# Interpretable Physics-Informed Hydrodynamic Derivative Identification for Autonomous Marine Vehicles Using Virtual Sea Trials and Symbolic Residuals


1st Paria Rezayan
*School of Eng. & Built Env.*
*Sheffield Hallam University*
Sheffield, UK
P.Rezayan@shu.ac.uk

2nd Yogang Singh
*School of Eng. & Built Env.*
*Sheffield Hallam University*
Sheffield, UK
Y.Singh@shu.ac.uk

3rd Konstantinos Domdouzis
*School of Comp. & Digital Tech.*
*Sheffield Hallam University*
Sheffield, UK
k.domdouzis@shu.ac.uk

4th Michele Martelli
*Dept. of Naval Arch.,*
*Electrical, Electronic & Telecomm. Eng.*
*University of Genoa*
Genoa, Italy
michele.martelli@unige.it

5th Fuat Kara
*School of Eng. & Built Env.*
*Sheffield Hallam University*
Sheffield, UK
Fuat.Kara@shu.ac.uk



***Abstract*—Accurate manoeuvring models are a prerequisite for autonomy in marine vehicles such as unmanned surface vehicles (USVs), enabling predictive planning, control, and navigation. However, identifying hydrodynamic derivatives populating these models remains expensive and time-consuming when relying on towing-tank facilities or full-scale sea trials, while purely data-driven approaches often suffer from poor interpretability and unstable long-horizon predictions. This paper presents a hybrid framework that combines standards-guided virtual sea trials in the Marine Robotics Unity Simulator (MARUS), a physics-informed neural network (PINN) with a Manoeuvring Modelling Group (MMG)-style core, and symbolic residual discovery with large language model (LLM)-assisted physical interpretation. The dataset used in this study is derived from turning-circle and zig-zag manoeuvres conducted within the high-fidelity simulation environment of MARUS, in accordance with International Towing Tank Conference (ITTC) and International Maritime Organization (IMO) guidelines. It consists of corresponding time series of body-fixed velocities, heading, and command–actuation signals from both manoeuvre types. A grey-box PINN estimates hydrodynamic-derivative-like parameters while learning a residual for unmodelled dynamics. Symbolic regression compresses this learned residual into compact closed-form terms, which are then interpreted through an LLM-assisted physics-audit layer and re-injected into the 3-DOF surge–sway–yaw dynamics using bounded adaptive gains to stabilise long-horizon rollouts. Experimental results show that symbolic residual re-injection notably improves recursive prediction performance. On an external held-out 15°/15° zig-zag test set, it reduces yaw-rate rollout error by 61.0% and heading-rollout error by 91.8% relative to the grey-box baseline, while a physics-only MMG core fails to reproduce the repeated yaw-rate reversals. In addition, first-order hydrodynamic derivatives identified in this study were compared with corresponding Smitt empirical estimates, yielding a sign-independent dimensionless mean absolute error (MAE) of 1.46 across $Y_v$, $Y_r$, $N_v$, and $N_r$. Overall, the resulting model is physically interpretable, exhibits improved rollout stability, and is designed to ingest the same signals as real USV logs, enabling transfer to sea-trial validation.***



***Index Terms*—Unmanned surface vehicles; hydrodynamic derivatives; system identification; physics-informed neural networks; symbolic regression; digital twin; virtual sea trials; MARUS**


## I. Introduction

USVs operate in disturbance-rich marine environments, where the manoeuvring model governing vessel navigation is a core component of the safety-critical guidance-navigation-control (GNC) system. Such models mathematically represent the hydrodynamic forces and moments acting on the hull as functions of time-varying vessel states and control inputs, and their predictive accuracy depends on the hydrodynamic derivatives embedded in the manoeuvring equations, whose order determines the number and coupling of derivatives required [1]. Low-order 3-DOF surge–sway–yaw models remain attractive for this objective because they contain the derivatives most directly involved in path following, trajectory tracking, and collision-avoidance planning. Within this family, MMG-style models are especially useful because they decompose manoeuvring forces into hull, propulsion, and rudder contributions, making them a physically interpretable basis for hydrodynamic-derivative estimation and manoeuvring prediction [2].

Obtaining these derivatives remains constrained by practical challenges. Conventional approaches derive them from captive model tests, free-running tests, full-scale sea trials, or CFD-based virtual captive tests. For free-running and sea-trial approaches, standard manoeuvring assessment includes straight-line, stopping, turning-circle (TC), and zig-zag (ZZ) tests, with TC and ZZ being particularly useful as IMO/ITTC-recommended assessments of turning ability, yaw response, and directional stability [3]–[5]. Nevertheless, the aforementioned derivative estimation methods are often costly, time-consuming, and highly vulnerable to modelling and measurement errors. Consequently, this work employs MARUS-generated IMO/ITTC-compliant TC/ZZ virtual sea-trial data, as simulation enables rapid and reproducible generation of high-fidelity manoeuvring datasets under controlled, risk-free conditions in a cost-effective way [6].

Once such manoeuvre records are available, the remaining difficulty is converting them into a model that is both predic-

tively accurate and physically interpretable. To address this, classical system identification (SI) reduces reliance on exhaustive physical experiments and manual derivative extraction by estimating model parameters directly from available manoeuvring data. Yet it remains sensitive to noise, excitation richness, and model-structure choices. Modern machine learning (ML) approaches have recently gained traction as an alternative due to their impressive predictive accuracy, but their purely black-box nature lacks the physical interpretability needed for safety-critical maritime operations. Physics-informed learning is therefore the natural next step in the accuracy–interpretability trade-off: a grey-box physics-informed neural network (PINN) can embed an MMG-style physical core while using a neural residual to capture nonlinear damping, actuator-dependent effects, and cross-coupled dynamics [7], [8].

Nonetheless, even physically constrained neural residuals can remain difficult to interpret. Symbolic regression (SR) offers a route to convert learned residual corrections into sparse closed-form expressions. Yet SR alone does not fully resolve the interpretability problem. Classical SR search procedures can become slow and unstable as the candidate equation space grows, particularly for hydrodynamic residuals arising from complex and highly nonlinear manoeuvring. Recent work has therefore explored the use of large language models (LLMs) to support symbolic discovery [9], [10]. In this work, however, the LLM is used only as a physics-aware interpretation layer, assigning hydrodynamically meaningful labels to the SR-discovered residual terms to facilitate human-in-the-loop (HIL) physical interpretation.

Therefore, this paper addresses the absence of an end-to-end framework connecting standards-aligned virtual sea trials, grey-box identification, symbolic residual discovery, physics-aware interpretation, bounded re-injection, and rollout validation. The proposed MARUS–PINN–SR+LLM workflow trains an MMG-based PINN on smoothed acceleration labels, compresses learned residuals using PySR, interprets accepted symbolic terms through an LLM-assisted audit layer, and re-injects them into the grey-box manoeuvring model through bounded adaptive gains. Although both TC and ZZ manoeuvres are used for training, evaluation uses external ZZ15°/15° yaw-rate and heading rollouts because ZZ manoeuvres directly test yaw reversals and heading drift, moving hydrodynamic-derivative estimation from local fitting towards stable, interpretable, and externally validated recursive manoeuvring prediction.

## II. Related Work

A recent review argues that accurate manoeuvring SI depends not only on selecting an appropriate manoeuvring model, but also on trial-data quality, environmental disturbances, parameter uncertainty, and coefficient identifiability [5]. In this context, classical Nomoto, Abkowitz, and MMG models remain valuable because they provide physically interpretable equations whose derivatives can be related to hydrodynamic damping, cross-coupling, and control effects. Nevertheless, reduced-order models may be too simple for trajectory-level prediction, while higher-order formulations can suffer from over-parameterisation and multicollinearity. Recent free-running SI workflows therefore focus on improving state reconstruction and coefficient estimation, for example through EKF/RTS smoothing or bounded inverse-dynamics regression [11], [12]. Nevertheless, these methods remain sensitive to preprocessing, excitation richness, and model-structure assumptions, so physically plausible derivatives may still accumulate prediction errors during recursive long-horizon rollout.

While classical SI seeks physically meaningful hydrodynamic derivatives, data-driven manoeuvring prediction shifts attention to learning nonlinear vessel motion directly from observations. Gaussian-process, recurrent-network, and hybrid feedforward neural network approaches have shown that data-driven corrections can improve accuracy, but their residual terms often remain difficult to audit, trust, and certify under sparse, noisy, or out-of-distribution manoeuvre data [13]–[15]. Hence, PINNs provide a physically constrained learning framework by simultaneously minimising data mismatch and enforcing hydrodynamic governing equations during training [7]. In marine applications, MMG-based PINNs and physics-informed regression methods have improved identification and generalisation. Yet low one-step prediction error does not necessarily imply accurate long-horizon motion prediction, so rollout-level validation remains essential [8], [16], [17]. Fig. 1 summarises this method landscape in terms of interpretability, extrapolation capability, and long-horizon rollout stability, motivating the grey-box approach adopted in this work.

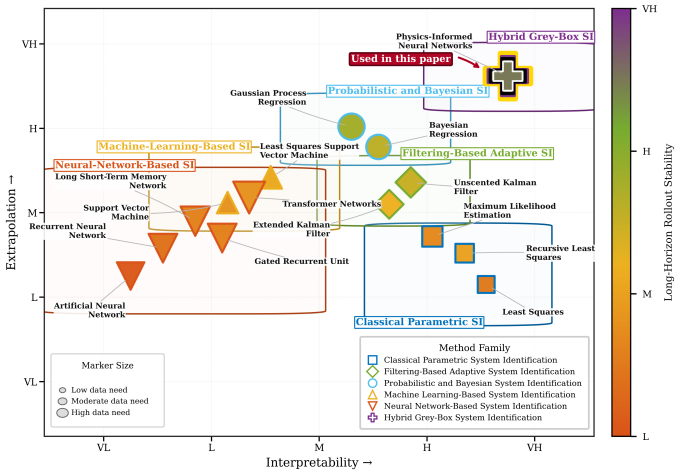



Fig. 1: Hydrodynamic SI method landscape, highlighting the trade-off between interpretability, predictive flexibility, and long-horizon rollout stability across classical, data-driven, and grey-box approaches [5], [8], [11], [13], [14], [18]–[23].

Although PINNs improve physical consistency, their learned residuals remain difficult to interpret; sparse equation discovery addresses this auditability gap by recovering explicit equations from data. Brunton et al. introduced SINDy for recovering compact governing equations from data, while recent USV studies show that physics-informed sparse libraries can improve validation behaviour over generic polynomial libraries [9], [24]. Symbolic regression has also been coupled with PINNs to translate learned unknown motion dynamics into closed-form terms, although the result still depends on the quality

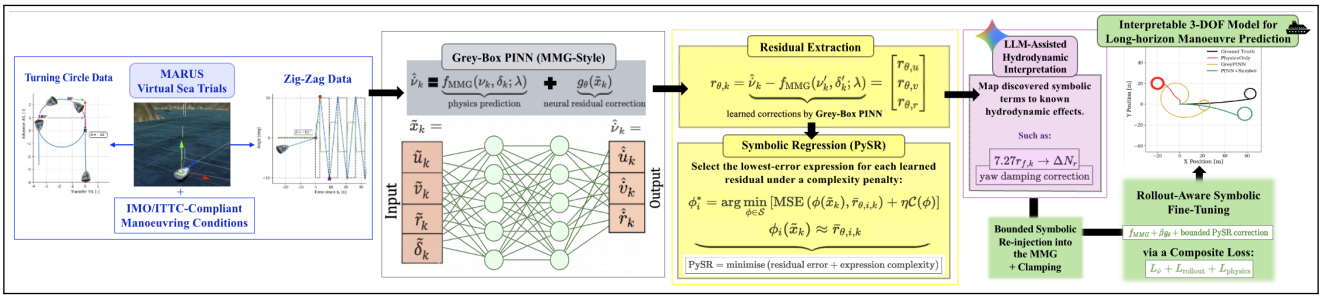



Fig. 2: Proposed MARUS–PINN–SR+LLM workflow for interpretable and stable long-horizon manoeuvring prediction.

of the learned residual and available data [25]. Recent LLM-assisted SR methods further use language models to inject scientific priors or guide equation-program search [10]. In contrast, the proposed framework does not use the LLM for equation generation or search; it is applied only after PySR selection as a bounded semantic audit layer before symbolic re-injection into the 3-DOF manoeuvring model.

## III. Methodology

This section details the implementation of the MARUS–PINN–SR+LLM workflow shown in Fig. 2.

### A. MARUS Virtual Sea Trials

The input data are obtained from the MARUS virtual sea-trial framework in [6]. The dataset consists of TC and ZZ manoeuvres because these trials adequately excite the surge, sway, and yaw dynamics needed for low-order hydrodynamic derivative identification. Each logged sample contains body-fixed velocities, heading, ordered and executed actuation signals, and the steering input. The SI-ready input vector is therefore written as $x_k = [u_k,\, v_k,\, r_k,\, \delta_k]^T$, where $u$, $v$, and $r$ denote surge velocity, sway velocity, and yaw rate, respectively, and $\delta$ is the executed steering input in radians.

Because acceleration labels obtained from direct finite differencing are highly sensitive to signal noise, the measured velocity channels are smoothed independently for each manoeuvre file using a Rauch–Tung–Striebel (RTS) smoother with $Q_{\mathrm{RTS}} = 10^{-3}$ and $R_{\mathrm{RTS}} = 10^{-2}$. For each velocity signal, the smoother estimates the latent velocity–acceleration state, from which $\dot{u}$, $\dot{v}$, and $\dot{r}$ are extracted. Inputs are then normalised using training-set mean and standard deviation only,

$$\tilde{x}_{k,j} = \frac{x_{k,j} - \mu_j}{\sigma_j}, \qquad j \in \{u, v, r, \delta\}, \tag{1}$$

while acceleration targets remain in physical units so that the learned model predicts physically consistent surge, sway, and yaw accelerations. Train, validation, and test splits are formed at file level to prevent any leakage.

### B. Grey-Box Physics-Informed Neural Network

The vessel is modelled as a rigid body in the horizontal plane using $\nu = [u, v, r]^T$, whose underlying surge–sway–yaw dynamics can be expanded from these compact equations:

$$m(\dot{u} - vr) = X, \qquad m(\dot{v} + ur) = Y, \qquad I_z\dot{r} = N, \tag{2}$$

where $m$ is the vessel mass, $I_z$ is the yaw inertia, and $X$, $Y$, and $N$ are the surge force, sway force, and yaw moment. Following the modular approach of MMG modelling, these hydrodynamic forces can be expressed as physically interpretable contributions:

$$X = X_H + X_P + X_R, \qquad Y = Y_H + Y_R, \qquad N = N_H + N_R, \tag{3}$$

where the subscripts $H$, $P$, and $R$ denote the hull, propulsion, and rudder components, respectively.

Rather than implementing a full standard MMG model, this work uses the MMG decomposition as a physically interpretable basis for constructing a compact acceleration core, denoted by $f_{\mathrm{MMG}}(\nu, \delta; \lambda)$, to reduce parameter complexity while retaining the main surge–sway–yaw coupling structure [2], [26]. The parameter vector $\lambda$ contains 23 trainable physical-core parameters that are interpreted as hydrodynamic-derivative-like terms. These include added-mass contributions $(X_{\dot{u}}, Y_{\dot{v}}, N_{\dot{r}}, Y_{\dot{r}})$, linear damping and sway–yaw coupling terms $(X_u, Y_v, Y_r, N_v, N_r)$, nonlinear damping terms $(X_{uu}, Y_{vv}, Y_{rr}, N_{vv}, N_{rr})$, and compact propulsion and rudder coefficients $(K_{T0}, K_{T1}, k_\delta, t_R, a_X, a_Y, a_N)$.

The resulting MMG core provides the interpretable part of the acceleration model, but it cannot capture all residual nonlinearities, cross-coupled effects, and actuator-dependent dynamics in the logged data. We address this issue using a grey-box PINN that augments the MMG physics core with a neural residual:

$$\hat{\dot{\nu}}_k = f_{\mathrm{MMG}}(\nu_k, \delta_k; \lambda) + g_\theta(\tilde{x}_k), \tag{4}$$

where the residual network $g_\theta$ is a multilayer perceptron with architecture $4 \to 128 \to 128 \to 3$, mapping the normalised state–actuation vector to surge, sway, and yaw acceleration corrections.

Grey-box PINN training is performed in two stages. The first stage fits the model to the RTS-smoothed acceleration labels using the local acceleration loss $\mathcal{L}_{\dot{\nu}} = \mathrm{MSE}(\hat{\dot{\nu}}, \dot{\nu})$. However, as discussed earlier, low one-step acceleration error does not necessarily guarantee stable recursive prediction. Therefore, the second stage introduces rollout stabilisation using ordered sequence segments sampled from individual manoeuvre files. Starting from the measured initial normalised state, predicted physical accelerations are converted into normalised increments and recursively integrated as

$$\tilde{\nu}^{\mathrm{pred}}_{h+1} = \tilde{\nu}^{\mathrm{pred}}_{h} + \Delta t \begin{bmatrix} \hat{\dot{u}}_h/\sigma_u \\ \hat{\dot{v}}_h/\sigma_v \\ \hat{\dot{r}}_h/\sigma_r \end{bmatrix}. \tag{5}$$

The resulting rollout sequence is compared with the measured normalised sequence through a rollout loss: $\mathcal{L}_{\mathrm{roll}} = \mathrm{MSE}_u + \mathrm{MSE}_v + \mathrm{MSE}_r$. The final grey-box PINN loss function

is therefore $\mathcal{L}_{\text{total}} = \mathcal{L}_{\dot{\nu}} + \mathcal{L}_{\text{roll}} + \mathcal{L}_{\text{phys},r}$, where the last term denotes an additional yaw-acceleration penalty introduced to prevent the neural residual from fully absorbing yaw dynamics that should remain represented by the MMG physical core.

### C. Symbolic Regression and LLM-Assisted Interpretation

To convert the grey-box PINN residual from an implicit neural correction into an explicit analytical form, PySR-based symbolic regression is applied to the isolated neural residual:

$$\rho_{\theta,k} = \hat{\dot{\nu}}_k - f_{\text{MMG}}(\nu_k, \delta_k; \lambda), \tag{6}$$

Separate residual datasets are constructed for the surge, sway, and yaw channels. For each channel, PySR receives $(\tilde{u}, \tilde{v}, \tilde{r}, \tilde{\delta})$ as inputs and the corresponding learned residual component as the target, giving the channel-wise symbolic residual model

$$\epsilon_i = s_i(\tilde{x}_k) \approx \rho_{\theta,i,k}, \qquad i \in \{u, v, r\}. \tag{7}$$

LLM-assisted hydrodynamic interpretation is then introduced as an auditability layer for the PySR-selected symbolic expressions. First, all 3 discovered expressions are transformed from normalised to physical variables using the stored training-set normalisation constants and residual scaling parameters. The expressions, variable definitions, operating conditions, and vessel context are passed to a physics-aware LLM. The LLM identifies plausible hydrodynamic mechanisms such as damping, cross-coupling, actuation-effectiveness, cross-flow, or nonlinear correction terms. The LLM then outputs structured interpretations containing effect names, descriptions, confidence scores for human review, and concise HIL-ready physical explanations, making the symbolic residuals inspectable and traceable to familiar manoeuvring terms, without influencing term acceptance.

### D. Bounded Symbolic Re-Injection and Rollout-Aware Fine-Tuning

The PySR-selected and LLM-audited symbolic expressions are compiled into bounded numerical modules and re-injected into the grey-box model as bounded residual corrections:

$$\hat{\dot{\nu}}_{\text{sym},k} = f_{\text{MMG}}(\nu_k, \delta_k; \lambda) + \beta g_\theta(\tilde{x}_k) + K_s \odot \bar{\epsilon}_k, \tag{8}$$

where $\bar{\epsilon}_k = \text{clip}(\epsilon(\tilde{x}_k), -c, c)$ is the bounded symbolic residual, $\beta$ scales the retained neural residual, $K_s = [k_u, k_v, k_r]^T$ contains the learned symbolic gains, $c = [c_u, c_v, c_r]^T$ defines the channel-wise saturation limits, and $\odot$ denotes element-wise multiplication. This bounded formulation allows symbolic terms to contribute interpretable corrections while preventing them from dominating the grey-box dynamics. The gain and residual-scale values redistribute bounded residual capacity between neural and symbolic branches rather than acting as fixed physical coefficients; their learned effective values remain below the prescribed caps. Symbolic fine-tuning uses acceleration loss, rollout loss, and mild physical-core regularisation, and is run for 500 epochs with gradient clipping at norm 0.5.

## IV. Experiments and Results

This section evaluates whether the proposed framework improves the quantity most relevant to long-horizon USV autonomy: recursive yaw-rate and heading rollout. Three configurations are benchmarked: PhysicsOnly, which uses only the compact MMG-inspired core; GreyPINN, which augments the MMG physical core with a neural residual; and PINN+Symbol, the final proposed model, which further incorporates bounded symbolic residual re-injection. Long-horizon performance is assessed through free recursive rollouts over active manoeuvre windows. The data are sampled at $\Delta t = 0.1\,\text{s}$ (10 Hz), with 20-step GreyPINN and 25-step symbolic rollout horizons, corresponding to 2.0 and 2.5 seconds. The reported errors are file-wise RMSEs, $Roll_r = \sqrt{K^{-1}\sum_{k=1}^{K}(\hat{r}_k - r_k)^2}$ and $Roll_\psi = \sqrt{K^{-1}\sum_{k=1}^{K}(\hat{\psi}_k - \psi_k)^2}$, where $\hat{\psi}_k = \sum_{\tau=1}^{k}\hat{r}_\tau \Delta t$. Results are aggregated over three seeds using mean values and descriptive 2.5–97.5 percentile intervals, with the external ZZ15°/15° results reported in Table I.

### A. Rollout Prediction on External Manoeuvres

TABLE I: Rollout-only prediction errors on the external ZZ15°/15° set. Lower is better; brackets denote seed ranges.

| Model | $\text{Roll}_\text{r}$ | $\text{Roll}_\psi$ |
|---|---|---|
| PhysicsOnly | 0.0731 [0.0731, 0.0731] | 0.1524 [0.1524, 0.1524] |
| GreyPINN | 0.0302 [0.0244, 0.0407] | 0.2961 [0.2085, 0.3783] |
| PINN+Symbol | **0.0118 [0.0117, 0.0119]** | **0.0243 [0.0238, 0.0248]** |

The out-of-distribution test results in Table I evaluate whether the learned model transfers beyond its training domain. The PhysicsOnly model remains physically interpretable, but its recursive prediction is unreliably weak, and the corresponding trajectory in Fig. 3 is a flat line resembling a straight steady-state run. This suggests that the compact MMG core lacks sufficient dynamic richness to reproduce the alternating rudder response and observed manoeuvre dynamics of repeated zig-zag manoeuvres.

GreyPINN improves the yaw-rate rollout error to $Roll_r = 0.0302$, showing that the neural residuals do capture part of the yaw response. However, its heading rollout error increases to $Roll_\psi = 0.2961$, which is worse than PhysicsOnly. This is an important negative result: better short-horizon yaw-rate tracking does not necessarily imply better integrated heading prediction, because small phase and bias errors can accumulate under recursive integration. Ultimately, PINN+Symbol achieves the best rollout performance, reducing $Roll_r$ to $0.0118$ and $Roll_\psi$ to $0.0243$, corresponding to improvements of $61.0\%$ and $91.8\%$, respectively, relative to GreyPINN. Together, these findings show that symbolic re-injection preserves the interpretability of the MMG-based grey-box structure while improving recursive rollout accuracy and reliability. Figure 3 supports the same conclusion visually.

### B. Identified Local Hydrodynamic Derivatives

Table II reports representative local linearised hydrodynamic derivatives identified at $(u_0, v_0, r_0) = (0.786, -0.008, 0.018)$ together with empirical Smitt estimates. The empirical values predict negative $Y_v$, $N_v$, and $N_r$. GreyPINN improves yaw-rate tracking but yields positive $N_v$ and $N_r$, indicating locally destabilising sway–yaw behaviour and explaining its heading drift. PINN+Symbol recovers the empirical signs of $Y_v$, $N_v$, and $N_r$, with only $Y_r$ differing from the estimates due to Unity

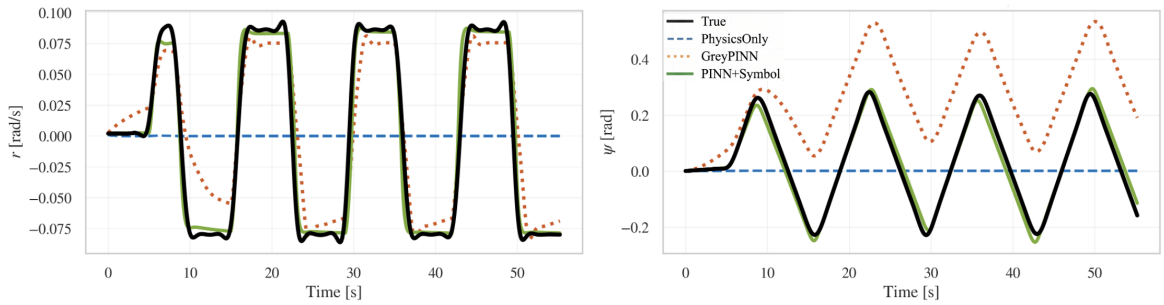



Fig. 3: External ZZ15°/15° rollout comparison. PhysicsOnly produces a flat response, GreyPINN captures part of the yaw-rate behaviour but accumulates heading drift, and PINN+Symbol maintains the lowest yaw-rate and heading rollout errors.

TABLE II: Hydrodynamic derivatives identified by each model and corresponding Smitt estimates (sign-convention dependent).

| **Model** | $\mathbf{X_u}$ | $\mathbf{Y_v}$ | $\mathbf{Y_r}$ | $\mathbf{N_v}$ | $\mathbf{N_r}$ |
|---|---|---|---|---|---|
| PhysicsOnly | $-0.121$ | $-1.844$ | $-1.880$ | $-1.482$ | $-7.641$ |
| GreyPINN | $-0.322$ | $-3.548$ | $-0.802$ | $1.647$ | $1.176$ |
| PINN+Symbol | $-0.219$ | $-2.827$ | $-0.886$ | $-1.227$ | $-1.566$ |
| Smitt (Emp.) | – | $-0.338$ | $0.217$ | $-0.412$ | $-0.138$ |

and nautical sign discrepancies, while also achieving the lowest Smitt-comparison MAE of 1.46 and the best rollout accuracy.

### *C. Symbolic Residual Compression and Interpretation*

PySR-selected channel-wise expressions that best balance complexity and accuracy are reported in Table III. The discovered expressions remain compact while preserving the dominant structure of the PINN residuals. In particular, the yaw residual reduces to an approximately linear function of the normalised yaw rate, which is consistent with the yaw/heading-dominated rollout instability observed in the external ZZ15°/15° test.

TABLE III: Best symbolic expressions identified by PySR for the learned PINN residuals.

| **Channel** | **Best PySR expression** | **RMSE** |
|---|---|---|
| Surge | $\epsilon_u \approx 1.16\tilde{u}/[\text{sign}(\lvert\tilde{r}\rvert + \lvert\tilde{\delta} - 0.309\rvert + \text{sign}(\tilde{u})) + 0.703] - 0.037$ | $7.07\times10^{-3}$ |
| Sway | $\epsilon_v \approx 0.160\tilde{u} - 0.160\tilde{v} + 0.928\tilde{r}$ | $4.38\times10^{-3}$ |
| Yaw | $\epsilon_r \approx \tilde{r}$ | $1.68\times10^{-2}$ |

The yaw symbolic model is especially effective at reproducing the behaviour of the learned neural residual. As shown in Fig. 4, the PySR-predicted yaw residual aligns strongly with the PINN residual, giving $R^2 = 0.998$ and RMSE $= 1.68 \times 10^{-2}$. This indicates that the main unmodelled yaw contribution learned by the neural branch can be represented by a sparse analytical correction rather than a complex black-box function. Interpreting the discovered relation in physical variables yields $\epsilon_r \approx 7.27 r_{\text{filtered}} - 0.01485$, corresponding to a linear yaw-rate correction plus a small steady yaw-bias term and suggesting that the compact MMG core mainly required a yaw-damping adjustment rather than a high-order nonlinear yaw residual.

The LLM is used exclusively for semantic interpretation after numerical identification; it neither influences model training

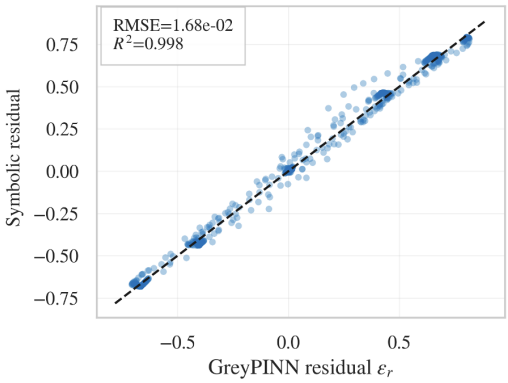



Fig. 4: Yaw-residual symbolic compression, showing that PySR closely reproduces the learned PINN yaw residual and captures the dominant yaw correction.

TABLE IV: LLM-assisted symbolic-yaw residual interpretation.

| **Term** | **Concept** | **Conf.** | **Interpretation** |
|---|---|---|---|
| $7.27\, r_{\text{filtered}}$ | Yaw-rate acceleration compensation $(\Delta N_r/I_z)$ | 0.90 | Suggests excessive yaw damping in the compact MMG core. |
| $-0.01485$ | Steady yaw-acceleration bias $(N_0/I_z)$ | 0.85 | Represents a small systematic yaw-acceleration bias or actuation imbalance. |

nor modifies the discovered equations or physical parameters. The LLM-assisted interpretation identifies and names the hydrodynamic effects represented by the symbolic yaw residual, as summarised in Table IV. It maps the $7.27 r_{\text{filtered}}$ term to a yaw-rate acceleration compensation $(\Delta N_r/I_z)$ with high confidence, interpreting it as a correction to overpredicted yaw damping in the compact MMG core. The negligible offset $-0.01485$ is mapped to a steady yaw-acceleration bias $(N_0/I_z)$, possibly caused by actuator imbalance, alignment bias, or residual asymmetry in the simulated dynamics.

Finally, having shown that symbolic re-injection improves rollout behaviour, a causal ablation isolates the contribution of the symbolic branch. Relative to an otherwise identical PINN+Symbol model with the symbolic correction disabled, symbolic re-injection reduces $Roll_r$ by 33.6% and $Roll_\psi$

by $44.8\%$. The symbolic branch therefore contributes not only compact, review-ready expressions that can be inspected, audited, and certified within HIL frameworks, but also a measurable improvement in long-horizon predictive accuracy.

## V. Conclusion

This paper established a MARUS-driven hydrodynamic derivative identification framework combining grey-box PINN learning, symbolic residual compression, and LLM-assisted interpretation. The results show that bounded symbolic re-injection improves recursive yaw-rate and heading prediction while preserving an interpretable MMG-based structure. This matters because compact symbolic corrections support auditable long-horizon manoeuvring models for autonomy. Future work will validate the framework using physical USV sea-trial data.

## Acknowledgment

This work was supported by Sheffield Hallam University Graduate Teaching Assistant funding. The author has applied a CC BY licence to any accepted manuscript; model code is available upon request.

## References

[1] T. I. Fossen, *Handbook of marine craft hydrodynamics and motion control*, 1st ed. Wiley, Apr. 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119994138

[2] H. Yasukawa and Y. Yoshimura, "Introduction of MMG standard method for ship maneuvering predictions," *Journal of Marine Science and Technology*, vol. 20, no. 1, pp. 37–52, Mar. 2015. [Online]. Available: https://link.springer.com/10.1007/s00773-014-0293-y

[3] International Maritime Organization, *Standards for ship manoeuvrability*, International Maritime Organization Std. Resolution MSC.137(76), Annex 6, 2002, adopted 4 December 2002.

[4] International Towing Tank Conference, *Recommended procedures and guidelines: Free running model tests*, International Towing Tank Conference Std. 7.5-02-06-01, 2008.

[5] H. Xu and C. Guedes Soares, "Review of system identification for manoeuvring modelling of marine surface ships," *Journal of Marine Science and Application*, vol. 24, no. 3, pp. 459–478, Jun. 2025. [Online]. Available: https://link.springer.com/10.1007/s11804-025-00681-w

[6] P. Rezayan, "Traceable virtual sea trials in the Marine Robotics Unity Simulator for manoeuvring assessment of unmanned surface vehicles," 2026, version Number: 2. [Online]. Available: https://arxiv.org/abs/2606.12349

[7] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, Feb. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0021999118307125

[8] G. An, G. Xiang, X. Xiang, and C. Guedes Soares, "Physics informed neural networks based identification modelling of ship maneuvering motion and associated optimal excitation design," *Engineering Applications of Computational Fluid Mechanics*, vol. 19, no. 1, p. 2566860, Dec. 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/19942060.2025.2566860

[9] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, Apr. 2016. [Online]. Available: https://pnas.org/doi/full/10.1073/pnas.1517384113

[10] P. Shojaee, K. Meidani, S. Gupta, A. B. Farimani, and C. K. Reddy, "LLM-SR: Scientific equation discovery via programming with large language models," in *The Thirteenth International Conference on Learning Representations*, 2025. [Online]. Available: https://openreview.net/forum?id=m2nmp8P5in

[11] M. Alexandersson, W. Mao, and J. W. Ringsberg, "System identification of vessel manoeuvring models," *Ocean Engineering*, vol. 266, p. 112940, Dec. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801822022235

[12] G. Chillcce and O. El Moctar, "Data-driven system identification of hydrodynamic maneuvering coefficients from free-running tests," *Physics of Fluids*, vol. 35, no. 5, p. 057122, May 2023. [Online]. Available: https://pubs.aip.org/pof/article/35/5/057122/2890752/Data-driven-system-identification-of-hydrodynamic

[13] Z.-L. Ouyang, G. Chen, and Z.-J. Zou, "Identification modeling of ship maneuvering motion based on local Gaussian process regression," *Ocean Engineering*, vol. 267, p. 113251, Jan. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801822025343

[14] R. E. Nielsen, D. Papageorgiou, L. Nalpantidis, B. T. Jensen, and M. Blanke, "Machine learning enhancement of manoeuvring prediction for ship digital twin using full-scale recordings," *Ocean Engineering*, vol. 257, p. 111579, Aug. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801822009453

[15] K. Hatayama, K. Wakita, Y. Ichinose, and A. Maki, "Proposing a hybrid ship maneuvering model through force/moment-level integration of MMG and FNN with output constraints," *Journal of Marine Science and Technology*, vol. 31, no. 2, pp. 277–297, Jun. 2026. [Online]. Available: https://link.springer.com/10.1007/s00773-026-01115-0

[16] A. Liu, Y. Xue, H. Qin, and Z. Zhu, "Physics-informed identification of marine vehicle dynamics using hydrodynamic dictionary library-inspired adaptive regression," *Ocean Engineering*, vol. 296, p. 117013, Mar. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801824003500

[17] L. A. Aguirre, B. H. Barbosa, and A. P. Braga, "Prediction and simulation errors in parameter estimation for nonlinear systems," *Mechanical Systems and Signal Processing*, vol. 24, no. 8, pp. 2855–2867, Nov. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0888327010001469

[18] G. Zhang, X. Zhang, and H. Pang, "Multi-innovation auto-constructed least squares identification for 4 DOF ship manoeuvring modelling with full-scale trial data," *ISA Transactions*, vol. 58, pp. 186–195, Sep. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0019057815001019

[19] E. Revestido Herrero and F. J. Velasco González, "Two-step identification of non-linear manoeuvring models of marine vessels," *Ocean Engineering*, vol. 53, pp. 72–82, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801812002429

[20] W. Ariza Ramirez, Z. Q. Leong, H. Nguyen, and S. G. Jayasinghe, "Non-parametric dynamic system identification of ships using multi-output Gaussian processes," *Ocean Engineering*, vol. 166, pp. 26–36, Oct. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801818314094

[21] W. Luo, L. Moreira, and C. Guedes Soares, "Manoeuvring simulation of catamaran by using implicit models based on support vector machines," *Ocean Engineering*, vol. 82, pp. 150–159, May 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801814000894

[22] Y. Jiang, X.-R. Hou, X.-G. Wang, Y. Liu, and W. Zhang, "Identification modeling and prediction of ship maneuvering motion based on LSTM deep neural network," *Journal of Marine Science and Technology*, vol. 27, no. 1, pp. 125–137, Mar. 2022. [Online]. Available: https://link.springer.com/article/10.1007/s00773-021-00819-9

[23] L. Dong, H. Wang, and J. Lou, "An attention mechanism model based on positional encoding for the prediction of ship maneuvering motion in real sea state," *Journal of Marine Science and Technology*, vol. 29, no. 1, pp. 136–152, Mar. 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00773-023-00978-x

[24] E. Kandemir, E. M. L. Coates, and A. Hasan, "SINDy-based modeling of an uncrewed surface vessel with experimental validation," *Applied Ocean Research*, vol. 165, p. 104825, Dec. 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0141118725004110

[25] E. Kiyani, K. Shukla, G. E. Karniadakis, and M. Karttunen, "A framework based on symbolic regression coupled with eXtended physics-informed neural networks for gray-box learning of equations of motion from data," *Computer Methods in Applied Mechanics and Engineering*, vol. 415, p. 116258, Oct. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0045782523003821

[26] M. Martelli, M. Viviani, M. Altosole, M. Figari, and S. Vignolo, "Numerical modelling of propulsion, control and ship motions in 6 degrees of freedom," *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, vol. 228, no. 4, pp. 373–397, 2014. [Online]. Available: https://journals.sagepub.com/doi/10.1177/1475090214544181