

Jornadas de Automática

Control tolerante a fallos en propulsores de ROV mediante SMC+TD3

Jose Ramón Llata^{a,*}, Jose Joaquín Sainz^a, Elias Revestido^a, Carlos Torre-Ferrero^a, Luciano Alonso^a, Sandra Robla^a

^aDepartamento de tecnología electrónica e ingeniería de sistemas y automática, Universidad de Cantabria. Av. de los Castros s/n, Santander, 39006. Spain.

To cite this article: Llata, J.R., Sainz, J.J., Revestido, H., Torre-Ferrero, C., Alonso, L., Robla, S. Fault-Tolerant Control of ROV Thrusters via SMC+TD3. Jornadas de Automática, 47.
<https://doi.org/10.17979/ja-cea.2026.47.13690>

Resumen

Este trabajo analiza una arquitectura de control tolerante a fallos producidos en los propulsores de un ROV tipo BlueROV2 que está fundamentada en la utilización combinada de un sistema de control basado en modos deslizantes (SMC) y de un controlador basado en aprendizaje profundo reforzado y, más concretamente, en Twin Delayed Deep Deterministic Policy Gradient (TD3). En este diseño, el algoritmo SMC establece un control dinámico robusto, siempre activado, y que toma constantemente las acciones necesarias para mitigar los efectos de las perturbaciones propias del medio, imprecisiones en el modelado, etc. Por otro lado, el agente TD3, una vez ha sido entrenado convenientemente, se encargará de realizar una corrección adicional puntual orientada a mitigar los efectos producidos por imprevistas pérdidas de eficiencia de los propulsores. El trabajo de validación realizado en simulación muestra reducciones del error de yaw ante fallos horizontales, manteniendo los valores de roll y pitch acotados y preservando el comportamiento de SMC ante fallos verticales.

Palabras clave: Vehículo subacuático autónomo, Arquitecturas de Control en sistemas marinos, Control tolerante a fallos, Control por aprendizaje reforzado, Control por modos deslizantes.

Fault-Tolerant Control of ROV Thrusters by using SMC+TD3

Abstract

This work examines a fault-tolerant control architecture for thruster faults in a BlueROV2-type ROV. The proposed architecture combines a sliding mode control strategy (SMC) with a deep reinforcement learning controller, specifically based on Twin Delayed Deep Deterministic Policy Gradient (TD3). In this design, the SMC provides a robust dynamic control action that remains active at all times, continuously compensating for environmental disturbances, modelling inaccuracies, and other sources of uncertainty. The TD3 agent, once properly trained, provides an additional targeted correction aimed at mitigating the effects of unexpected losses of thruster efficiency. The simulation-based validation shows that the proposed approach reduces yaw error under horizontal thruster faults, while keeping roll and pitch bounded and preserving the nominal SMC behaviour under vertical thruster faults.

Keywords: Autonomous underwater vehicles, Control architectures in marine systems, Fault-tolerant control, Reinforcement learning control, Sliding mode control.

1. Introducción

Es indudable el hecho de que la sociedad actual se dirige cada vez más hacia una automatización del conjunto de los procesos, tareas y labores relacionadas con los ámbitos de la actividad humana. Uno de estos ámbitos es, sin duda, el de

la actividad subacuática, en la que los especialistas humanos se ven muchas veces realizando tareas en situaciones en las que ya el propio entorno es peligroso debido tanto a las condiciones del medio (presión, corrientes, baja visibilidad, humedad, etc.), como a las propias de la actividad que se esté desa-

*Autor para correspondencia: llataj@unican.es

rollando (exploración, actividades de mantenimiento, inspección submarina, soldadura y reparación, etc.). De esta forma, tanto los vehículos operados remotamente (ROVs) como los vehículos autónomos subacuáticos (UUVs) están experimentando un amplio despliegue y han pasado a ser una herramienta de enorme interés.

Desde el punto de vista del control, un vehículo subacuático autónomo continúa siendo un reto debido a su propia complejidad y al entorno en el que trabaja, de forma que, en la actualidad, existe una investigación constante en diversas líneas de control, como por ejemplo (Sainz et al. (2023)), (Sainz et al. (2024)), así como en control tolerante a fallos (Hosseinajad y Loueipour (2023)), (Yin et al. (2024)), (Liu et al. (2025)), objeto principal de este trabajo.

Dentro de este contexto, los fallos en propulsores representan uno de los problemas más críticos para la operación segura de un ROV. Una pérdida parcial o total de eficiencia en uno de los actuadores puede romper la simetría de actuación del vehículo y producir desviaciones importantes en la respuesta transitoria. En este sentido, el control tolerante a fallos busca mantener un comportamiento aceptable del sistema incluso cuando alguno de sus componentes presenta degradación. En el presente trabajo se considera una situación en la que se produce un fallo en los propulsores, de forma que afecta a la eficiencia de los mismos. Esta eficiencia de los propulsores se representa mediante lo que denominamos el vector de salud f_k , que presenta tantos valores como impulsores tenga el ROV, de forma que cada uno de estos valores representa el nivel de eficiencia de uno de los propulsores. Es importante señalar que en el presente trabajo se asume que este vector f_k está disponible, procedente de un modulo externo de diagnóstico o estimación que no es objeto de este trabajo.

La arquitectura del algoritmo de control presentada en este trabajo está formada por dos elementos clave. En primer lugar, un controlador de modos deslizantes (SMC) que siempre va a estar activo y que ha sido seleccionado debido a su robustez ante incertidumbres y perturbaciones acotadas. Este SMC proporcionará una acción de control para los seis grados de libertad del vehículo, esto es, actuará en las siguientes fuerzas y momentos $[X, Y, Z, K, M, N]$, expresados en el sistema de coordenadas fijo o NED. En segundo lugar, el algoritmo de control utiliza una acción de control adicional basada en un agente de inteligencia artificial obtenido a través del procedimiento de aprendizaje profundo por refuerzo, cuya misión es generar señales de actuación complementarias al SMC con objeto de reducir al máximo los efectos producidos debido a un fallo, total o parcial, en los propulsores del ROV. Más concretamente, para este segundo elemento del algoritmo de control presentado, está formado por un algoritmo Twin Delayed Deep Deterministic Policy Gradient, o TD3, que es adecuado para entornos de control continuo (Lillicrap et al. (2016)) y que presenta como características principales el emplear una arquitectura actor-crítico, con dos críticos para reducir la sobreestimación, actualización retardada del actor y un suavizado de la política objetivo.

2. Modelo Dinámico

Para el desarrollo del presente trabajo se considera un ROV modelo Bluerov2, en la versión heavy que consta de 8

impulsores, y que es modelado como un vehículo subacuático de seis grados de libertad. El modelo se formula siguiendo la representación habitual de vehículos marinos, distinguiendo entre las variables de posición y orientación expresadas en el sistema inercial NED y las velocidades lineales y angulares expresadas en el sistema cuerpo. Además, se incorpora un modelo de propulsores y un modelo explícito de pérdida de eficiencia mediante un vector de salud de actuadores.

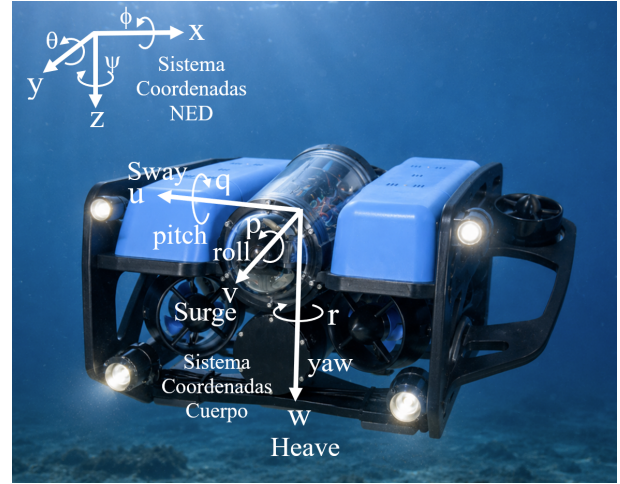


Figura 1: Sistemas de Referencia NED y body.

2.1. Modelo Dinámico del Bluerov2

El modelo cinemático y dinámico del vehículo está basado en las bien conocidas ecuaciones cinemáticas y dinámicas de (Fossen (2021)) y de (Benzon et al. (2022)):

$$\dot{\eta} = J(\eta)v \quad (1)$$

$$M\dot{v} + C(v)v + D(v_r)v_r + g(\eta) = \tau \quad (2)$$

donde $\eta = [x, y, z, \phi, \theta, \psi]^T$ es el vector de posición y orientación expresado en el sistema de coordenadas NED o fijo, $v = [u, v, w, p, q, r]^T$ son las velocidades lineales y angulares expresadas en el sistema de coordenadas móvil ligado al vehículo, también denominado cuerpo o BODY. En la fig. 1 se muestran los sistemas de coordenadas asociados a este sistema. $J(\eta)$ es la matriz de transformación entre el sistema cuerpo y el sistema inercial, que puede escribirse como:

$$J(\eta) = \begin{bmatrix} R_b^n(\phi, \theta, \psi) & 0_{3 \times 3} \\ 0_{3 \times 3} & T(\phi, \theta) \end{bmatrix} \quad (3)$$

En la que la expresión $R_b^n(\phi, \theta, \psi)$ transforma velocidades lineales desde el sistema cuerpo al sistema NED y la expresión $T(\phi, \theta)$ relaciona las velocidades angulares del cuerpo con las derivadas de los ángulos de Euler. De esta forma, la expresión (1) permite relacionar las velocidades expresadas en el sistema de coordenadas cuerpo con las velocidades expresadas en el sistema de coordenadas fijo o NED. Por otro lado, la ecuación (2), que modeliza el comportamiento dinámico del vehículo, está compuesta por los siguientes términos. M es la matriz de masa e inercia total, $C(v)$ agrupa los términos de Coriolis y fuerzas centrípetas, $D(v)$ representa el amortiguamiento hidrodinámico dependiente de la velocidad relativa, $g(\eta)$ contiene

los términos de restauración debidos al peso y la flotabilidad, τ es el vector de fuerzas y momentos generalizados aplicados al vehículo, expresado como $[X, Y, Z, K, M, N]$ donde X, Y, Z son las fuerzas generalizadas en surge, sway y heave, mientras que K, M, N son los momentos generalizados de roll, pitch y yaw. Por otro lado, $v_r = v - v_c$ es la velocidad relativa del vehículo, esto es, la diferencia entre la velocidad del vehículo y la velocidad de la corriente del agua, ambos expresados en el sistema de coordenadas cuerpo.

En el caso de la existencia de corrientes, la acuación (2) puede expresarse de forma más detallada como:

$$M_{RB}\dot{v} + C_{RB}(v)v + M_A\dot{v}_r + C_A(v_r)v_r + D(v_r)v_r + g(\eta) = \tau \quad (4)$$

donde M_{RB} representa la masa e inercia del cuerpo rígido, M_A representa los efectos de masa añadida inducidos por el fluido y $M = M_{RB} + M_A$. $C_{RB}(v)$ representa los términos de Coriolis y centrípetos asociados al cuerpo rígido, mientras que $C_A(v_r)$ representa los términos asociados a la masa añadida. $D(v_r)$ representa el amortiguamiento hidrodinámico dependiente de la velocidad relativa. En este trabajo, esta formulación se emplea como base física del entorno de simulación del ROV.

2.2. Modelo Dinámico de los Propulsores

El BlueROV2, en configuración heavy, dispone de ocho propulsores, por lo que el sistema es sobreactuado respecto a los seis grados de libertad del modelo dinámico. La relación entre los empujes individuales de los propulsores y el vector de fuerzas y momentos generalizados se expresa mediante una matriz de asignación $\tau = TF$, donde T es la matriz de asignación de propulsores y F es el vector de empujes individuales (Benzon et al. (2022)):

$$F = [F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8]^T \quad (5)$$

La matriz T recoge la orientación y posición de cada propulsor respecto al centro de masa del vehículo. Dado que existen ocho actuadores para generar un vector generalizado de dimensión seis, pueden existir distintas combinaciones de empujes individuales capaces de producir una misma acción generalizada τ . Esta redundancia es relevante para la asignación de control, la gestión de saturaciones y la compensación ante fallos de actuadores.

En el modelo de simulación, el controlador genera una acción generalizada deseada. A partir de dicha acción se calculan los empujes individuales requeridos por los propulsores. De forma genérica, esta asignación inversa puede expresarse como $F_{cmd} = T^{\dagger}\tau_{cmd}$, donde $T^{\dagger}$ representa una inversa generalizada o pseudoinversa de la matriz de asignación, τ_{cmd} es la acción generalizada solicitada por el controlador y F_{cmd} es el vector de empujes comandados. Cada propulsor está sujeto a límites físicos de empuje. Por tanto, el empuje individual aplicado viene definido como $F_i = \text{sat}(F_{cmd,i}, F_{i,\min}, F_{i,\max})$, siendo $F_{i,\min}$ y $F_{i,\max}$ los límites inferior y superior del empuje del propulsor i . En forma vectorial se puede escribir como $F = \text{sat}(F_{cmd}, F_{\min}, F_{\max})$. Estas saturaciones son importantes porque, en presencia de fallos severos, el controlador puede demandar fuerzas y momentos que no sean físicamente realizables por el conjunto de propulsores disponible. Finalmente, los empujes efectivos

generan la acción generalizada aplicada al vehículo mediante $\tau_{aplicada} = TF$.

Indicar finalmente que el modelo cinemático y dinámico del BlueROV2 se ha programado en el entorno de Matlab-Simulink, utilizando los parámetros especificados por Benzon et al. (2022)), buscando una ejecución eficiente y rápida que permita facilitar el proceso de aprendizaje del algoritmo TD3.

3. Fallo en Propulsores

3.1. Modelo de Fallo en Propulsores

Para representar pérdidas de eficiencia en los propulsores se introduce un vector de salud:

$$f_k = [f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8]^T \quad (6)$$

donde cada componente f_i representa la eficiencia remanente del propulsor i . Se impone

$$0 \leq f_i \leq 1, \quad i = 1, \dots, 8 \quad (7)$$

El caso $f_i = 1$ representa un propulsor sano, mientras que $f_i = 0$ representa un fallo total. Los valores intermedios $0 < f_i < 1$ modelan pérdidas parciales de eficiencia. El fallo se aplica sobre el empuje físico generado por cada propulsor después de la saturación. Así, el vector de empujes efectivos bajo fallo se calcula como $F_{fault} = f_k \odot F$, donde $\odot$ denota el producto elemento a elemento. Por tanto, el vector de fuerzas y momentos realmente aplicado al ROV bajo fallo queda definido como $\tau_{fault} = TF_{fault}$. Sustituyendo la expresión anterior, $\tau_{fault} = T(f_k \odot F)$. Este modelo permite representar fallos totales, fallos parciales y, de forma directa, fallos múltiples simultáneos. No obstante, en las validaciones principales de este trabajo se consideran fallos simples, es decir, escenarios en los que un único propulsor presenta una pérdida total de eficiencia. La expresión (8) modeliza un fallo en el propulsor j , siendo α el parámetro de eficiencia del propulsor.

$$f_i = \begin{cases} \alpha, & i = j \\ 1, & i \neq j \end{cases}, \quad 0 \leq \alpha < 1 \quad (8)$$

3.2. Supuesto de disponibilidad del vector de salud

En este trabajo se asume que el vector de salud f_k está disponible para el controlador. Esta hipótesis permite centrar el estudio en el problema de compensación de fallos, es decir, en cómo utilizar información sobre la pérdida de eficiencia de los propulsores para mejorar la respuesta dinámica del ROV. El controlador recibe, por tanto, una estimación o medida idealizada del estado de salud de los propulsores, dado por (6).

4. Controlador SMC

El sistema desarrollado en este trabajo utiliza un controlador SMC que constituye la acción principal del sistema y es responsable de mantener el seguimiento de referencia en condiciones nominales y bajo perturbaciones acotadas. El agente TD3 propuesto posteriormente no sustituye a este controlador, sino que añade una acción residual limitada cuando se detectan fallos horizontales de propulsores.

4.1. Seguimiento de referencia en posición y yaw

En las maniobras consideradas, el objetivo principal es el seguimiento de referencias en posición y yaw, manteniendo roll y pitch estabilizados alrededor de cero. Por tanto, la referencia empleada tiene la forma

$$\eta_{ref} = [x_{ref}, y_{ref}, z_{ref}, 0, 0, \psi_{ref}]^T \quad (9)$$

Esta elección responde a una operación típica de ROV, donde se desea controlar la posición espacial y la orientación de guiñada, evitando inclinaciones significativas en roll y pitch. Por tanto, los errores de roll y pitch se interpretan como variables de estabilización, no como referencias de seguimiento no nulas, esto es: $\phi_{ref} = 0$ y $\theta_{ref} = 0$. El objetivo de seguimiento puede expresarse como

$$\lim_{t \rightarrow \infty} [e_x(t), e_y(t), e_z(t), e_\psi(t)]^T = 0 \quad (10)$$

manteniendo simultáneamente, $\phi(t) \approx 0$ y $\theta(t) \approx 0$.

En el caso de yaw, el error debe calcularse con envolvimento angular para evitar saltos artificiales en torno a $\pm\pi$. Por tanto,

$$e_\psi = \text{atan2}(\sin(\psi - \psi_{ref}), \cos(\psi - \psi_{ref})). \quad (11)$$

La acción generada por el SMC actúa sobre los seis canales generalizados,

$$\tau_{SMC,NED} = [X_{SMC}, Y_{SMC}, Z_{SMC}, K_{SMC}, M_{SMC}, N_{SMC}]^T \quad (12)$$

Aunque las referencias no imponen valores no nulos en roll y pitch, los canales K_{SMC} y M_{SMC} son necesarios para estabilizar la actitud del vehículo. En la arquitectura propuesta posteriormente, esta responsabilidad se deja exclusivamente al SMC, evitando que el agente TD3 actúe directamente sobre los momentos de roll y pitch.

4.2. Papel del SMC como controlador base robusto

El controlador SMC se utiliza en este trabajo como controlador nominal y como controlador base robusto frente al que se evalúa la contribución del compensador residual TD3. La acción total aplicada en el caso nominal (sin fallo) es

$$\tau_{total,NED} = \tau_{SMC,NED} \quad (13)$$

y, cuando se incorpora el compensador residual TD3, la acción total pasa a ser

$$\tau_{total,NED} = \tau_{SMC,NED} + \Delta\tau_{TD3,NED} \quad (14)$$

Donde $\Delta\tau_{TD3,NED}$ es el incremento de fuerzas y pares que genera el controlador TD3 en el caso de existencia de fallo. Por tanto, el TD3 no sustituye al SMC, sino que añade una corrección residual. Esta elección es importante porque mantiene una acción de control base físicamente interpretable y robusta. El SMC sigue siendo responsable de la estabilización general del ROV, mientras que el TD3 se limita a compensar degradaciones transitorias bajo fallos de propulsores. Bajo condiciones nominales o ante fallos verticales para los cuales el agente TD3 se desactiva, el sistema opera únicamente con el controlador SMC:

$$\Delta\tau_{TD3,NED} = 0 \implies \tau_{total,NED} = \tau_{SMC,NED} \quad (15)$$

Esta propiedad permite que el sistema conserve el comportamiento del controlador nominal en escenarios donde el TD3 no aporta ventaja o puede resultar perjudicial. En particular, los resultados obtenidos muestran que para fallos verticales puros el uso del SMC sin acción residual es preferible al empleo indiscriminado del agente TD3.

El papel del SMC como base robusto puede resumirse en tres puntos. Primero, proporciona una respuesta estable y físicamente interpretable para el ROV en condiciones nominales. Segundo, mantiene la estabilización de roll y pitch incluso cuando el TD3 actúa sobre los canales de posición y yaw. Tercero, sirve como controlador de respaldo cuando el selector fault-aware desactiva la acción residual aprendida.

En este trabajo, el interés principal no es demostrar que el TD3 sustituye al SMC, sino que una compensación residual limitada y activada selectivamente puede mejorar el comportamiento transitorio bajo fallos horizontales sin degradar la estabilidad de actitud. Por ello, la comparación principal se realiza entre el SMC puro y la arquitectura híbrida SMC+TD3 residual 4DOF.

4.3. Control SMS Implementado

El controlador SMC empleado se basa en la formulación clásica de modos deslizantes para sistemas de segundo orden, siguiendo la estructura definida en (Utkin (1992)) y (Sutton y Barto (2018) e implementado tal y como se señala en (Benzon et al. (2022))). De esta forma se han implementado un conjunto de 6 controladores SMC independientes, uno por cada grado de libertad del ROV, en los que la superficie de deslizamiento viene definida por:

$$\sigma_i = \dot{e}_i + c_{0,i}e_i \quad (16)$$

donde $c_{0,i} > 0$ fija la dinámica sobre la superficie de deslizamiento y $\dot{e}_i$ es la derivada del error de cada grado de libertad, estando definido el error como $e_i = \eta_i - \eta_{ref,i}$. La ley de control implementada en cada eje viene definida por la ecuación

$$\tau_{SMC,NED,i} = \ddot{\eta}_{ref,i} - c_{0,i}\dot{e}_i - \rho_i \frac{\sigma_i}{|\sigma_i| + \epsilon_i} \quad (17)$$

donde $\rho_i = L_i + \frac{\alpha_i}{\sqrt{2}}$ es la ganancia robusta de cada eje y se ha introducido una aproximación a la función $\text{sgn}(\sigma_i)$, en la que ϵ_i define la capa límite. La tabla 1 muestra los diferentes parámetros utilizados en el controlador SMC de este trabajo, donde η_i representa cada grado de libertad i , $c_{0,i}$ es la ganancia de la superficie de deslizamiento, que determina la velocidad de convergencia del error, α_i es el margen adicional de convergencia para calcular la ganancia robusta, ϵ_i es el espesor de la capa límite suavizada para reducir el chattering, L_i es la cota de diseño de los términos inciertos no compensados explícitamente en el eje i y ρ_i es la ganancia robusta efectiva en el eje i .

Tabla 1: Parámetros del SMC para cada grado de libertad.

DOF	η_i	$c_{0,i}$	α_i	ϵ_i	L_i	ρ_i
Surge	x	3.0	0.5	0.03	1.17	1.5236
Sway	y	3.0	0.5	0.03	1.04	1.3936
Heave	z	2.0	0.1	0.05	0.70	0.7707
Roll	ϕ	0.4	0.1	0.10	2.01	2.0807
Pitch	θ	0.4	0.1	0.10	1.83	1.9007
Yaw	ψ	2.0	0.1	0.10	2.05	2.1207

5. Aprendizaje por refuerzo y algoritmo TD3

El objetivo de este trabajo no es desarrollar un controlador puramente basado en aprendizaje, sino utilizar aprendizaje reforzado (Sutton y Barto (2018)), (Goodfellow et al. (2016)), y más concretamente TD3, para aprender una acción correctora residual limitada que complemente al controlador SMC nominal bajo condiciones de fallo.

5.1. Motivación para usar TD3 en el control residual

El uso de TD3, para realizar esta acción correctora residual, se justifica, en primer lugar, porque la acción residual que debe aprenderse es continua. En segundo lugar, la relación entre fallo, estado del vehículo y corrección óptima no es trivial debido a los acoplamientos dinámicos. En tercer lugar, el agente puede aprender estrategias de compensación a partir de la interacción directa con el entorno de simulación, sin necesidad de establecer un modelo matemático del comportamiento del sistema en condición de fallo de los propulsores.

En este trabajo, el TD3 no genera la acción total de control. En su lugar, aprende una acción residual

$$\Delta\tau_{TD3,NED} \quad (18)$$

que se suma a la acción nominal del SMC dado por la ecuación (14). Esta formulación tiene una ventaja importante: el SMC mantiene el papel de controlador principal y el TD3 únicamente corrige el comportamiento bajo condiciones de fallo. Por tanto, el aprendizaje no parte de la necesidad de controlar completamente el ROV, sino de mejorar una respuesta base ya estabilizada. Así, en el presente trabajo, su uso se limita mediante los siguientes tres mecanismos. El TD3 solo actúa en caso de detección de fallo en propulsores, solo se le permite actuar ante fallos en los impulsores horizontales y su actuación se limita a cuatro grados de libertad. De esta forma, y como se verá más adelante, se ha demostrado experimentalmente que se obtiene una arquitectura de control más segura y eficaz.

6. Entrenamiento del agente TD3

Antes de disponer de un agente TD3 operativo, es necesario entrenarlo. El objeto de este entrenamiento es que aprenda la acción correcta, y complementaria al SMC, que debe proporcionar en el caso de fallos horizontales de los propulsores, para que la respuesta transitoria del ROV mejore frente a la respuesta que se obtiene utilizando el SMC como controlador único. El agente se entrena en simulación usando el modelo dinámico del ROV y el modelo de fallo mediante vector de salud f_k .

6.1. Vector de observación

El agente TD3 recibe en cada instante de muestreo un vector de observación que contiene información sobre el error de seguimiento, la actitud del vehículo, sus velocidades, la acción nominal generada por el SMC y el vector de salud de propulsores.

El vector de observación se define como

$$o_k = [e_x, e_y, e_z, e_\psi, \phi, \theta, v^T, \tau_{SMC,NED}^T, f_k^T]^T \quad (19)$$

El vector, con una dimensión de 26 señales, contiene los siguientes bloques:

e_x, e_y, e_z, e_ψ	errores de seguimiento en posición y yaw,
ϕ, θ	ángulos de roll y pitch,
v	velocidades lineales y angulares del ROV,
$\tau_{SMC,NED}$	acción nominal generada por el SMC,
f_k	vector de salud de propulsores.

6.2. Espacio de acción

El agente TD3 no genera directamente una acción 6DOF completa. Su espacio de acción, de dimensión 4, se restringe a los siguientes componentes:

$$a_{TD3,k} = [\Delta X_k, \Delta Y_k, \Delta Z_k, \Delta N_k]^T \quad (20)$$

La acción se encuentra acotada por límites simétricos:

$$-a_{\max} \leq a_{TD3,k} \leq a_{\max}, \quad (21)$$

6.3. Función de recompensa

Se emplea una recompensa instantánea de la forma:

$$r_k = -e_{pos,k}^T Q_p e_{pos,k} - q_\psi e_{\psi,k}^2 - q_{\phi\theta} (\phi_k^2 + \theta_k^2) - v_k^T Q_v v_k - \Delta\tau_{TD3,k}^T R_\Delta \Delta\tau_{TD3,k} \quad (22)$$

donde $e_{pos,k} = [e_{x,k}, e_{y,k}, e_{z,k}]^T$.

De esta forma, el término $e_{pos,k}^T Q_p e_{pos,k}$ penaliza el error de posición, $q_\psi e_{\psi,k}^2$ penaliza el error de yaw, $q_{\phi\theta} (\phi_k^2 + \theta_k^2)$ penaliza las desviaciones en roll y pitch, $v_k^T Q_v v_k$ penaliza las velocidades elevadas y $\Delta\tau_{TD3,k}^T R_\Delta \Delta\tau_{TD3,k}$ penaliza el esfuerzo de control realizado por el agente TD3.

Los valores utilizados en la función de recompensa fueron $Q_p = \text{diag}(10, 10, 12)$, $q_\psi = 12$, $q_{\phi\theta} = 25$, $Q_v = \text{diag}(0,5, 0,5, 0,7, 0,2, 0,2, 0,3)$ y $R_\Delta = \text{diag}(0,05, 0,05, 0,06, 0,08, 0,08, 0,04)$. El agente TD3 recibe una observación de dimensión 26 y genera una acción de dimensión 4, definida como $[\Delta X, \Delta Y, \Delta Z, \Delta N]^T$, con límites $[15, 15, 20, 8]^T$. Esta acción se expande posteriormente a $[\Delta X, \Delta Y, \Delta Z, 0, 0, \Delta N]^T$, evitando que el agente actúe directamente sobre los momentos de roll y pitch. El actor se implementó como un perceptrón multicapa totalmente conectado, con arquitectura 26-256-256-128-4, activaciones ReLU en las capas ocultas, salida tanh y una capa final de escalado para ajustar la acción a sus límites físicos. Los críticos se implementaron como dos redes independientes de tipo $Q(o, a)$, con ramas separadas para observación y acción, posteriormente combinadas para estimar el valor acción-estado, siguiendo la estructura habitual del algoritmo TD3.

7. Resultados

Del conjunto de simulaciones llevadas a cabo se puede decir que los resultados muestran que la arquitectura SMC+TD3 cumple el objetivo principal presentado al inicio de este trabajo, esto es, reducir de forma significativa los transitorios de yaw bajo fallos horizontales sin excitar los movimientos de roll y pitch. Por otro lado, la principal limitación está en el error posicional de la maniobra combinada, donde la mejora en yaw se obtiene a costa de un mayor error acumulado de posición.

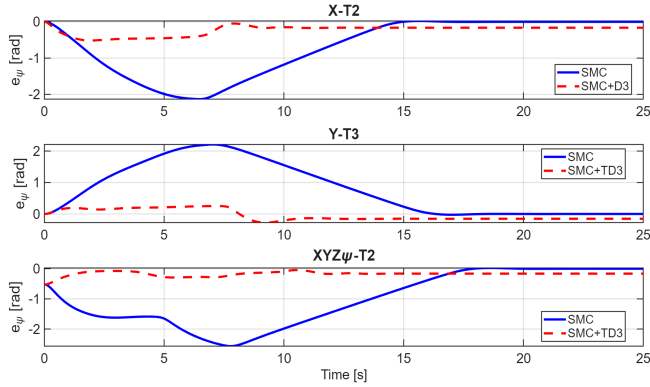


Figura 2: Error de Yaw en fallos horizontales

En la fig.2 se puede observar como la adición de la actuación del TD3 reduce el transitorio de yaw provocado por fallos horizontales de propulsores. En esta figura se muestra la evolución temporal del error de yaw para tres situaciones de fallo horizontal. En todos los casos, cuando el controlador SMC es el único presente, se produce un transitorio de yaw significativo debido a la pérdida de simetría en la actuación. La arquitectura SMC+TD3 reduce de forma clara dicho transitorio, especialmente en los casos X-T2 (Valor de referencia deseada $[1, 0, 0, 0, 0, 0]^T$ y fallo total en T2) e Y-T3 (Valor de referencia deseada $[0, 1, 0, 0, 0, 0]^T$ y fallo total en T3). Por otro lado, y de una forma más cuantitativa, señalar que el error absoluto integral de yaw se reduce un 74,14 % en X-T2, un 98,74 % en Y-T3 y un 72,32 % en la maniobra combinada XYZψ-T2 (Valor de referencia deseada $[1, 1, 0, 5, 0, 0, \pi/6]^T$ y fallo total en T2). Estos resultados indican que la actuación adicional aportada por el TD3 es capaz de compensar en gran medida la degradación de yaw inducida por fallos horizontales de propulsores.

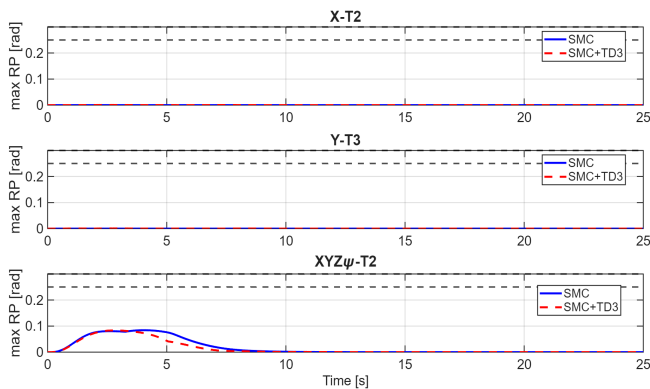


Figura 3: Roll y pitch en los fallos horizontales

La fig.3 se puede ver la evolución en el tiempo de la desviación máxima de roll y de pitch para los mismos fallos horizontales que los indicados para la figura anterior, y se puede ver que la mejora que se experimenta en yaw, no es a costa de producir un efecto significativo en los movimientos de roll y pitch. Así, esta figura muestra que la reducción significativa del error de yaw no introduce transitorios relevantes ni en roll ni en pitch. Esto se debe a que la acción residual del agente se restringe a los canales $[\Delta X, \Delta Y, \Delta Z, \Delta N]$, imponiendo $\Delta K = \Delta M = 0$. En los casos X-T2 e Y-T3, la desviación de actitud permanece prácticamente nula, mientras que en la maniobra combinada se mantiene por debajo de la obtenida con el SMC nominal. Este resultado confirma que la compensación adicional del TD3 propuesta mejora yaw sin comprometer la estabilidad de actitud.

8. Agradecimientos

Este trabajo fue parcialmente financiado por el proyecto Control Inteligente y Colaborativo de Vehículos Submarinos No Tripulados para el Posicionamiento Dinámico de Estructuras Marinas Flotantes A Escala (Ayuda Financiada Contrato Programa del Gobierno de Cantabria -UC).

Referencias

- Benzon, M. V., Sørensen, F. F., Uth, E. T., Jouffroy, J., Liniger, J. y Pedersen, S. (2022). An open-source benchmark simulator: Control of a BlueROV2 underwater robot. *Journal of Marine Science and Engineering*, vol. 10, no. 12, 1898.
- Edwards, C. y Spurgeon, S. K. (1998). *Sliding mode control: Theory and applications*. Taylor & Francis.
- Fossen, T. I. (2021). *Handbook of marine craft hydrodynamics and motion control* (2nd ed.). John Wiley & Sons.
- Goodfellow, I., Bengio, Y. y Courville, A. (2016). *Deep learning*. MIT Press.
- Hosseinnajad, A. y Loueipour, M. (2023). Fault tolerant control system for an ROV based on a novel integral sliding mode control and a state and fault observer in the presence of thruster limitations. *Ocean Engineering*, vol. 280, 114687. <https://doi.org/10.1016/j.oceaneng.2023.114687>
- Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. y Wierstra, D. (2016). Continuous control with deep reinforcement learning. *International Conference on Learning Representations*.
- Liu, M., Yu, S. y Xu, H. (2025). Fault-Tolerant Control of ROV Based on HJI Integral Sliding Mode. En L. Liu, Y. Niu, W. Fu, y Y. Qu (Eds.), *Proceedings of 4th ICAUS 2024 International Conference on Autonomous Unmanned Systems (4th ICAUS 2024)*, Springer Nature Singapore, pp. 232–241.
- Sainz, J. J., Revestido, E., Llata, J. R., & Velasco, F. J. (2023). Auto-tuning of Non-Linear Model Predictive Controllers for a Remote Operated Vehicle. *OCEANS 2023 - Limerick*. <https://doi.org/10.1109/OCEANSLimerick52467.2023.10244375>
- Sainz, J. J., Becerra, V., Revestido, E., Llata, J. R., Alonso, L., & Torre, C. (2024). Adaptive ROV controller for monitoring marine structures. *Jornadas de Automática*, 45. <https://doi.org/10.17979/ja-cea.2024.45.10807>
- Sutton, R. S. y Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Utkin, V. I. (1992). *Sliding Modes in Control and Optimization*. Springer-Verlag, Berlin.
- Yin, Z., Gan, W., Ou, J. y Wang, Z. (2024). Fault-Tolerant Control of Remotely Operated Vehicle Thrusters Based on the Dung Beetle Optimizer. *2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE)*, pp. 1186–1189. <https://doi.org/10.1109/ICAACE61206.2024.10548829>