



## Nâng cao chất lượng điều khiển hệ thống nâng thân giàn khoan dựa trên giải pháp học tăng cường

## Reinforcement learning approach for enhancing the control quality of the jacking system

Đặng Xuân Kiên<sup>1,\*</sup>, Đỗ Việt Dũng<sup>1</sup>, Lý Sợi<sup>1,2</sup>, Vũ Hải Long<sup>3</sup>

<sup>1</sup>Nhóm nghiên cứu AIT, Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Hồ Chí Minh

<sup>2</sup>Khoa Công nghệ điện, Trường Cao đẳng Công nghệ cao Đồng An, Bình Dương

<sup>3</sup>Liên doanh Vietsovetro, Bà Rịa - Vũng Tàu

Từ khóa:

Giàn khoan tự  
nâng  
Hệ thống nâng hạ  
Học tăng cường  
Mạng nơ-ron hồi  
quy  
Tác động môi  
trường

TÓM TẮT

Hệ thống nâng hạ thân giàn (JS) cho giàn khoan tự nâng (JuR) tại Việt Nam hầu hết đang vận hành bán tự động. Trong bài báo này, nhóm tác giả tập trung phát triển giải thuật học tăng cường (RL) nâng cao chất lượng điều khiển cho JS, với giá trị tham số điều khiển cơ sở Tỷ lệ-Tích phân-Vi phân (PID) được chỉnh định thích nghi qua các lớp mạng học tăng cường. Hai kịch bản thử nghiệm cho giải pháp đề xuất được thực hiện trong điều kiện ngoài khơi có tác động từ các thành phần sóng, gió, và dòng chảy, trên môi trường tính toán Matlab 2023b. Các kết quả cho thấy rằng hành trình nâng thân giàn luôn bám sát giá trị mong muốn. Bên cạnh đó, giải pháp được so sánh và đánh giá với mạng nơ-ron hồi quy (RNN) mang lại kết quả có thời gian đáp ứng nhanh hơn, độ vọt lố và sai số xác lập nhỏ hơn, đã xác nhận hiệu quả của giải pháp đề xuất.

Keywords:

Jack-up rig  
Jacking system  
Reinforcement  
learning  
Recurrent neural  
network  
Environmental  
Impact

ABSTRACT

The Jacking System (JS) for Jack-up Rig (JuR) in Vietnam mostly operates semi-automatically. In this paper, the authors focus on developing a reinforcement learning (RL) algorithm to enhance control quality for JS by adaptively adjusting the basic Proportional - Integral - Derivative (PID) control parameters through reinforcement learning network layers. Two testing scenarios for the proposed solution were performed in offshore conditions with the impact of wave, wind, and current components, on the Matlab 2023b computing environment. The results show that the JS lift the hull follows the desired value. Additionally, the solution was compared to a Recurrent Neural Network (RNN), which demonstrated faster response times, reduced overshoot, and lower steady-state error. This comparison confirms the effectiveness of the proposed solution.

\* Đặng Xuân Kiên. Nhóm nghiên cứu AIT, Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh

Email: [kien.dang@ut.edu.vn](mailto:kien.dang@ut.edu.vn)

[https://www.doi.org/10.55228/JTST.14\(3\).1-13](https://www.doi.org/10.55228/JTST.14(3).1-13)

Ngày nhận bài: 9/04/2025; Ngày nhận bài sửa: 05/05/2025; Ngày chấp nhận đăng: 9/05/2025

Ngày xuất bản trực tuyến: 15/05/2025

pISSN: 1859-4263; eISSN: 3030-4261

## 1. Giới thiệu

Giàn khoan tự nâng là công trình phổ biến được sử dụng trong khai thác dầu khí ngoài khơi nhờ khả năng di chuyển linh động giữa các địa điểm khoan. Tuy nhiên, việc thiết lập và cố định kết cấu thân giàn chuẩn bị cho các công tác khoan thăm dò hay khai thác bao gồm nhiều quy trình phức tạp, đòi hỏi tuân thủ nghiêm ngặt các quy định. Trong đó, quá trình nâng hạ đóng vai trò then chốt giúp giàn ổn định trong quá trình hoạt động ngoài khơi. Nếu thân giàn không cân bằng, trước tiên sẽ dẫn tới hư hỏng các kết cấu dẫn động cho hệ thống nâng hạ cũng như nguy cơ gãy đổ thân giàn [1]. Cụ thể, theo tiêu chuẩn ISO 19905-1:2023, góc lệch trục dọc và trục ngang không lớn hơn 1 độ [2]. Do đó, quá trình điều khiển phải đảm bảo tin cậy và ổn định cao. Đã có nhiều nghiên cứu được thực hiện nhằm đáp ứng yêu cầu trên trong quá trình vận hành giàn khoan. Một số phương pháp điều khiển được đề xuất như thông minh [3], lai ghép [4], thích nghi [5], và mạng nơ-ron [6] cũng đã cho thấy tính hiệu quả và ổn định trong môi trường hoạt động thực tế.

Do những yêu cầu khắt khe của hệ thống động lực, các thuật toán được áp dụng phải đạt độ tin cậy cao. Trong đó, thuật toán kinh điển như PID có những ưu điểm đã được chứng minh trong các hệ thống điều khiển thực. Tuy nhiên, các thuật toán kinh điển chỉ mang hiệu quả tốt với hệ thống tuyến tính có ít sự nhiễu động từ bên ngoài. Các thuật toán có khả năng hiệu chỉnh các thông số bộ điều khiển nhằm gia tăng khả năng thích nghi, đặc biệt là ứng dụng trong điều khiển các đối tượng phi tuyến [7]. Tuy nhiên, việc triển khai thực tế lại phát sinh nhiều vấn đề do các cấu trúc hệ thống phức tạp, có nhiều tham số không chắc chắn biến đổi theo thời gian. Do đó, yêu cầu các thuật toán không những có khả

năng thích nghi cao mà còn phản ứng tốt với tác động của thành phần nhiễu động ngoại vi.

Các kỹ thuật điều khiển nâng cao ứng dụng mạng nơ-ron ngày càng phát triển mạnh mẽ. Trong đó, kỹ thuật học máy (ML) có ưu thế trong công nghiệp với thời gian đáp ứng nhanh, trong khi đó kỹ thuật học sâu (DL) mang lại hiệu quả với những hệ thống có cấu trúc phức tạp và tham số biến đổi theo thời gian. Bên cạnh đó, kỹ thuật học tăng cường (RL) lại hiệu quả cao cho các đối tượng có tính chất phi tuyến cao, đòi hỏi khả năng thích nghi và yêu cầu độ chính xác cao [8]. Tiêu biểu, kỹ thuật RL được áp dụng trong điều khiển tàu [9] cũng như điều khiển rô-bốt bám hành trình [10].

Quá trình nâng hạ thân giàn khoan được thực hiện trên nền cấu trúc phức tạp dễ bị ảnh hưởng bởi các yếu tố vật lý, cũng như tác động của môi trường cũng gây ra tính phi tuyến cao. Bài báo này phát triển RL trong điều khiển quá trình nâng hạ thân giàn với các đóng góp chính như sau: *i)* Mô hình hoá giàn khoan tự nâng trên môi trường tính toán Matlab; *ii)* Xây dựng và huấn luyện mô hình RL cho quá trình điều khiển nâng hạ thân giàn khoan có xét đến nhiễu động của môi trường; *iii)* Thực hiện các kịch bản thử nghiệm và so sánh với các giải pháp tương đồng để minh chứng tính khả thi của đề xuất.

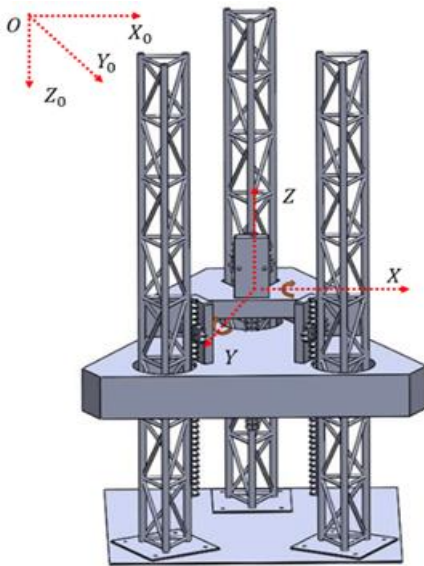
Bài báo được cấu trúc gồm các phần sau: Phần II trình bày mô hình động lực học của JuR. Phần III mô tả quy trình xây dựng bộ điều khiển RL cho quá trình nâng hạ thân giàn. Các kịch bản mô phỏng và thảo luận kết quả sẽ được trình bày ở Phần IV. Phần V đưa ra các kết luận và hướng pháp triển cho nghiên cứu.

## 2. Thiết lập vấn đề

### 2.1. Mô hình hóa JuR

JuR được thiết kế với ba chân trụ kết nối với thân giàn nơi đặt các thiết bị vận hành phục vụ

công việc thăm dò và khai thác. Quá trình di chuyển lên xuống của thân giàn được cung cấp bởi các động cơ được gắn trực tiếp lên thân giàn. Đầu ra của động cơ là trục bánh răng ăn khớp với các thanh răng được bố trí tại chân giàn. Khi 2 bộ phận kết hợp với nhau tạo thành truyền động bánh răng giúp thân giàn di chuyển lên chiều cao  $H$  tương ứng chuyển động theo trục  $Z$ , được mô tả tại **Hình 1**. Thân giàn khoan di chuyển lên xuống cần phải đồng bộ hóa giữa 3 chân trụ. Trong thực tế, việc giữ cân bằng luôn gặp khó khăn khi điều khiển đồng bộ cả 3 chân trụ, do sai lệch giữa các chân dẫn đến mặt sàn bị nghiêng. Các góc nghiêng của giàn bao gồm góc nghiêng trục dọc (Pitch =  $\theta_P$ ), và góc nghiêng trục ngang (Yaw =  $\theta_Y$ ).



**Hình 1.** Tổng quan mô hình giàn khoan tự nâng.

Cơ cấu động lực chuyển đổi tín hiệu điều khiển thành các lệnh điều khiển (như hướng quay, tốc độ và mô-men xoắn) để phân phối cho các bộ phận động lực. Các động cơ ở chân giàn sẽ chuyển động xoay theo lệnh điều khiển nhằm nâng và hạ thân giàn. Hệ phương trình vi phân mô tả động học JuR bởi [11], [12]:

$$M \frac{d^2 x}{dt^2} + C \frac{dx}{dt} + Kx(t) = \tau(t) \quad (1)$$

Trong đó,

$$\begin{aligned} [M] &= [[M]_1 [M]_2 [M]_3] \\ [K] &= [[K]_1 [K]_2 [K]_3] \\ \{x\} &= \{\{x\}_1 \{x\}_2 \{x\}_3\} \\ \{\tau\} &= \{\{\tau\}_1 \{\tau\}_2 \{\tau\}_3\} \end{aligned} \quad (2)$$

Với  $M$  là tổng trọng lượng của JuR,  $C$  và  $K$  là hệ số giảm chấn và độ cứng của hệ thống. Trong nghiên cứu này, mô hình JuR có cấu trúc 3 chân trụ nên các thông số hoạt động của giàn được tổng hợp theo công thức số (2). Bên cạnh đó,  $\tau(t)$  biểu thị lực nâng thân giàn sinh ra bởi JS, phát xuất từ chuyển động quay trên trục động cơ thông qua cơ cấu thanh răng bánh răng làm dịch chuyển thân giàn. Độ dịch chuyển của thân giàn được xác định bởi

$$x(t) = R\theta(t) \quad (3)$$

Trong đó  $R$  là bán kính của bánh răng, và  $\theta(t)$  là góc của bánh răng. Từ (1) và (3), phương trình động lực học thân giàn JuR được viết lại là

$$Ms^2 R\theta(s) + CsR\theta(s) + KR\theta(s) = \tau(s) \quad (4)$$

Hàm truyền của mô hình động học JuR là

$$G_{JuR} = \frac{\theta(s)}{\tau(s)} = \frac{k_e}{Ms^2 + Cs + K} \quad (5)$$

Với  $k_e = 1/R$  là hệ số chuyển đổi trong chuyển động quay thành tịnh tiến. Thân giàn khoan được nâng lên qua giá đỡ và hệ thống bánh răng. Hệ động lực sử dụng các động cơ DC cung cấp mô-men xoắn cho hệ thống truyền động được tính toán như sau:

$$\begin{aligned} L_a \dot{I}_a + R_a I_a + K_b W_m &= V \\ \tau_m &= k_t I_a \end{aligned} \quad (6)$$

Trong đó,  $K_t$  là hằng số mô-men xoắn và  $K_b$  là hằng số phản điện động.  $V$  là điện áp đầu vào,  $I_a$  là dòng điện phần ứng,  $R_a$  là điện trở phần ứng,  $W_m$  là vận tốc góc và  $L_a$  là điện cảm.

Mặt khác,  $J$  là quán tính rôto và  $D_a$  là hằng số ma sát nhớt, hàm truyền của động lực học nâng được trình bày như sau:

$$G_m = \frac{k_t}{JL_a s^2 + (R_a J + L_a D_a) s + (R_a D_a + K_b K_t)} \quad (7)$$

Từ phương trình (5) và (7), hàm truyền của hệ thống nâng hạ được xác định

$$G = G_{Th} G_{Jur} = \frac{k_e}{Ms^2 + Cs + K} \dots \frac{k_t}{JL_a s^2 + (R_a J + L_a D_a) s + (R_a D_a + K_b K_t)} \quad (8)$$

*Ghi chú 1:* Kết cấu nâng hạ thân giàn có tính chất phi tuyến cao bao gồm các hệ số không chắc chắn phát sinh từ kết cấu cơ khí. Ngoài ra, các yếu tố xuất phát từ sự sai lệch của hệ thống cảm biến trong quá trình đo lường trạng thái của giàn cũng gây ra các yếu tố không chắc chắn trong quá trình vận hành.

## 2.2. Tác động môi trường

Các tác động môi trường bao gồm sóng, gió, dòng chảy, đặc điểm đáy biển và độ sâu gây ra các sai lệch cho quá trình điều khiển. Trong đó, các ảnh hưởng môi trường là nguyên nhân chính gây ra sự mất ổn định được mô tả bởi

$$\tau_d(t) = \tau_{sóng}(t) + \tau_{gió}(t) + \tau_{dòng}(t) \quad (9)$$

Thành phần sóng ảnh hưởng đến hệ thống nâng được xác định [13] như sau:

$$\tau_{sóng}(t) = [x_H, y_H, \psi_H]^T \quad (10)$$

$$\delta'_x = x_H \quad (11)$$

$$x'_H = -2\omega_0 x_H - \omega_0^2 \delta_x + K_\omega \omega_x$$

$$\delta'_y = y_H \quad (12)$$

$$y'_H = -2\omega_0 y_H - \omega_0^2 \delta_y + K_\omega \omega_y$$

$$\delta'_\psi = \psi_H \quad (13)$$

$$\psi'_H = -2\omega_0 \psi_H - \omega_0^2 \delta_\psi + K_\omega \omega_\psi$$

Trong đó,  $\omega_0$  là tần số sóng với  $\delta$  là hệ số bù liên quan và  $K_\omega$  là hệ số mật độ sóng, và  $\omega_x, \omega_y, \omega_\psi$  thể hiện nhiễu trắng Gausse. Thành phần gió ảnh hưởng đến độ ổn định

nổi của giàn và phụ thuộc vào diện tích chắn gió, vận tốc, hình dạng kết cấu và chiều cao. Lực gió được tính theo công thức sau [14]:

$$\begin{aligned} \tau_{gió}(t) &= [X_{gió}, Y_{gió}]^T \\ X_{gió} &= 0.5C_x \cdot g_r \cdot \rho_w \cdot V_r \cdot 2A_T \\ Y_{gió} &= 0.5C_y \cdot g_r \cdot \rho_w \cdot V_r \cdot 2A_T \end{aligned} \quad (14)$$

Lực tác động của gió  $A_T$  được biểu diễn bằng  $C_x$  và  $C_y$ .  $V_r$  và  $g_r$  là tốc độ và hướng gió tác động lên giàn. Dòng chảy có vận tốc thay đổi theo thời gian, không gian, và độ sâu. Tốc độ tương đối của dòng chảy được mô tả

$$\begin{aligned} u_c &= V_c \cos \beta_c - \Phi_L - \Phi_H \\ v_c &= V_c \sin \beta_c - \Phi_L - \Phi_H \\ \tau_{dòng} &= [u_c, v_c, 0]^T \end{aligned} \quad (15)$$

Với  $u_c$  và  $v_c$  biểu diễn thành phần tốc độ dòng chảy,  $\Phi_L$  và  $\Phi_H$  thể hiện cấu hình góc bị ảnh hưởng bởi các giá trị tần số thấp và cao.

*Ghi chú 2:* Trong điều kiện môi trường nhiễu loạn biến đổi theo thời gian, việc thiết kế bộ điều khiển cho hệ thống giàn khoan nâng hạ thường bỏ qua ảnh hưởng của các yếu tố nhiễu động, dẫn đến chất lượng điều khiển chưa tối ưu.

Để cải thiện khả năng thích nghi, các giải pháp điều khiển được huấn luyện trực tiếp từ tập dữ liệu có ảnh hưởng môi trường trở nên cấp thiết, giúp hệ thống nhận dạng chính xác thành phần nhiễu động và phản ứng kịp thời. Thiết lập giải pháp điều khiển đề xuất sẽ trình bày chi tiết trong phần sau.

## 3. Điều khiển nâng hạ thân giàn JuR sử dụng RL

### 3.1. Các giai đoạn và mục tiêu nghiên cứu

Bộ xử lý trung tâm chịu trách nhiệm thu thập, và xử lý thông tin nhằm đưa ra các lệnh điều khiển cho hệ động lực đẩy. Các tham số trạng thái của giàn được đo lường từ hệ cảm biến và gửi về trung tâm xử lý. Giá trị này kết hợp với giá trị mong muốn

được thiết lập theo người vận hành, từ đó làm đầu vào cho bộ điều khiển sẽ tính toán các lệnh hoạt động phù hợp theo yêu cầu. Lệnh điều khiển này được gửi trực tiếp đến hệ thống lực đẩy, bao gồm hệ thống các động cơ được bố trí tại mỗi chân để nâng hạ thân giàn. Tính toán các tín hiệu điều khiển phù hợp và chính xác giúp ổn định hệ thống, và gia tăng khả năng thích nghi đảm bảo tính ổn định cho kết cấu ngay cả khi có sự biến động của tải trọng và tác động từ môi trường. Trong nghiên cứu này, tác giả tập trung phát triển thuật toán RL nhằm điều khiển trạng thái nâng hạ thân giàn dưới tác động môi trường. Giải pháp đề xuất được xây dựng trực tiếp từ tham số hoạt động của mô hình JuR nhằm hiệu chuẩn chính xác các thông số của mạng nơ-ron bên trong bộ điều khiển, quá trình được mô tả như **Hình 2** bao gồm 3 giai đoạn như sau:

- *Giai đoạn 1:* Xây dựng mô hình hệ thống nâng hạ thân giàn khoan hoạt động dưới ảnh hưởng của môi trường;
- *Giai đoạn 2:* Xây dựng và huấn luyện giải thuật RL cho quá trình nâng hạ JuR trên môi trường tính toán Matlab;
- *Giai đoạn 3:* Thử nghiệm giải pháp đề xuất trong hai kịch bản và so sánh với giải thuật RNN, qua đó minh chứng hiệu quả của giải pháp đề xuất.

### 3.2. Thiết kế bộ điều khiển nâng hạ JuR với RL

Cơ chế hoạt động của RL dựa trên đánh giá tín hiệu phản hồi trạng thái  $S_t$  của thân giàn với hàm *Reward* so với giá trị điều khiển  $a_t$  thời điểm hiện tại, các tín hiệu này được tạo ra từ quyết định *Agent* thông qua việc tương tác với hệ thống. Quá trình này bắt đầu khi *Actor* quan sát trạng thái của môi trường thông qua bộ *Observation*, thu thập sai số giữa giá trị tham chiếu  $H_{ref}$  và đầu ra của đối tượng  $y(t)$  tại các vị trí. Ngoài ra, việc giám sát vi phân sai số theo thời gian còn giúp RL kiểm soát được tốc

độ thay đổi, kiểm soát vận tốc di chuyển tại các chân giàn. Dựa vào chiến lược *Policy* đã định trước, *Actor* thực hiện xuất tín hiệu điều khiển tác động hệ thống nâng hạ thân giàn, và phản hồi trạng thái qua hàm *Reward* được mô tả bởi

$$Reward = -0.01u^2(t) - \begin{cases} 200, e(t) < 0 \\ e(t)^2, e(t) > 0 \end{cases} \quad (16)$$

Với  $e(t)$  là sai số giữa vị trí mong muốn và hiện tại của thân giàn. Thông qua đánh giá hàm *Reward* giúp *Agent* nhận biết được mức độ sai số của hệ thống, thông qua đó *Agent* sẽ chỉnh định *Policy* tối ưu hoá sai số tiệm cận về 0. Ngoài ra, thành phần  $-0.01u^2$  biểu thị tác động của  $u(t)$  giúp *Agent* biết rằng với lực tác động hiện tại sẽ thu được giá trị phần thưởng (*Reward*) tương ứng, qua đó *Agent* hiệu chỉnh các *Policy* tối ưu hơn qua từng tập huấn luyện. Hàm *Reward* đóng vai trò đánh giá giúp *Actor* nhận biết mức độ hiệu quả của tín hiệu điều khiển đã thực hiện. Dựa vào những thông tin phản hồi từ môi trường, tác nhân *Actor* điều chỉnh chiến lược của mình để tối ưu hóa các giá trị *Reward* nhận được trong dài hạn. Mục tiêu cuối cùng tối ưu hoá bài toán giữ cân bằng thân giàn JuR cần đảm bảo giải pháp hoạt động ổn định và tin cậy thông qua kiểm nghiệm dưới tác động môi trường giúp hiệu chỉnh thích nghi các tham số bộ điều khiển cơ sở, được mô tả bởi

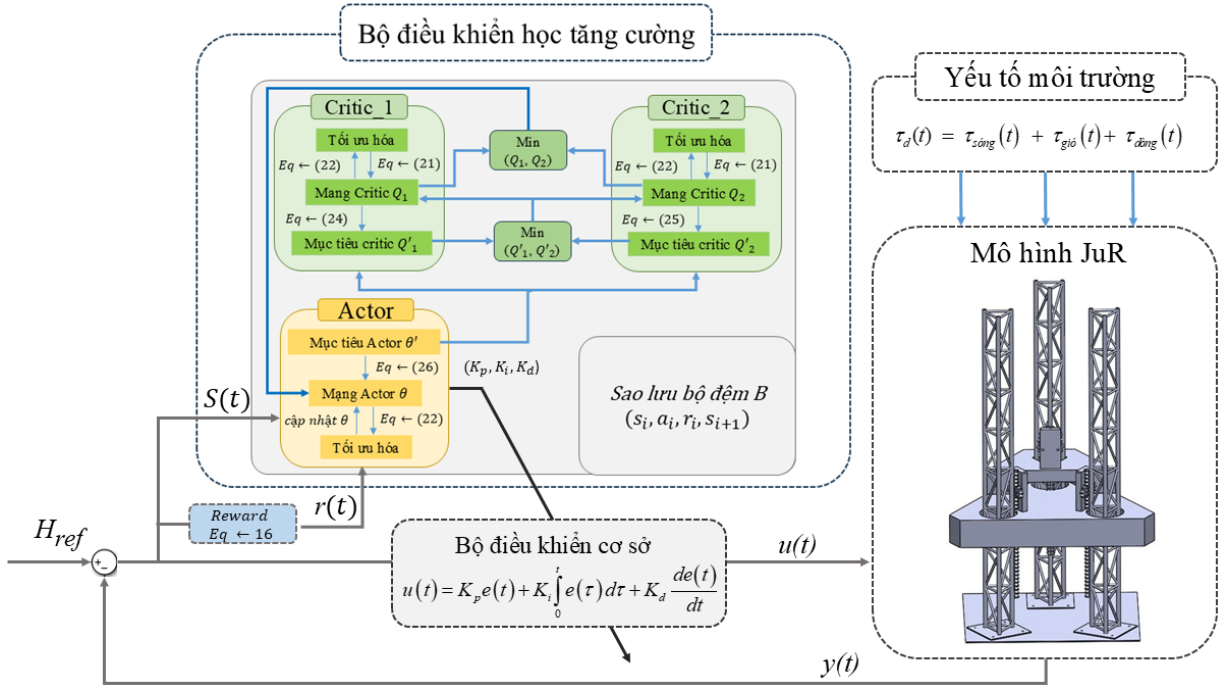
$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt} \quad (17)$$

Trong nghiên cứu này, bộ điều khiển PID được sử dụng là giải pháp điều khiển cơ sở cho quá trình nâng hạ thân giàn. Các hệ số của PID được tinh chỉnh thông qua giải pháp RL giúp tăng hiệu quả đáng kể. Đầu ra của *Agent* tác động bao gồm 3 hệ số tương ứng với hệ số  $K_p$ ,  $K_i$ , và  $K_d$ . Cấu trúc của *Agent* bao gồm hai thành phần *Actor* và *Critic*. Trong đó,

Actor quyết định hành động dựa trên trạng thái của môi trường. Mục tiêu của Actor là học cách tối ưu hóa chính sách để tối đa hóa tổng giá trị Reward mà tác nhân nhận được từ môi trường theo thời gian. Đồng thời, Actor nhận phản hồi từ Critic để cập nhật và cải thiện chính sách một cách hiệu quả.

Qua quá trình học và điều chỉnh liên tục, Actor giúp tác nhân đưa ra những quyết

định tối ưu trong môi trường nhiễu loạn. Critic đóng vai trò đánh giá và tối ưu hóa giá trị hành động thông qua áp dụng hai hàm Critic thay vì chỉ một như giải thuật RL khác. Mục tiêu của việc sử dụng hai Critic là để giảm thiểu sự đánh giá quá cao giá trị tác động, vấn đề phổ biến trong các giải thuật RL. Quá trình hoạt động của Critic như sau:



**Hình 2.** Sơ đồ tổng quan hệ thống điều khiển JS sử dụng giải thuật RL.

• **Đánh giá hành động:** Mỗi Critic ước tính giá trị Q của trạng thái và hành động, sử dụng mục tiêu từ hàm Critic thứ hai để đảm bảo tính chính xác cao hơn. Nghiên cứu này áp dụng phương trình Bellman cập nhật giá trị trạng thái cho hàm Critic dựa trên đánh giá Reward cho các hoạt động theo thời gian, nhằm lựa chọn các giá trị tối ưu cho bộ điều khiển đề xuất. Bên cạnh đó, phương trình Bellman cũng giúp giảm phương sai trong quá trình huấn luyện mô hình qua đó gia tăng tốc độ hội tụ và nâng cao hiệu quả huấn luyện, được mô tả như sau [15]:

$$Q_w(s_k, a_k) \leftarrow Q_w(s_k, a_k) + \alpha \left[ r_k + \gamma \max_{a \in A} Q_w(s_{k+1}, a) - Q_w(s_k, a_k) \right] \quad (18)$$

Với  $\alpha$  là tỉ lệ học đặc trưng cho khả năng thích ứng với sự biến động của môi trường. Sau khi đánh giá đã được tính theo (18) sẽ đánh giá được mức độ hiệu quả của trạng thái hiện tại so với mục tiêu. Mục tiêu của Critic  $Q_{w1}$  được tham số hoá bởi  $\hat{w}$  và mục tiêu của Actor  $\pi_\theta$  được tham số hoá bởi  $\theta$  [16].

• **Giá trị mục tiêu:** Critic sử dụng chính sách làm trơn để làm giảm khả năng bị ảnh hưởng bởi nhiễu hoặc giá trị bất thường.

$$\varphi_k = r_k + \gamma \min \left\{ Q_{w1}(s_{k+1}, \pi_\theta(s_{k+1})), Q_{w2}(s_{k+1}, \pi_\theta(s_{k+1})) \right\} + \eta \quad (19)$$

• **Cập nhật:** Giá trị Q của Critic được cập nhật dựa trên sai số giữa giá trị Q hiện tại và

giá trị mục tiêu, sử dụng hàm đánh giá để điều chỉnh mô hình. Được mô tả

$$\begin{aligned} L_1 &\leftarrow \frac{1}{N} \sum (\varphi_k - Q_{w_1}(s_k, a_k))^2 \\ L_2 &\leftarrow \frac{1}{N} \sum (\varphi_k - Q_{w_2}(s_k, a_k))^2 \end{aligned} \quad (20)$$

Quá trình huấn luyện mô hình được thực hiện nhằm tối đa hoá giá trị *Reward* cũng như tối thiểu hoá giá trị hàm đánh giá. Khi đó giá trị được cập nhật lại.

$$\begin{aligned} w_1 &\leftarrow \operatorname{argmin} \frac{1}{n} \sum (\varphi_k - Q_{w_1}(s_k, a_k))^2 \\ w_2 &\leftarrow \operatorname{argmin} \frac{1}{n} \sum (\varphi_k - Q_{w_2}(s_k, a_k))^2 \end{aligned} \quad (21)$$

Tham số mục tiêu của *Actor* được cập nhật theo chính xác hàm Gradient được mô tả [17]

$$\nabla \theta_j = \frac{1}{N} \sum_i (\nabla a Q_{w_1}(s_k, a_k) \nabla \theta \pi_\theta(s_k)) \quad (22)$$

---

#### **Thuật toán 1:** Điều khiển JS với giải thuật RL

**Đầu vào:** Tham số mô hình JuR, giá trị đặt, và thông số huấn luyện.

**Đầu ra:** Mô hình điều khiển tối ưu.

- 1 **Khởi tạo:** Mạng critic  $Q_{w_1}, Q_{w_2}$  và mạng *Actor*  $\pi_\theta$  với thông số  $w_1, w_2, \theta$ .
- 2 **Khởi tạo:** Thông số mạng mục tiêu  
 $w_1 \leftarrow w_1, w_2 \leftarrow w_2, \theta \leftarrow \theta$
- 3 **Khởi tạo:** bộ đệm  $B$ , hệ số  $K_p, K_I, K_D$
- 4 **While**  $k = 1$  to  $T$  **do**  
 Xuất tín hiệu tác động  $a \leftarrow \pi_\theta + \eta$  và thu thập dữ liệu hồi tiếp phần thưởng  $r_k$  và trạng thái mới  $s_{k+1}$ .
- 5 Cập nhật và lưu giá trị  $s_k, a_k, s_{k+1}, r_k \rightarrow B$ .  
 Tính toán các tham số [18]  
 $a_k \leftarrow \pi_\theta(s_{(i+1)}) + \eta$
- 6  $\eta \leftarrow \operatorname{clip}(N(\theta, \sigma), -c, c)$  (23)  
 $\varphi_k \leftarrow r_k + \gamma \min\{Q_{w_1}, Q_{w_2}\} + \eta$
- 7 Cập nhật mạng Critic  $\leftarrow (21)$
- 8 Cập nhật  $\theta \leftarrow (22)$
- 9 Cập nhật mục tiêu *Actor-Critic*  
 $w_1 \leftarrow \tau w_1 + (1 - \tau) w_1$  (24)  
 $w_2 \leftarrow \tau w_2 + (1 - \tau) w_2$  (25)  
 $\theta \leftarrow \tau \theta + (1 - \tau) \theta$  (26)
- 10
- 11 **end while**
- 12 **end**

Khi đó giữa giá trị  $Q$  hiện tại và  $Q$  mục tiêu, được mô tả thông qua hàm đánh giá. Quá trình huấn luyện mô hình được thực hiện nhằm tối đa hoá phần thưởng cũng như tối thiểu hoá giá trị hàm đánh giá. Quá trình được mô tả trong Thuật toán 1.

## 4. Mô phỏng và đánh giá

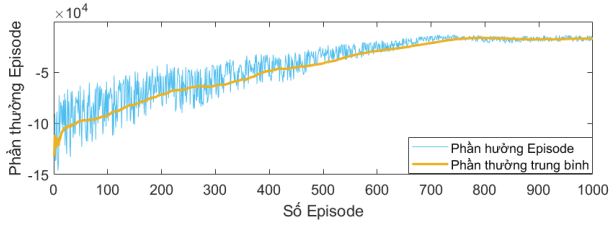
### 4.1. Thiết lập tham số

**Bảng 1.** Thông số giàn Tam Đảo 05 theo tỉ lệ 1:100.

Mô tả	Thông số
Loại	Giàn tự nâng
Chiều dài chân	1.62 m
Chiều cao thân	0.23 m
Số động cơ	6
Dòng điện định mức	6 A
Tốc độ định mức	468 RPM
Công suất định mức	60 W
Điện áp định mức	24 V
Mô-men định mức	15 kg/cm

Thông số mô phỏng được tham chiếu theo giàn khoan Tam Đảo 05 với các tham số số tỉ lệ 1:100 so với giá trị thực, chi tiết trình bày trong Bảng 1 [5]. Triển khai thử nghiệm bao gồm xây dựng mô hình và giải pháp thuật toán đề xuất được thực hiện trên môi trường tính toán Matlab 2023b.

Quá trình huấn luyện *Agent* RL-TD3 được thực hiện với các thông số thiết lập trong Bảng 2. Cụ thể, mạng *Actor* khởi tạo với thông số tốc độ học là 0.001 và áp dụng thuật toán tối ưu hóa Gradient descent với ngưỡng gradient bằng 1. Mạng *Critic* cũng được huấn luyện với tốc độ học 0.001 và tối ưu hóa bằng Gradient descent có cùng ngưỡng gradient là 1. Trong quá trình huấn luyện, dữ liệu được lấy mẫu với chu kỳ 0.1 giây, kích thước dữ liệu là 128 mẫu và tổng dung lượng bộ đệm phát lại được thiết lập là 106 mẫu.



**Hình 3.** Quá trình huấn luyện *Agent*.

Trong mỗi quá trình huấn luyện gồm 1000 tập mẫu, với mỗi tập kéo dài 100 bước. Ban đầu, giá trị hàm *Reward* biến động mạnh, nhưng sau quá trình tích lũy kinh nghiệm, giá trị dần ổn định và tăng đều trong suốt quá trình huấn luyện. Kết quả **Hình 3** cho thấy giá trị *Reward* trung bình mỗi 100 tập hội tụ ổn định xấp xỉ  $-1.3 \times 10^4$ . Ngoài ra, nhằm minh chứng hiệu quả của giải pháp đề xuất, tác giả so sánh kết quả đạt được với giải pháp RNN.

**Bảng 2.** Thông số huấn luyện *Agent*.

Mô tả	Thông số
Kích thước tập trích xuất	106 mẫu
Kích thước bộ đệm phát lại	500.000 mẫu
Biên độ nhiễu Gaussian	0.1
Thời gian thực hiện 1 bước	1 giây
Số bước 1 tập	100 bước
Số tập tối đa	1000 tập
Ngưỡng Gradient	1
Hệ số cắt giảm	0.97
Khởi tạo mạng <i>Actor</i>	[0.001,2,2]
Tốc độ học mạng <i>Actor</i>	0.001
Tốc độ học mạng <i>Critic</i>	0.001

#### 4.2. So sánh với mạng nơ-ron hồi quy

Mạng nơ-ron hồi quy (RNN) là dạng mạng nơ-ron nhân tạo được thiết kế để xử lý dữ liệu tuần tự. Điểm đặc trưng của RNN so với các loại mạng nơ-ron khác là khả năng duy trì thông tin trạng thái từ các bước trước đó trong chuỗi, nhờ vào cơ chế hồi quy. Điều này giúp RNN cập nhật được các mối tương quan và sự phụ thuộc giữa các phần tử trong chuỗi. Thiết lập một mạng RNN gồm ba lớp: một lớp đầu vào, một lớp

ẩn, và một lớp đầu ra. Trong đó, có một vòng phản hồi ở một lớp ẩn [19], trạng thái tại lớp ẩn có hồi quy tại thời điểm  $t$  được tính theo công thức sau

$$h_t = \text{Tanh}(W_h \cdot h_{(t-1)} + W_x \cdot x_t + b_h) \quad (27)$$

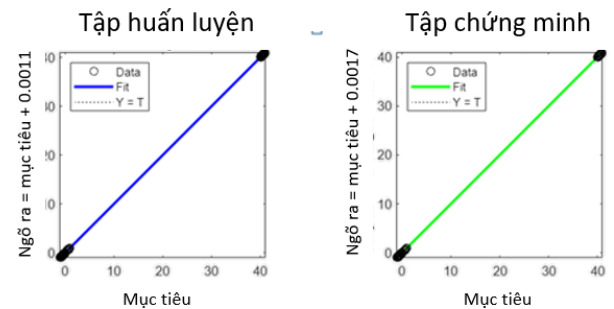
Trong đó  $h_t$  là trạng thái ẩn tại thời điểm  $t$ ,  $h_{t-1}$  là trạng thái ẩn tại thời điểm  $t-1$ , và  $x_t$  là đầu vào tại thời điểm  $t$ . Mặt khác,  $W_h$  là ma trận trọng số kết nối trạng thái ẩn,  $W_x$  là ma trận trọng số kết nối đầu vào, và  $b_h$  là hệ số bias [20]. Ngõ ra tại thời điểm  $t$  được tính toán như sau:

$$y_t = \text{ReLU}(W_y \cdot h_t + b_y) \quad (28)$$

Với  $y_t$  là đáp ứng tại thời điểm  $t$  và  $W_y$  là ma trận trọng số kết nối đầu ra với hệ số bias  $b_y$ . Mạng trên sẽ được huấn luyện với tập đầu vào là các thông số đáp ứng và đầu ra là các hệ số  $K_p$ ,  $K_i$ , và  $K_d$ . Mục tiêu của quá trình huấn luyện là tối thiểu hóa sai số giữa giá trị đầu ra  $y_i$  với giá trị mẫu  $s_i$  [21] như sau:

$$\Delta_i = y_i - s_i \quad (29)$$

Tập dữ liệu huấn luyện gồm 1000 mẫu (được tổng hợp tại **Bảng 3**). Các tập mẫu được tập hợp từ ba bộ số liệu tương ứng với  $K_p$ ,  $K_i$ , và  $K_d$  kiểm nghiệm trên mô hình mô phỏng để thu thập các đáp ứng về sai số xác lập, độ vọt lố, thời gian đáp ứng, và thời gian xác lập.



**Hình 4.** Tương quan giá trị thực - phỏng đoán.

Sau huấn luyện, các mẫu hội tụ ở các tất cả các tập huấn luyện được minh chứng như **Hình 4**, cho thấy mạng nơ-ron đã được huấn luyện thành công với các thông số  $K_{P-RNN} = 44.02$ ,  $K_{I-RNN} = 1.105$  và  $K_{D-RNN} = -1.290$ .

**Bảng 3.** Tập dữ liệu sử dụng trong huấn luyện.

Đầu vào				Đầu ra		
Sai số xác lập	Vọt lỗ	Đáp ứng	Ổn định	$K_p$	$K_I$	$K_D$
0.012	0.005	2.778	4.351	40.60	0.436	-0.369
0.000	0.000	2.697	4.918	40.21	0.001	-0.847
0.006	0.004	2.795	4.594	40.67	0.219	-0.274
0.011	0.005	2.764	4.418	40.32	0.388	-0.437
0.020	0.004	2.789	4.102	40.63	0.724	-0.318
0.021	0.004	2.820	4.103	40.35	0.767	-0.125
0.007	0.004	2.728	4.588	40.15	0.240	-0.640
0.001	0.001	2.723	4.889	40.08	0.035	-0.695
0.025	0.003	2.689	3.896	40.73	0.095	-0.930
0.014	0.005	2.707	4.212	40.83	0.53	-0.779

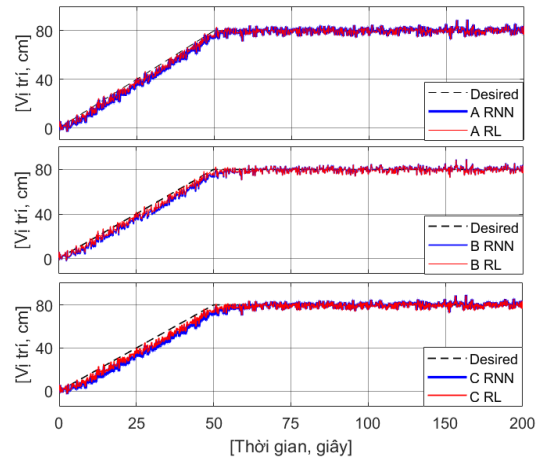
### 4.3. Kết quả mô phỏng

#### 4.3.1. Kịch bản thử nghiệm 1

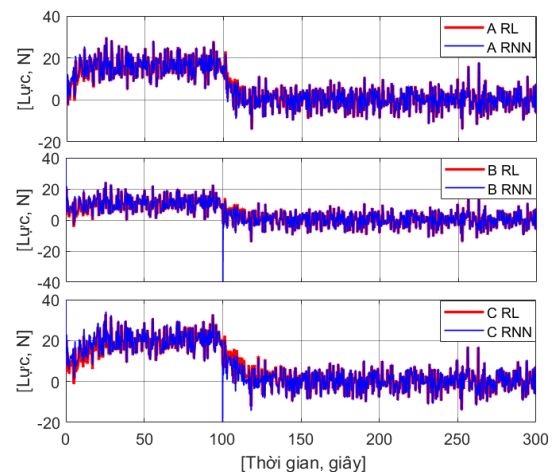
Kịch bản 1 thiết lập thử nghiệm nâng thân giàn lên độ cao xác định và giữ ổn định nhằm chuẩn bị cho quá trình khai thác. Trong kịch bản này, các giải pháp điều khiển được áp dụng nâng thân giàn lên độ cao 80cm từ vị trí tham chiếu là 0cm với tốc độ nâng dưới 1.5cm/s, sau đó giàn được giữ ở vị trí cố định nhằm đánh giá tính ổn định của hệ thống JS trong quá trình điều khiển. Các kết quả khả thi với giải pháp đề xuất RL (đường đỏ) so sánh với bộ điều khiển RNN (đường xanh) được thể hiện trên các Hình 5, Hình 6 và Hình 7. Chi tiết so sánh các tham số đáp ứng được tổng hợp tại Bảng 4.

Hình 5 chỉ ra kết quả thử nghiệm theo kịch bản 1 với 2 giải pháp RL và RNN trong cùng một môi trường thử nghiệm. Tổng quan, giải pháp thực nghiệm RL luôn có kết quả đáp ứng vượt trội so với giải pháp RNN, đặc biệt tại chân B đạt trạng thái ổn định nhất trong 3 chân. Trong thời điểm nâng thân giàn 0s đến 50s, RL có xu hướng bám theo giá trị đặt tốt hơn so với RNN trong tất cả 3 chân. Khi ở chế

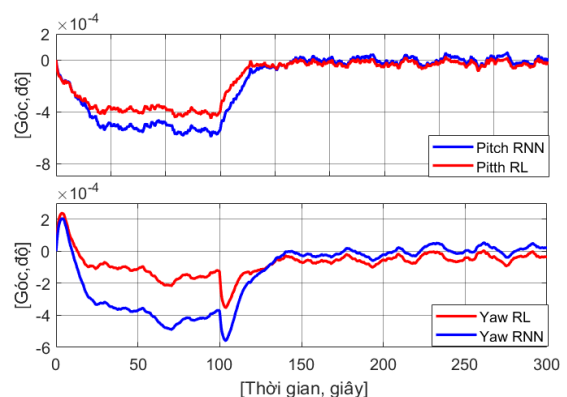
độ giữ vị trí từ 50s đến 200s, giải pháp RL có biên độ dao động thấp hơn so với RNN, các giá trị này ảnh hưởng bởi các yếu tố môi trường. Ngoài ra, Hình 6 mô tả giá trị lực điều khiển theo thời gian của 2 giải pháp, có thể thấy giá trị điều khiển của giải pháp RL luôn nằm ở mức ổn định. Tuy nhiên, với RNN xảy ra hiện tượng giá trị vọt lỗ bất thường, có thể gây nguy hại cho các cơ cấu cơ khí của JuR.



**Hình 5.** Đáp ứng RNN và RL ở ba chân.



**Hình 6.** Lực điều khiển do RNN và RL ở ba chân.



**Hình 7.** Kết quả đáp ứng theo góc lệch.

Kết quả trên **Bảng 4** cho thấy giải pháp RL yêu cầu lực khởi động tối đa lần lượt là 29/23.5/33N, giảm thiểu đáng kể năng lượng tiêu hao so với 30/24/36N của RNN. Độ chính xác của RL cũng được khẳng định qua góc lệch phương dọc tối đa  $3.6 \times 10^{-4}$  độ và phương ngang tối đa  $4 \times 10^{-4}$  độ, thấp hơn so với góc lệch tương ứng  $5.7 \times 10^{-4}$  độ và  $5 \times 10^{-4}$  độ của RNN. Trong khi đó, thời gian đáp ứng của RL chỉ 43 giây nhanh hơn RNN khi mất 46 giây, đảm bảo hệ thống phản ứng tức thì tốt hơn. Cuối cùng, sai số xác lập của RL là  $-2.1 \pm 0.5\%$ , thấp hơn nhiều so với  $-4.5 \pm 0.5\%$  của RNN, thể hiện sự chính xác tốt hơn trong quá trình hoạt động và thỏa mãn yêu cầu nêu tại Ghi chú 1.

**Bảng 4.** Kết quả các giải pháp của Kịch bản 1.

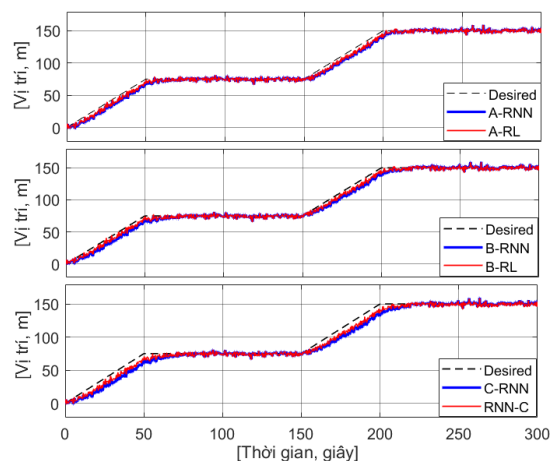
Mô tả	RNN [16]	RL
Lực tối đa	30/24/36N	29/23.5/33N
Góc lệch ngang	$5 \times 10^{-4}$ độ	$4 \times 10^{-4}$ độ
Góc lệch dọc	$5.7 \times 10^{-4}$ độ	$3.6 \times 10^{-4}$ độ
Thời gian đáp ứng	43s	46s
Độ vọt lố	6%	4%
Sai số xác lập	$-4.5 \pm 0.5\%$	$-2.1 \pm 0.5\%$

#### 4.3.1. Kịch bản thử nghiệm 2

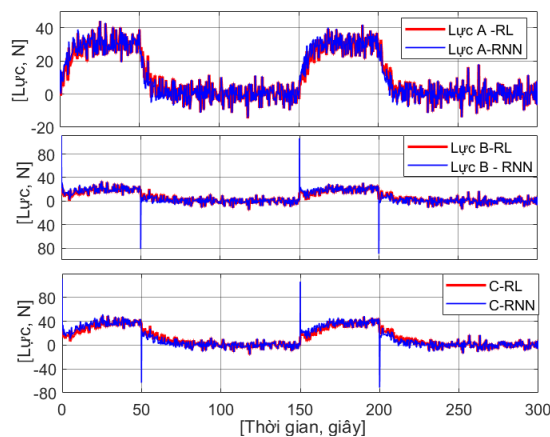
Bên cạnh đó, kịch bản 2 mô phỏng hoạt động JuR khi thay đổi chế độ làm việc. Áp dụng các bộ điều khiển JS thực hiện nâng thân giàn đến vị trí 80cm và giữ cố định trong khoảng thời gian làm việc, tiếp theo thân giàn tiếp tục di chuyển lên vị trí 150cm nhằm thiết lập trạng thái làm việc mới. Kết quả mô phỏng thể hiện chi tiết trên **Hình 8**, **Hình 9** và **Hình 10**.

Kết quả trên **Hình 8** cho thấy tại thời gian 0s đến thời điểm 50s và 150s đến 200s, khi giàn khoan đang ở chế độ nâng thân giàn, bộ điều khiển RL cho đáp ứng bám theo giá trị đặt tốt hơn so với RNN. Khi chuyển sang chế độ giữ vị trí cố định, giải pháp RL mang lại kết quả ít dao động hơn so với giải thuật RNN trên tất cả 3 chân, qua đó thể hiện tính hiệu quả

của giải pháp dưới sự tác động của môi trường. Ngoài ra, trong thời điểm chuyển chế độ 50s, 150s và 200s, giải thuật RNN luôn có khoảng dao động lớn, trong thực tế việc xuất tín hiệu lớn đột ngột ảnh hưởng rất lớn đến hệ thống truyền động và kết cấu động cơ. Ngoài ra, khả năng kiểm soát góc lệch dọc và lệch ngang thể hiện chi tiết trên **Hình 10**. Giải pháp điều khiển RL giữ vị trí tốt hơn so với giải thuật RNN, với biên độ góc lệch và khả năng nhanh chóng điều khiển giàn về trị trí cân bằng, tại thời điểm 50s và 200s trong khi RNN có thời gian phản ứng chậm hơn.



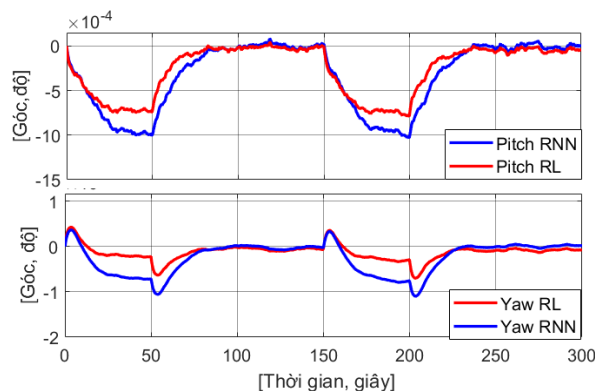
**Hình 8.** Đáp ứng RNN và RL ở ba chân.



**Hình 9.** Lực điều khiển do RNN và RL ở ba chân.

Trong kịch bản 2, bộ điều khiển RL yêu cầu lực tối đa thấp hơn ở 43/45/58N, so với 45/102/92N của RNN, với các thông số chi tiết thể hiện tại **Bảng 5**. Điều này cho thấy khả năng tiết kiệm năng lượng và hiệu suất điều khiển của RL cao hơn. Với góc lệch phương dọc là  $-10 \times 10^{-4}$  độ và phương ngang là  $-8 \times 10^{-4}$

độ, RL mang lại độ chính xác tốt hơn, trong khi RNN có góc lệch lớn hơn với  $-1.1 \times 10^{-4}$  độ tại trục dọc và  $-0.6 \times 10^{-4}$  độ tại phương ngang. Trong Hình 10 cho thấy trong suốt thời gian hạ xuống, các góc nghiêng của RL gần như không đổi trong khi RNN có xu hướng tăng dần cho thấy RL chống rung lắc tốt hơn RNN.



Hình 10. So sánh góc lệch giữa các trục.

Thời gian đáp ứng của RL và RNN lần lượt là 43 giây và 41 giây, tương tự Kịch bản 1. Tuy nhiên, độ chênh lệch của cả hai giải pháp chỉ chênh lệch 2 giây. Cả hai phương pháp đều có độ vọt lố rất thấp, nhưng kết quả với RL (1.1%) đạt được tốt hơn so với với RNN (1.3%). Bên cạnh đó, về sai số xác lập thì RL vẫn đạt kết quả tốt hơn ở mức  $-1.3 \pm 0.5\%$  so với RNN là  $-1.5 \pm 0.5\%$ , qua đó thể hiện độ chính xác cao hơn trong việc duy trì trạng thái ổn định, thỏa mãn các vấn đề nêu tại Ghi chú 2.

Bảng 5. So sánh đáp ứng trong Kịch bản 2.

Mô tả	RNN [16]	RL
Lực tối đa	45/102/92N	43/45/58N
Góc lệch ngang	$-1 \times 10^{-3}$ độ	$-0.8 \times 10^{-3}$ độ
Góc lệch dọc	$-1.1 \times 10^{-3}$ độ	$-0.6 \times 10^{-3}$ độ
Thời gian đáp ứng	41 giây	43 giây
Độ vọt lố	1.3%	1.1%
Sai số xác lập	$-1.5 \pm 0.5\%$	$-1.3 \pm 0.5\%$

## 5. Kết luận

Bài báo này phát triển giải pháp điều khiển RL cho quá trình nâng hạ thân giàn. Giàn khoan JuR được mô hình hóa trên môi

trường tính toán Matlab, nhằm phát triển giải pháp điều khiển RL nâng cao chất lượng tham số điều khiển chuyển động nâng hạ thân giàn dưới các tác động môi trường. Các kết quả khả thi trong quá trình thực nghiệm so sánh với giải pháp RNN đã chứng minh hiệu quả của giải pháp đề xuất. Trong tương lai, nghiên cứu này sẽ phát triển học tăng cường đa tác nhân (multi-agent) để điều khiển riêng biệt 36 động cơ nâng hạ trong hệ thống thực tế.

## Đóng góp của các tác giả trong bài báo

**Đặng Xuân Kiên:** Phương pháp, phân tích chính thức, biên soạn – đánh giá và biên tập; **Đỗ Việt Dũng:** Đánh giá và biên tập, phân tích dữ liệu, xác thực; **Lý Sợi:** Tính toán dữ liệu số, quản lý dữ liệu, viết – bản thảo gốc; **Vũ Hải Long:** Phân tích, tính toán dữ liệu số.

## Tuyên bố không xung đột lợi ích và cam kết bản quyền

Các tác giả tuyên bố về sự không xuất hiện những xung đột tiềm ẩn từ nghiên cứu này, và cam kết bài báo chưa từng được công bố trước đây.

## Chia sẻ dữ liệu theo yêu cầu

Dữ liệu sẽ được cung cấp theo yêu cầu.

1<sup>st</sup> Đặng Xuân Kiên\*. Nhóm nghiên cứu AIT, Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Hồ Chí Minh

2<sup>nd</sup> Đỗ Việt Dũng. Nhóm nghiên cứu AIT, Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Hồ Chí Minh

3<sup>rd</sup> Lý Sợi. Nhóm nghiên cứu AIT, Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Hồ Chí Minh; Khoa Công nghệ điện, Trường Cao đẳng Công nghệ cao Đồng An, Bình Dương

4<sup>th</sup> Vũ Hải Long. Liên doanh Vietsovpetro, Bà Rịa - Vũng Tàu

\*Corresponding author: kien.dang@ut.edu.vn

## Tài liệu tham khảo

- [1] J. T. Yi, F. Liu, T. Zhang, Z. Z. Qiu, and Y. Y. Zhang, "Determination of the ultimate consolidation settlement of jack-up spudcan footings embedded in clays," *Ocean Eng.*, vol. 236, no. 4, pp. 1–13, 2021, doi: [10.1016/j.oceaneng.2021.109509](https://doi.org/10.1016/j.oceaneng.2021.109509).
- [2] ISO 19905-1:2023: Oil and gas industries including lower carbon energy - Site-specific assessment of mobile offshore units, 3rd ed. International Organization for Standardization, 2023.
- [3] X. K. Dang, L. A. H. Ho, and V. D. Do, "Analyzing the sea weather effects to the ship maneuvering in Vietnam's Sea from Binh Thuan Province to Ca Mau Province based on fuzzy control method," *TELKOMNIKA*, vol. 16, no. 2, pp. 533–543, 2016, doi: [10.12928/telkomnika.v16i2.7753](https://doi.org/10.12928/telkomnika.v16i2.7753).
- [4] M. Pashna, R. Yusof, Z. H. Ismail, T. Namerikawa, and S. Yazdani, "Autonomous multi-robot tracking system for oil spills on sea surface based on hybrid fuzzy distribution and potential field approach," *Ocean Eng.*, vol. 207, pp. 1–11, 2020, doi: [10.1016/j.oceaneng.2020.107238](https://doi.org/10.1016/j.oceaneng.2020.107238).
- [5] V. D. Do, X. K. Dang and A. T. Le, "Fuzzy adaptive interactive algorithm for rig balancing optimization," in 2017 International Conference on Recent Advances in Signal Processing, Telecommunications & Computing, Da Nang, Vietnam, pp. 143-148, 2017, doi:[10.1109/SIGTELCOM.2017.7849812](https://doi.org/10.1109/SIGTELCOM.2017.7849812).
- [6] S. Wang, Y. Yin, P. Li, Y. Zhang, Y. Wang, and S. Tong, "Cognitive control using adaptive RBF neural networks and reinforcement learning for networked control system subject to time-varying delay and packet losses," *Arab. J. Sci. Eng.*, vol. 46, no. 10, pp. 10245–10259, 2021, doi: [10.1007/s13369-021-05752-y](https://doi.org/10.1007/s13369-021-05752-y).
- [7] K. J. Aström, and T. Hägglund, "Advanced PID Control," *IEEE Control Systems*, vol. 461, 2006, doi: [10.1109/MCS.2006.1580160](https://doi.org/10.1109/MCS.2006.1580160).
- [8] G. Bujgoi, and D. Sendrescu, "Tuning of PID Controllers Using Reinforcement Learning for Nonlinear System Control," *Processes*, vol. 13, no. 3, pp. 735, 2025, doi: [10.3390/pr13030735](https://doi.org/10.3390/pr13030735).
- [9] O. Dogru, K. Velswamy, F. Ibrahim, Y. Wu, A. S. Sundaramoorthy, B. Huang, S. Xu, M. Nixon, and N. Bell, "Reinforcement learning approach to autonomous PID tuning," *Computers & Chemical Engineering*, vol. 161, pp. 107760-107777, 2022, doi: [10.1016/j.compchemeng.2022.107760](https://doi.org/10.1016/j.compchemeng.2022.107760).
- [10] N. Mizuno, and T. Koide, "Application of Reinforcement Learning to Generate Non-linear Optimal Feedback Controller for Ship's Automatic Berthing System," *IFAC-PapersOnLine*, vol. 56, pp. 162–168, 2023, doi: [10.1016/j.ifacol.2023.02.028](https://doi.org/10.1016/j.ifacol.2023.02.028).
- [11] Y. Hu, W. Wang, H. Liu and L. Liu, "Reinforcement Learning Tracking Control for Robotic Manipulator With Kernel-Based Dynamic Model," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3570-3578, 2020, doi: [10.1109/TNNLS.2019.2945019](https://doi.org/10.1109/TNNLS.2019.2945019).
- [12] F. Wang, W. Yao, Y. Liu, and C. Li, "An analytical procedure to predict transverse vibration response of Jack-Up Riser under the random wave load," *Shock.Vib*, vol. 2020, pp. 5072989-5072998, doi: [10.1155/2020/5072989](https://doi.org/10.1155/2020/5072989).
- [13] Q. Yin, J. Yang, G. Xu, R. Xie, M. Tyagi, L. Li, X. Zhou, N. Hu, G. Tong, C. Fu, and D. Pang, "Field experimental investigation of punch-through for different operational conditions during the jack-up rig spudcan penetration in sand overlying clay," *Journal of Petroleum Science and Engineering*, vol. 195, pp. 107823-107844, 2020, doi: [10.1016/j.petrol.2020.107823](https://doi.org/10.1016/j.petrol.2020.107823).
- [14] Y. Yie, J. Huang, Y. Li, Y. Tian, G. Liu, and D. Leng, "Experimental study on hydrodynamic characteristics of three truss-type legs of jack-up offshore platform," *Ocean Eng.*, vol. 234, pp. 1–15, 2021, doi: [10.1016/j.oceaneng.2021.109305](https://doi.org/10.1016/j.oceaneng.2021.109305).
- [15] J. S. Wu, and C. Y. Chang, "Structural simplification of jack-up rig and its dynamic responses in regular waves," *Journal of Ship Research*, vol. 32, no. 2, pp.134-153, 1988, doi: [10.5957/jsr.1988.32.2.134](https://doi.org/10.5957/jsr.1988.32.2.134).
- [16] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, no. 1, pp. 1–7, 2016, doi: [10.1609/aaai.v30i1.10295](https://doi.org/10.1609/aaai.v30i1.10295).
- [17] E. Mahmoud, A. A. -R. Hourani and S. Mammam, "PID Control by Safe Reinforcement Learning for Vehicle Lateral Control," in *2024 International Conference on Networking, Sensing and Control*, Hangzhou, China, 2024, pp. 1-6, doi: [10.1109/ICNSC62968.2024.10760015](https://doi.org/10.1109/ICNSC62968.2024.10760015).
- [18] Z. Zhang, Y. S. Ong, D. Wang and B. Xue, "A Collaborative Multiagent Reinforcement Learning Method Based on Policy Gradient Potential," *IEEE Transactions on Cybernetics*, vol. 51, no. 2, pp. 1015-1027, 2021, doi: [10.1109/TCYB.2019.2932203](https://doi.org/10.1109/TCYB.2019.2932203).
- [19] R. Zhang, D. Liu, T. Li, D. Li, C. Tian and Z. Zheng, "A Novel Attitude Control Strategy of Reusable Launch Vehicle via an Improved TD3 Reinforcement Learning," 2023 42<sup>nd</sup> Chinese Control Conference, Tianjin, China, 2023, pp.

- 2264-2269, doi: [10.23919/CCC58697.2023.10240656](https://doi.org/10.23919/CCC58697.2023.10240656).
- [20] S. Silvestrini, and M. Lavagna, "Deep Learning and Artificial Neural Networks for Spacecraft Dynamics," *Navigation and Control*, vol. 6, no. 10, pp. 270, doi: [10.3390/drones6100270](https://doi.org/10.3390/drones6100270).
- [21] Z. Ma, H. Zhang and J. Liu, "DB-RNN: An RNN for Precipitation Nowcasting Deblurring," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 5026-5041, 2024, doi: [10.1109/JSTARS.2024.3365612](https://doi.org/10.1109/JSTARS.2024.3365612).