

# XÂY DỰNG MÔ HÌNH XU THẾ DỰ BÁO SỐ VỤ CHÁY TẠI VIỆT NAM GIAI ĐOẠN 2020 – 2024

Đại úy, ThS. ĐẶNG VĂN TRỌNG

Khoa Khoa học cơ bản và ngoại ngữ, Trường Đại học PCCC

**Tóm tắt:** Nghiên cứu này tập trung phân tích dữ liệu thống kê về số vụ cháy ở Việt Nam các năm trong giai đoạn 1981 – 2019, từ đó xây dựng mô hình xu thế phù hợp với tập dữ liệu để dự báo số vụ cháy ở Việt Nam các năm trong giai đoạn 2020 – 2024. Sau khi kiểm định sự phù hợp của mô hình, đánh giá sai số, tác giả đã chọn được mô hình với hàm xu thế bậc hai phù hợp với tập dữ liệu này và đưa ra các dự báo điểm, dự báo bằng khoảng tin cậy 90% cho số vụ cháy ở Việt Nam các năm trong giai đoạn 2020 - 2024. Nghiên cứu được thực hiện với sự hỗ trợ của phần mềm phân tích thống kê R.

**Từ khóa:** Mô hình, dự báo, hàm xu thế, chuỗi thời gian, số vụ cháy, phần mềm phân tích thống kê R.

**Abstract:** This study focuses on analyzing statistical data on the number of fires in Vietnam in the period 1981 - 2019, thereby developing a suitable trend model for the data set to forecast the number of fires in Vietnam in years of period 2020 - 2024. After testing the model's suitability and evaluating the error, the author has selected a model with a suitable quadratic trend function for this data set, proposing point forecast and the forecast with confidence interval of 90% for the number of fires in Vietnam in each year over the period from 2020 to 2024. This study was conducted with the application of R statistical analysis software.

**Keywords:** Model, forecast, trend function, time series, number of fires, R statistical analysis software.

**Ngày nhận bài:** 18/11/2020; **Ngày thẩm định:** 23/02/2021; **Ngày duyệt đăng:** 05/3/2021.

## 1. Giới thiệu

Ngày nay, khi xã hội phát triển con người càng phải đối mặt với nhiều hiểm họa. Một trong những vấn đề lớn đang được quan tâm hiện nay là tình hình cháy, nổ đang diễn ra ngày một tăng. Mỗi vụ cháy xảy ra đã tiêu hủy nhiều tài sản, có nhiều vụ cháy còn cướp đi sinh mạng của nhiều người. Việc cháy, nổ chịu sự tác động của rất nhiều yếu tố do đó số lượng các vụ cháy xảy ra hàng năm là một đại lượng ngẫu nhiên (biến ngẫu nhiên). Mặc dù là ngẫu nhiên nhưng nếu ta nghiên cứu một cách hệ thống trên một tập dữ liệu lớn có thể tìm được quy luật từ đó giúp các nhà quản lý có những quyết định đúng đắn trong việc hoạch định chiến lược, dự báo tình hình cháy, nổ cho tương lai. Với sự phát triển khoa học thống kê, sự hỗ trợ của công nghệ thông tin, rất nhiều bài toán phân tích thống kê trên dữ liệu lớn đã được xử lý từ đó tìm ra các mô hình dự báo phù hợp.

Với dữ liệu về số vụ cháy ở Việt Nam trong các năm từ 1981 đến 2019 được thu thập từ các Báo cáo tổng kết công tác năm của Cục Cảnh sát Phòng cháy, chữa cháy và cứu nạn, cứu hộ [1], với sự hỗ trợ của phần mềm phân tích thống kê R, tác giả thử nghiệm xây dựng mô hình xu thế trong phân tích chuỗi thời gian để dự báo số vụ cháy cho các năm trong giai đoạn 2020 – 2024 [4].

## 2. Mô hình xu thế và phần mềm phân tích thống kê R

Chuỗi thời gian là một dãy gồm  $n$  giá trị được ghi lại về sự biến động của sự vật, hiện tượng,... tại  $n$  kỳ thời gian (ở đây kỳ thời

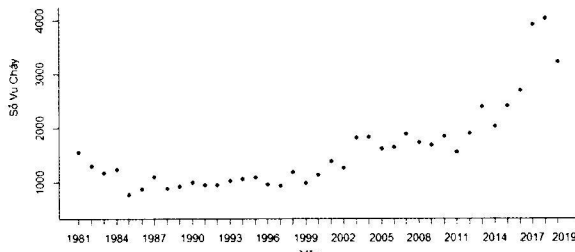
gian có thể là giờ, ngày, tháng, quý, 6 tháng, năm,...)[2]. Phân tích, dự báo chuỗi thời gian được thực hiện trong hầu như tại mọi tổ chức có thao tác dữ liệu lượng hóa được. Các phương pháp phân tích chuỗi thời gian rất đa dạng nhưng có thể phân loại thành ba nhóm phương pháp chính: nhóm phương pháp định hướng dữ liệu; nhóm phương pháp dựa vào mô hình; nhóm phương pháp sử dụng mạng thần kinh [4]. Mô hình xu thế thuộc nhóm phương pháp dựa vào mô hình. Thực chất mô hình xu thế là một hàm hồi quy mô tả sự vận động của biến cần dự báo ( $Y$ ) theo biến thời gian ( $t$ ), có dạng tổng quát  $Y = f(t) + \varepsilon$ , ở đây  $\varepsilon$  là sai số ngẫu nhiên, giá trị  $y=f(t)$  gọi là giá trị ước lượng của  $Y$  dựa theo mô hình. Tùy theo hàm  $f(t)$  mà mô hình được gọi là mô hình tuyến tính hay phi tuyến. Các tham số của mô hình được ước lượng từ tập dữ liệu thống kê bằng phương pháp bình phương cực tiểu (cực tiểu tổng bình phương sai số giữa giá trị tính theo mô hình và giá trị thực tế). Bằng việc sử dụng kiến thức giải tích, việc ước lượng các tham số dẫn tới việc giải các hệ phương trình. Ngày nay, với sự phát triển của các phần mềm thống kê đã giúp cho việc ước lượng các tham số này được dễ dàng hơn, đặc biệt là trong trường hợp dữ liệu thống kê lớn. Trong nghiên cứu này, tác giả sử dụng các gói lệnh trong phần mềm phân tích thống kê R để hỗ trợ tính toán [3].

Phần mềm phân tích thống kê R là một phần mềm mã nguồn mở, được các nhà nghiên cứu thống kê trên thế giới hợp tác phát

triển để sử dụng thống nhất, hoàn toàn miễn phí. Khởi đầu của ý tưởng xây dựng phần mềm R là hai nhà thống kê học Ross Ihaka và Robert Gentleman vào năm 1996, thống nhất đặt tên phần mềm là R. Cho đến nay, càng ngày càng có nhiều nhà thống kê học, toán học, nhà nghiên cứu trong mọi lĩnh vực đã chuyển sang sử dụng R để phân tích dữ liệu khoa học.

### 3. Phân tích dữ liệu số vụ cháy trong giai đoạn 1981 - 2019 và lựa chọn mô hình

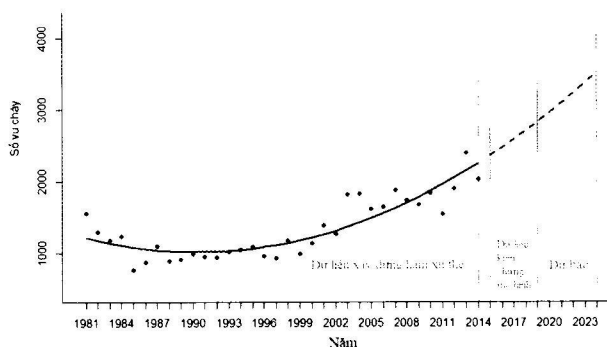
Tiến hành trực quan hóa số liệu tổng hợp về số vụ cháy tại Việt Nam từ năm 1981 đến năm 2019 bằng biểu đồ tán xạ (Hình 1).



Hình 1: Biểu đồ tán xạ của số vụ cháy ở Việt Nam giai đoạn 1981 – 2019.

Từ biểu đồ tán xạ của số vụ cháy ở Việt Nam giai đoạn 1981 – 2019 cho thấy xu hướng phát triển của số vụ cháy có thể tuân theo quy luật hàm xu thế bậc 2 hoặc hàm xu thế dạng mũ với biến độc lập là chỉ số thời gian (t). Để đánh giá sai số cũng như kiểm tra sự phù hợp của mô hình dự báo ta sử dụng dữ liệu về số vụ cháy của các năm từ 1981 đến 2014 để xây dựng mô hình và sử dụng dữ liệu về số vụ cháy từ năm 2015 đến 2019 để kiểm chứng mô hình, đánh giá sai số. Mô hình tối ưu nhất theo nghĩa có sai số nhỏ hơn và khoảng ước lượng cho số vụ cháy ở cùng một độ tin cậy hẹp hơn sẽ được sử dụng để dự báo về số vụ cháy cho các năm giai đoạn từ 2020 đến 2024.

Qua kết quả phân tích bằng phần mềm phân tích thống kê R đối với các mô hình theo dự đoán, cho thấy xu hướng phát triển của số vụ cháy phù hợp với hàm xu thế bậc hai (Hình 2).



Hình 2: Đường xu thế và khoảng dự báo số vụ cháy giai đoạn 2020 – 2024.

Trong hình 2, đường cong liền là đường xu thế bậc hai được xây dựng từ tập dữ liệu về số vụ cháy của các năm trong giai đoạn 1981 - 2014, các điểm biểu thị số vụ cháy trong thực tế, đường nét đứt màu xanh và vùng màu xám tương ứng biểu thị cho dự báo điểm và khoảng tin cậy 90% của số vụ cháy các năm trong giai đoạn 2015 – 2024 tính theo mô hình. Kết quả phân tích sau khi đã lựa chọn hàm xu thế bậc hai thể hiện ở bảng dưới đây (Bảng 1).

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1262.0244 108.0654 11.678 6.96e-13 ***
trend -46.4425 14.2362 -3.262 0.00269 **
I(trend^2) 2.2195 0.3946 5.625 3.58e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 197.8 on 31 degrees of freedom
Multiple R-squared: 0.7851, Adjusted R-squared: 0.7713
F-statistic: 56.64 on 2 and 31 DF, p-value: 4.452e-11
```

Bảng 1: Kết quả ước lượng các tham số của mô hình.

Từ kết quả phân tích cho thấy, mô hình được lựa chọn có thể giải thích được 78,51% sự biến động của số vụ cháy thông qua mô hình (Multiple R-squared: 0.7851), các hệ số hồi quy được ước lượng (Estimate) của mô hình được kiểm định đều có ý nghĩa thống kê thể hiện ở các giá trị  $Pr(>|t|)$  đều nhỏ hơn 0,01 tức có thể khẳng định các hệ số của mô hình khác 0 với độ tin cậy đến 99%, giá trị F - statistic: 56.64 với p-value rất nhỏ cho thấy mô hình phù hợp với dữ liệu dùng xây dựng mô hình (dữ liệu về số vụ cháy giai đoạn 1981 – 2014). Từ kết quả phân tích xác định được đường xu thế bậc hai,

$$So\ vu\ chay(t) = 1262,0244 - 46,4425t + 2,2195t^2$$

Ở đây t là thứ tự thời gian của chuỗi, tương ứng với  $t = i$  là năm  $1981 + i - 1$ ,  $i = (1, 34)$

Kết quả dự báo số vụ cháy giai đoạn 2015 – 2019 được tính thông qua mô hình đã xây dựng được thể hiện ở bảng dưới đây (Bảng 2).

	Point Forecast	Lo 90	Hi 90
2015	2355.479	1973.307	2737.651
2016	2466.624	2073.381	2859.867
2017	2582.208	2176.056	2988.361
2018	2702.232	2281.258	3123.206
2019	2826.694	2388.935	3264.454

Bảng 2: Kết quả dự báo số vụ cháy giai đoạn 2015 – 2019.

Kết quả dự báo gồm có 4 cột tương ứng là năm dự báo, giá trị dự báo điểm (Point Forcaste) và hai cột cuối là cận dưới (Lo 90) và cận trên (Hi 90) của khoảng dự báo với độ tin cậy 90%. Đối chiếu với dữ liệu thực tế, số vụ cháy năm 2015 là 2404, năm 2016 là 2688, năm 2019 là 3217 đều là các giá trị nằm trong khoảng tin cậy 90% của dự báo, chỉ riêng hai năm 2017 (có 3924 vụ cháy), năm 2018 (có 4037 vụ cháy) nằm ngoài khoảng tin cậy 90% của dự báo. Điều đáng chú ý là hai năm 2017, 2018 cả nước có số vụ cháy tăng đột biến so với các năm còn lại trong giai đoạn 2013 – 2019 có thể coi đó là những giá trị bất thường, điều này có thể do một số nguyên nhân gây ra chẳng hạn do thay đổi cách thức thống kê, nguyên nhân đặc biệt nào đó về kinh tế, xã hội,... Chính các giá trị bất thường này trong thống kê số vụ cháy của hai năm 2017, 2018 đã làm tăng các sai số của mô hình trên kỳ kiểm chứng.

Kết quả tính toán sai số của mô hình trên kỳ kiểm chứng được phần mềm phân tích thống kê R tính toán (Bảng 3):

```
> accuracy(fire.lm.pred$mean, test.ts)
      VE      RMSE      MAE      MPE      MAPE
Test set 667.3525 870.1392 667.3525 17.9289 17.9289
```

Bảng 3: Các loại sai số của mô trên kỳ kiểm chứng 2015 – 2019.

Trong kết quả phân tích này, giá trị  $ME = 667,3525$  là sai số trung bình,  $RMSE = 870,1392$  là sai số căn bình phương trung bình,  $MAE = 667,3525$  là sai số tuyệt đối trung bình,  $MPE$ ,

MAPE tương ứng là sai số phần trăm trung bình và sai số phần trăm tuyệt đối trung bình, trong mô hình này hai giá trị này bằng nhau và bằng 17,9289%. Như đã trình bày ở trên, nguyên nhân để gây ra sai số lớn trên kỳ kiểm chứng là do sự bất thường về số vụ cháy ở các năm 2017, 2018, tuy nhiên đây là mô hình hàm xu thế phù hợp nhất có thể với dữ liệu đã cho. Trong các nghiên cứu tiếp theo, tác giả có thể thử nghiệm các mô hình như: mô hình tự hồi quy (AR), mô hình trung bình trượt tích hợp tự hồi quy (ARIMA), mô hình mạng thần kinh để phân tích với hy vọng tìm ra được mô hình phù hợp hơn nữa với số liệu thực tế.

## 4. Dự báo số vụ cháy cho giai đoạn 2020 - 2024

Ứng dụng mô hình dự báo đã xây dựng, giá trị dự báo điểm và khoảng tin cậy 90% cho số vụ cháy trong cả nước trong các năm giai đoạn 2020 – 2024 là (Bảng 4):

> fire.lm.pred				
	Point Forecast	Lo 90	Hi 90	
2020	2955.596	2499.058	3412.134	
2021	3088.937	2611.614	3566.259	
2022	3226.716	2726.606	3726.826	
2023	3368.935	2844.050	3893.820	
2024	3515.593	2963.970	4067.216	

Bảng 4: Kết quả dự báo số vụ cháy giai đoạn 2020 – 2024.

Kết quả dự báo bằng mô hình cho thấy, với độ tin cậy 90% trong năm 2020 có khoảng 2499 đến 3412 vụ cháy và dự báo điểm là 2956 vụ cháy, năm 2021 dự báo có khoảng 2612 đến 3566 vụ cháy và dự báo điểm là 3089 vụ cháy, năm 2022 dự báo có khoảng 2727 đến 3727 vụ cháy và dự báo điểm là 3227 vụ cháy, năm 2023 dự báo có khoảng 2844 đến 3894 vụ cháy và dự báo điểm 3369 vụ cháy, năm 2024 dự báo có khoảng 2964 đến 4067 vụ cháy và dự báo điểm 3516 vụ cháy.

## 5. Kết luận

Nghiên cứu này bước đầu phân tích, đối chiếu sai số giữa một số mô hình xu thế có thể sử dụng và đã lựa chọn được mô hình với đường xu thế bậc hai phù hợp với tập dữ liệu về số vụ cháy của Việt Nam trong các năm ở giai đoạn 1981 – 2019. Từ dữ liệu thực tế đã tiến hành ước lượng các hệ số của mô hình, tính toán các sai số của mô hình dựa trên tập giá trị kiểm chứng và từ đó sử dụng mô hình đưa ra các dự báo bao gồm dự báo điểm, dự báo bằng khoảng tin cậy 90% cho số vụ cháy ở Việt Nam trong các năm ở giai đoạn 2020 – 2024. Kết quả nghiên cứu là một cơ sở khoa học để các đơn vị chiến đấu của lực lượng PCCC, các nhà quản lý tham khảo và sử dụng phục vụ cho việc xây dựng kế hoạch công tác của đơn vị, hoạch định chính sách cho phù hợp. Để các kết quả phân tích, dự báo được chính xác, tác giả mong muốn nhận được các số liệu thống kê được thu thập theo từng ngày, từng tuần, từng tháng để có thể xem xét nghiên cứu sự tác động của yếu tố mang tính mùa vụ đến số lượng vụ cháy, các số liệu liên quan phải được ghi nhận, thu thập trung thực và mẫu số liệu càng lớn thì phân tích, đánh giá, dự báo càng chính xác. Qua đây cho thấy, vai trò hết sức quan trọng của công tác lưu trữ, tham mưu, tổng hợp, báo cáo. Để ngành PCCC hoạt động ngày càng hiệu quả thì việc làm cho các số liệu biết nói đóng góp một phần không nhỏ.■

## TÀI LIỆU THAM KHẢO

1. Cục Cảnh sát PCCC & CNCH, các Báo cáo tổng kết công tác năm 1981 đến 2019.
2. Đỗ Ngọc Cẩn (2011), *Thống kê phòng cháy chữa cháy*, NXB Giao thông vận tải.
3. Nguyễn Văn Tuấn (2020), *Phân tích dữ liệu với R*, NXB Tổng hợp Thành phố Hồ Chí Minh.
4. Galit Shmueli, Kenneth C. Lichtendahl Jr (2016), *Practical time series for casting with R*, Axelrod Schnall Publishers.

# NGUY CƠ CHÁY NỔ TỪ TĨNH ĐIỆN... (Tiếp theo trang 32)

thể. Tuy nhiên việc tạo độ ẩm chỉ có thể thực hiện trong không gian kín. Còn không gian mở thì biện pháp này khó khả thi.

- Tạo Ion (Ionizer)

Với những vật liệu cách điện bị nhiễm điện thì có thể sử dụng phương pháp này. Nhờ thiết bị tạo ion trong môi trường sẽ làm trung hòa các điện tích trong các vật bị nhiễm điện.

- Sử dụng các thiết bị chống tĩnh điện

Một số thiết bị chống tĩnh điện thông dụng như: thanh khử tĩnh điện, vòi phun chống tĩnh điện, súng chống tĩnh điện, quạt tĩnh điện...

## 4. Kết luận

Hiện tượng nhiễm điện là một hiện tượng thường xuyên xảy ra và khá phức tạp. Trong các môi trường nguy hiểm về cháy nổ thì đó là một trong các nguy cơ luôn rình rập và khó lường.

Các tiêu chuẩn về phòng cháy chữa cháy ở Việt Nam cũng đã có các quy định về việc này. Bên cạnh việc chấp hành tốt các quy định hiện hành, chúng ta cần có những nghiên cứu, áp dụng các công nghệ tiên tiến để phòng ngừa các sự cố do hiện tượng nhiễm điện gây ra.■

## TÀI LIỆU THAM KHẢO

1. NFPA77 (2014). Recommended Practice on Static Electricity.
2. Vahid Ebadat (2019). Static Electricity and Chemical Safety. Stonehouse Process Safety, Inc. Princeton, New Jersey
3. Kiyoshi OTA (2018). Electrostatic Risk Assessment for Chemical Plants: Fire and Explosion Prevention. Sumitomo Kagaku 2018.