

ỨNG DỤNG MÔ HÌNH HỒI QUY LOGISTIC DỰ BÁO ỚNG BỂ TRÊN HỆ THỐNG PHÂN PHỐI NƯỚC

Phạm Thị Minh Lành

ĐH Bách Khoa Tp. Hồ Chí Minh

Phạm Hà Hải

ĐH Kiến Trúc Tp. Hồ Chí Minh

Vũ Thị Vân Anh

ĐH Tài nguyên và Môi trường TP HCM

Trương Thị Thu Hằng

ĐH Thủy Lợi, số 2 Trường Sa

Nguyễn Quang Trường và Lê Đình Hồng

ĐH Bách Khoa Tp. Hồ Chí Minh

Tóm tắt: Hiện tượng bể ống cấp nước thường xuyên xảy ra trên các hệ thống phân phối không những gây nên những tổn thất về mặt kinh tế cho các công ty cấp nước mà còn là nguy cơ để các chất ô nhiễm xâm nhập vào đường ống cấp nước. Một số mô hình đã được phát triển để xác định khả năng xảy ra ống hỏng từ số liệu thống kê hoặc thực nghiệm, tuy nhiên đa số các mô hình đều giới hạn cho vật liệu ống kim loại, trong khi số lượng ống nhựa như uPVC, HDPE được sử dụng rất nhiều trên mạng lưới lại bị bỏ qua. Trong bài báo này sẽ đề xuất mô hình hồi quy logistics để xác định mối liên quan giữa hiện tượng ống bể và các đặc điểm của đường ống trên mạng lưới phân phối, từ đó đưa ra dự báo về xác suất bể ống cho tất cả các loại vật liệu ống dẫn trên mạng lưới cấp nước. Kết quả của mô hình được kiểm chứng cho mạng lưới phân phối nước quận Hải Châu thành phố Đà Nẵng.

Từ khóa: ống cấp nước; mô hình hồi quy logistic; hệ thống phân phối nước; phần mềm R.

Abstract: Pipe failure phenomenon often have on water distribution systems, it is not only causes economic losses to water supply companies, but also the risk of pathogen contaminant intrusion into the pipe. Some models were applied to define probability of pipe failure by statistic data or experimental data, but most models are limited to metal materials, however the number of plastic as UPVC, HDPE is used popular on the network again ignored. In this paper, we propose a logistic regression model to determine the relationship between the likelihood of pipe failure and pipe characteristics on the water distribution systems for all pipe materials. Results of model will be apply for water distribution system in Hai Chau district, Da Nang city.

Keywords: water supply pipe; Logistic Regression Model; water distribution systems; R software.

1. GIỚI THIỆU

Mỗi ngày, các công ty cấp nước đều ghi nhận các sự cố bể vỡ trên tuyến ống đặc biệt là các tuyến ống dịch vụ và vẫn tồn thêm chi phí để tìm rò bể trên các tuyến ống cấp I, II [1]. Đôi khi chi phí để sửa chữa đoạn ống nhiều lần sẽ cao hơn so với việc lắp đặt ống mới [2]. Bên

cạnh đó hoạt động sửa chữa cũng làm ảnh hưởng đáng kể đến chất lượng phục vụ của hệ thống, chưa kể đến khả năng làm gián đoạn giao thông của khu vực đặt ống. Để giảm thiểu chi phí quản lý đồng thời ngăn chặn nguy cơ lan truyền bệnh qua hệ thống phân phối nước, cần phải có một kế hoạch bảo dưỡng hợp lý và thay thế kịp thời các đường ống có nguy cơ giảm hiệu suất làm việc hay không còn khả năng phục vụ. Một trong những cơ sở quan trọng để xây dựng quy trình này là dự báo khả

Ngày nhận bài: 27/6/2017

Ngày thông qua phản biện: 22/7/2017

Ngày duyệt đăng: 28/7/2017

năng ống bê trên hệ thống phân phối nước. Vậy nên, nghiên cứu được thực hiện với các mục tiêu sau:

- Đánh giá các yếu tố liên quan tới khả năng hỏng của ống dẫn nước cấp.
- Đề xuất mô hình dự báo ống bê trên hệ thống phân phối.

Nghiên cứu được giới hạn trong các nội dung sau:

- Bỏ qua đặc điểm hình học của điểm bê (bê do vết nứt dọc, vết nứt vòng hay vết nứt ngang).
- Ống dẫn làm việc trong điều kiện bình thường, không xét tới sự cố bất thường gây nên bê ống và các trường hợp bê tại phụ tùng đầu nối.
- Khu vực áp dụng là dự báo khả năng ống hỏng trên mạng lưới cấp nước quận Hải Châu, thành phố Đà Nẵng.

2. TỔNG QUAN CÁC NGHIÊN CỨU

Ống cấp nước bị bê là tập hợp của các yếu tố đã phân tích ở trên, để đánh giá mức độ ảnh hưởng của các yếu tố này bằng phương pháp thống kê tác giả thường sử dụng các mô hình số mũ, hồi quy tuyến tính, mô hình Bayesian, Poisson. Mỗi mô hình có một ưu điểm riêng và phù hợp với số liệu khảo sát.

Nhận định ban đầu cho rằng khả năng ống hỏng phụ thuộc vào thời gian làm việc, tác giả [3] đã đưa ra mô hình số mũ để mô tả hiện tượng này và mô hình này tiếp tục được phát triển bởi Ossman năm 2011 [4], từ lịch sử bê ống để dự báo xác suất hỏng trong tương lai tuy nhiên dữ liệu khảo sát chưa đề cập tới các yếu tố liên quan tới môi trường làm việc của ống như điều kiện áp suất, môi trường trong và ngoài ống. Một số nghiên cứu tiếp cận theo phương pháp mạng ANN để xác định hiệu suất làm việc của hệ thống cũng như tỉ lệ hỏng ống trong thời gian làm việc [5], ngoài ra còn đề xuất đưa phân tích phân tầng vào mô hình

ANN để tăng hiệu suất mô hình [6], kết quả cho thấy tuổi ống ảnh hưởng đến điều kiện đánh giá là 20,95%, vật liệu là 17,49 % và tỉ lệ bê ống 13,13%, mục đích cung cấp nước của từng đoạn ống cũng được cho là ảnh hưởng tới khả năng bê ống là 2,85%. Một trong những nhược điểm của mô hình ANN là thời gian chạy mô hình lớn và chi phí cho khảo sát số liệu đầu vào khá cao, vậy nên mô hình mới chỉ đáp ứng cho nghiên cứu học thuật.

Các mô hình hồi quy tuyến tính được áp dụng phổ biến cho dự báo khả năng làm việc của ống do nội dung phương pháp đơn giản và được Andreou, 1987 đánh giá đạt kết quả chính xác đến 70% [7], đồng thời tác giả cũng nhận định ống hỏng phụ thuộc vào đường kính, vị trí đặt ống, nguyên nhân gây bê và không phụ thuộc vào độ tuổi. Những kết luận đánh giá về tỉ lệ bê bị ảnh hưởng bởi lần vỡ thứ hai cũng được đưa ra từ mô hình tuyến tính, những ống bê sớm sau giai đoạn cài đặt sẽ có tỉ lệ nguy cơ cao hơn những ống bê ở giai đoạn sau này [8].

So sánh hiệu quả các mô hình hồi quy tuyến tính, Poisson, số mũ tác giả [9] đề xuất mô hình số mũ thời gian và Poisson là mô hình dự báo chính xác nguy cơ bê, đặc biệt là đếm được số lần không bê. Đường kính ống, vật liệu, khu vực bao phủ là rừng đất nông nghiệp, giao thông, và biến đo đếm được như nhiệt độ, lượng mưa, độ ẩm của đất là một trong những nguyên nhân gây nên bê vỡ trong hệ thống. Nghiên cứu này lại đưa ra kết luận khác với tác giả [8]: tỉ lệ bê vỡ tăng theo thời gian tính từ lần bê cuối cùng.

Hồi quy Bayesian là mô hình mới được áp dụng trong các nghiên cứu gần đây, với những ưu điểm của mô hình đã giúp bài toán có kết quả chính xác hơn các phương pháp trước. Đánh giá các yếu tố ảnh hưởng là các biến không chắc chắn [10] chạy mô hình tính được hệ số ảnh hưởng của điều kiện ống và môi trường xung quanh cho từng loại vật liệu. Kết quả cho thấy ảnh hưởng của các tính chất vật lý ống

dẫn quan trọng hơn tác động của môi trường đất ống, đặc biệt các vật liệu CI và DI có khả năng bền cao hơn do tác động của nhiệt độ và lượng mưa.

Phân tích các nghiên cứu trước đây, đề tài đưa ra một số kết luận sau:

- Để xác định được tỉ lệ ăn mòn trên các ống thực tế rất khó thực hiện nên các nghiên cứu hiện nay mới dừng ở hai vật liệu là ống gang và ống thép. Bên cạnh đó, đa số các đường ống cấp nước được chôn dưới mặt đất nên chi phí lấy mẫu sẽ rất tốn kém, các mẫu ống thường chỉ khảo sát trên một tuyến đường như vậy tính chất đại diện của mẫu nghiên cứu chưa cao. Vậy nên các kết luận về bền vỡ ống do ăn mòn hay không phải do ăn mòn cần phải được nghiên cứu nhiều hơn trong tương lai.

- Thông số đại diện cho ăn mòn là độ tuổi ống dẫn và đã được chứng minh là có mối tương quan giữa tuổi bền vỡ và khả năng xuất hiện bền ống, không những thế những ống bền sớm thì tuổi thọ ống dẫn sẽ ngắn hơn những ống bền muộn hơn. Yếu tố vật liệu ống, môi trường làm việc và vị trí đặt quan trọng hơn là độ tuổi khi xem xét khả năng bền vỡ.

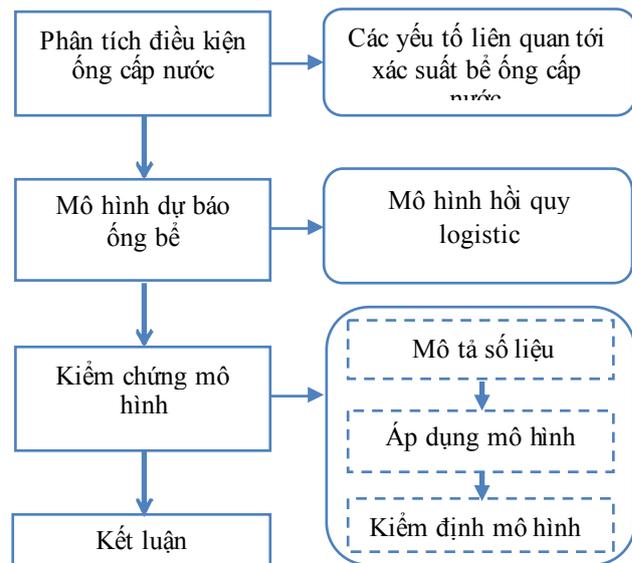
- Đường kính ống dẫn là một trong những thông tin đáng tin cậy để dự báo khả năng bền vỡ, các tác giả cho thấy tỉ lệ ống hỏng tăng theo đường kính với những ống có đường kính nhỏ 300mm, tuy nhiên kết quả này được lưu ý là còn phụ thuộc vào diện tích thành phố khảo sát. Áp suất làm việc trong ống cũng cần xem xét cụ thể hơn về mức độ ảnh hưởng tới khả năng bền vỡ.

Cải thiện kết quả của các nghiên cứu trước đây và phát triển mô hình dự báo ống bền trong điều kiện Việt Nam, tác giả đặt ra nội dung chính cần nghiên cứu như sau:

Nội dung 1: mức độ ảnh hưởng của các yếu tố tới sự kiện ống hỏng trên hệ thống cấp nước và trong các yếu tố đó thì yếu tố nào ảnh hưởng nhiều nhất tới sự kiện ống hỏng.

Nội dung 2: Dự báo xác suất hỏng của ống (P) làm việc trong môi trường thực tế.

Nghiên cứu được thực hiện qua các bước như hình 1. Nội dung nghiên cứu sẽ tập trung nhiều hơn tới đặc điểm hình học của mạng lưới ống dẫn cũng như vị trí tuyến đường đặt ống, tổng chiều dài trong một DMA của mạng.



Hình 1. Nội dung nghiên cứu dự báo ống hỏng trên HTPPN

Hệ thống cấp nước của Việt Nam bắt đầu số hóa khoảng từ năm 2013 và vẫn đang trong giai đoạn hoàn chỉnh, đồng bộ hóa thông tin, vậy nên trong quá trình thu thập dữ liệu có những số liệu không có khả năng thu thập cũng như thời gian thực hiện không cho phép nên giới hạn đề tài đặt ra là phân tích trên những số liệu thu thập được từ hệ thống thông tin GIS của hệ thống cấp nước. Vì vậy kết quả của đề tài có thể không đánh giá chi tiết được các điều kiện liên quan như chỉ số dẫn điện, độ ẩm của môi trường đất bên ngoài hay chỉ số oxy hóa tan của nước trong đường ống, giá trị lực tác dụng từ nền đường cho từng đoạn ống. Tuy nhiên mô hình đề xuất trong nghiên cứu lại có tính áp dụng thực tế tốt hơn, từ những dữ liệu GIS của công ty cấp nước hoàn toàn có thể ước lượng khả năng bền vỡ mà không phải

thu thập hay khảo sát cụ thể từng vị trí đặt ống trên mạng lưới.

Các yếu tố ảnh hưởng tới ống hỏng được xem xét trong đề tài bao gồm các đặc trưng vật lý của hệ thống (D, L, Mat), chỉ số độ tuổi (A) đại diện cho tỉ lệ ăn mòn và giá trị áp suất trung bình làm việc của ống (P). Ngoài ra các đại lượng đặc trưng cho đặc điểm hình học của mạng lưới bao gồm vị trí tuyến đường đặt ống (R), số đoạn ống kết nối trên một tuyến ống (N0), tổng chiều dài ống trong một DMA và số lần bẻ vỡ trước đây (Prior) cũng được khảo sát để đưa vào mô hình thống kê.

Độ tuổi ống dẫn (A)

Tuổi ống không phải là yếu tố ảnh hưởng trực tiếp tới nguy cơ ống bẻ nhưng là giá trị hiển thị dễ nhất để nhận biết khả năng làm việc của ống. Với cùng một độ tuổi nhưng điều kiện vật lý khác nhau và môi trường làm việc khác nhau thì xác suất xảy ra ống hỏng cũng không giống nhau. Nhưng nếu dùng độ tuổi để xem xét thời điểm xảy ra bẻ vỡ nhiều đột ngột trong một giai đoạn nhất định thì cũng có thể phát hiện ra những yếu tố ảnh hưởng đặc biệt hoặc dự báo một sự kiện bất thường trong tương lai. Vậy nên cần thiết phải đưa biến độ tuổi vào trong dữ liệu khảo sát.

Số lượng nút đầu nối trên một tuyến ống dẫn (N0)

Tồn thất cục bộ xuất hiện tại các vị trí chuyển hướng nước chảy và tồn thất này thường được bỏ qua trên mạng lưới đường ống dẫn nước và trong tính toán cũng cho thấy giá trị này không làm ảnh hưởng nhiều tới áp lực làm việc trên đường ống. Tuy nhiên nếu xét về mặt xáo trộn dòng chảy trong ống thì qua mỗi nút dòng chảy có sự xáo trộn nhất định. Nếu đánh giá tỉ lệ ăn mòn bề mặt trong của ống dẫn ngoài tác động của nồng độ hóa học và tính chất vật lý của nước thì xáo trộn dòng chảy giúp hóa chất được trộn đều hơn trên đường ống và giảm hiện tượng lắng cặn. Vậy nên số lượng điểm đầu nối trên một tuyến ống cũng ảnh hưởng tới

sự kiện ống hỏng.

Vật liệu ống dẫn (Mat)

Tính chất cơ học của vật liệu là khác nhau, vật liệu kim loại được cho là dễ ăn mòn hơn trong môi trường ẩm ướt trong khi vật liệu nhựa lại dễ bị lão hóa và khả năng chịu tải được cho là kém hơn [20]. Trong mạng lưới phân phối nước đô thị, những ống truyền tải thường sử dụng vật liệu kim loại như SS, ST đặc biệt là vật liệu gang xám (CI) trong thời gian trước đây được sử dụng phổ biến, nên đa số các nghiên cứu về ống hiện nay cũng tập trung xem xét vật liệu này. Tuy nhiên, nghiên cứu ống hỏng cần có dự báo rộng hơn trong tương lai nên hướng tiếp cận cần đa dạng hơn cho tất cả các loại vật liệu. Trong nghiên cứu này, các vật liệu trong số liệu thống kê đều được xem xét, bao gồm SS, ST, DI, CI, uPVC, HDPE và cả những ống AC (xi măng amiăng).

Đường kính ống dẫn D(mm)

Là đại lượng đặc trưng cho khả năng vận chuyển nước của đường ống, với nhận định ban đầu cho thấy xác suất ống hỏng sẽ giảm khi đường kính ống tăng. Theo yêu cầu sản xuất, chiều dày thành ống cũng tăng theo độ lớn của đường kính.

Áp lực làm việc trung bình trong DMA (m)

Áp lực làm việc thay đổi theo cấp ống phân phối và thời gian làm việc, không những thế áp lực ống dẫn cũng có thể thay đổi khi xảy ra áp va hoặc khu vực lắp đặt sửa chữa ống thường xuyên. Ngoài lực tác động lên thành ống, giá trị áp lực nước trong ống dẫn còn làm ảnh hưởng tới mức độ xáo trộn hóa chất và lắng cặn. Áp lực yếu hoặc bằng 0 sẽ tạo điều kiện cho cặn bám dính vào thành ống làm giảm khả năng làm việc, khi áp lực dao động lại gây nên hiện tượng bào mòn vật liệu ống [17]. Quá trình lặp lại nhiều lần theo thời gian làm bề mặt ống dẫn không còn trơn nhẵn như lúc mới lắp đặt, một số vị trí bị lồi do lắng cặn, một số vị trí khác lại bị bào mòn. Bề mặt vật liệu

ngày càng khả năng chịu tải khả năng bề ống

Chiều dài ống dẫn $L(km)$

Có nhiều quan điểm khác nhau khi dự báo xác suất hỏng theo chiều dài ống, nghiên cứu gần đây nhất cho rằng ống hỏng tăng khi chiều dài giảm nhưng có nghiên cứu lại cho rằng ống hỏng không tăng theo chiều dài mà tăng theo chiều dài bình phương. Trên phương diện thủy lực cho thấy, dọc theo chiều dài đường ống áp lực dòng chảy trong ống giảm dần như vậy nguy cơ hỏng ống do tải trọng nước bên trong ống cũng giảm. Tuy nhiên nếu xét tới các tác động từ bên ngoài thì đoạn ống càng ngắn, ngoại lực tác dụng lên thành ống càng nhỏ khi đó nguy cơ ống hỏng lại tỉ lệ thuận với chiều dài.

Tuyến đường đặt ống R

Mỗi tuyến đường đặt ống là một đại lượng đặc trưng cho môi trường đất xung quanh ống cũng như tải trọng tác dụng bên ngoài ống dẫn nước.

Tổng chiều dài đường ống trong một DMA

Mỗi DMA đặc trưng cho số lượng đối tượng phục vụ và khoanh vùng khu vực thủy lực, từ đó được sử dụng để kiểm soát rò rỉ trên hệ thống phân phối. Quy mô của mỗi DMA sẽ quyết định công suất phục vụ của từng đường ống, với cùng một loại đường kính và chiều dài ống nhưng số lượng đối tượng đầu nối nhiều thì chế độ thủy lực trên đường ống sẽ thay đổi nhiều hơn so với những ống có đối tượng đầu nối ít và điều này có ảnh hưởng đáng kể tới khả năng xảy ra ống hỏng.

Lịch sử bề ống (Prior)

Theo công bố các nghiên cứu trước đây, lịch sử bề vỡ có ảnh hưởng đáng kể tới khả năng ống hỏng trong tương lai, với những ống đã bị bề thì khả năng xảy ra lần bề tiếp theo sẽ cao hơn những ống chưa bề lần nào.

Như vậy cần xác định tương quan giữa 9 biến ảnh hưởng tới 1 biến ống hỏng :

$F \sim A + N0 + D + Mat + P + L + La + R + Prior$

Từ đó dự báo xác suất xảy ra ống hỏng P_f trên

mỗi đoạn ống khảo sát.

3. ĐỀ XUẤT MÔ HÌNH DỰ BÁO ỐNG HỎNG

Giới thiệu phần mềm R

Được phát triển từ năm 1996 bởi Ross Ihaka và Robert Gentleman [11] phần mềm R là một ngôn ngữ phân tích thống kê mới thay thế cho các phần mềm thống kê thương mại như SPSS, SAS, Stata, Statistica. Ngôn ngữ R nhanh chóng được các nhà thống kê học hưởng ứng và tham gia xây dựng các công cụ hỗ trợ cho phần mềm, sau hơn 10 năm R đã trở thành ngôn ngữ thống kê phổ biến cho các nhà nghiên cứu.

R là phần mềm có mã nguồn mở sử dụng cho thống kê và biểu diễn biểu đồ, có thể chạy trên môi trường Window, MacOS và UNIX. Phần mềm có thể đọc dữ liệu trực tiếp hoặc đọc file từ Excel (.csv); SAS; SPSS; Stata; text;.... Môi trường R có thể làm việc với các biến số cũng như biến ký tự chữ cái, mỗi đối tượng được tạo ra sẽ lưu trong bộ nhớ cho đến khi thoát khỏi môi trường R. Phần mềm linh hoạt trong quá trình xử lý số liệu thống kê bằng cách phân loại thành các nhóm, thay thế biến từ ký tự sang số và ngược lại, nhập tách dữ liệu hoặc loại bỏ các số liệu trống.

Với các ưu điểm trên, đề tài lựa chọn ngôn ngữ R để xây dựng hai mô hình hồi quy logistic từ đó dự báo xác suất ống hỏng và so sánh với dữ liệu thực tế.

Mô hình hồi quy logistic

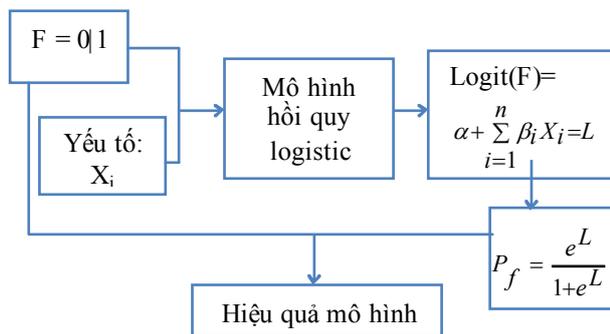
Trong các nghiên cứu về dự báo ống hỏng chủ yếu là phân tích tương quan giữa nguy cơ ống hỏng và các yếu tố ảnh hưởng tới nguy cơ hỏng. Đối tượng phân tích thường được thể hiện qua biến số nhị phân như hỏng/không hỏng, xảy ra/không xảy ra. Yếu tố ảnh hưởng là biến liên tục (tuổi ống, đường kính, chiều dài, số điểm đầu nối,...). Trong các nghiên cứu trước đây thường áp dụng mô hình tuyến tính bằng cách quy đổi biến ống hỏng về thành biến liên tục là số lần hỏng trên 1km ống trong

1 năm [12], tuy nhiên cách quy đổi này chưa xác định được xác suất hỏng cho từng đường ống mà chỉ dừng lại ở việc ước tính độ tương quan giữa yếu tố ảnh hưởng và tỉ lệ hỏng hỏng trên mạng lưới.

Hồi quy logistic là một thuật toán phân lớp được sử dụng để dự báo biến nhị phân, có thể coi đây là một trường hợp đặc biệt của hồi quy tuyến tính. Trong hồi quy logistic, biến Y chỉ có hai trạng thái 0 (ống không hỏng) và 1 (ống hỏng). D.R. Cox là người đầu tiên đưa ra mô hình hồi quy logistic vào năm 1969, mô hình được sử dụng phổ biến trong các nghiên cứu tự nhiên cũng như xã hội. Ý nghĩa mô hình này bao gồm:

- Mô tả mối liên quan giữa biến tiên lượng (A, N0, Mat, D, P, J, R, La, Prior là biến liên tục) và biến phụ thuộc (xác suất ống hỏng F là biến nhị phân), không tuân theo luật phân phối chuẩn.
- Kiểm soát biến nhiễu (Prior)
- Phát triển mô hình tiên lượng. Mô hình không sử dụng bình phương tối thiểu (OLS) để dự đoán tham số mà sử dụng ước lượng cực đại (MLE).

Có thể khái quát mô hình hồi quy logistic được áp dụng cho bài toán dự báo ống hỏng như sơ đồ sau:



Hình 2. Mô hình hồi quy logistic

Các thông số cơ bản của mô hình hồi quy logistic đó là yếu tố nguy cơ, chỉ số Odds và

Logit. Yếu tố nguy cơ là xác suất xảy ra một sự kiện theo thời gian và chỉ số khả năng Odds cho biết xác suất biến cố xảy ra trên xác suất biến cố không xảy ra

$$Odds = \frac{P}{1-P} = e^{\alpha + \sum_{i=1}^n \beta_i x_i}$$

Chỉ số “logit (P) = log (Odds)” là đại lượng đặc trưng của mô hình logistic. Mô hình này được viết cho một biến như sau:

$$\text{logit}(P) = \alpha + \sum_{i=1}^n \beta_i x_i$$

$$\text{hay } P = \frac{e^{\alpha + \sum_{i=1}^n \beta_i x_i}}{1 + e^{\alpha + \sum_{i=1}^n \beta_i x_i}}$$

Trong đó: α - Tỉ lệ hiện hành

β_i - Tỉ số liên quan tới một đơn vị tăng của X_i

Tuy nhiên trong thực tế, giá trị dự báo không chỉ phụ thuộc vào một yếu tố mà là tập hợp của nhiều biến nên phương trình của mô hình hồi quy logistic đa biến được áp dụng cho trường hợp này:

$$\text{logit}(P) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p = L \quad \text{vậy } \mu = \frac{e^L}{1 + e^L}$$

Dấu mũ μ để chỉ số ước tính chứ không phải là xác suất quan sát. Nếu mô hình mô tả số liệu đầy đủ thì khác biệt giữa P và μ nhỏ, nếu không thích hợp thì độ khác biệt này (deviance) có thể cao. Đánh giá mô hình hồi quy logistic bằng hai chỉ số :

- Tương tự như hệ số tương quan bình phương thì AIC là hệ số phạt để đo mô hình hồi quy logistic tối ưu. Mô hình nào có AIC nhỏ nhất là mô hình tốt nhất.

- Chỉ số Null deviance hay Residual deviance càng thấp thì cho mô hình càng tốt.

Kiểm chứng mô hình

Số liệu khảo sát được thu thập từ mạng lưới phân phối của quận Hải Châu (HC) thuộc

thành phố Đà Nẵng. Tổng chiều dài mạng lưới là 304,27 km ống trong đó ống uPVC có chiều dài lớn nhất là 80,6 km ống với tổng số lần bể vỡ ghi nhận là 162 cho tất cả các loại đường ống có đường kính lớn hơn hoặc bằng 100mm. Vật liệu có chiều dài ống ngắn nhất là SS (0,81km) với độ tuổi trung bình là 5,75 tuổi.

Số liệu thống kê bao gồm các yếu tố ảnh hưởng A, N0, Mat, D, P, L, R, La, Prior và số liệu ống hỏng F của quận HC là (1979 hàng

x10 cột). Nguyên nhân dẫn đến bể ống từ hồ sơ sửa chữa ống bể của công ty bao gồm bể do bản thân vật liệu ống, do yếu tố bên ngoài tác động và bể tại các nút đầu nối. Mục tiêu đề tài tập trung phân tích hỏng do ống bị lão hóa theo thời gian nên F=1 trong Bảng 1 chỉ bao gồm các số liệu ống hỏng do nguyên nhân này.

- Nếu ống hỏng tại độ tuổi A thì F=1, nếu không hỏng thì F=0 lúc này A là tuổi ống tính đến thời

Bảng 1. Dữ liệu mẫu sử dụng cho mô hình dự báo

F	A	N0	Mat	D	P	L	R	La	Prior
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	16.07	8	uPVC	100	8.9	0.002	1299	31.85	No
0	5.75	8	uPVC	100	8.9	0.000	1299	31.85	No
0	10.76	10	uPVC	150	8.9	0.000	1267	31.85	No
1	6.46	11	DI	100	16.27	0.074	1179	62.56	No
1	13.28	35	ST	150	8.06	0.012	1259	18.36	No
1	15.44	15	uPVC	300	12.81	0.137	1140	22.94	No
1	16.84	15	uPVC	300	11.83	0.137	1140	22.94	1st

- N0, L, R lần lượt là số điểm đầu nối, chiều dài ống dẫn, tuyến đường được đặt ống lấy từ dữ liệu GIS của công ty.

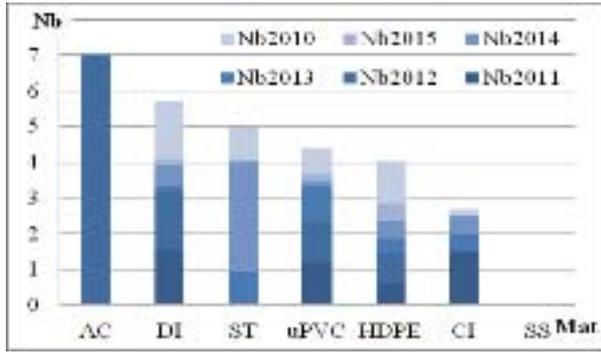
- Mat là vật liệu ống dẫn, bao gồm AC, CI, DI, HDPE, SS, ST và uPVC

- D là đường kính ống dẫn. Đề tài tập trung nghiên cứu các đường ống có đường kính $D \geq 100\text{mm}$ vì số liệu bể vỡ ghi nhận trên các ống có đường kính nhỏ hơn đa số là do khách hàng hoặc các công trình thi công gần tuyến ống.

- P là giá trị áp lực trung bình làm việc của đoạn ống được nội suy từ dữ liệu cảm biến áp lực trên mạng lưới.

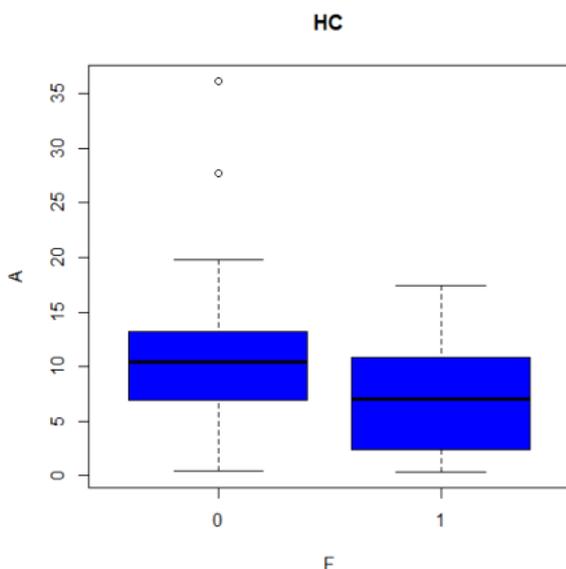
- Cột Prior ghi nhận nếu ống hỏng lần thứ nhất là No, nếu ống hỏng lần thứ 2 thì lịch sử bể vỡ Prior=1st tương tự cho ống hỏng lần thứ 3 thì Prior=2nd.

Số liệu khảo sát ống hỏng được ghi nhận trong 7 năm từ cuối năm 2010 đến tháng 4 năm 2016 cho thấy tỉ lệ ống bể trung bình của các loại vật liệu có sự khác nhau như Hình 3. Qua biểu đồ thống kê tỉ lệ bể vỡ qua các năm cho thấy giá trị này có xu thế giảm theo từng loại vật liệu ống, trong đó ống AC có tỉ lệ bể vỡ lớn nhất còn ống CI lại có số lượng bể vỡ trên mạng lưới nhỏ nhất. Số lần hỏng ống ghi nhận được trên mạng lưới lớn nhất là ống uPVC (79 lần bể trong 6 năm) tuy nhiên do tổng chiều dài vật liệu uPVC là lớn nhất trên mạng lưới nên tỉ lệ bể đối với vật liệu này chỉ đứng thứ 3, số liệu bể không được ghi nhận trên ống sắt SS. Các nghiên cứu trước đây cho thấy ống có vật liệu DI có khả năng bể cao hơn các vật liệu khác và kết luận này cũng đúng cho dữ liệu quận Hải Châu, Đà Nẵng.

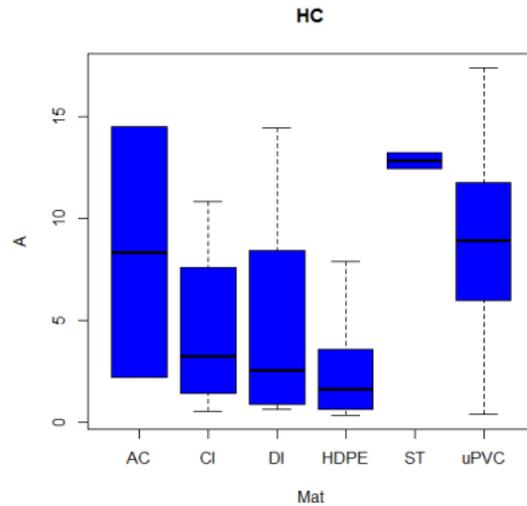


Hình 3. Thống kê số lần bể vỡ theo chiều dài ống trên mạng lưới

Từ số liệu thống kê, so sánh độ tuổi ống dẫn trong nhóm hỏng và không hỏng Hình 4 cho thấy tuổi ống trong nhóm hỏng ($F=1$) nhỏ hơn nhóm không hỏng ($F=0$), số liệu $F=1$ có phân diện tích 25% phía dưới trung vị lớn hơn, điều này chứng tỏ xu hướng ống hỏng ở độ tuổi trẻ nhiều hơn độ tuổi già, còn trong tập $F=0$ thì trung vị chia hai phần tương đối đồng đều. Số liệu ống không hỏng của quận HC xuất hiện hai giá giới hạn hộp không nằm trong giới hạn hộp, để xem xét hai giá trị này có phải giá trị ngoại vi (Outlier) hay không cần kiểm định số liệu để biết rõ hơn.



Hình 4. Biểu đồ so sánh độ tuổi (A) ống hỏng và không hỏng



Hình 5. Biểu đồ quan hệ giữa A~Mat quận HC

Những ống có thời gian làm việc lâu nhất so với các ống khác trên mạng lưới là vật liệu uPVC (HC:17,42. Được cho là vật liệu có độ bền cao và linh hoạt trong quá trình thi công lắp đặt, HDPE mới được đưa vào sử dụng phổ biến trong những năm gần đây, nhưng số liệu khảo sát lại cho thấy tuổi ống nhỏ nhất (0,36 tuổi) lại xuất hiện trên vật liệu này, bên cạnh đó xét tổng quát thì thời gian làm việc thực tế của ống cũng nhỏ hơn các vật liệu khác. Một phần nguyên nhân là do kỹ thuật thi công mỗi hàn không đúng kết hợp với tải trọng khu vực đặt ống và môi trường đất gây nên bể ống.

Số liệu bể cũng không tuyến tính với chiều dài mạng lưới, năm 2012 tỉ lệ bể vỡ ống tăng đột biến, nhưng đến năm 2015 lại giảm xuống mức thấp nhất trong các năm. Lượng ống bể giảm do mạng lưới được cải thiện về công nghệ cũng như phương pháp quản lí. Tuổi ống khảo sát lớn nhất tại quận HC là 36,5 tuổi và nhỏ nhất là 0,36 tuổi trong đó tuổi hỏng ống trung bình khoảng 10 tuổi.

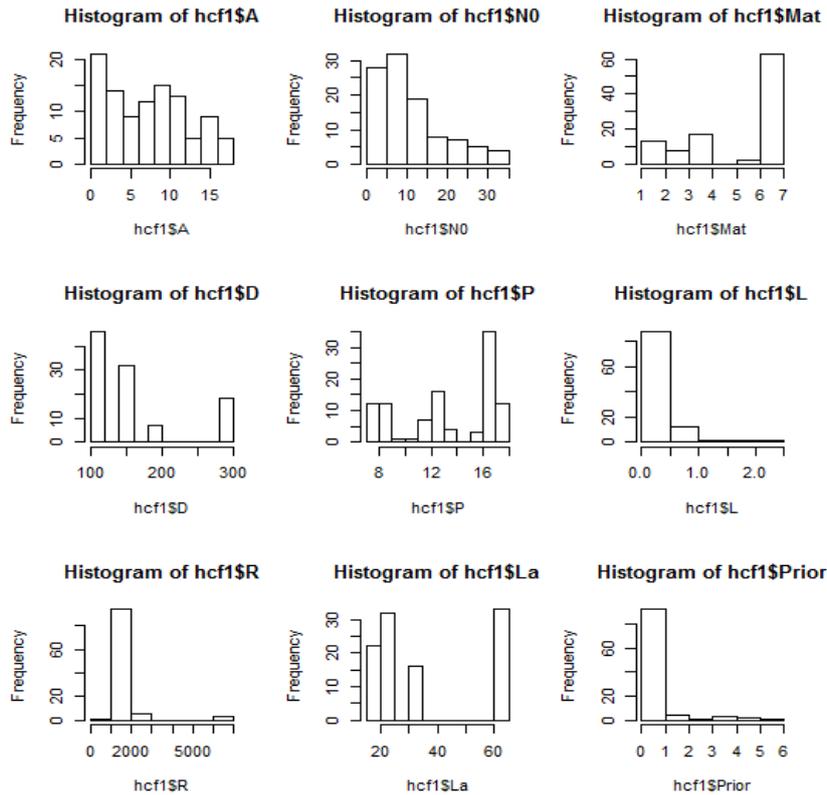
Biểu đồ phân bố mật độ dùng để đo tần số xuất hiện của một biến số từ đó thấy rõ sự thay đổi hay biến động của dữ liệu, qua đó có thể phát hiện sai sót trong số liệu đầu vào. Hình 6 cho thấy số lượng ống bể giảm từ đường kính 100-

200 mm nhưng lại tăng ở D300mm.

Kiểm chứng mô hình hồi quy logistic

Sử dụng hàm glm trong phần mềm R để phân tích số liệu ống hỏng và không hỏng. Nghiên cứu tiến hành phân tích hai biến đại diện cho

biến liên tục là độ tuổi (A) của ống dẫn và biến chữ (Mat) là vật liệu ống để phân tích ý nghĩa kết quả của mô hình đề xuất, sau đó ứng dụng mô hình hồi quy logistic cho tổ hợp các yếu tố ảnh hưởng và dự báo xác suất hỏng cho từng ống trong tập số liệu.



Hình 6. Biểu đồ phân bố các yếu tố của hệ thống cấp nước Hải Châu

Phân tích mức độ ảnh hưởng của biến tuổi thọ đến nguy cơ hỏng ống

Với F=0 thì độ tuổi ống dẫn được tính đến thời điểm khảo sát, số liệu ống hỏng F=1 thì độ tuổi A được ghi nhận tại thời điểm ống hỏng. Xem xét tương quan giữa biến F và A trong mô hình hồi quy logistic có kết quả như Bảng 2. Như vậy, khi độ tuổi tăng một độ lệch chuẩn thì khả năng ống bể ống giảm 37%. Qua mô hình có thể biểu diễn mối tương quan giữa tuổi ống và nguy cơ ống hỏng theo phương trình sau:

$$\text{Logit}(p) = -1,47 - 0,16.(A) \text{ Vậy } p = \frac{e^{-1,47-0,16(A)}}{1+e^{-1,47-0,16(A)}}$$

Theo đó, khi A=5 thì p=0,094 và A= 5 - sd(A)= 0,48 (giảm 1 độ lệch chuẩn) thì p=0,18 vậy xác suất hỏng ống tăng 0,48/0,094 = 5,13%.

Bảng 2. Kết quả mô hình logistic F~A

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.6242	-0.3768	-0.2872	-0.2271	2.9428

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.47323	0.21103	-6.981	2.93e-12 ***

A -0.16324 0.02471 -6.605 3.97e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 809.40 on 1978 degrees of freedom
Residual deviance: 762.47 on 1977 degrees of freedom
AIC: 766.47

Number of Fisher Scoring iterations: 6

Phân tích mức độ ảnh hưởng của biến vật liệu đến nguy cơ hỏng ống

Xem xét ảnh hưởng của vật liệu tới khả năng bể ống cấp nước được thực hiện và cho kết quả như Bảng 3.

Bảng 3. Kết quả mô hình logistic F~Mat

Deviance Residuals:

Min 1Q Median 3Q Max
-0.6039 -0.3686 -0.3686 -0.2126 3.0214

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.6094	0.7746	-2.078	0.0377
MatCI	-0.6371	0.8370	-0.761	0.4466
MatDI	-2.1691	0.8531	-2.542	0.0110
MatHDPE	-1.3381	0.8136	-1.645	0.1000
MatSS	-12.9566	624.1943	-0.021	0.9834
MatST	-2.9444	1.0513	-2.801	0.0051
MatuPVC	-1.0454	0.7855	-1.331	0.1832

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 809.40 on 1978 degrees of freedom
Residual deviance: 782.16 on 1972 degrees of freedom
AIC: 796.16

Number of Fisher Scoring iterations: 13

Chỉ số Pr >0.05 chỉ ra rằng sự khác biệt về nguy cơ bể vỡ ống của vật liệu CI, HDPE, SS và uPVC so với vật liệu AC. Tương tự như biến độ tuổi thì biến vật liệu cũng có quan hệ nghịch biến với khả năng bể vỡ ống. Mô hình so sánh khả năng hỏng của các loại vật liệu thấp hơn (ΔP) so với vật liệu AC bao

nhiều phần trăm, có thể tóm tắt kết quả trong bảng 4.

Bảng 4. Mô hình hồi quy logistic quận Hải Châu

Mat	Ước lượng hồi quy (β)	$\Delta P = 1 - e^{\beta}$
CI	-0.6371	47.12%
DI	-2.1691	88.57%
HDPE	-1.3381	73.77%
SS	-12.9566	100%
ST	-2.9444	94.74%
uPVC	-1.0454	64.84%

Phân tích tổng hợp ảnh hưởng của các biến đến nguy cơ hỏng ống

Ống hỏng là tổng hợp của nhiều yếu tố tác động nên không chỉ đánh giá tác động riêng lẻ của từng yếu tố mà cần một đánh giá tổng hợp để xem xét vấn đề toàn diện hơn. Kết quả chạy mô hình hồi quy logistic cho biến phụ thuộc F và các biến độc lập được thể hiện trong Hình 7.

Kết quả cho thấy độ lệch ước số hồi quy của biến Prior và N0 nhiều nhất và chỉ số ống hỏng tuyến tính với P, La là khác nhau. Tuy nhiên để đưa ra mức độ ảnh hưởng cụ thể của các biến số này tương đối phức tạp, cần có một phương pháp tốt hơn để đánh giá khả năng hỏng ống do yếu tố nào là chính, yếu tố nào là phụ.

Phương trình hồi quy cho quận Hải Châu được viết như sau:

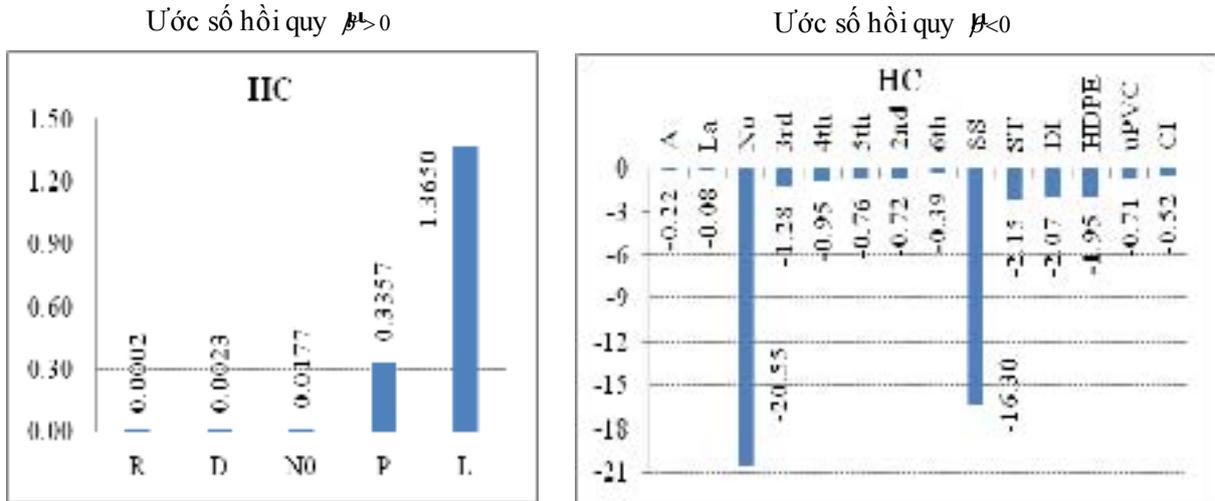
$$F = 18,13 + 0,0002.R + 0,002.D + 0,018.N0 + 0,34P + 1,37L - 0,22A - 0,08.La - 20,55.Prior(No/1st) - 1,28.Prior(2nd/1st) - 0,95.Prior(3rd/1st) - 0,76.Prior(4th/1st) - 0,72.Prior(5th/1st) - 0,39.Prior(6th/1st) - 16,30.Mat(SS/AC) - 2,15.Mat(ST/AC) - 2,072,15.Mat(DI/AC) - 2,15.Mat(HDPE/AC) - 0,712,15.Mat(uPVC/AC) - 0,522,15.Mat(CI/AC)$$

Áp dụng phương trình hồi quy cho các số liệu khảo sát để dự báo khả năng xảy ra ống hỏng

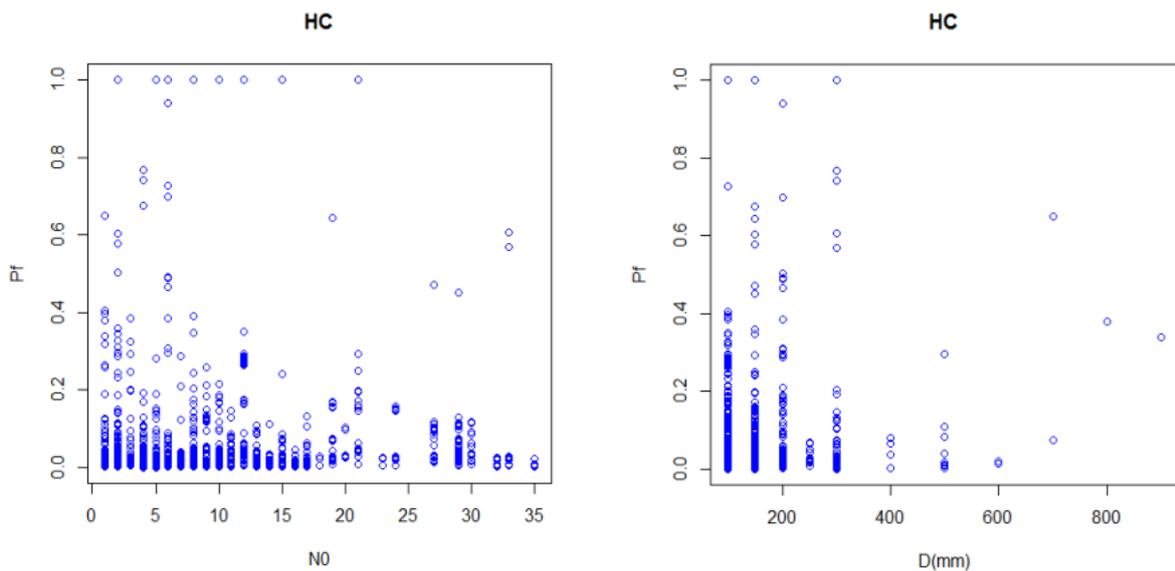
trên mạng lưới cấp nước, từ đó so sánh với số liệu hỏng thực tế để kiểm chứng hiệu quả của mô hình đề xuất. Dùng lệnh predict trong R cho mô hình hồi quy, kết quả dự báo hỏng cho từng đoạn ống như Hình 8.

Xác suất dự báo ống hỏng được tính cho từng số liệu trong dữ liệu khảo sát và dao động từ 0%-100%, kết quả được thể hiện trong Hình 8,

khả năng ống hỏng sẽ giảm trên các tuyến dài có số lượng điểm đầu nối ít. Xác suất xuất hiện hỏng ống tập trung tại các ống có đường kính nhỏ hơn hoặc bằng 400mm, cũng có một vài giá trị xuất hiện ở cỡ đường kính lớn nhưng đa số nguy cơ ống hỏng <50% (tính đến thời điểm khảo sát).



Hình 7. Biểu đồ so sánh ước số hồi quy của các yếu tố



Hình 8. Xác suất dự báo ống hỏng theo NO và D

Kiểm tra sự phù hợp giữa kết quả mô hình và số liệu thực tế

Bảng 5. Kết quả kiểm tra hỗn hợp xác suất ống hỏng dự báo và thực tế

Quận Hải Châu		Dữ liệu thực tế	
		Ống không hỏng (0)	Ống hỏng (1)
Mô hình dự báo	Ống không hỏng (0)	1871	73
	Ống hỏng (1)	5	30

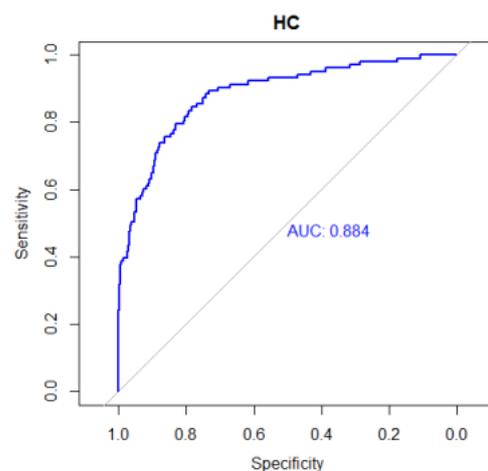
Trong số ống hỏng dự báo bằng mô hình hồi quy logistic thì bao nhiêu ống hỏng thật ở ngoài thực tế và bao nhiêu ống không hỏng lại được mô hình dự báo là hỏng, vấn đề này cần phải được kiểm tra để đánh giá hiệu suất của mô hình dự báo. Giới hạn được cho là mô hình dự báo chính xác là $P_f = 50\%$, kết quả được thể hiện trong Bảng 5.

Mô hình dự báo cho quận Hải Châu có 1871 ống không hỏng và dữ liệu thực tế cũng cho thấy các ống này không hỏng, có 5 ống không hỏng nhưng mô hình lại dự báo ống hỏng, dự báo sai 0,27%.

Các số liệu dự báo sai này của mô hình có được chấp nhận hay không sẽ được xác định theo hai tiêu chí là đường cong ROC (Receiver Operating Characteristic Curve) và chỉ số AUC.

Đường cong ROC là đường cong thể hiện độ nhạy (sensitivity) và độ đặc hiệu (specificity) của kết quả dự báo, một dự báo tốt khi diện tích dưới đường cong là lớn nhất. Chỉ số AUC (Area Under the Curve) là một giá trị biểu diễn diện tích dưới đường cong ROC được sử dụng

để đánh giá phân định giữa giá trị ống hỏng và không hỏng. Giá trị này được đánh giá theo giới hạn từ 0,5-1 như Bảng 6. AUC được tính toán bằng hàm auc trong phần mềm R cho thấy kết quả dự báo ống hỏng cho quận Hải Châu được chấp nhận ở mức độ tốt. Qua đó cho thấy mô hình hồi quy logistics áp dụng cho dự báo ống hỏng trên hệ thống phân phối là phù hợp.



Hình 8. Đường cong ROC cho HTTPN Hải Châu

Bảng 6. Diện tích dưới đường cong(AUC) để đánh giá hiệu suất của mô hình

AUC	Ý nghĩa diễn giải	AUC. HC
>0,9	Rất tốt	0.8843
0,8-0,9	Tốt	
0,7-0,8	Trung bình	
0,6-0,7	Không tốt	
0,5-0,6	Kết quả không sử dụng được	

KẾT LUẬN

Mô hình dự báo sử dụng cho tập số liệu quận Hải Châu cho thấy các biến liên quan tới bể ống bao gồm độ tuổi (A), vật liệu (Mat), đường kính (D), chiều dài ống trong một DMA (La), lịch sử bể vỡ (Prior), chiều dài ống dẫn (L), áp lực làm việc trung bình của từng đoạn ống (P) và số điểm đầu nối trên một tuyến (N0).

Mô hình hồi quy logistic áp dụng cho biến F với hai giá trị 1(hồng) và 0(không hồng) đã đưa ra mối quan hệ tuyến tính của biến này với các biến còn lại. Quan hệ này đồng biến với thông số chiều dài, áp lực làm việc, đường kính, vị trí đặt ống, số điểm đầu nối trên tuyến ống, tổng chiều dài ống trong một DMA và nghịch biến với các giá trị còn lại. Từ đó xác định xác suất xảy ra hồng ống trên mô hình, các giá trị dự báo được so sánh với số liệu thực tế cho thấy dự báo đúng 70%-84%.

Đề tài đã đưa ra kết quả dự báo ống hồng theo dữ liệu khảo sát thực và kiểm chứng độ chính xác bằng đường cong ROC và tiêu chuẩn AUC, mô hình logistic cho kết quả nằm trong giới hạn tốt

Để giảm khả năng ống bể cần tác động tới yếu tố độ tuổi, chiều dài ống, tổng chiều dài trong mộ DMA và áp lực làm việc. Với những ống đã bể vỡ trong lịch sử thì cần thiết phải xem xét phương án thay thế mới trong các lần hồng sau vì giá trị này có ảnh hưởng lớn nhất đến khả năng xảy ra bể ống.

Sử dụng số liệu thống kê thu thập được từ GIS của mạng lưới cấp nước quận Hải Châu, áp dụng mô hình dự báo nghiên cứu đã đề xuất được mối quan hệ giữa các thông số trong điều kiện làm việc của ống và sự cố ống hồng. Tuy nhiên, do hạn chế trong quá trình thu thập thông tin nên các biến độc lập chưa mô tả hết được các yếu tố liên quan đến biến phụ thuộc như các tải trọng xe chạy trên đường là bao nhiêu thì ảnh hưởng tới xác suất ống hồng hay chỉ số về điều kiện địa chất thủy văn tại từng vị trí đặt ống. Bên cạnh đó mô hình cũng không đánh giá cụ thể chiều dày còn lại hay tỉ lệ ăn mòn của từng đoạn ống theo độ tuổi vì giá trị này không có số liệu khảo sát thực tế.

TÀI LIỆU THAM KHẢO

- [1] A. Wood and B. J. Lence, "Using Water Main Break Data to Improve Asset Management for Small and Medium Utilities: District of Maple Ridge, B.C.," *J. Infrastruct. Syst.*, vol. 15, no. 2, pp. 111–119, 2009.
- [2] T. M. W. A. Pelliccia, "Water main repair/replacement for Binghamton, N. Y.," Vicksburg, Miss. 39180, 1981.
- [3] N. S. Bubbis, "Maintenance and Operating Problems of Winnipeg," *J. Am. Water Work. Assoc.*, vol. 41, no. 5, pp. 429–436, 1948.
- [4] H. O. and K. Bainbridge, "Comparison of Statistical Deterioration Models for Water Distribution Networks," *J. Perform. Constr. Facil.*, vol. 25, no. May, pp. 259–266, 2011.
- [5] H. Al-barqawi and T. Zayed, "Infrastructure Management : Integrated AHP / ANN Model to Evaluate Municipal Water Mains ' Performance," no. December, pp. 305–318, 2008.
- [6] D. Achim; F. Ghotb; and K. J. McManus, "Prediction of water pipe asset life using neural networks," *J. Infrastruct. Syst.*, vol. 13, no. 1, pp. 26–30, 2007.

- [7] T. Wengström, “Comparative analysis of Pipe Break Rates A Literature review,” *Chalmers Univ. Technol.*, 1993.
- [8] B. I. C. Goulter, “Analysis of water distribution pipe failure type in Winnipeg, Canada,” *J. Transp. Eng.*, vol. 115, no. 2, pp. 95–111, 1989.
- [9] S. Yamijala, “Statisticcal estimation of water distribution system pipe break risk,” Texas A&M University, 2007.
- [10] G. Kabir, S. Tesfamariam, and R. Sadiq, “Bayesian Model Averaging for the Prediction of Water Main Failure for Small to Large Candian Muncialities,” *Can. J. Civ. Eng.*, vol. 43, no. 3, pp. 233–240, 2016.
- [11] R. I. Gentleman, “R-paper.pdf,” *J. Comput. Graph. Statistics*, vol. 5, no. 3, pp. 299–314, 1996.
- [12] G. Kabir, S. Tesfamariam, J. Loeppky, and R. Sadiq, “Integrating Bayesian Linear Regression with Ordered Weighted Averaging: Uncertainty Analysis for Predicting Water Main Failures,” *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.*, vol. 1, no. 3, p. 4015007, 2015.