

## NGHIÊN CỨU ỨNG DỤNG THUẬT TOÁN HỌC MÁY ĐỂ DỰ BÁO KHAI THÁC CHO ĐỐI TƯỢNG MÓNG NÚT NẸ, MỎ BẠCH HỔ

Trần Đăng Tú<sup>1</sup>, Đinh Đức Huy<sup>1</sup>, Phạm Trường Giang<sup>1</sup>, Lê Quang Duyên<sup>2</sup>, Trần Xuân Quý<sup>1</sup>, Lê Thế Hùng<sup>1</sup>, Lưu Đình Tùng<sup>1</sup>

<sup>1</sup>Viện Dầu khí Việt Nam

<sup>2</sup>Đại học Mô Địa chất

Email: tutd@vpi.pvn.vn

<https://doi.org/10.47800/PVJ.2020.12-05>

### Tóm tắt

Các công cụ đang được sử dụng để dự báo sản lượng khai thác truyền thống cho đối tượng móng nút nẻ (như mô hình mô phỏng thủy động lực và phương pháp hệ đường cong suy giảm...) có độ tin cậy và hiệu quả dự báo chưa cao, mang tính ngắn hạn, ảnh hưởng đến kế hoạch phát triển, điều hành mỏ cũng như tối ưu hiệu quả thu hồi dầu.

Bài báo giới thiệu khả năng ứng dụng thuật toán học máy để dự báo khai thác cho đối tượng móng mỏ Bạch Hổ. Kết quả nghiên cứu cho thấy mô hình mạng neural nhân tạo (artificial neural network - ANN) sử dụng thuật toán lan truyền ngược và mô hình tăng trưởng logistic (logistics growth model - LGM) sử dụng thuật toán tối ưu đã nâng cao khả năng dự báo khai thác với mức độ chính xác cao.

**Từ khóa:** Trí tuệ nhân tạo, học máy, dự báo khai thác, mỏ Bạch Hổ, mô hình tăng trưởng.

### 1. Đặt vấn đề

Dự báo sản lượng đặc biệt cần thiết cho công tác quản lý - điều hành khai thác mỏ. Các phương pháp truyền thống sử dụng trong dự báo khai thác như phân tích đường cong suy giảm (DCA) và mô hình mô phỏng thủy động lực cho thấy hiệu quả rõ rệt đối với các đối tượng trầm tích hạt vụn như cát kết. Tuy nhiên, dự báo khai thác sử dụng các phương pháp trên cho thấy một số nhược điểm đối với đối tượng móng nút nẻ do đòi hỏi nhiều thời gian cho việc xây dựng, hiệu chỉnh từ mô hình địa chất tới thủy động lực học, xây dựng các kịch bản phát triển khác nhau. Hơn nữa, móng nút nẻ là đối tượng địa chất phức tạp, khó dự báo các đặc điểm địa chất, do vậy kết quả dự báo có thể không tin cậy do phụ thuộc nhiều thông số không chắc chắn.

Với mong muốn tiếp cận theo hướng đi mới và hạn chế các nhược điểm của phương pháp truyền thống và ứng dụng sự phát triển mạnh mẽ của các thuật toán hiện đại, nhóm tác giả thực hiện xây dựng công cụ dự báo sử dụng các thuật toán học máy. Bản chất của phương pháp

này là sử dụng các thuật toán học máy nhằm xác lập các mối quan hệ phi tuyến phức tạp giữa các thông số vận hành mỏ và sau đó thực hiện dự báo. Tuy nhiên, giải pháp này vẫn còn có nhược điểm như: không thể áp dụng cho các đối tượng mỏ chưa hoặc mới khai thác, được khuyến nghị áp dụng với các mỏ dầu khí trưởng thành, có dữ liệu tin cậy.

Một số các thuật toán hiện đại được nhóm tác giả sử dụng trong các mô hình mạng trí tuệ nhân tạo và mô hình tăng trưởng như: thuật toán lan truyền ngược và thuật toán tối ưu nhằm tăng năng lực cho dự báo khai thác ngắn hạn và dự báo khai thác dài hạn:

- Dự báo khai thác ngắn hạn:
  - + Phục vụ các công tác sản xuất, xây dựng kế hoạch khai thác định kỳ hàng năm của nhà điều hành;
  - + Đề xuất thực hiện các giải pháp can thiệp giếng (xử lý vùng cận đáy giếng, nứt vỉa thủy lực,...) nhằm duy trì gia tăng sản lượng khai thác.
- Dự báo khai thác dài hạn:
  - + Xây dựng phương án sản lượng, định hướng và tối ưu vận hành khai thác;
  - + Xây dựng kế hoạch phát triển mỏ (phương án can



Ngày nhận bài: 1/12/2020. Ngày phân biên đánh giá và sửa chữa: 1 - 22/12/2020.

Ngày bài báo được duyệt đăng: 22/12/2020.

thiệt vĩa: bơm ép nước, bơm ép hóa phẩm,...), tối ưu khai thác, hệ số thu hồi và gia tăng hiệu quả kinh tế.

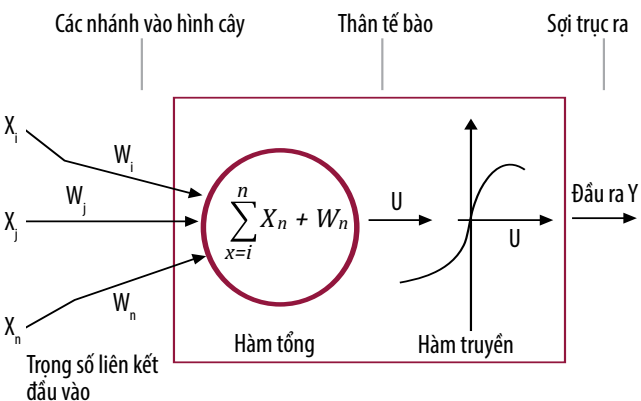
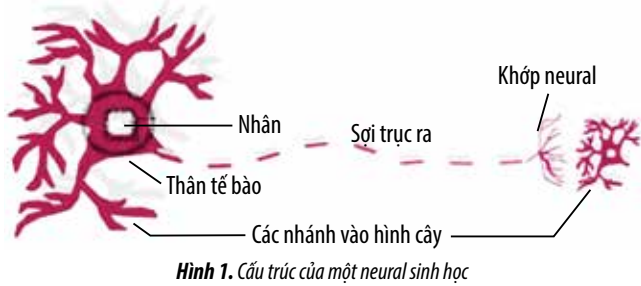
**2. Mô hình mạng trí tuệ nhân tạo và mô hình tăng trưởng logistic (LGM)**

**2.1. Mô hình mạng trí tuệ nhân tạo**

Trong khoa học máy tính, mô hình mạng trí tuệ nhân tạo là mô hình tính toán được xây dựng dựa trên các mạng neural sinh học gồm có một nhóm các neural nhân tạo (nút) nối với nhau và xử lý thông tin bằng cách truyền theo các kết nối và tính giá trị mới tại các nút. Trong nhiều trường hợp, ANN là hệ thống thích ứng tự thay đổi cấu trúc của mạng dựa trên các thông tin về dữ liệu lịch sử hoặc các kế hoạch tương lai trong quá trình học.

Trong thực tế sử dụng, nhiều mạng neural là các công cụ mô hình hóa dữ liệu thống kê phi tuyến, bản chất của hầu hết các bài toán trong thực tế. Thông thường, các nhà toán học sẽ tuyến tính hóa các bài toán phi tuyến để thu được kết quả gần đúng. Nhờ năng lực của máy tính, ANN có thể mô hình hóa các mối quan hệ dữ liệu phi tuyến phức tạp và tìm ra kết quả có độ chính xác cao cũng như tìm kiếm các dạng/mẫu của mỗi quan hệ trong dữ liệu. Hình 1 và 2 lần lượt biểu diễn cấu trúc mạng neural sinh học cơ bản và cấu trúc neural nhân tạo.

Mô hình ANN phổ biến nhất là cấu trúc mạng đa lớp (multi-layer perceptron, MLP) sử dụng thuật toán lan



truyền ngược. Cấu trúc MLP bao gồm ít nhất 3 lớp trong bộ xử lý được liên kết thông qua các kết nối có trọng số. Lớp đầu tiên bao gồm các vector đầu vào và lớp cuối chứa vector đầu ra. Các lớp ẩn mô tả các chuỗi neural và hiệu chỉnh dữ liệu đầu vào thông qua việc gán trọng số.

Có 3 giai đoạn chính để huấn luyện mạng với thuật toán lan truyền ngược. Trong giai đoạn 1, vector đầu vào hiển thị một mạng, được kích hoạt thông qua quá trình tính toán trực tiếp. Quá trình tạo ra sai số giữa dữ liệu đầu vào và dữ liệu đầu ra mong muốn của mạng. Trong giai đoạn 2, các sai số đầu ra được tính toán trở lại thông qua thuật toán lan truyền ngược. Đến giai đoạn 3, các trọng số kết nối được điều chỉnh bằng phương pháp tổng sai số bình phương bắt đầu từ lớp đầu ra, thông qua các lớp ẩn tới lớp đầu vào. Quá trình được lặp lại cho đến khi đạt được kết quả đầu ra mong muốn.

Lựa chọn một mô hình cấu trúc tối ưu là nhiệm vụ khó khăn yêu cầu một quá trình thử và tìm lỗi liên tục. Do đó, các mạng với nhiều lớp ẩn, thuật toán huấn luyện, các hàm kích hoạt sẽ được thử để dự báo các sai số tổng quát cho mỗi mạng. Mạng có sai số tổng quát dự báo nhỏ nhất sẽ được chọn.

**2.2. Mô hình tăng trưởng logistic**

Mô hình tăng trưởng logistic (LGM) được phát triển bởi nhà toán học người Bỉ Pierre Verhulst vào năm 1830 [7]. Đường cong tăng trưởng logistic là một tập hợp các mô hình toán học được sử dụng để dự báo dân số. Verhulst căn cứ vào ý tưởng của Malthus [8] - người tin rằng dân số của một quốc gia hoặc một khu vực cụ thể chỉ có thể tăng lên một mức nhất định. Verhulst đã lấy ý tưởng này bằng cách thêm một hệ số nhân vào phương trình tăng trưởng lũy tiến tạo ra mô hình tăng trưởng logistic.

Các LGM sau đó được ứng dụng trên nhiều lĩnh vực khác nhau như: vật lý, địa lý, hóa học... Bên cạnh mô hình tăng trưởng dân số, mô hình này còn được mô hình hóa sự tăng trưởng của nấm men, tái tạo các cơ quan và sự thâm nhập của các sản phẩm mới vào thị trường (Tsoularis và Wallace) [9]. Mô hình này còn được sử dụng trước đó trong ngành dầu khí dưới dạng mô hình Hubbert [10] để dự báo khai thác cho toàn mỏ hoặc vùng khai thác riêng biệt.

Mô hình được đề xuất sau đây là trường hợp đặc biệt của mô hình tăng trưởng logistic tổng quát. Mô hình này rất linh hoạt và có thể thích ứng với nhiều dạng đường cong khác nhau. Với mục đích để dự báo khai thác các giếng dầu và khí, mô hình được hiệu chỉnh có dạng:

$$Q(t) = \frac{Kt^n}{a + t^n} \quad (1)$$

Trong đó:

- Q: Sản lượng khai thác cộng dồn;
- K: Trữ lượng thu hồi cuối cùng (EUR);
- a: Hằng số;
- n: Số mũ hyperbolic;
- t: Thời gian.

### 3. Dữ liệu khai thác của đối tượng móng mỏ Bạch Hổ

Tại thời điểm bắt đầu đưa vào khai thác năm 1988, áp suất vỉa ban đầu của tầng móng mỏ Bạch Hổ tại độ sâu tuyệt đối 3.650 mSS đạt 417 at, theo kết quả đo khảo sát áp suất (2, 401, 402, 417). Ở giai đoạn đầu khai thác áp suất vỉa suy giảm mạnh, cơ chế suy giảm năng lượng tự nhiên và đàn hồi ảnh hưởng chính đến thân dầu khai thác. Do vậy, giải pháp bơm ép nước được áp dụng khi áp suất vỉa trung bình đạt 280 at tại phần đáy của thân dầu nhằm duy trì năng lượng và gia tăng hiệu quả thu hồi. Từ năm 1995, sau 2 năm bơm ép, tốc độ suy giảm áp suất vỉa dần dần ổn định. Tính đến ngày 31/5/2018, tổng lượng dầu khai thác từ đá móng là 217 triệu m<sup>3</sup> (180 triệu tấn) chiếm 86% tổng sản lượng dầu đã khai thác của Vietsovpetro, lưu lượng khai thác trung bình khoảng 6.000 tấn/ngày, độ ngập nước 60%.

### 4. Cấu trúc ANN và dự báo khai thác

Mạng neural có 5 thông số đầu vào và 3 thông số đầu ra. Các thông số đầu vào là lưu lượng khai thác dầu trung bình (FOPR) tại thời điểm t, lưu lượng khai thác chất lưu trung bình (FLPR) tại thời điểm t, áp suất vỉa trung bình (FPR) tại thời điểm t, lưu lượng bơm ép nước (FWIR) tại thời điểm t+1 và số lượng giếng khai thác (NP) tại thời điểm t+1. Thông số đầu ra là lưu lượng dầu trung bình tại thời điểm t+1, lưu lượng khai thác chất lưu trung bình và áp suất vỉa trung bình tại thời điểm t+1. Lựa chọn một mạng neural có cấu trúc tối ưu bằng phương pháp thử và phát hiện lỗi. Hình 3 là biểu đồ lưu lượng dầu từ tháng 9/1988 - tháng 5/2018.

### 4.1. Dự báo khai thác ngắn hạn

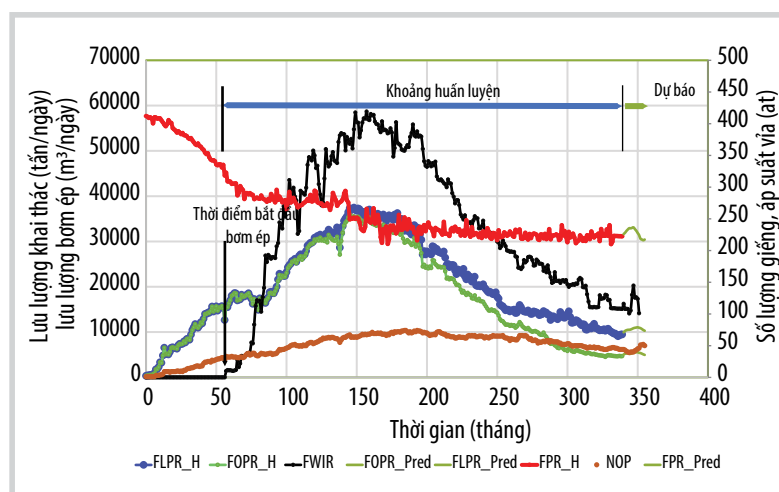
#### 4.1.1. Tiền xử lý dữ liệu

Thông thường, 1 mô hình mạng không thể chính xác nếu dữ liệu đầu vào không đầy đủ. Vì thế, trước khi huấn luyện mạng, dữ liệu khai thác phải đảm bảo độ tin cậy để tránh sai số lớn. Tuy nhiên, phụ thuộc vào các vấn đề gặp phải, một phần dữ liệu đầu vào có thể được sử dụng để kiểm tra chất lượng mô hình mạng. Một cách để kiểm tra chất lượng mô hình mạng là quan sát biểu đồ sản lượng đầu vào để lựa chọn giai đoạn phù hợp nhằm loại bỏ các dữ liệu nhiễu. Hình 3 thể hiện sản lượng khai thác dầu theo thời gian và được chia thành các giai đoạn. Tập dữ liệu I sử dụng 284 tháng dữ liệu (từ tháng 5/1993 - tháng 12/2016) để xây dựng mô hình cấu trúc. Tập dữ liệu II sử dụng 15 tháng dữ liệu (từ tháng 1/2017 - tháng 5/2018) để dự báo lưu lượng khai thác dầu, dự báo lưu lượng khai thác chất lưu và dự báo áp suất vỉa.

Để tránh hiện tượng mô hình mạng quá khớp hoặc không khớp lịch sử với tập huấn luyện và nâng cao chất lượng mô hình ANN, dữ liệu ban đầu được phân chia làm 3 giai đoạn: huấn luyện, xác thực và kiểm tra. Phần huấn luyện sử dụng 190 tháng dữ liệu (67%) để tính toán độ dốc, cập nhật trọng số và sai số của mô hình mạng. Phần xác thực sử dụng 47 tháng dữ liệu (16,5%) để kiểm tra chất lượng mạng trong quá trình huấn luyện. Quá trình huấn luyện có thể dừng khi kết quả mô hình mạng trong quá trình xác thực cho sai số nhỏ nhất. Phần kiểm tra sử dụng 47 tháng dữ liệu (16,5%) để tinh chỉnh mô hình mạng, không được sử dụng để huấn luyện và xác thực, chỉ để xác thực cấu trúc tối ưu, lựa chọn mô hình mạng phù hợp và đánh giá hiệu suất hoạt động mô hình mạng.

#### 4.1.2. Cấu trúc ANN

Nhóm tác giả lựa chọn kết quả tốt nhất từ mô hình ANN bao gồm 2 lớp ẩn, mỗi lớp ẩn có 50 neural. Các nút trong lớp ẩn và lớp đầu ra được kích hoạt thông qua hàm chuyển Sigmoid và được huấn luyện bằng thuật toán lan truyền ngược (back propagation neural network, BPNN).



Hình 3. Biểu đồ lưu lượng khai thác dầu từ tháng 9/1988 - tháng 5/2018

## 4.2. Dự báo khai thác dài hạn

### 4.2.1. Tiền xử lý dữ liệu

Tập dữ liệu I sử dụng 236 tháng dữ liệu (từ tháng 5/1993 - tháng 12/2012) để xây dựng mô hình cấu trúc mạng. Tập dữ liệu II sử dụng 60 tháng dữ liệu (từ tháng 1/2013 - tháng 12/2017) để dự báo lưu lượng khai thác dầu, dự báo lưu lượng khai thác chất lưu và dự báo áp suất vỉa.

Phần huấn luyện sử dụng 160 tháng dữ liệu (67%) để tính toán độ dốc (Gradient) và cập nhật trọng số và sai số của mạng. Phần xác thực sử dụng 38 tháng dữ liệu (16,5%) để đánh giá chất lượng mạng trong quá trình huấn luyện. Quá trình huấn luyện có thể dừng khi kết quả mô hình mạng trong quá trình xác thực cho sai số nhỏ nhất. Phần kiểm tra sử dụng 38 tháng dữ liệu (16,5%) để tinh chỉnh mô hình mạng, lựa chọn mô hình mạng phù hợp và đánh giá hiệu suất hoạt động mô hình mạng.

### 4.2.2. Cấu trúc ANN

Nhóm tác giả lựa chọn kết quả tốt nhất từ mô hình ANN bao gồm 1 lớp ẩn có 60 neural. Các nút trong lớp ẩn và lớp đầu ra được kích hoạt thông qua hàm chuyển Sigmoid và được huấn luyện bằng thuật toán lan truyền ngược.

## 5. Đánh giá kết quả dự báo khai thác sử dụng mô hình ANN trong dự báo khai thác ngắn hạn và dài hạn

### 5.1. Đánh giá kết quả dự báo khai thác của mô hình ANN để dự báo ngắn hạn

Các thông số thống kê sử dụng để tính toán khả năng dự báo của ANN đạt được từ quá trình huấn luyện, xác thực, kiểm tra và dự báo được tóm tắt trong Bảng 1. Bảng 1 cho thấy kết quả đánh giá sai số tuyệt đối (absolute error, AE) và sai số tương đối trung bình (average relative error, ARE) của 3 thông số: lưu lượng dầu khai thác, lưu lượng chất lưu khai thác, áp suất vỉa trung bình như sau:

+ Quá trình huấn luyện:

AE: 526 tấn/ngày, 637 tấn/ngày, 6 at;

ARE: 3,11%, 3,13%, 2,47%;

+ Quá trình xác thực:

AE: 998 tấn/ngày, 1112 tấn/ngày, 6,67 at;

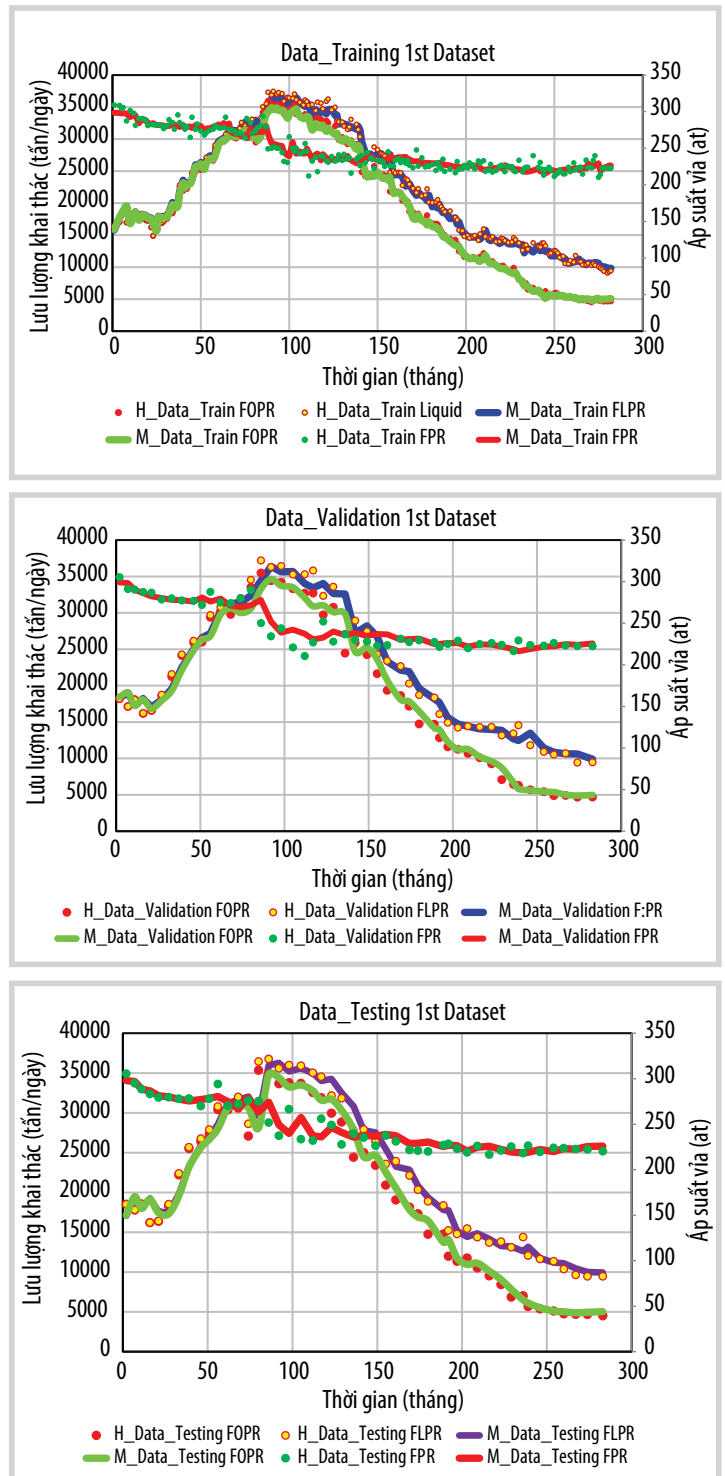
ARE: 5,51%, 5,26%, 2,76%;

+ Quá trình kiểm tra:

AE: 1157 tấn/ngày, 1165 tấn/ngày, 6,12 at;

ARE: 6,46%, 5,54%, 2,5%.

Các sai số này được đánh giá là thấp và dưới giới hạn cho phép. Kết quả các quá trình huấn luyện, xác thực và kiểm tra được biểu diễn trên Hình 4.



Hình 4. Kết quả quá trình huấn luyện, xác thực và kiểm tra

Để nghiên cứu, đánh giá mức độ chính xác và chất lượng của mô hình mạng, tập dữ liệu II được sử dụng để dự báo sản lượng khai thác dầu. Lưu lượng khai thác dầu được dự báo khớp với giá trị lịch sử cho thấy mô hình mạng được huấn luyện có thể sử dụng như một công cụ quản lý khai thác hiệu quả và thực tế (Hình 4). Sai số tuyệt đối của lưu lượng dầu khoảng 255 tấn/ngày và sai số tương đối trung bình là 4,82% (Bảng 1).

**5.2. Đánh giá kết quả dự báo khai thác của mô hình ANN để dự báo dài hạn**

Các thông số thống kê sử dụng để tính toán khả năng dự báo của ANN đạt được từ quá trình huấn luyện, xác thực, kiểm tra và dự báo được tóm tắt trong Bảng 2. Bảng 2 cho thấy các kết quả đánh giá sai số tuyệt đối và sai số tương đối trung bình của 3 thông số: lưu lượng dầu khai thác, lưu lượng chất lưu khai thác, áp suất vỉa trung bình như sau:

+ Quá trình huấn luyện:

AE: 553 tấn/ngày, 644 tấn/ngày, 5,25 at;

ARE: 2,79%, 2,78%, 2,1%.

+ Quá trình xác thực:

AE: 1001 tấn/ngày, 1025 tấn/ngày, 6,34 at;

ARE: 4,91%, 4,4%, 2,52%.

+ Quá trình kiểm tra:

AE: 1215 tấn/ngày, 1261 tấn/ngày, 7,69 at;

ARE: 5,6%, 5,43%, 3,13%.

Các sai số này được đánh giá là thấp và dưới giới hạn cho phép. Kết quả các quá trình huấn luyện, xác thực và kiểm tra được biểu diễn trên Hình 6.

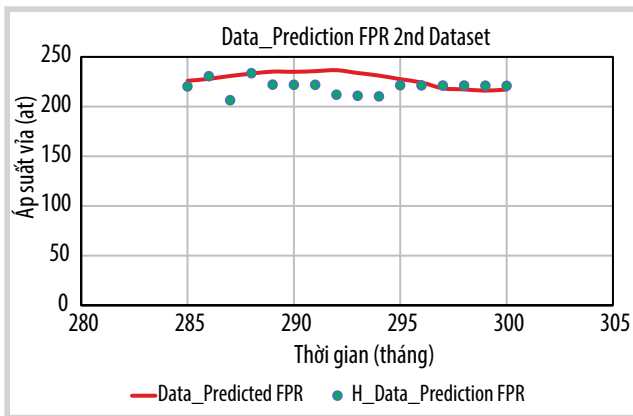
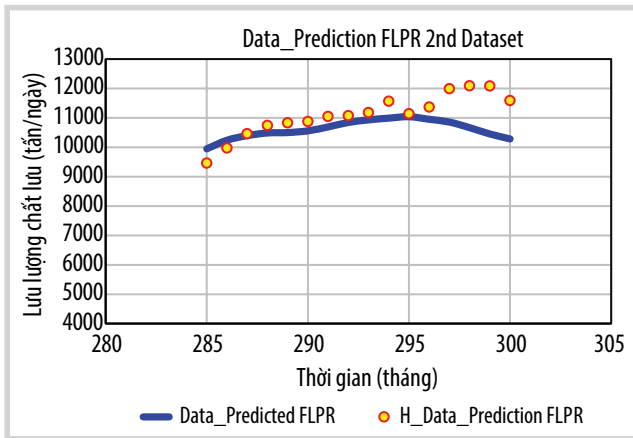
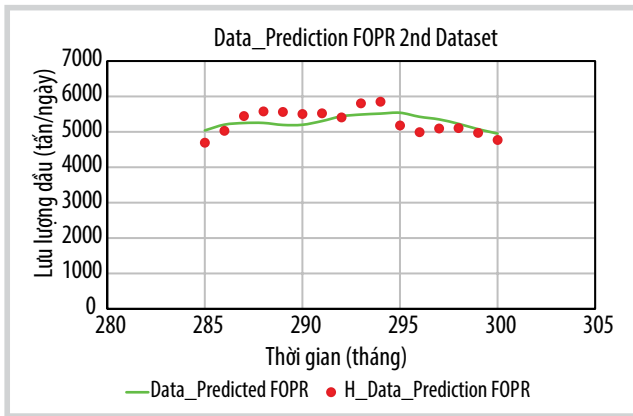
Để nghiên cứu, đánh giá mức độ chính xác và chất lượng của mô hình mạng, tập dữ liệu II được sử dụng để dự báo sản lượng khai thác dầu. Lưu lượng khai thác dầu được dự báo khớp với giá trị lịch sử cho thấy mô hình mạng được huấn luyện có thể sử dụng như một công cụ quản lý khai thác hiệu quả và thực tế (Hình 7). Sai số tuyệt đối của lưu lượng dầu khoảng 698 tấn/ngày và sai số tương đối trung bình là 12,61% (Bảng 2).

**6. Kết quả tái lập lịch sử của mô hình LGM**

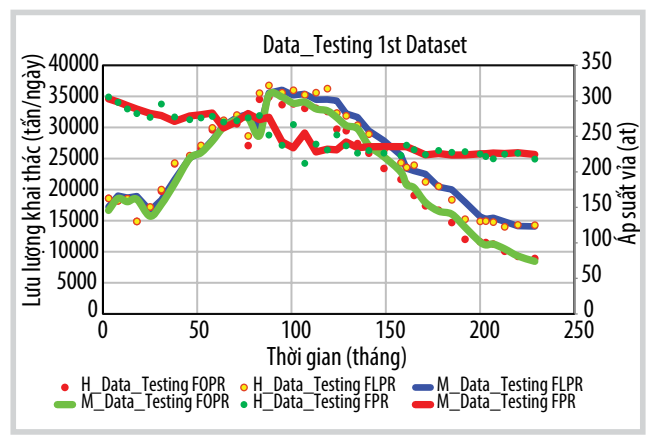
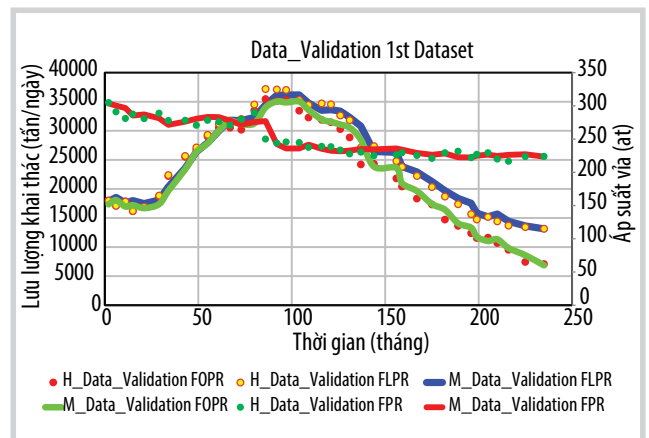
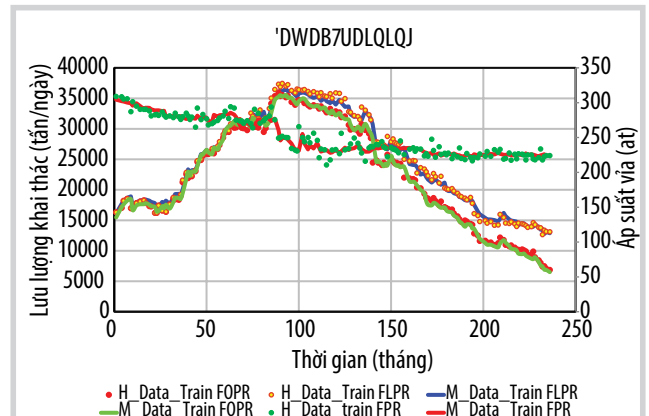
Năm 2019, Trần Đăng Tú và nnk đã “Nghiên cứu ứng dụng mô hình tăng trưởng logistic để dự báo khai thác cho tầng Miocene dưới mỏ Bạch Hổ” [6] sử dụng thuật toán tối ưu để tự động tái lập lịch sử khai thác dựa trên

*Bảng 1. Bảng thống kê đánh giá sai số của mô hình mạng trí tuệ nhân tạo dự báo ngắn hạn*

		FOPR_H	FOPR_ANN	AE1	ARE1 (%)	FLPR_H	FLPR_ANN	AE2	ARE2 (%)	FPR_H	FPR_ANN	AE3	ARE3 (%)
<b>Dữ liệu I</b>	<b>Huấn luyện</b>												
	Trung bình	19.523	19.421	526	3,11	22.410	22.273	637	3,13	245	245	6	2,47
	Độ lệch chuẩn	10.034	9.914	485	2,79	8.815	8.656	522	2,63	26	25	5,37	2,23
	Nhỏ nhất	4.521	4.765	4	0,01	9.081	9.719	10	0,04	210	216	0,08	0,03
	Lớn nhất	35.959	34.902	2.496	17,84	37.452	36.707	2.765	14,81	309	298	33,11	15,62
	<b>Xác thực</b>												
	Trung bình	19.469	19.558	998	5,51	22.289	22.270	1.112	5,26	245	247	6,67	2,76
	Độ lệch chuẩn	9.973	9.649	995	4,92	8.772	8.411	1.020	4,45	28	26	7,8	3,02
	Nhỏ nhất	4.669	4.901	15	0,23	9.438	9.902	20	0,14	211	216	0,16	0,06
	Lớn nhất	35.478	34.648	5.407	23,2	3.7204	36.382	5.480	20,22	305	299	27,86	12,69
	<b>Kiểm tra</b>												
	Trung bình	19.380	1.9523	1.157	6,46	22.216	22.250	1.165	5,54	246	247	6,12	2,5
	Độ lệch chuẩn	10.015	9.794	1.281	4,98	8.765	8.531	1.218	4,72	27	25	6,47	2,59
	Nhỏ nhất	4.525	4.900	6	0,11	9.464	9.947	65	0,19	217	218	0,02	0,01
	Lớn nhất	35.367	34.801	7.271	20,56	36.778	36.267	6.883	18,88	306	298	26,7	10,18
	<b>Dữ liệu II</b>	<b>Dự báo</b>											
Trung bình		5.280	5.277	254,5	4,82	11.095	10.615	573	5	220	228	10,38	4,83
Độ lệch chuẩn		350	173	111,78	2,13	726	312	504	4,13	7	7	8,68	4,15
Nhỏ nhất		4.692	4.951	30,54	0,56	9.464	9.947	66	0,63	206	216	0,35	0,15
Lớn nhất		5.848	5.538	434,59	8,71	12.097	11.041	1.637	13,54	234	237	24,78	11,83



Hình 5. Kết quả dự báo lưu lượng dầu, lưu lượng chất lỏng và áp suất vỉa trung bình (từ tháng 1/2017 - tháng 4/2018)



Hình 6. Kết quả quá trình huấn luyện, xác thực và kiểm tra

tập dữ liệu lịch sử khai thác. Kết quả dự báo 71 giếng khai thác đối tượng Miocene dưới mỏ Bạch Hổ cho thấy sai số tương đối trung bình giữa mô hình LGM và dữ liệu khai thác thực tế là 0,6%. Kết quả nghiên cứu cho thấy mô hình tăng trưởng logistic đã cải thiện khả năng dự đoán với độ tin cậy cao cho đối tượng Miocene. Từ nghiên cứu trên, nhóm tác giả tiếp tục thử nghiệm mô hình tăng trưởng logistic để tái lập lịch sử và dự báo khai thác cho đối tượng móng nứt nẻ mỏ Bạch Hổ.

Kết quả tái lập lịch sử đối tượng tầng móng mỏ Bạch Hổ như Hình 8 và 9.

Hình 8 và 9 cho thấy kết quả tái lập lịch sử khai thác trong 340 tháng và trong 292 tháng ở tầng móng tương đối tốt. Kết quả tái lập lịch sử trong 340 tháng cho sai số thấp và xu hướng đúng hơn kết quả khớp lịch sử trong 292 tháng. Từ kết quả đó, tiếp tục áp dụng mô hình LGM để dự báo khai thác ngắn hạn (16 tháng) và dài hạn (60 tháng).

**7. Đánh giá, so sánh kết quả dự báo khai thác sử dụng mô hình ANN, LGM và phần mềm OFM, mô hình thủy động lực học**

Từ Bảng 3 và 4 nhóm tác giả có một số nhận xét sau:

**Bảng 2.** Bảng thống kê đánh giá sai số của mô hình mạng trí tuệ nhân tạo dự báo dài hạn

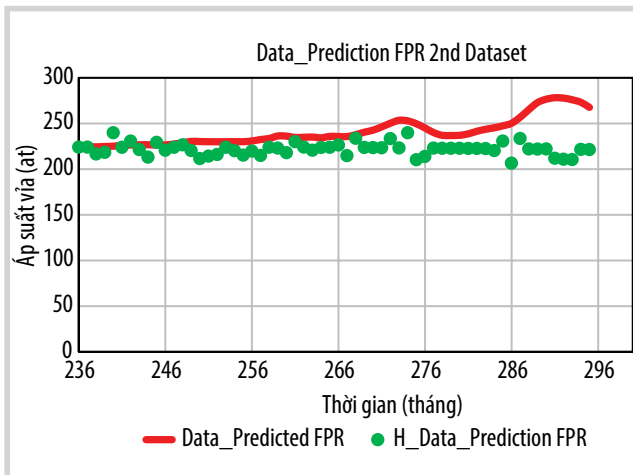
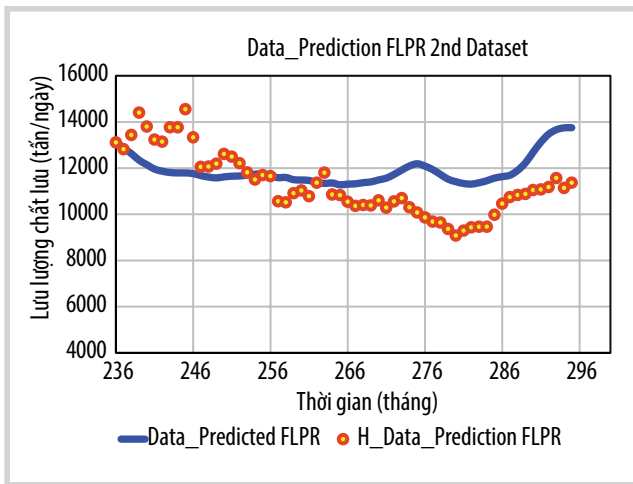
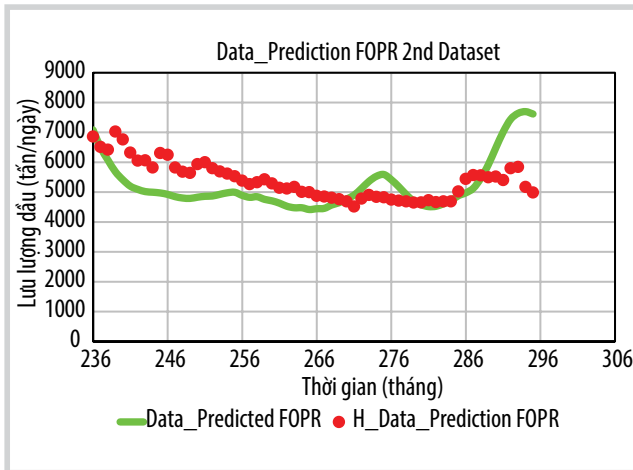
		FOPR_H	FOPR_ANN	AE1	ARE1 (%)	FLPR_H	FOPR_ANN	AE2	ARE2 (%)	FPR_H	PR_ANN	AE3	ARE3 (%)
<b>Dữ liệu I</b>	<b>Huấn luyện</b>												
	Trung bình	22.342	22.302	553	2,79	24.599	24.498	644	2,78	250	249	5,25	2,1
	Độ lệch chuẩn	8.365	8.571	478	2,45	7.820	7.475	551	2,45	27	26	5,41	2,14
	Nhỏ nhất	6.864	6.498	6	0,06	12.650	12.748	9	0,04	211	223	0,02	0,01
	Lớn nhất	35.959	35.569	3.941	16,11	37.452	36.606	3.693	13,62	309	304	34,37	12,94
	<b>Xác thực</b>												
	Trung bình	22.430	22.633	1.001	4,91	24.633	24.668	1.025	4,4	250	252	6,34	2,52
	Độ lệch chuẩn	8.729	8.666	844	3,97	8.149	76.44	794	3,45	27	27	5,01	1,95
	Nhỏ nhất	7.083	6.852	58	0,18	13.167	13.137	12	0,08	217	222	0,24	0,11
	Lớn nhất	22.430	22.633	1.001	4,91	24.633	24.668	1.025	4,4	250	252	6,34	2,52
	<b>Kiểm tra</b>												
	Trung bình	22.563	22.753	1.215	5,6	24.779	24.796	1.261	5,43	250	251	7,69	3,13
	Độ lệch chuẩn	8.236	8.300	1.307	5,68	7.804	7.367	1.216	5,53	27	26	9,02	3,89
	Nhỏ nhất	8.935	8.453	30	0,27	14.021	14.080	95	0,38	212	224	0,25	0,11
	Lớn nhất	34.899	35.255	5.887	23,91	36.778	36.052	5.963	27,26	306	303	42,64	20,11
	<b>Dữ liệu II</b>	<b>Dự báo</b>											
Trung bình		5.405	5.238	697,84	12,61	11.302	11.899	1.254	11,44	222	241	19,6	8,94
Độ lệch chuẩn		630	863	559,05	10,08	1.369	638	761	7,29	7	15	16,88	7,9
Nhỏ nhất		4.521	4.419	16,73	0,26	9.081	11.281	9	0,08	206	224	0	0
Lớn nhất		7.031	7.697	2628,28	52,67	14.553	13.755	2.758	25,71	240	278	66,92	31,7

- Mô hình ANN dự báo khai thác ngắn hạn có nhiều dữ liệu (284 tháng dữ liệu) được đưa vào xây dựng mô hình cấu trúc mạng sẽ có kết quả dự báo lưu lượng dầu, lưu lượng chất lưu, áp suất vỉa chính xác hơn mô hình ANN dự báo dài hạn có ít dữ liệu (236 tháng dữ liệu) được đưa vào xây dựng mô hình cấu trúc mạng (Hình 5, 7). Kết quả dự báo đường lưu lượng dầu ngắn hạn và dài hạn sử dụng mô hình ANN cho sai số tương đối trung bình lần lượt 10% và 5%. Kết quả dự báo đường lưu lượng dầu ngắn hạn và dài hạn được dự báo theo đúng xu hướng và có phản ánh được ảnh hưởng của các thông số vận hành như lưu lượng khai thác dầu, lưu lượng bơm ép, áp suất vỉa... Hơn nữa, mô hình mạng ANN có thể dự báo khai thác dựa trên tập dữ liệu đầu vào mà không phụ thuộc vào kinh nghiệm chủ quan của các chuyên gia nhờ chủ động xác định dựa trên các tập trọng số sau quá trình huấn luyện mạng. Việc dự báo sử dụng mô hình mạng ANN tự động xác định mối quan hệ liên quan giữa các tập dữ liệu đầu vào cho thấy kết quả dự báo này đáng tin cậy hơn so với phương pháp dự báo sử dụng mô hình LGM và phương pháp truyền thống. Bên cạnh đó, kết quả dự báo khai thác sử dụng mô hình ANN là quá trình huấn luyện không phải lúc nào cũng hội tụ và có khả năng rơi vào cực tiểu địa phương (local minimum). Nếu rơi vào trường hợp này, cần

phải huấn luyện mạng lại hoặc thay đổi số neural của lớp ẩn, điều này khiến cho mạng neural không áp dụng được cho các bài toán yêu cầu độ chính xác cao trong thời gian tối thiểu.

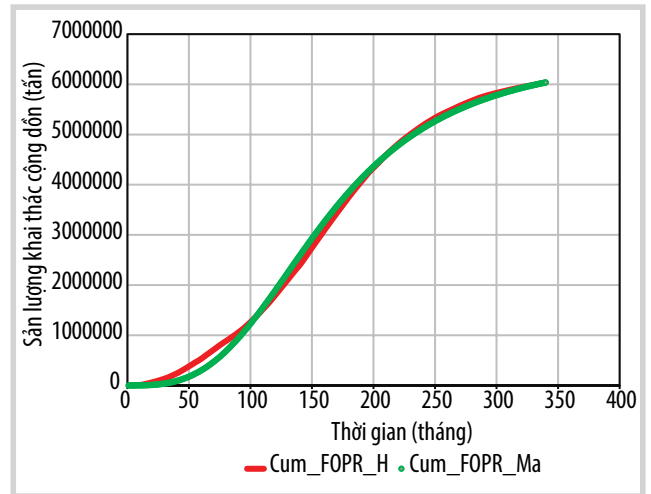
- Kết quả dự báo đường lưu lượng dầu ngắn hạn và dài hạn sử dụng mô hình LGM cho sai số tương đối trung bình lần lượt là 16% và 4%. Kết quả dự báo đường lưu lượng dầu dài hạn cho sai số tương đối trung bình lớn hơn kết quả dự báo đường lưu lượng ngắn hạn. Nguyên nhân là do mô hình LGM dự báo ngắn hạn có nhiều tháng dữ liệu (340 tháng) được đưa vào để tái lập lịch sử thì kết quả tái lập lịch sử và kết quả dự báo chính xác hơn mô hình LGM dự báo dài hạn với ít tháng dữ liệu hơn (292 tháng). Nói cách khác, nếu dữ liệu lịch sử đủ lớn thì hiệu suất dự báo mô hình LGM sẽ cải thiện rất nhiều. Bên cạnh đó, mô hình LGM chỉ là công cụ hỗ trợ dự báo nhanh và chính xác hơn công cụ dự báo bằng phần mềm OFM. Mô hình LGM cũng không thể dự báo chính xác được khi mở/giếng thay đổi cơ chế vận hành: mở côn, bơm ép gaslift, bơm ép nước, đóng giếng...

- Kết quả dự báo đường lưu lượng dầu ngắn hạn và dài hạn của mô hình thủy động lực học cho sai số tương đối trung bình lần lượt là 19% và 81%. Nguyên nhân chính do lưu lượng dầu có độ lệch lớn so với thực tế ngay khi bắt

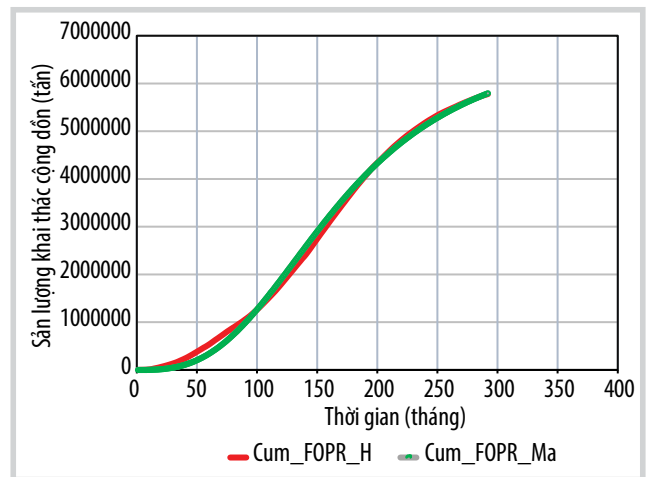


Hình 7. Kết quả dự báo lưu lượng dầu, lưu lượng chất lưu và áp suất vỉa trung bình (từ tháng 1/2013 - 31/1/2017)

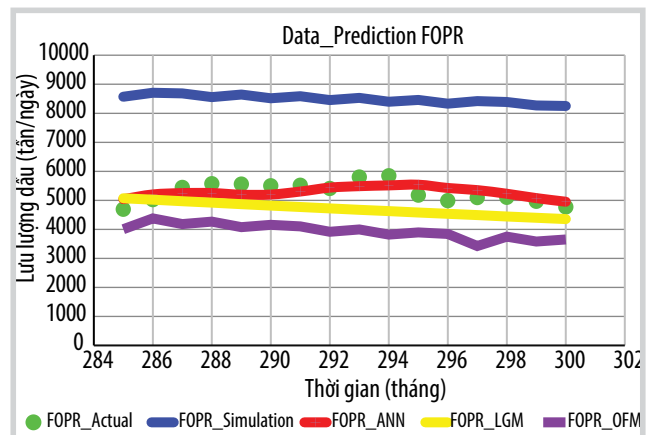
đầu thực hiện dự báo, điều này cho thấy những phức tạp về địa chất phân bố thuộc tính đá chứa và mạng lưới khe nứt cũng như tính liên thông thủy động phức tạp của tầng chứa móng nứt nẻ. Thực tế hiện nay vẫn chưa có phương pháp xây dựng mô hình mô phỏng đối tượng móng chính xác, đáng tin cậy và được công nhận rộng rãi.



Hình 8. Kết quả tái lập lịch sử trong 340 tháng



Hình 9. Kết quả tái lập lịch sử trong 292 tháng



Hình 10. Kết quả dự báo đường lưu lượng dầu ngắn hạn

- Kết quả dự báo đường lưu lượng dầu ngắn hạn và dài hạn sử dụng phần mềm OFM cho sai số tương đối trung bình lần lượt là 16% và 32%. Phương pháp dự báo đường cong suy giảm sử dụng phần mềm OFM cho thấy kết quả dự báo đường lưu lượng dầu ngắn hạn cho sai số tương đối trung bình thấp hơn kết quả dự báo đường

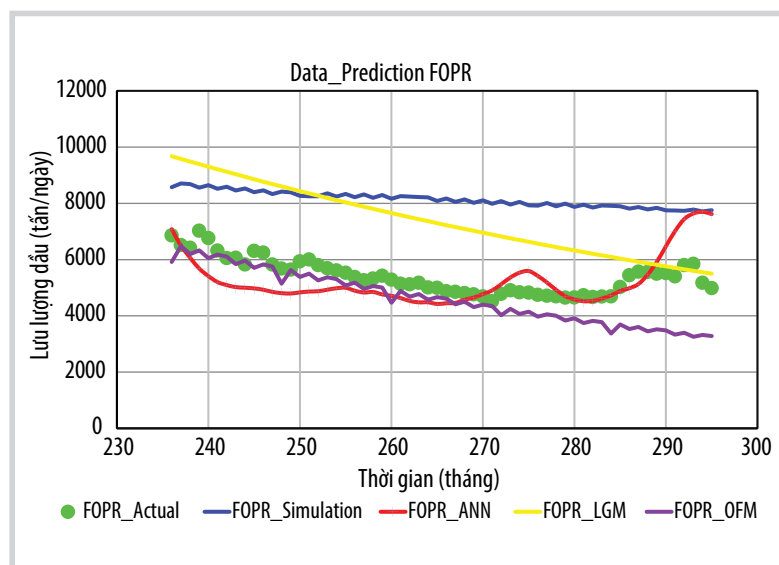


**Bảng 3.** Bảng thống kê đánh giá sai số của các mô hình dự báo khai thác ngắn hạn

Dự báo dài hạn	AE_ANN	ARE_ANN (%)	AE_Simulation	ARE_Simulation (%)	AE_LGM	ARE_LGM (%)	AE_OFM	ARE_OFM (%)
Trung bình	665	9,7	1.287	19	1.087	15,84	2.197	32
Độ lệch chuẩn	503	7,3	270	4	761	11,18	932	13
Nhỏ nhất	17	0,2	850	12	24	0,35	389	6
Lớn nhất	2.521	36,7	1.842	27	2.813	40,97	3.612	53

**Bảng 4.** Bảng thống kê đánh giá sai số của các mô hình dự báo khai thác dài hạn

Dự báo ngắn hạn	AE_ANN	ARE_ANN (%)	AE_Simulation	ARE_Simulation (%)	AE_LGM	ARE_LGM (%)	AE_OFM	ARE_OFM (%)
Trung bình	254	5,4	3.794	81	191	4,1	751	16
Độ lệch chuẩn	112	2,4	139	3	114	2,4	258	5
Nhỏ nhất	31	0,7	3.562	76	17	0,4	313	7
Lớn nhất	435	9,3	4.014	86	382	8,1	1.264	27



**Hình 11.** Kết quả dự báo đường lưu lượng dầu dài hạn

lưu lượng dài hạn. Tuy nhiên, kết quả dự báo đường lưu lượng dầu sử dụng phần mềm OFM chủ yếu mang tính chủ quan của người dự báo và không thể dự báo chính xác được khi mở/giếng thay đổi cơ chế vận hành: mở côn, bơm ép gaslift, bơm ép nước, đóng giếng...

**8. Kết luận**

Nghiên cứu cung cấp một số phương pháp mới dự báo khai thác trên tập dữ liệu lịch sử khai thác. Kết quả nghiên cứu cho thấy khả năng tổng quát hóa bài toán dự báo trên mô hình ANN thành công cụ hữu hiệu để có thể giải quyết hiệu quả nhiều bài toán khác nhau trong kỹ thuật khai thác mỏ. Mô hình ANN với nhiều đặc trưng: khả năng học từ dữ liệu, tính thích nghi, chịu lỗi khi dữ liệu không đầy đủ hoặc có nhiễu

là những lợi thế so với phương pháp dự báo sử dụng mô hình LGM và các phương pháp dự báo truyền thống.

Mô hình ANN sử dụng thuật toán lan truyền ngược đã chứng tỏ khả năng rất tốt cho nhiều bài toán phức tạp. Tuy nhiên, không có một mô hình chung về số lượng neural và sự hội tụ của mạng cho tất cả các bài toán. Để có khả năng ứng dụng hiệu quả cần có thời gian để đào tạo, điều chỉnh các tham số mạng.

Kết quả nghiên cứu ứng dụng LGM sử dụng thuật toán tối ưu để tự động tái lập lịch sử khai thác cho đối tượng móng nứt nẻ là tương đối tốt. LGM là một công cụ dự báo nhanh có thể thay thế phần mềm OFM với dữ liệu lịch sử đủ lớn. Tuy nhiên, LGM cũng không thể dự báo chính xác được khi mở/giếng thay đổi cơ chế vận hành: mở côn, bơm ép gaslift, bơm ép nước, đóng giếng...

**Tài liệu tham khảo**

[1] Q. Cao, R. Banerjee, S. Gupta, J. Li, W. Zhou, and B. Jeyachandra, "Data driven production forecasting using machine learning", *SPE Argentina Exploration and Production of Unconventional Resources Symposium, Buenos Aires, Argentina, 1 - 3 June 2016*. DOI: 10.2118/180984-MS.

[2] Yanan Li and Yifu Han, "Decline curve analysis for production forecasting based on machine learning", *SPE Symposium: Production Enhancement and Cost Optimisation, Kuala Lumpur, Malaysia, 7 - 8 November 2017*. DOI: 10.2118/189205-MS.

[3] A. Mirzaei-Paiamna and S. Salavati, "The application of artificial neural networks for the prediction of oil production flow rate", *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, Vol. 34, No. 19, pp. 1834 - 1843, 2012. DOI: 10.1080/15567036.2010.492386.

[4] David Fulford, "Machine learning for Production forecasting: Accuracy through uncertainty", *12<sup>th</sup> Annual Ryder Scott Reservoir Conference, Houston, TX, 14 September 2016*.

[5] Trần Văn Hồi, Nguyễn Văn Đức và Phạm Xuân Sơn, "Tìm kiếm thăm dò và phát triển dầu trong đá móng mỏ Bạch Hổ: Tư liệu, sự kiện và bài học kinh nghiệm", *Hội nghị khoa học kỷ niệm 30 năm khai thác dầu từ đá móng mỏ Bạch Hổ, Vũng Tàu, 6/9/2018*.

[6] Trần Đăng Tú, Đinh Đức Huy, Trần Xuân Quý, Phạm Trường Giang, Lê Vũ Quân, Lê Thế Hùng, Lê Quốc Trung và

Trần Nguyên Long, "Nghiên cứu ứng dụng mô hình tăng trưởng logistic để dự báo khai thác cho tầng Miocene dưới mỏ Bạch Hổ", *Tạp chí Dầu khí*, số 9, tr. 16 - 22, 2019.

[7] Pierre-François Verhulst, "Notice sur la loi que la population poursuit dans son accroissement", *Correspondance Mathématique et Physique*, Vol. 10, pp. 113 - 121, 1838.

[8] Thomas Robert Malthus, *An essay on the principle of population: or, a view of its past and present effects on human happiness; with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions*. Biodiversity Heritage Library (BHL), 1872. DOI: 10.5962/bhl.title.49216.

[9] A. Tsoularis and J. Wallace, "Analysis of logistic growth models", *Mathematical Biosciences*, Vol. 179, No. 1, pp. 21 - 55, 2002. DOI: 10.1016/S0025-5564(02)00096-2.

[10] M.King Hubbert, *Nuclear energy and the fossil fuel*. Drilling and Production Practice, New York. 1956.

## RESEARCH ON APPLICATION OF MACHINE LEARNING ALGORITHM TO FORECAST PRODUCTION FOR FRACTURE BASEMENT - BACH HO FIELD

**Tran Dang Tu<sup>1</sup>, Dinh Duc Huy<sup>1</sup>, Pham Truong Giang<sup>1</sup>, Le Quang Duyen<sup>2</sup>, Tran Xuan Quy<sup>1</sup>, Le The Hung<sup>1</sup>, Luu Dinh Tung<sup>1</sup>**

<sup>1</sup>Vietnam Petroleum Institute

<sup>2</sup>Hanoi University of Mining and Geology

Email: tutd@vpi.pvn.vn

### Summary

Conventional tools that are currently used to forecast production for fracture basement (such as simulation model and decline curve analysis) are still not highly reliable and their forecasting performance is still short-term, affecting the plan for field development, field operation and optimisation of oil recovery.

The paper introduces the applicability of machine learning algorithm to predict oil production for basement reservoirs of Bach Ho field. The research results show that the artificial neural network (ANN) model using reverse propagation algorithm and the logistic growth model (LGM) using optimisation algorithm have improved the ability to predict production with high accuracy.

**Key words:** Artificial neural network, machine learning, forecasting production, Bach Ho field, logistic growth model.