

SỬ DỤNG MẠNG NEURAL NHÂN TẠO ĐỂ DỰ ĐOÁN HỆ SỐ Z CHO KHÍ HYDROCARBON THIÊN NHIÊN

Trần Khả Tiến, TS. Hoàng Thịnh Nhân

Đại học Dầu khí Việt Nam

Email: tientk01@pvu.edu.vn

Tóm tắt

Hệ số nén khí hay hệ số lệch khí (hệ số Z) là thông số quan trọng để xác định tính chất nhiệt động của vỉa khí, tỷ trọng khí, độ nhớt khí, độ nén khí, mô phỏng vỉa khí, tính toán cân bằng vật chất, dự đoán PVT cho giếng dầu và khí... Dựa trên dữ liệu từ biểu đồ Standing-Katz, một số phương pháp đã được phát triển để tính hệ số Z [1 - 3]. Dựa trên phương pháp của Mohammadreza Kamyab [4], nhóm tác giả đã tính toán, xác định hệ số Z bằng mạng neural nhân tạo (ANN) với thông số đầu vào gồm áp suất giả giảm và nhiệt độ giả giảm của 5.940 điểm dữ liệu thực nghiệm [5]. Kết quả nghiên cứu cho thấy mô hình này có thể dự đoán chính xác hệ số Z hơn so với các phương pháp khác và có khả năng áp dụng trên phạm vi nhiệt độ giả giảm $1,05 \leq T_{pr} \leq 3$ và áp suất giả giảm $0,2 \leq P_{pr} \leq 15$.

Từ khóa: Hệ số nén khí, hệ số Z, MATLAB, mạng neural nhân tạo.

1. Giới thiệu

Khí thiên nhiên là một hỗn hợp đa thành phần, trong đó chủ yếu là methane (CH_4), ngoài ra còn có ethane (C_2H_6), propane (C_3H_8), carbon dioxide (CO_2), nitrogen (N_2) và thành phần hydrocarbon nặng hơn [6]. Dự đoán chính xác hệ số Z giúp giải quyết các vấn đề về kỹ thuật dầu khí, thiết kế, phân tích khai thác khí tự nhiên và xử lý hệ thống [7]. Do đó, cần phát triển phương pháp tin cậy để dự đoán hệ số Z từ các thông số như nhiệt độ, áp suất của hỗn hợp khí thiên nhiên.

Trong phương trình khí thực (khí thiên nhiên), hệ số Z có thể nhận được bởi công thức sau [8]:

$$PV = ZnRT \quad (1)$$

Với:

P: Áp suất (psia);

V: Thể tích (ft^3);

n: Số lượng mol khí tổng (mol);

R: Hằng số khí = $10.732 \text{ psia}\cdot\text{ft}^3/\text{lb}\cdot\text{mole}\cdot^\circ\text{R}$.

Lý thuyết về trạng thái tương đương [9, 10] chỉ ra, hệ số Z có thể được định nghĩa như một hàm số của áp suất và nhiệt độ giả giảm. Áp suất và nhiệt độ giả giảm được định nghĩa bởi những phương trình sau [8]:

$$T_{pr} = \frac{T}{T_{pc}} \quad (2)$$

$$P_{pr} = \frac{P}{P_{pc}} \quad (3)$$

Nhiệt độ giả tới hạn (pseudo critical) T_{pc} và áp suất giả tới hạn P_{pc} được tính như sau:

$$P_{pc} = \sum_{i=1}^n P_{ci} y_i \quad (4)$$

$$T_{pc} = \sum_{i=1}^n T_{ci} y_i \quad (5)$$

Trong đó:

P_{ci} : Áp suất tới hạn của thành phần i (psia);

T_{ci} : Nhiệt độ tới hạn của thành phần i ($^\circ\text{R}$);

y_i : Phần mol của thành phần i.

Có thể sử dụng biểu đồ Standing và Katz hoặc các phương pháp thực nghiệm để xác định hệ số Z nếu biết thành phần của khí tự nhiên, áp suất và nhiệt độ của khí. Một số nhà nghiên cứu đã tạo ra phương trình toán học để tính hệ số Z phù hợp với biểu đồ Standing và Katz như: Hall và Yarborough [8]; Dranchuk và Abou Kassem [1]; Dranchuk, Purvis và Robinson [2]; Brill và Beggs [3]. Tuy nhiên, những phương trình đó có những hạn chế như: chưa dự đoán chính xác hệ số Z so với dữ liệu thực nghiệm hoặc chỉ dự đoán trong một phạm vi nhất định.

Bài báo này giới thiệu phương pháp tính hệ số Z chính xác bằng cách sử dụng mạng neural nhân tạo với 5.940 điểm dữ liệu thực nghiệm sử dụng phần mềm MATLAB. Mạng neural có thể tính chính xác hệ số Z từ những thông số giả giảm với phạm vi nhiệt độ giả giảm ($1,05 \leq T_{pr} \leq 3$) và áp suất giả giảm ($0,2 \leq P_{pr} \leq 15$), như được chứng minh trong nghiên cứu này bởi so sánh kết quả nhận được với những phương pháp khác.

2. Một số phương pháp tính trực tiếp hệ số Z

Có 3 phương pháp ẩn (implicit) phổ biến nhất để tính hệ số Z như sau:

- Phương pháp Hall và Yarborough [8];
- Phương pháp Dranchuk và Abou-Kassem [1];
- Phương pháp Dranchuk, Purvis và Robinson [2].

2.1. Phương pháp Hall và Yarborough

Phương pháp này là một sự biến đổi của phương trình trạng thái Carnahan-Starling, với các hệ số được phát triển qua sự hồi quy và 1.500 điểm dữ liệu lấy từ biểu đồ hệ số Z ban đầu của Standing và Katz [8]:

$$Z = \frac{A_1 P_{pr}}{y} \quad (6)$$

Với y là nghiệm của phương trình sau:

$$-A_1 P_{pr} + \frac{y + y^2 + y^3 - y^4}{(1-y)^3} - A_2 y^2 + A_3 y^{A_4} = 0 \quad (7)$$

Với:

$$\begin{aligned} A_1 &= 0,06125te^{-1,2[t-t]^2} \\ A_2 &= 14,7t - 9,76t^2 + 4,58t^3 \\ A_3 &= 90,7t - 242,2t^2 + 42,4t^3 \\ A_4 &= 2,18 + 2,82t \\ t &= \frac{1}{T_{pr}} \end{aligned}$$

Phương trình (7) là một phương trình không tuyến tính và cần sử dụng kỹ thuật lặp Newton-Raphson hoặc phương pháp chia đôi để giải. Hall và Yarborough chỉ ra rằng phương pháp này không nên áp dụng nếu nhiệt độ giả giảm < 1 [8].

2.2. Phương pháp Dranchuk và Abou-Kassem

Phương pháp này là một biến đổi 11 hệ số của phương trình trạng thái Benedict-Wedd-Rubin. Những hệ số được tính sử dụng một phương pháp hồi quy với 1.500 điểm dữ liệu lấy từ biểu đồ của Standing và Katz [1]:

$$Z = \frac{0,27P_{pr}}{yT_{pr}} \quad (8)$$

Trong đó:

P_{pr} : Áp suất giả giảm (pseudo-reduced pressure);

T_{pr} : Nhiệt độ giả giảm (pseudo-reduced temperature);

y là nghiệm của phương trình sau:

$$\left[R_5 y^2 (1 + A_{11} y^2) e^{(-A_{11} y^2)} \right] + R_1 y - \frac{R_2}{y} + R_3 y^2 - R_4 y^5 + 1 = 0 \quad (9)$$

Với:

$$R_1 = A_1 + \frac{A_2}{T_{pr}} + \frac{A_3}{T_{pr}^3} + \frac{A_4}{T_{pr}^4} + \frac{A_5}{T_{pr}^5}$$

$$R_2 = \frac{0,27P_{pr}}{T_{pr}}$$

$$R_3 = A_6 + \frac{A_7}{T_{pr}} + \frac{A_8}{T_{pr}^2}$$

$$R_4 = A_9 \left(\frac{A_7}{T_{pr}} + \frac{A_8}{T_{pr}^2} \right)$$

$$R_5 = \frac{A_{10}}{T_{pr}^3}$$

Trong đó $A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11}$ là những hệ số như trong Bảng 1.

Bảng 1. Những hệ số của phương pháp Dranchuk và Abou-Kassem

A_1	A_2	A_3	A_4	A_5	A_6
0,3265	-1,070	-0,5339	0,01569	-0,05165	0,5475
A_7	A_8	A_9	A_{10}	A_{11}	
0,7361	0,1844	0,1056	0,6134	0,7210	

Phương trình (9) cần áp dụng phương pháp lặp Newton-Raphson hoặc phương pháp chia đôi để giải. Phương pháp này được báo cáo là dự đoán hệ số Z từ biểu đồ Standing và Katz với một sai số tuyệt đối trung bình là 0,585% và có thể áp dụng trên phạm vi: $0,2 < P_{pr} < 15$ và $1 < T_{pr} < 3$ [8].

2.3. Phương pháp Dranchuk, Purvis và Robinson

Phương pháp này là một biến đổi xa hơn của phương pháp Dranchuk và Abou-Kassem nhận được ban đầu [2]. Phương pháp Dranchuk, Purvis và Robinson có 8 hệ số và cần ít khối lượng tính toán để nhận được hệ số Z.

$$Z = \frac{0,27 P_{pr}}{y T_{pr}} \quad (10)$$

Trong đó:

y là nghiệm của phương trình sau:

$$\left[T_4 y^2 (1 + A_8 y^2) e^{(-A_8 y^2)} \right] + 1 + T_1 y + T_2 y^2 + T_3 y^5 - \frac{T_5}{y} = 0 \quad (11)$$

Với:

$$T_1 = A_1 + \frac{A_2}{T_{pr}} + \frac{A_3}{T_{pr}^3}$$

$$T_2 = A_4 + \frac{A_5}{T_{pr}}$$

$$T_3 = \frac{A_5 A_6}{T_{pr}}$$

$$T_4 = \frac{A_7}{T_{pr}^3}$$

$$T_5 = \frac{0,27 P_{pr}}{T_{pr}}$$

Trong đó $A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8$ là những hệ số như trong Bảng 2.

Bảng 2. Những hệ số của phương pháp Dranchuk, Purvis và Robinson

A_1	A_2	A_3	A_4
0,31506237	-1,04670990	-0,57832720	0,53530771
A_5	A_6	A_7	A_8
-0,61232032	-0,10488813	0,68157001	0,68446549

Cách giải phương trình (11) tương tự với phương pháp Dranchuk và Abou-Kassem. Phương pháp này phù hợp trong phạm vi áp suất và nhiệt độ giả giảm sau [8]:

$$1,05 < T_{pr} < 3 \text{ và } 0,2 < P_{pr} < 3$$

2.4. Các phương pháp khác

Ngoài 3 phương pháp ẩn (implicit) đã nêu còn có những phương pháp hiện (explicit) để tính hệ số Z. Những phương pháp hiện không cần dùng phép lặp để tính hệ số Z nên tránh được vấn đề hội tụ so với phương pháp ẩn. Một trong những phương pháp tốt nhất hiện nay để tính hệ số Z là phương pháp Beggs và Brill [3]. Gần đây nhiều nhà nghiên cứu đã tìm ra những phương pháp mới để tính hệ số Z như:

- Phương pháp Heidaryan, Salarabadi, Moghadasi [7];
- Phương pháp Azizi, Behbahani và Isazadeh [11];
- Phương pháp Sanjari và Lay [12];
- Phương pháp Lateef A.Kareem [13].
- ❖ Phương pháp Brill và Beggs [3]:

$$Z = A + \frac{1-A}{e^B} + CP_{pr}^D \quad (12)$$

Với:

$$A = 1,39(T_{pr} - 0,92)^{0,5} - 0,36T_{pr} - 0,10$$

$$B = \left(0,62 - 0,23T_{pr}\right)P_{pr} + \left(\frac{0,066}{T_{pr} - 0,86} - 0,037\right)P_{pr}^2 + \frac{0,32P_{pr}^2}{10^E}$$

$$C = 0,132 - 0,32 \log\left(T_{pr}\right), \quad D = 10^F$$

$$E = 9(T_{pr} - 1), \quad F = 0,3106 - 0,49T_{pr} + 0,1824T_{pr}^2$$

2.5. Hạn chế của các phương pháp xác định hệ số Z

2.5.1. Hạn chế của phương pháp hiện

Phương pháp Beggs và Brill khá chính xác trong một phạm vi nào đó, tuy nhiên phương pháp này không được áp dụng khi $T_{pr} < 0,92$. Với một vài giá trị của T_{pr} và P_{pr} có thể tạo ra giá trị âm của hệ số Z [14].

Phương pháp Heidaryan, Salarabadi, Moghadasi có hệ số hồi quy 0,99963 nhưng lại không liên tục tại $P_{pr} = 3$. Do đó, phương pháp này không được sử dụng để tính hệ số Z tại áp suất giả giảm $P_{pr} = 3$. Phương pháp Azizi, Behbahani, Isazadeh chỉ tính hệ số Z trong phạm vi của nhiệt độ giả giảm $1,1 < T_{pr} \leq 2$ và áp suất giả giảm $0,2 \leq P_{pr} \leq 11$. Phương pháp Sanjari và Lay ít hiệu quả so với các phương pháp khác vì hệ số hồi quy là 0,8757 và tỷ lệ sai số tại một điểm nào đó có thể cao tới 90% [13].

2.5.2. Hạn chế của phương pháp ẩn

Các phương pháp ẩn sử dụng phép lặp để giải và đòi hỏi một giả sử trước khi xuất hiện sự hội tụ. Ngoài ra, có một vài trường hợp khi sử dụng những giá trị ban đầu nào đó cho hệ số Z sẽ tạo ra giá trị âm của hệ số Z. Mặc dù có độ chính xác cao nhưng những phương pháp ẩn này có thể không được sử dụng vào một chương trình thiết kế vì nó không thể dự đoán hoặc xác định khi những giá trị sai số đó có thể tạo thành [14].

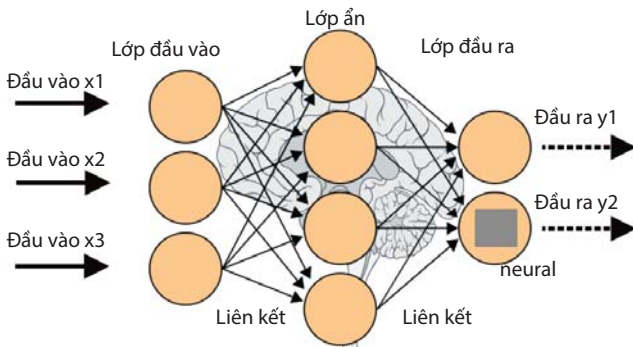
3. Mô hình mạng neural nhân tạo sử dụng phần mềm MATLAB

3.1. Phần mềm MATLAB

Phần mềm MATLAB được sử dụng trong tất cả những lĩnh vực của toán học ứng dụng, trong giáo dục và nghiên cứu tại các trường đại học và trong công nghiệp. MATLAB (viết tắt của MATrix LABoratory) được xây dựng xung quanh những vector và ma trận. MATLAB là phần mềm hữu ích cho đại số tuyến tính, là một công cụ tuyệt vời cho việc giải phương trình đạo hàm, phương trình đại số và giải tích số. MATLAB có thể cung cấp những bức ảnh đẹp trong 2D và 3D và là một trong những ngôn ngữ lập trình đơn giản nhất cho việc viết những chương trình toán học. Ngoài ra, MATLAB cũng hữu ích cho xử lý tín hiệu và ảnh, truyền thông, thiết kế điều khiển tự động, đo lường kiểm tra, phân tích mô hình tài chính, hay tính toán sinh học [15]. Với hàng triệu kỹ sư và nhà khoa học làm việc trong môi trường hàn lâm, MATLAB là ngôn ngữ của tính toán khoa học.

3.2. Mạng neural nhân tạo

Mạng neural nhân tạo là một hệ thống xử lý thông tin và thử bắt chước chức năng và sự phức tạp của hệ thống neural sinh học [16]. Neural sinh học là một khối cơ sở của hệ thống neural bao gồm 3 phần



Hình 1. Cấu trúc mạng neural nhân tạo [23]

chính: thân neural (soma), sợi nhánh (dendrite) và sợi trục (axon). Thông tin ở dạng tín hiệu điện được nhận bởi các nhánh, được xử lý bởi thân neural sau đó truyền qua sợi trục. Tương tự, những neural nhân tạo là những khối cơ sở cho mạng neural nhân tạo. Thông tin đến neural nhân tạo qua những đầu vào (input), mỗi đầu vào được nhân bởi một hàm trọng số trước khi đi vào neural. Neural sau đó tổng hợp tất cả dữ liệu đầu vào đã được gia trọng, sai số và xử lý tổng qua một hàm chuyển đổi trước khi truyền thông tin tới những đầu ra [17]. Hình 1 trình bày cấu trúc mạng neural nhân tạo.

Phương trình sau mô tả hoạt động toán học xảy ra bên trong neural nhân tạo:

$$y_i(k) = F\left(\sum_{i=0}^m w_i(k)x_i(k) + (b)\right) \quad (13)$$

Trong đó:

$x_i(k)$ là giá trị đầu vào, $w_i(k)$ là giá trị trọng số; trong thời gian rời rạc k (i bắt đầu từ 0 tới m), F là hàm chuyển đổi, b là sai số và $y_i(k)$ là giá trị đầu ra trong thời gian rời rạc k [17].

Số lượng những ứng dụng sử dụng ANN rất lớn do ANN có khả năng để xấp xỉ mọi hàm số theo một cách ổn định và hiệu quả. Trong ngành công nghiệp dầu khí, không có một hàm số chính xác hoặc phù hợp để liên kết những thông số đầu vào tới những thông số đầu ra, vì vậy ANN xuất hiện như ứng viên phù hợp để nhận được những mối liên hệ hàm số giữa những thông số đó và kết quả mong muốn [18]. Một số lĩnh vực của công nghiệp dầu khí đã sử dụng thành công ANN như: địa chất và địa vật lý, đánh giá thành hệ và công nghệ mỏ.

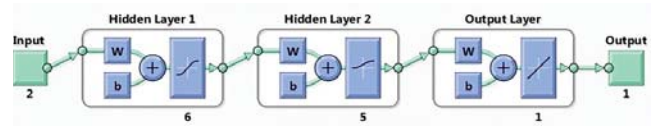
3.3. Phương pháp thực hiện

3.3.1. Chuẩn bị dữ liệu cho mạng neural

Phải tập hợp và chuẩn bị dữ liệu trước khi bắt đầu quá trình thiết kế mạng. Những dữ liệu đầu vào trong nghiên cứu này là P_{pr} và T_{pr} và dữ liệu đầu ra là hệ số Z . Tổng số 5.940 điểm dữ liệu được chọn từ sách công nghệ mỏ khí

Bảng 3. Phạm vi của các thông số đầu vào và ra của mạng neural

Thông số	Giá trị nhỏ nhất	Giá trị lớn nhất
P_{pr}	0,2	15
T_{pr}	1,05	3
Z	0,2992	1,753



Hình 2. Cấu trúc mạng tiến tiếp (feedforward) 2-6-5-1

thiên nhiên được xuất bản bởi SPE [5]. 2/3 dữ liệu được chọn từ tập hợp dữ liệu tổng để thành lập tập dữ liệu đào tạo (training). Một nửa những điểm dữ liệu còn lại được chọn để thiết lập tập dữ liệu phê chuẩn (validation) và phần còn lại là của tập dữ liệu kiểm tra (testing). Những điểm dữ liệu cho mỗi tập hợp được chọn từ tập hợp tổng được phân bố sao cho vẫn giữ được sự đồng nhất giống với tập dữ liệu tổng [15].

3.3.2. Tạo mạng

Sau khi dữ liệu được tập hợp, bước tiếp theo trong đào tạo một mạng là tạo ra một đối tượng mạng. Trong nghiên cứu này, nhóm tác giả sử dụng mạng tiến tiếp (feedforward), một số mạng 1 lớp ẩn (hidden layer) được kiểm tra không đưa ra ứng xử như mong đợi trong toàn bộ tập dữ liệu với mức độ sai số trên tập phê chuẩn và kiểm tra khá lớn. Qua quá trình nghiên cứu và thử nghiệm, nhóm tác giả đã chọn được mạng 2 lớp ẩn cung cấp kết quả chính xác hơn nhiều so với mạng 1 lớp ẩn. Ngoài ra, những mạng với 3 hoặc nhiều lớp ẩn hơn sẽ chỉ tăng thời gian tính toán mà không cải thiện độ chính xác. Nhóm tác giả sử dụng 6 neural cho lớp ẩn đầu tiên và 5 neural cho lớp ẩn thứ 2 (Hình 2).

3.3.3. Định hình mạng

Tiếp theo là chọn số lượng neural trong mỗi lớp. Số lượng neural nhỏ sẽ đào tạo mạng nhanh hơn nhưng không cho kết quả chính xác. Trong khi đó, tăng số lượng neural có thể tăng thời gian xử lý. Nhóm tác giả sử dụng cấu trúc mạng với 6 neural trong lớp ẩn thứ nhất và 5 neural trong lớp ẩn thứ hai (mạng 2-6-5-1). Hình 2 biểu diễn cấu trúc mạng mô phỏng. Những hàm hoạt động cho những lớp ẩn là hàm tag-sigmoid và log-sigmoid và hàm pure linear được sử dụng cho lớp đầu ra. Những hàm đó được biểu diễn như sau:

$$\text{tansig}(x) = \frac{2}{1 + e^{-2x}} - 1$$

$$\log \operatorname{sig}(x) = \frac{1}{1 + e^{-x}}$$

$$\operatorname{purelin}(x) = x$$

3.3.4. Khởi chạy trọng số và sai số

Trước khi đào tạo một mạng tiến tiếp phải khởi động (thủ công hoặc tự động) những trọng số và những sai số.

3.3.5. Đào tạo mạng

Khi những trọng số và sai số của mạng được khởi động, mạng đã sẵn sàng cho đào tạo. Quá trình đào tạo mạng bao gồm việc điều chỉnh những giá trị trọng số và sai số của mạng để tối ưu hiệu suất mạng. Có 2 cách đào tạo mạng: phương pháp tăng và phương pháp theo đợt. Trong phương pháp tăng, gradient được tính và những trọng số được cập nhật sau khi mỗi đầu vào được áp dụng tới mạng. Trong phương pháp theo đợt, tất cả dữ liệu đầu vào trong tập đào tạo được áp dụng tới mạng trước khi trọng số được cập nhật [15].

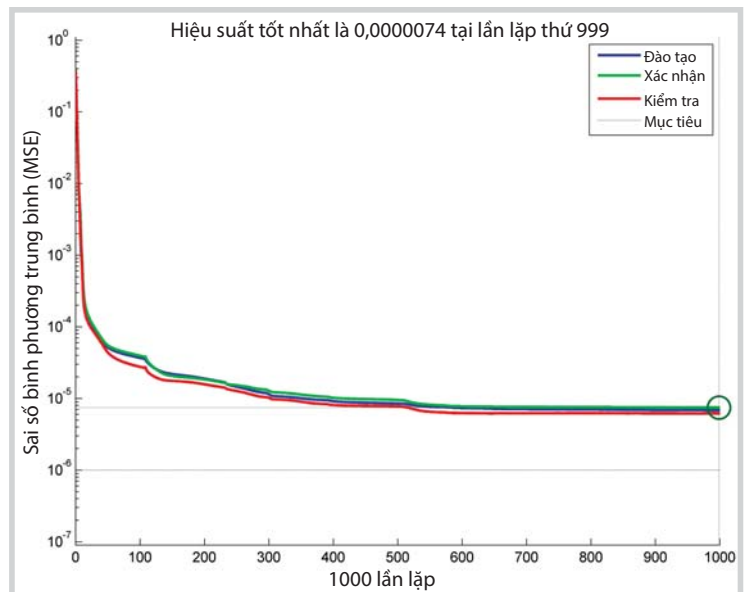
Một vài phương pháp đào tạo lan truyền ngược đã tồn tại như thuật toán Levenberg-Marquardt Algorithm (LMA), Scaled Conjugate Gradient (SCG), Pola-Ribiere conjugate gradient (PCG) và những thuật toán khác. Một số nghiên cứu chỉ ra rằng LMA đủ mạnh và cung cấp ANN chính xác [19, 20]. Trong quá trình đào tạo, trọng số được thay đổi bởi thuật toán cho tới khi tổng của những sai số bình phương đủ nhỏ.

3.3.6. Phân tích hiệu suất mạng neural sau khi đào tạo

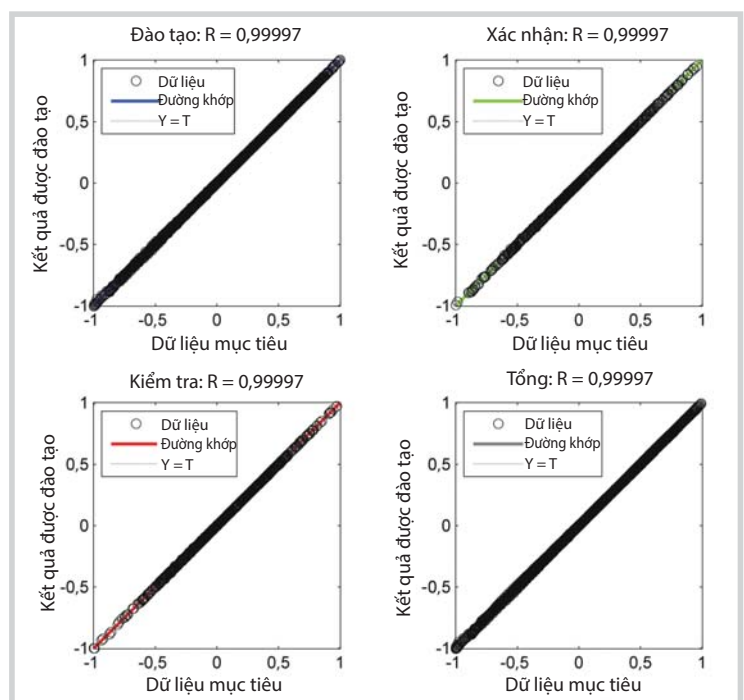
Sau khi mạng được đào tạo và đạt tới 1.000 phép lặp có thể thu được đồ thị biểu diễn hiệu suất của quá trình đào tạo như Hình 3.

Hình 3 biểu diễn tổng sai số bình phương cho 3 tập dữ liệu với số lần lặp cho mạng có cấu trúc 2-6-5-1 sử dụng thuật toán đào tạo lan truyền ngược LMA (Levenberg-Marquardt Algorithm). Ứng xử của đường cong thể hiện quá trình đào tạo thành công và 3 tập dữ liệu được chọn phù hợp từ tập dữ liệu ban đầu. Hiệu suất tốt nhất là $MSE = 7.444 \times 10^{-6}$ tại lần lặp thứ 999.

Hình 4 trình bày dữ liệu đào tạo, xác nhận, kiểm tra và dữ liệu tổng. Đường nét đứt trong



Hình 3. Đồ thị biểu diễn hiệu suất của mạng neural nhân tạo 2-6-5-1



Hình 4. Đồ thị hồi quy của mạng neural nhân tạo

mỗi đồ thị của Hình 4 trình bày kết quả sau khi đào tạo trùng với dữ liệu mục tiêu. Đường nét liền trình bày sự hồi quy tuyến tính tốt nhất giữa kết quả đào tạo và mục tiêu. Giá trị R biểu thị một mối quan hệ giữa kết quả và mục tiêu. Nếu $R = 1$ thì đó là một mối quan hệ tuyến tính chính xác giữa kết quả đào tạo và mục tiêu [15].

Kết quả của dữ liệu được đào tạo bởi mô hình mạng neural nhân tạo trong nghiên cứu này chỉ ra đây là một mối quan hệ chính xác: kết quả đào tạo, kết quả xác nhận, kiểm tra và kết quả tổng cộng đều có cùng một giá trị $R = 0,99997$.

3.3.7. Sử dụng mạng

Sau khi mạng được đào tạo với hiệu suất mong muốn có thể sử dụng mô hình đó để xác định hệ số Z. Mô hình dự báo hệ số Z có dạng như sau:

$$Z = \frac{[(A_7 + 3,16096829287339) + 1] \times 1,4538}{2} + 0,2992 \quad (14)$$

Với:

$$A_1 = \begin{bmatrix} -0,58807461290394 & -0,162590864228887 \\ -0,985567895097498 & -3,58914907839306 \\ -0,184285983142621 & 0,750043752672046 \\ 1,67732474376959 & -0,437482063248194 \\ -1,68657765389108 & 0,509333187371346 \\ -4,52723604209889 & -2,40353331303964 \end{bmatrix} \begin{bmatrix} P_{pr-conv} \\ T_{pr-conv} \end{bmatrix}$$

$$P_{pr-conv} = \frac{2 \times (P_{pr} - 0,2)}{14,8} - 1; \quad T_{pr-conv} = \frac{2 \times (T_{pr} - 1,05)}{1,95} - 1$$

$$A_2 = A_1 + \begin{bmatrix} 0,75762447365063 \\ -6,62896287950176 \\ 1,82006252951295 \\ 0,937881909714968 \\ -0,95290888785043 \\ -6,92056775590929 \end{bmatrix}; \quad A_3 = \frac{2}{1 + \exp(-2 \times A_2)} - 1;$$

$$A_4 = \begin{bmatrix} 3,897465 & 61,968783 & -5,519129 & -10,500512 & -10,460376 & -4,506669 \\ 3,931415 & 350,500652 & -19,127087 & 70,103013 & 66,408963 & -18,291279 \\ 0,607552 & -103,887722 & 6,452176 & 4,356828 & 4,216874 & 7,756330 \\ 5,129133 & -226,530302 & -23,920222 & 9,423814 & 3,993938 & 13,144008 \\ -20,412276 & 128,752019 & -5,208601 & -74,161628 & -62,072784 & -270,518599 \end{bmatrix} [A_3]$$

$$A_5 = A_4 + \begin{bmatrix} 57,5700294875709 \\ 349,825769447068 \\ -101,442975357111 \\ -199,331507457456 \\ -128,320898888629 \end{bmatrix}; \quad A_6 = \frac{1}{1 + \exp(-A_5)};$$

$$A_7 = [-4,48841980 \quad -0,53755610 \quad -2,88707125 \quad -0,19033986 \quad -0,21902471] [A_6]$$

Trong đó:

$P_{pr-conv}$: Áp suất giả giảm chuyển đổi trong phạm vi thông số đầu vào;

$T_{pr-conv}$: Nhiệt độ giả giảm chuyển đổi trong phạm vi thông số đầu vào.

Đồ thị hệ số Z nhận được từ phương trình (14) được biểu diễn như Hình 5.

4. So sánh sai số của các phương pháp xác định hệ số Z

Để đánh giá sự chính xác của mô hình ANN so với các phương pháp thực nghiệm trước cần sử dụng một số thông số độ lệch thống kê gồm: tỷ lệ sai

số tương đối trung bình (APRE - Average percent relative error), sai số bình phương trung bình (MSE - Mean squared error), căn bậc hai của sai số bình phương trung bình (RMSE - Root mean squared error), hệ số xác định (R^2 - Coefficient of determination).

➤ Phần trăm sai số tương đối trung bình (APRE hoặc ARE%) đánh giá độ lệch tương đối của dữ liệu hệ số Z được dự đoán và dữ liệu thực tế [21]:

$$ARE\% = \frac{100}{N} \sum_{i=1}^N \left(\frac{Z_i^{Pred} - Z_i^{exp}}{Z_i^{exp}} \right) \quad (15)$$

Trong đó:

N: Tổng số các phần tử (5.940 phần tử);

i: Chỉ số phần tử thứ i;

Z_i^{Pred} : Hệ số Z được dự đoán từ các công thức thực nghiệm hoặc mạng neural nhân tạo;

Z_i^{exp} : Hệ số Z thực nghiệm được SPE công bố.

➤ Sai số bình phương trung bình (MSE) là một công thức ước lượng đo giá trị bình phương của sai số. Giá trị MSE càng nhỏ thể hiện sự dự đoán càng chính xác [22]:

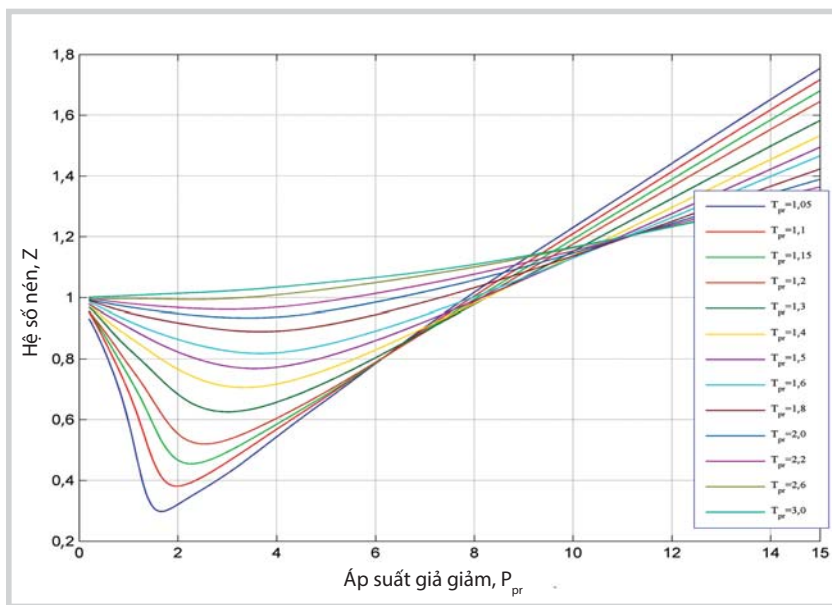
$$MSE = \frac{1}{N} \sum_{i=1}^N (Z_i^{Pred} - Z_i^{exp})^2 \quad (16)$$

➤ Căn bậc hai sai số bình phương trung bình (RMSE) đo dữ liệu rời rạc xung quanh độ lệch 0 [21]:

$$RMSE = \left(\frac{\sum_{i=1}^N (Z_i^{Pred} - Z_i^{exp})^2}{N} \right)^{\frac{1}{2}} \quad (17)$$

Hệ số xác định (R^2) là thông số độ lệch thống kê đơn giản trình bày sự phù hợp với dữ liệu [22]:

$$R^2 = 1 - \frac{\sum_{i=1}^N (Z_i^{Pred} - Z_i^{exp})^2}{\sum_{i=1}^N \left(Z_i^{Pred} - \frac{1}{N} \sum_{i=1}^N Z_i^{exp} \right)^2} \quad (18)$$



Hình 5. Đồ thị hệ số Z được tạo ra bởi phương trình (14)

Bảng 4. So sánh những sai số thống kê giữa mô hình ANN và các phương pháp khác

TT	Phương pháp	ARE%	MSE	RMSE	R ²
1	Hall và Yarborough	-0,53099	0,014815	0,121717	0,784172
2	Dranchuk và Abou-Kassem	0,01431	0,000023	0,004794	0,999638
3	Dranchuk, Purvis và Robinson	-0,87809	0,009029	0,095023	0,877296
4	Brill và Beggs	-1070,9057	4909,0871	70,0649	0,00094
5	ANN	0,0002053	0,0000036	0,0019	0,999944

So sánh chất lượng giữa các phương pháp được thể hiện trong Bảng 4 với 5.940 điểm dữ liệu của SPE được sử dụng. Kết quả cho thấy, phương pháp mạng neural nhân tạo có giá trị hệ số xác định (R^2) cao nhất: 0,999944 và các sai số thống kê ARE, MSE và RMSE nhỏ nhất. Phương pháp Dranchuk và Abou-Kassem cũng dự đoán hệ số Z khá chính xác với $R^2 = 0,9996$. Phương pháp Brill và Beggs dự báo kém chính xác nhất và không đáng tin cậy khi dự đoán trong phạm vi rộng lớn của áp suất và nhiệt độ giả giảm.

5. Kết luận

Hệ số Z là một thông số rất quan trọng để xác định tỷ trọng khí, độ nhớt khí, độ nén khí, mô phỏng vỉa khí, tính toán cân bằng vật chất và dự đoán PVT cho giếng dầu và khí. So sánh phương pháp xác định hệ số Z theo mô hình dựa trên mạng neural nhân tạo và các phương pháp thực nghiệm khác cho thấy sự vượt trội của phương pháp mới qua các chỉ số như: hệ số xác định (R^2), sai số bình phương trung bình (MSE), căn bậc hai sai số bình phương trung bình (RMSE), tỷ lệ sai số tương đối trung bình (ARE) lần lượt là 0,999944, 0,0000036, 0,0019 và 0,0002053. Kết quả này đã chứng minh mô hình mới có thể là công cụ dự đoán chính xác hệ số Z và có khả năng áp dụng trên phạm vi nhiệt độ và áp suất rộng hơn so với các phương pháp khác.

Tài liệu tham khảo

1. P.M.Dranchuk, J.H.Abou-Kassem. *Calculation of Z-factors for natural gases using equations of state*. Journal of Canadian Petroleum Technology. 1975; 14(3): p. 34 - 36.
2. R.A.Dranchuk, D.B.Purvis, P.M.Robinson. *A reduced equation of state applied to generalized compressibility factor tables*. Journal of Canadian Petroleum Technology. 1971.
3. D.H.Beggs, J.P.Brill. *An experimental study of two-phase flow in inclined pipes*. Journal of Petroleum Technology. 1973; 25(5): p. 607 - 617.
4. Mohammadreza Kamyab, Jorge H.B Sampaio Jr., Farhad Qanbaria, Alfred W.Eustes III. *Using artificial neural networks to estimate the z-factor for naturalhydrocarbon gases*. Elsevier. 2010.
5. D.L.Katz, D.Cornell. *Handbook of natural gas engineering*. New York: McGraw-Hill. 1959.
6. E.Sanjari, E.N.Lay. *Estimation of natural gas compressibility factors using artificial neural network approach*. Journal of Natural Gas Science and Engineering. 2012; 9: p. 220 - 226.
7. E.Heidaryan, A.Salarabadi, J.Moghadas. *A novel correlation approach for prediction of natural gas compressibility factor*. Journal of Natural Gas Chemistry. 2010; 19(2): p. 189 - 192.
8. T.Ahmed. *Reservoir engineering handbook*. Gulf professional publishing. 2010.
9. A.Danesh. *PVT and phase*. Elsevier. 1998.
10. Y.A.Cengel, M.A. Boles. *Thermodynamics: An engineering approach*. McGraw Hill. 2007.
11. N.Azizi, R.Behbahani, M.A.Isazadeh. *An efficient correlation for*

calculating compressibility factor of natural gases. Journal of Natural Gas Chemistry. 2010; 19(6): p. 642 - 645.

12. E.Sanjari, E.N.Lay. *An accurate empirical correlation for predicting natural gas compressibility factors*. Journal of Natural Gas Chemistry. 2012; 21(2): p. 184 - 188.

13. L.A.Kareem. *New explicit correlation for the compressibility factor of natural gas: linearized Z-factor isotherms*. Journal of Petroleum Exploration and Production Technology. 2015; 6(19): p. 1 - 12.

14. K.K.Dune, B.N.Oriji. *A new computerized approach to Z-factor determination*. Transnational Journal of Science and Technology. 2012; 2(7): p. 64 - 80.

15. M.H.Beale, M.T.Hagan, H.B.Demuth. *Neural network toolbox user's guide*. The MathWorks. 2015.

16. S.Mohaghegh. *Virtual-intelligence applications in petroleum engineering: Part 1- Artificial neural networks*. Journal of Petroleum Technology. 2000; 52(9): p. 64 - 73.

17. K.Suzuki, A.Krenker, J.Bester, A.Kos. *Introduction to the artificial neural networks in Artificial neural networks - Methodological advances and biomedical applications*. In-Tech. 2011: p. 3 - 18.

18. S.Mohaghegh. *Neural network: What it can do for petroleum engineers*. Society of Petroleum Engineers. 1995; 47(1).

19. H.Esen, M.Inalli. *Modelling of a vertical ground coupled heat pump system by using artificial neural networks*. Expert Systems with Applications. 2009; 36(7): p. 10229 - 10238.

20. A.Sozen, E.Arcaklioglu, E.G.Kanit. *Use of artificial neural networks for mapping of solar potential in Turkey*. Applied Energy. 2004; 77(3): p. 273 - 286.

21. A.H.Mohammadi, A.Kamari, F.Gharagheizi. *A corresponding states-based method for the estimation of natural gas compressibility factors*. Journal of Molecular Liquids. 2016; 216: p. 25 - 34.

22. M.B.Mohamad, A.Reza, O.Shahriar, Z.Zeinab. *Prediction of gas compressibility factor using intelligent models*. Natural Gas Industry B. 2015; 2(4): p. 283 - 294.

23. Hoàng Anh. *Sự trở dậy của những cỗ máy*. PC World VN. 8/2015.

Applying artificial neural networks to predict Z-factor for natural hydrocarbon gas

Tran Kha Tien, Hoang Thinh Nhan
Petrovietnam University
Email: tientk01@pvu.edu.vn

Summary

The natural gas compressibility factor or Z-factor is an important parameter to determine thermodynamic properties of gas reservoir, gas density, gas viscosity, gas compression, gas reservoir simulation, calculate the material balance equation, and estimate PVT for oil and gas wells. Based on the data from the Standing-Katz chart, several methods have been developed to calculate the Z-factor [1 - 3]. Based on the method of Mohammadreza Kamyab [4], the authors have calculated and determined the Z-factor using the artificial neural networks (ANN) with input parameters being the pseudo-reduced pressure and temperature of 5,940 experimental data points [5]. The results of research show that this model is able to predict the Z-factor more accurately than other methods and can be applied over the pseudo-reduced temperature range of $1.05 \leq T_{pr} \leq 3$ and the pseudo-reduced pressure of $0.2 \leq P_{pr} \leq 15$.

Key words: Natural gas compressibility factor, Z-factor, MATLAB, Artificial Neural Networks.