài báo này chúng tôi hướng dẫn các bước thực hiện phân tích chuỗi thời gian để đánh giá tác động tức thời của ô nhiễm không khí. Chúng tôi hy vọng sẽ cung cấp các bước cơ bản để người đọc có thể thực hiện được phép phân tích và hiểu được một số khái niệm khi xây dựng mô hình.

KẾT QUẢ

Bài báo này sử dụng số liệu là số ca nhập viện do các bệnh đường hô hấp tại bệnh viện Nhi Trung Ương từ năm 2007 đến năm 2016. Bệnh hô hấp được định nghĩa là những trẻ nhập viện có mã phân loại bệnh tật quốc tế (ICD)10 là J00-J99. Nồng độ ô nhiễm trung bình ngày do PM_{10} (bụi có đường kính nhỏ hơn $10\mu\text{m}$) tại Hà Nội được trích xuất từ hai trạm quan trắc là trạm 556 Nguyễn Văn Cừ và trạm Láng Hạ. Câu hỏi nghiên cứu của phân tích này là ***“Liệu có mối liên quan giữa sự biến thiên hàng ngày của nồng độ PM_{10} (biến độc lập) và số ca nhập viện do bệnh hô hấp ở trẻ em Hà Nội (biến phụ thuộc) không?”***. Trong nghiên cứu này trẻ em Hà Nội được định nghĩa là những trẻ em được tự báo cáo trong hồ sơ bệnh án là sinh sống tại Hà Nội tại thời điểm nhập viện. Trong nghiên cứu cũng sử dụng các chỉ số: trung bình nhiệt độ, độ ẩm và tốc độ gió tại bốn trạm quan trắc khí tượng: Hà Đông, Láng Hạ, Ba Vì và Sơn Tây. Đây là những chỉ số đóng vai trò là những biến nhiễu tiềm tàng của mối quan hệ giữa bệnh và phơi nhiễm.

Bảng 1. Bảng định nghĩa biến số

Tên biến	Định nghĩa		
nres	Tổng số ca nhập viện do bệnh hô hấp theo từng ngày		
dow	Thứ tự ngày trong tuần	0: Chủ nhật 1: Thứ 2 2: Thứ 3 3: Thứ 4	4: Thứ 5 5: Thứ 6 6: Thứ 7
weekend	Ngày cuối tuần	0: Thứ 7 hoặc chủ nhật 1: Các ngày còn lại	
holiday	Ngày lễ	0: Không 1: Có	
dmpm10	Trung bình nồng độ bụi có kích thước nhỏ hơn 10μm theo từng ngày		

dmwisp	Trung bình tốc độ gió theo từng ngày (m/s)
dmtemp	Trung bình nhiệt độ trong từng ngày (°C)
dmhum	Trung bình độ ẩm trong ngày (%)
season	Mùa nóng/lạnh

Có một số điểm cần lưu ý trước khi tiến hành phân tích chuỗi thời gian:

a. “Chuỗi thời gian” đơn giản là một chuỗi các bản ghi về thời gian trong các khoảng thời gian xác định. Trong bộ số liệu này, có bốn chuỗi số liệu theo thời gian: PM₁₀, nhiệt độ, độ ẩm, tốc độ gió và số trường hợp nhập viện do các bệnh hô hấp ở trẻ em Hà Nội theo ngày. Tuy nhiên, phân tích chuỗi thời gian có thể phân tích theo ngày, tuần, tháng, hoặc giờ.

b. Đơn vị phân tích là đơn vị thời gian (sự biến thiên theo ngày về nồng độ ô nhiễm không khí và số ca nhập viện do hô hấp ở trẻ em Hà Nội), không phải từng đối tượng nghiên cứu riêng lẻ. Đây là một điểm quan trọng khi chúng ta cần nhắc những yếu tố nhiễu tiềm

tàng để đưa vào phân tích. Có các biến nhiễu cũng là biến thay đổi theo thời gian như nhiệt độ, độ ẩm. Những biến cá nhân như giới tính, tình trạng sức khỏe hay hành vi hút thuốc lá không thay đổi hằng ngày nên không phải là biến nhiễu trong trường hợp này.

c. Biến phụ thuộc là biến có đo lường dạng số đếm, đây là cấu trúc số liệu thường gặp trong phân tích chuỗi thời gian, điển hình như là số ca nhập viện theo ngày hoặc số người mắc sốt xuất huyết dengue theo tháng. Trong phân tích này, chúng tôi quan tâm tới mô hình hóa sự biến thiên về số trường hợp mắc bệnh theo từng ngày.

Số liệu được định dạng theo “chuỗi thời gian” như Bảng 2

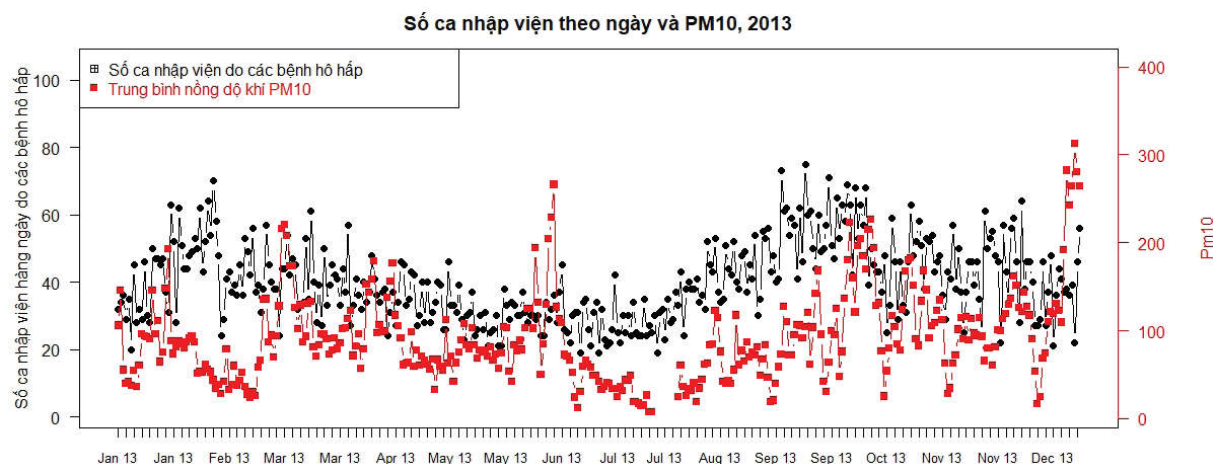
Bảng 2. Bộ số liệu mẫu với các chỉ số môi trường và số trường hợp nhập viện theo ngày do các bệnh đường hô hấp tại bệnh viện Nhi Trung ương, 2007 - 2016

Ngày	Số ca nhập viện (nres)	Thứ tự ngày trong tuần (dow)	Ngày cuối tuần (weekend)	Ngày lễ (holiday)	PM10* (dmpm10)	Tốc độ gió (dmwisp)	Nhiệt độ (dmtemp)	Độ ẩm (dmhum)	Mùa (season)
1-Jan-07	21	1	0	1	253,8	0,591304	18,1	89,75	lạnh
2-Jan-07	27	2	0	0	208,2083	0,591667	20,475	87,75	lạnh
3-Jan-07	14	3	0	0	85,08334	1,125	22,025	84,75	lạnh
4-Jan-07	16	4	0	0	54,47619	2,095652	16,225	75,25	lạnh
5-Jan-07	19	5	0	0	74,91666	1,5125	14,975	77,25	lạnh
6-Jan-07	16	6	1	0	93,20834	1,7625	15,65	69,25	lạnh
7-Jan-07	12	0	1	0	92,375	1,166667	15,35	51,5	lạnh
8-Jan-07	22	1	0	0	143,0417	1,495833	15,4	53,75	lạnh
9-Jan-07	22	2	0	0	143,7391	1,070833	14,75	62,25	lạnh
10-Jan-07	16	3	0	0	179,875	0,858333	14,5	69,75	lạnh

Bước 1: Mô tả bộ số liệu- mô tả đối tượng nghiên cứu

Trong bộ số liệu này, đầu tiên chúng ta sẽ xem xét mối quan hệ giữa số trường hợp nhập viện theo ngày và nồng độ chất gây ô nhiễm không

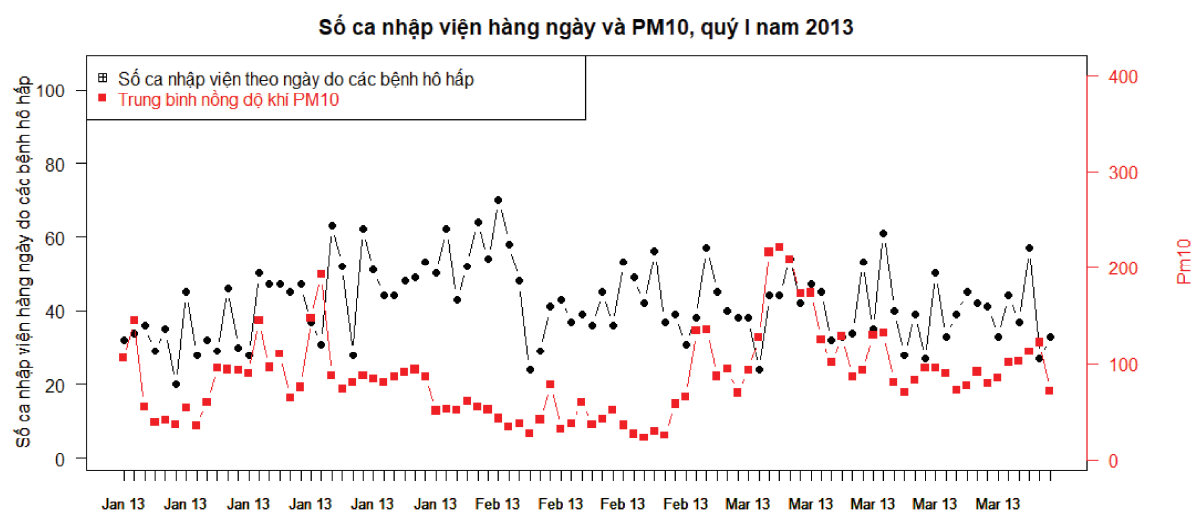
khí. Hình 1 biểu diễn mối liên quan giữa khí PM_{10} và số lượng nhập viện do bệnh đường hô hấp năm từ năm 2013 theo thời gian. Biểu đồ này cho thấy tại các thời điểm nồng độ khí PM_{10} tăng thì số ca nhập viện do các bệnh đường hô hấp cũng tăng lên.



Hình 1. Mối liên quan giữa số ca nhập viện hàng ngày do các bệnh đường hô hấp tại Hà Nội và nồng độ khí PM_{10} năm 2013

Thêm vào đó, số ca nhập viện có xu hướng tương tự nhau ở những khoảng thời gian gần nhau. Ví dụ tại Hình 2, trong những ngày

gần nhau giữa tháng 1 và đầu tháng 3, số ca nhập viện có xu hướng tăng tương tự nhau (Hình 2).



Hình 2. Mối liên quan giữa số ca nhập viện hàng ngày do các bệnh đường hô hấp tại Hà Nội và nồng độ khí PM_{10} , quý I/2013

Một số chỉ số mô tả khác cũng được tóm tắt trong giai đoạn khám phá bộ số liệu và định hướng chiến lược phân tích sau này, bao gồm

số ca nhập viện, các chất ô nhiễm không khí và yếu tố khí hậu.

Bảng 3. Mô tả số ca nhập viện do bệnh hô hấp, nồng độ ô nhiễm không khí và các yếu tố khí hậu tại Hà Nội được ghi nhận trong giai đoạn 2007 - 2016

	n	Trung bình	Độ lệch chuẩn	Giá trị Nhỏ nhất	Giá trị Lớn nhất	Khoảng phân vị (IQR)
Số ca nhập viện	3.651	31,9	12,4	1	105	16,0
PM ₁₀ (µg/m ³)	3.350	82,30	56,73	6,12	403,48	61,49
Nhiệt độ (°C)	3.653	24,07	5,39	6,9	34,88	8,38
Độ ẩm (%)	3.653	82,09	7,81	50,25	98	9,75
Tốc độ gió (m/s)	3.653	1,31	0,55	0,16	4,46	0,69

Bảng 3 cho thấy có tất cả 3.651 ca nhập viện do bệnh hô hấp ở trẻ em Hà Nội trong giai đoạn 2007 – 2016 với trung bình là $31,9 \pm 12,4$ ca. Như vậy, có thể thấy rằng phương sai số ca nhập viện lớn hơn nhiều so với giá trị trung bình. Trong khoảng thời gian này, nồng độ PM₁₀ trung bình ngày là $82,30 \pm 56,73$ µg/m³. Liên quan yếu tố thời tiết, trung bình ngày của nhiệt độ và độ ẩm lần lượt là $24,07 \pm 5,39$ °C và $82,09 \pm 7,81$ %.

Bước 2: Xây dựng mô hình

Đầu tiên chúng tôi trình bày một số điểm cần lưu ý trước khi xây dựng mô hình:

Mục tiêu của các mô hình hồi quy

Mục tiêu chính của việc xây dựng các mô hình hồi quy là đánh giá liệu sự thay đổi số lượng bệnh nhân trong một thời gian ngắn có thể giải thích thông qua sự thay đổi của biến đầu ra chính. Trong bộ số liệu này, chúng tôi đánh giá liệu sự khác nhau số trường hợp nhập viện do bệnh hô hấp ở trẻ em giữa các ngày có thể được giải thích thông qua sự thay đổi về nồng độ PM₁₀. Trong các mô hình hồi quy cũng kiểm soát các yếu tố nhiễu tiềm tàng như thời

tiết và các yếu tố khác có ảnh hưởng lên số ca nhập viện cũng thay đổi theo thời gian.

Kiểm tra các giả định

Biến đầu ra ở đây là biến có đo lường dạng số đếm (số lượng nhập viện mỗi ngày). Một trong những phương pháp hồi quy thường được cân nhắc lựa chọn để phân tích cho biến đầu ra dạng này là hồi quy Poisson. Tuy nhiên, cần cân nhắc một số điểm nổi bật trong dữ liệu chuỗi thời gian:

a. Trong bộ số liệu ví dụ, xu hướng dài hạn (ví dụ: mùa lạnh/nóng) cũng có khả năng đóng vai trò quan trọng trong bộ số liệu này. Tuy nhiên, trong ví dụ này, chúng tôi quan tâm tới mối liên quan ngắn hạn, với mục tiêu loại bỏ hay kiểm soát tác động của xu hướng dài hạn này), và để xem liệu sự biến thiên theo ngày của nồng độ khí gây ô nhiễm không khí có thể giải thích được sự thay đổi ngắn hạn của số ca nhập viện không?

b. Các quan sát không độc lập, tính tự tương quan (autocorrelation): những quan sát xảy ra trong khoảng thời gian gần nhau thường tương tự nhau hơn so với những quan sát ở

những thời điểm cách xa nhau. Tuy nhiên, tính tự tương quan không thường xuyên xảy ra trong nội tại giữa một loạt các đầu ra (số ca nhập viện), nhưng thường là do tính tự tương quan giữa các biến phơi nhiễm là những biến dự báo cho biến đầu ra.

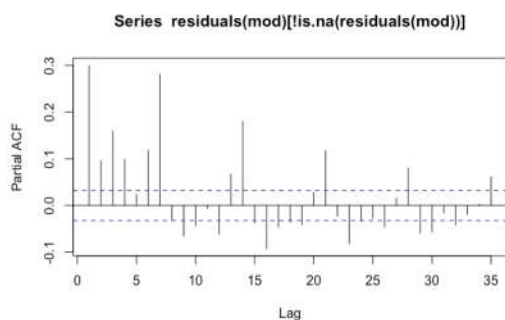
c. Số liệu có xu hướng biến thiên quá mức (overdispersion), có nghĩa là phương sai của số kết quả đầu ra cao hơn giá trị ước tính theo phân bố Poisson (trong đó phương sai bằng giá trị trung bình là giả định của phân bố Poisson).

Sau đây chúng tôi sẽ giới thiệu những phương

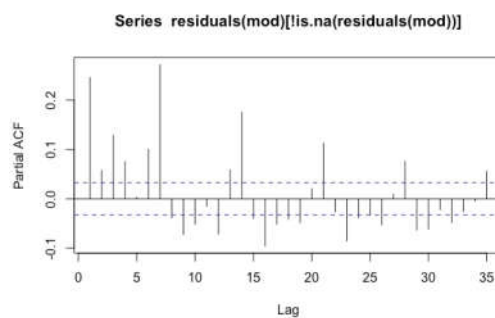
pháp để xử lý những vấn đề trên trong các giai đoạn tiến hành xây dựng mô hình:

1. Kiểm soát tính tự tương quan (autocorrelation)

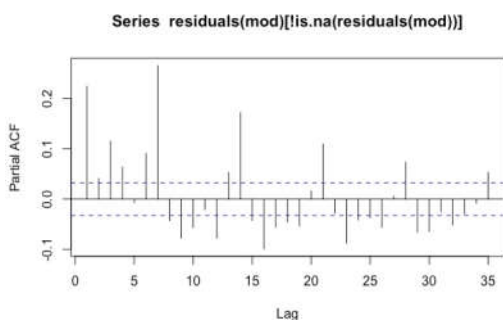
Trong phạm vi bộ số liệu tại Hà Nội, do các quan sát không độc lập và những quan sát xảy ra trong khoảng thời gian gần nhau có xu hướng tương tự nhau. Cụ thể, ngày hôm nay số ca nhập viện lớn thì những ngày sau đó (trong những ngày gần nhau thì số ca nhập viện sẽ cao giống nhau) (Hình 1). Chúng ta có thể kiểm tra tính tự tương quan bằng biểu đồ AFC với các bậc tự do khác nhau (Hình 3).



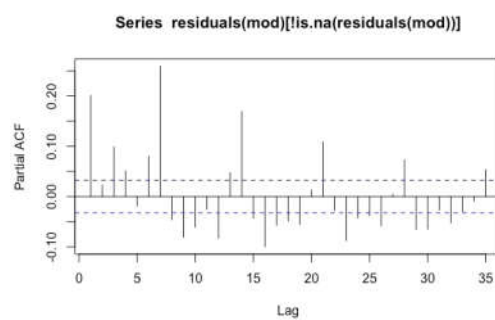
Biểu đồ AFC với bậc tự do = 5



Biểu đồ AFC với bậc tự do = 6



Biểu đồ AFC với bậc tự do = 7



Biểu đồ AFC với bậc tự do = 8

Hình 3. Biểu đồ AFC với các bậc tự do khác nhau

Hình 3 thể hiện biểu đồ AFC với các bậc tự do 5, 6, 7, và 8 để kiểm soát tính tự tương quan. Từ 4 biểu đồ này cho thấy, biểu đồ AFC với bậc tự do bằng 8 phù hợp để kiểm soát yếu tố này hay nói cách khác sự thay đổi về chiều hướng tự tương quan đều diễn ra tại ngày thứ

8 (câu lệnh là $k=a*b$ trong đó a là số năm và b là bậc tự do) (Chi tiết câu lệnh tại phụ lục 1).

2. Tính biến thiên quá mức (overdispersion)

Như đã đề cập ở trên, số liệu dạng cấu trúc này có xu hướng biến thiên quá mức, nên phương

sai lớn hơn rất nhiều so với giá trị trung bình của số kết quả đầu ra (số ca nhập viện). Do đó, kỹ thuật phù hợp sẽ là hồi quy giả Poisson (quasi-Poisson regression). Diễn hình trong bộ số liệu này, kết quả ở Bảng 2 cho thấy phương sai số ca nhập viện do bệnh đường hô hấp cao hơn nhiều so với trung bình số ca nhập viện. Vì vậy, trong ví dụ này chúng tôi sử dụng hồi quy giả Poisson (**family=quasipoisson**) (Chi tiết câu lệnh tại phụ lục 1).

3. Tính mùa (seasonality) và xu hướng dài hạn (long-term trend)

Một số yếu tố nhiễu không đo lường được cũng thay đổi theo thời gian làm ảnh hưởng mối quan hệ giữa ô nhiễm không khí và nhập viện. Sự thay đổi số ca nhập viện có thể tăng hoặc giảm hàng năm do sự gia tăng dân số hoặc khả năng tiếp cận y tế của quần thể và tình hình sức khỏe của cộng đồng. Ngoài ra, sự thay đổi ngắn hạn số ca nhập viện biến đổi theo chu kỳ mùa, ví dụ số ca nhập viện do bệnh cúm trong năm có xu hướng tăng vào các tháng có nhiệt độ thấp (mùa lạnh), thay vì vào các tháng có nhiệt độ cao (mùa nóng). Trong nghiên cứu này, chúng tôi chỉ quan tâm đến tác động ngắn hạn (short-term effect) hay tác động tức thời (acute effect)-đó là tác động ô nhiễm không khí lên thay đổi số ca nhập viện hằng ngày cho đến 7 ngày. Vì vậy, chúng tôi phải loại bỏ tác động dài hạn cũng như tính mùa (season) trong mô hình. Theo tác giả Roger D. Peng (9) trong bài viết “*Model choice in time series studies of air pollution and mortality*”, các phương pháp sử dụng để loại bỏ là sử dụng các mô hình bán tham số cho biến thời gian.

Trên thực tế, có nhiều hàm có thể dùng để mô hình hóa để loại bỏ yếu tố nhiễu không đo lường được (tác động dài hạn, tính mùa). Ví dụ hàm LOESS của thời gian (4) hay các dạng hàm spline khác như là smoothing splines, penalized splines và (natural) splines (5-8). Trong mô hình này, chúng tôi sử dụng hàm spline (**s(trend, k=a*b, bs="ad")**). Chi tiết việc lựa chọn mô hình, bậc

tự do cho các mô hình này, độc giả có thể tham khảo trong bài báo của Roger D. Peng (9)

Như vậy, trong ví dụ này chúng tôi sẽ tiến hành xây dựng mô hình “lỗi” trước gồm biến đầu ra- số ca nhập viện và thời gian đầu tiên. Câu lệnh sẽ có hình thái như sau:

```
mod<-gam(nres~s(trend, k=9*8, bs="ad")
(Chi tiết câu lệnh tại phụ lục 1).
```

Chúng tôi sử dụng hàm acf sau để đánh giá tính tự tương quan, tính mùa và tác động dài hạn: `acf(residuals(mod)[!is.na(residuals(mod))], type="p")` (Chi tiết câu lệnh tại phụ lục 1).

Bước 3: Hiệu chỉnh theo các biến nhiễu

Trong phân tích chuỗi thời gian, nhiễu do các yếu tố thay đổi theo thời gian. Do vậy các yếu tố nhiễu phổ biến, bao gồm tuổi, giới, chỉ số khối cơ thể, tình trạng hút thuốc hay sử dụng rượu/bia... không thay đổi theo ngày nên không thể xem là yếu tố nhiễu. Thay vào đó yếu tố nhiễu tiềm tàng sẽ là các biến có sự thay đổi theo ngày, và có mối liên quan với sự thay đổi theo ngày của biến phơi nhiễm chính (PM_{10}) cũng như biến đầu ra chính (số ca nhập viện). Trong bộ số liệu ví dụ, nhiệt độ và độ ẩm bên ngoài, cũng như tốc độ gió được xem là những biến nhiễu tiềm tàng do đây là chỉ số thay đổi theo ngày. Chúng ta sẽ đưa lần lượt các biến nhiễu tiềm tàng vào mô hình để đánh giá sự tác động của các biến này lên mối quan hệ thực sự giữa biến. Những biến có chỉ số AIC thấp và có ý nghĩa thống kê thường được ưu tiên giữ lại trong mô hình. Trong ví dụ này chúng tôi hiệu chỉnh yếu tố nhiễu là hàm đa thức các yếu tố khí hậu (bao gồm nhiệt độ trung bình ngày, độ ẩm trung bình ngày, tốc độ gió trung bình ngày).

Sau khi hoàn thành việc đánh giá các biến nhiễu, chúng ta sẽ có mô hình cuối cùng:

Đây cũng là mô hình cuối cùng dùng để đánh giá mối liên quan.

Bảng 3: Mô hình đa biến mối liên quan giữa PM₁₀ và nhập viện do bệnh hô hấp ở trẻ em tại Hà Nội, 2008 - 2016

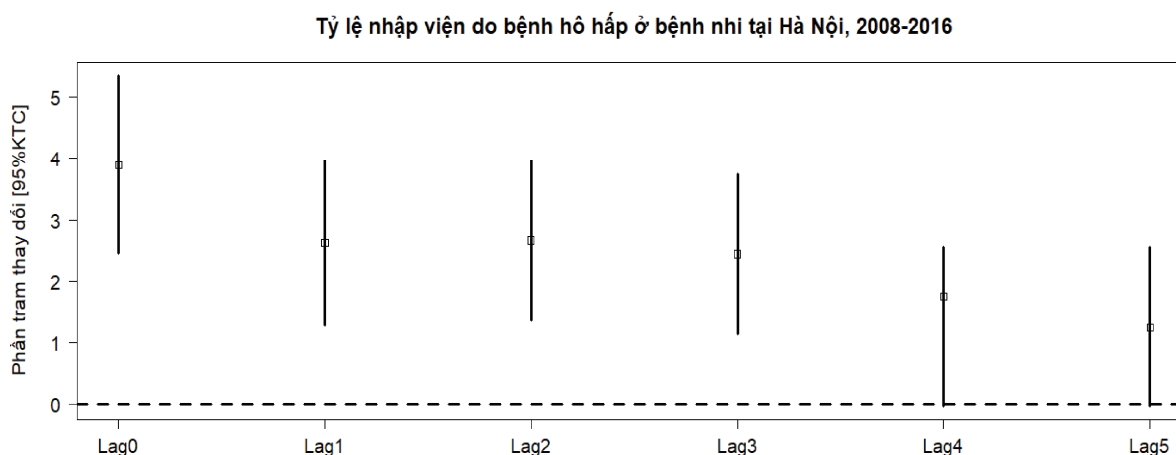
	RR	% nhập viện thay đổi 100* (1-RR)	95%CI	
			Nhỏ nhất	Lớn nhất
PM ₁₀	1,039	3.888	2,4423	5,354
Nhiệt độ (°C)	1,005	0.523	0,331	0,716
Độ ẩm (%)	1,006	0.611	0,386	0,836
Tốc độ gió (m/s)	1,0004	0.043	0,028	0,059

Kết quả mô hình này chúng ta có thể xác định mối liên quan giữa phơi nhiễm với PM₁₀ và tình trạng nhập viện. Trong ví dụ này, kết quả nghiên cứu cho thấy tại thời điểm nhập viện (lag0), khi nồng độ PM₁₀ tăng thêm 1 khoảng tứ phân vị (61,4µg/m³) thì tỷ lệ nhập viện do bệnh hô hấp của trẻ em Hà Nội tăng khoảng 3,9% (95%CI=2,4% - 5,4%) (Bảng 3).

Bước 4: Đánh giá độ trễ của tác động

Mối liên quan giữa nồng độ PM₁₀ và số ca

nhập viện sẽ khác nhau giữa các ngày. “Độ trễ” (delayed/lagged effect) trong mối quan hệ giữa ô nhiễm không khí và tình trạng nhập viện được định nghĩa là khoảng cách thời gian. Cụ thể, sự gia tăng nồng độ ô nhiễm trước 1, 2, 3,... hay thậm chí 7 ngày trước khi nhập viện có tác động đến tình trạng/số ca nhập viện. Do đó, việc ước tính độ trễ (lag) trong mô hình được thực hiện bằng cách tạo ra các mức độ phơi nhiễm ở các lag khác nhau và cập nhật lại mô hình được chọn trước đó.

**Hình 4: Sự thay đổi tỷ lệ nhập viện do bệnh hô hấp ở bệnh nhi tại Hà Nội tại các thời điểm phơi nhiễm với khí PM₁₀ khác nhau (lag), 2008 - 2016**

Hình 4 cho thấy mối liên quan giữa việc phơi nhiễm với khí PM₁₀ tại thời điểm nhập viện và 4 ngày trước đó và tình trạng nhập

viện do các bệnh hô hấp ở trẻ em tại Hà Nội giai đoạn 2008 – 2016. Cụ thể, tại thời điểm trước 1, 2, và 3 ngày nhập viện, khi nồng độ

PM_{10} tăng thêm 1 khoảng tứ phân vị thì tỷ lệ nhập viện cũng tăng. Tuy nhiên, không tìm thấy mối liên quan giữa phơi nhiễm tại thời điểm trước 4 và 5 ngày nhập viện. Hay nói một cách khác mức độ tác động giảm dần khi thời gian tăng lên. Thậm chí, sau 4 đến 5 ngày thì mối liên quan không còn có ý nghĩa thống kê. Hay nói một cách khác mối liên quan giữa ô nhiễm không khí do PM_{10} và số ca nhập viện giảm đi khi số ngày tăng thêm.

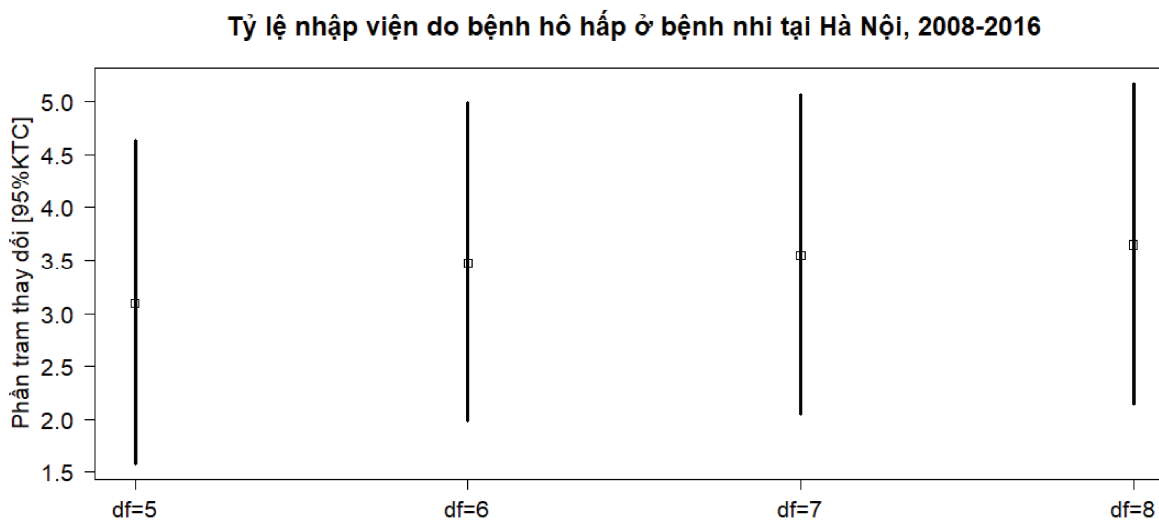
```
mod<-gam(nres~dmpm10_1+s(trend, k=9*8,
bs="ad") + paper$dmtemp+ paper$dmhum +
I(paper$dmhum^2) + I(paper$dmwisp^2)+
I(paper$dmtemp*paper$dmhum) +
I(paper$dmtemp*paper$dmwisp)+
I(paper$dmhum*paper$dmwisp), family=qua
sipoisson,data=paper,na.action=na.exclude)
```

(Chi tiết câu lệnh tại phụ lục 2)

Bước 5: Kiểm tra độ nhạy và sự ổn định của mô hình:

Kiểm tra sự ổn định của mô hình là một điều kiện bắt buộc khi chạy mô hình chuỗi thời gian. Trong phần này quan trọng nhất là chứng minh được mối quan hệ/ tác động bạn đang khảo sát không thay đổi đáng kể khi thay đổi cách mô hình.

Sau đây chúng tôi sẽ trình bày một cách đơn giản để kiểm tra độ ổn định của mô hình là thay đổi bậc tự do trong hàm spline (Chi tiết câu lệnh thay đổi bậc tự do tại phụ lục 3). Bên cạnh đó bạn có thể thay đổi cách mô hình hóa biến nhiễu. Ví dụ vì dùng hàm đa thức nhiệt độ và độ ẩm như chúng tôi làm thì bạn có thể dùng hàm spline cho biết nhiệt độ, độ ẩm, tốc độ gió, hoặc là thay đổi hàm thời gian từ spline sang spline tự nhiên (ns).



Hình 5: Sự thay đổi về tỷ lệ nhập viện do các bệnh hô hấp ở bệnh nhi tại Hà Nội khi thay đổi các bậc tự do, 2008 – 2016

Hình 5 biểu thị sự thay đổi về tỷ lệ phần trăm tích lũy của nhập viện do các bệnh hô hấp khi phơi nhiễm với khí PM_{10} trong vòng 4 ngày (trước 3 ngày và tại thời điểm nhập viện). Kết quả cho thấy có sự ổn định của mô hình khi

thay đổi nhẹ khi mô hình áp dụng các bậc tự do khác nhau. Cụ thể, Mỗi tác động ở $df=6$, 7 cao hơn ít so với các giá trị $df=5$ hoặc $df=8$ (Chi tiết câu lệnh tại phụ lục 3)

Bước 6: Một số phân tích khác

Trong các phần trên, chúng tôi chỉ trình bày mối liên quan giữa một chất ô nhiễm không khí với sức khỏe. Bên cạnh đó, khi tiến hành đánh giá mối liên quan giữa ô nhiễm không khí và tình trạng sức khỏe, một số phân tích khác có thể cân nhắc bổ sung để tìm hiểu đầy đủ về mối quan hệ này. Cụ thể là:

- Tác động của nhiều chất cùng lúc nên vấn đề sức khỏe: chúng ta đánh giá ảnh hưởng của các chất ô nhiễm khác như là $PM_{2.5}$, NO_2 hay CO tới mối quan hệ giữa tình trạng nhập viện và phơi nhiễm với PM_{10} bằng cách đưa thêm các chất này vào mô hình đã trình bày phía trên để đánh giá sự tác động.

- Đánh giá mỗi tác động: tuổi hay mùa (lạnh, nóng) có thể là yếu tố tác động (interaction term) của mối quan hệ giữa PM_{10} và số ca nhập viện. Để làm điều này, người đọc có thể phân chia số liệu làm hai nhóm (ví dụ: tuổi dưới 1, tuổi từ 1-5). Kiểm định khi bình phương có thể sử dụng để so sánh sự khác biệt giữa yếu tố nguy cơ ở các tầng.

- Trong nghiên cứu này chỉ đánh giá mối liên quan ở một cộng đồng (Hà Nội). Nếu muốn đánh giá mối quan hệ giữa ô nhiễm không khí và sức khỏe ở nhiều vùng khác nhau, mối quan hệ ở từng vùng cần được đánh giá theo phương pháp này, sau đó các đại lượng này sẽ được tổng hợp bằng phép phân tích gộp (meta-analysis).

KẾT LUẬN

Trong bài báo này, chúng tôi đã đưa ra một số lưu ý và các bước cơ bản khi tiến hành phân tích chuỗi thời gian. Diễn hình trong bộ số liệu ví dụ, chúng tôi đã tiến hành phân tích chuỗi thời gian để đánh giá tác động ngắn hạn của ô nhiễm không khí do PM_{10} và tình trạng nhập viện do bệnh hô hấp ở trẻ em tại Hà Nội giai đoạn 2008 – 2016. Điểm cần lưu ý với

phân tích chuỗi thời gian chính là mô hình hóa tính mùa và tác động dài hạn, mối quan hệ tính tự tương quan và xử lý được các tác động của biến nhiễu. Do vậy phân tích chuỗi thời gian có thể hạn chế được tính ngẫu biến sinh thái trong thiết kế nghiên cứu sinh thái. Mô hình này có thể tận dụng được các số liệu ghi nhận hằng ngày để đánh giá những tác động theo thời gian hoặc can thiệp. Tuy nhiên tính chính xác của mô hình phụ thuộc vào việc lựa chọn các tham số trong mô hình cũng như sự chính xác của số liệu ghi nhận

Đơn vị tài trợ: Nghiên cứu này được tài trợ bởi Quỹ phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số 105.08-2019.331

TÀI LIỆU THAM KHẢO

1. Tapia V, Steenland K, Sarnat SE, Vu B, Liu Y, Sánchez-Ccoyllo O, et al. Time-series analysis of ambient $PM_{2.5}$ and cardiorespiratory emergency room visits in Lima, Peru during 2010-2016. *Journal of exposure science & environmental epidemiology*. 2019.
2. Le TG, Ngo L, Mehta S, Do VD, Thach TQ, Vu XD, et al. Effects of short-term exposure to air pollution on hospital admissions of young children for acute lower respiratory infections in Ho Chi Minh City, Vietnam. *Research report (Health Effects Institute)*. 2012(169):5-72; discussion 3-83.
3. Nhung NTT, Schindler C, Dien TM, Probst-Hensch N, Perez L, Kunzli N. Acute effects of ambient air pollution on lower respiratory infections in Hanoi children: An eight-year time series study. *Environment international*. 2018;110:139-48.
4. Schwartz J. Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Canadian Journal of Statistics*. 1994;22(4):471-87.
5. Dominici F, McDermott A, Zeger SL, Samet JM. On the Use of Generalized Additive Models in Time-Series Studies of Air Pollution and Health. *American Journal of Epidemiology*. 2002;156(3):193-203.
6. Ramsay TO, Burnett RT, Krewski D. The effect

- of concurvity in generalized additive models linking mortality to ambient particulate matter. *Epidemiology*. 2003;14(1):18-23.
7. Schwartz J, Zanobetti A, Bateson T. Morbidity and mortality among elderly residents in cities with daily PM measurements. In *Revised Analyses of the National Morbidity, Mortality, and Air Pollution Study, Part II*. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health* 2003. 25-58 p.
 8. Touloumi G, Atkinson R, Tertre AL, Samoli E, Schwartz J, Schindler C, et al. Analysis of health outcome time series data in epidemiological studies. *Environmetrics*. 2004;15(2):101-17.
 9. Roger D. Peng FDaTAL. Model choice in time series studies of air pollution and mortality. *J R Statist Soc A*. 2006;169:179-203.

Time series regression to identifying the relationship between air pollution and health in Vietnam

Nguyễn Thùy Linh¹, Lê Tu Hoàng¹, Nguyễn Thị Kim Ngân¹, Nguyễn Thị Bích Liên²,
Nguyễn Thị Trang Nhung¹, Trần Minh Dien²

¹Hanoi University of Public Health

²Vietnam National Children's hospital

Abstract: Time series has been widely used in environmental epidemiology; especially in identifying the short-term associations between ambient air pollution and health outcomes. For both exposure and outcome, data are available at regular time intervals (daily hospital admissions and pollution levels) to explore short-term associations between them. In this article, we described main steps to conduct time series regression and highlighted some key ideas when applying this technique. A sample data was used to investigate short-term association between PM₁₀ and daily hospital admission among children in Hanoi between 2008 and 2016. This analysis was conducted with R software.

Keywords: Time series, air pollution, short-term effect

Phụ lục 1: Câu lệnh mô hình hoàn chỉnh đánh giá mối liên quan giữa PM10 và số ca nhập viện tại bệnh viện Nhi 2007 – 2013

```
mod<-gam(nres~dmpm10+s(trend, k=9*8, bs="ad") + paper$dmtemp+ paper$dmhum + I(paper$dmhum^2)+I(paper$dmwisp^2)+I(paper$dmtemp*paper$dmhum)+I(paper$dmtemp*paper$dmwisp)+I(paper$dmhum*paper$dmwisp),family=quasipoisson,data=paper,na.action=na.exclude)
```

Phụ lục 2: Đánh giá độ trễ của tác động

Phơi nhiễm với chất gây ô nhiễm không khí tại thời điểm nhập viện: mod<-gam(nres~**dmpm10**+s(trend, k=9*8, bs="ad") + paper\$dmtemp+ paper\$dmhum + I(paper\$dmhum^2) + I(paper\$dmwisp^2)+ I(paper\$dmtemp*paper\$dmhum) + I(paper\$dmtemp*paper\$dmwisp)+ I(paper\$dmhum*paper\$dmwisp), family=quasipoisson,data=paper,na.action=na.exclude)

Phơi nhiễm với chất gây ô nhiễm không khí trước 1 ngày nhập viện: mod<-gam(nres~**dmpm10_1**+s(trend, k=9*8, bs="ad") + paper\$dmtemp+ paper\$dmhum + I(paper\$dmhum^2) + I(paper\$dmwisp^2)+ I(paper\$dmtemp*paper\$dmhum) + I(paper\$dmtemp*paper\$dmwisp)+ I(paper\$dmhum*paper\$dmwisp), family=quasipoisson,data=paper,na.action=na.exclude)

Phơi nhiễm với chất gây ô nhiễm không khí trước 2 ngày nhập viện: mod<-gam(nres~**dmpm10_2**+s(trend, k=9*8, bs="ad") + paper\$dmtemp+ paper\$dmhum + I(paper\$dmhum^2) + I(paper\$dmwisp^2)+ I(paper\$dmtemp*paper\$dmhum) + I(paper\$dmtemp*paper\$dmwisp)+ I(paper\$dmhum*paper\$dmwisp), family=quasipoisson,data=paper,na.action=na.exclude)

Phơi nhiễm với chất gây ô nhiễm không khí trước 3 ngày nhập viện: mod<-gam(nres~**dmpm10_3**+s(trend, k=9*8, bs="ad") + paper\$dmtemp+ paper\$dmhum + I(paper\$dmhum^2) + I(paper\$dmwisp^2)+ I(paper\$dmtemp*paper\$dmhum) + I(paper\$dmtemp*paper\$dmwisp)+ I(paper\$dmhum*paper\$dmwisp), family=quasipoisson,data=paper,na.action=na.exclude)

Phơi nhiễm với chất gây ô nhiễm không khí trước 4 ngày nhập viện: mod<-gam(nres~**dmpm10_4**+s(trend, k=9*8, bs="ad") + paper\$dmtemp+ paper\$dmhum + I(paper\$dmhum^2) + I(paper\$dmwisp^2)+ I(paper\$dmtemp*paper\$dmhum) + I(paper\$dmtemp*paper\$dmwisp)+ I(paper\$dmhum*paper\$dmwisp), family=quasipoisson,data=paper,na.action=na.exclude)

Phơi nhiễm với chất gây ô nhiễm không khí trước 5 ngày nhập viện: mod<-gam(nres~**dmpm10_5**+s(trend, k=9*8, bs="ad") + paper\$dmtemp+ paper\$dmhum + I(paper\$dmhum^2) + I(paper\$dmwisp^2)+ I(paper\$dmtemp*paper\$dmhum) + I(paper\$dmtemp*paper\$dmwisp)+ I(paper\$dmhum*paper\$dmwisp), family=quasipoisson,data=paper,na.action=na.exclude)

Phụ lục 3: Câu lệnh kiểm tra độ nhạy và tính ổn định của mô hình

Bậc tự do (df=5): `mod<-gam(nres~dmpm10p3+s(trend, k=9*5, bs="ad") + paper$dmtemp+
paper$dmhum + I(paper$dmhum^2) + I(paper$dmwisp^2)+ I(paper$dmtemp*paper$dmhum)
+ I(paper$dmtemp*paper$dmwisp)+ I(paper$dmhum*paper$dmwisp), family=quasipoisson,d
ata=paper,na.action=na.exclude)`

Bậc tự do (df=6): `mod<-gam(nres~dmpm10p3+s(trend, k=9*6, bs="ad") + paper$dmtemp+
paper$dmhum + I(paper$dmhum^2) + I(paper$dmwisp^2)+ I(paper$dmtemp*paper$dmhum)
+ I(paper$dmtemp*paper$dmwisp)+ I(paper$dmhum*paper$dmwisp), family=quasipoisson,d
ata=paper,na.action=na.exclude)`

Bậc tự do (df=7): `mod<-gam(nres~dmpm10p3+s(trend, k=9*7, bs="ad") + paper$dmtemp+
paper$dmhum + I(paper$dmhum^2) + I(paper$dmwisp^2)+ I(paper$dmtemp*paper$dmhum)
+ I(paper$dmtemp*paper$dmwisp)+ I(paper$dmhum*paper$dmwisp), family=quasipoisson,d
ata=paper,na.action=na.exclude)`

Bậc tự do (df=8): `mod<-gam(nres~dmpm10p3+s(trend, k=9*8, bs="ad") + paper$dmtemp+
paper$dmhum + I(paper$dmhum^2) + I(paper$dmwisp^2)+ I(paper$dmtemp*paper$dmhum)
+ I(paper$dmtemp*paper$dmwisp)+ I(paper$dmhum*paper$dmwisp), family=quasipoisson,d
ata=paper,na.action=na.exclude)`