

Localized Generative Design of Ship Hulls with Diffusion Models and High-Fidelity Assessment

Hadi Keramati^{a,*}, Mohammed A. Hamman^a, Patrick Kirchen^a, Sina Tousi^b, Joonas Riekkinen^b, Jim Pyra^c, Rajeev K. Jaiman^a

^a*Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada*

^b*Elomatic Consulting Inc., Vancouver, BC, Canada*

^c*InnovMarine Inc., Lévis, QC, Canada*

Abstract

The ship design process is highly iterative, requiring repeated cycles of geometric modification, performance evaluation, and expert judgment. Recent advances in generative diffusion models show considerable promise for automating hull-form generation. However, most existing approaches focus on global, conceptual redesign and provide limited support for practical engineering workflows, which require localized refinement of specific hull features and high-fidelity performance evaluations to establish confidence in the generated designs. This paper presents a reward-directed diffusion framework for localized ship hull optimization, enabling targeted modification of specific geometric regions while preserving global hull form continuity. The proposed framework introduces a parametric masking mechanism in dataset generation that constrains the generative process to user-selected subsets of design variables, such as the bulbous bow and forebody, allowing controlled and physically smooth geometric refinement within a full-hull representation. A surrogate-assisted sequential Monte Carlo sampling strategy guides the diffusion process toward hull geometries with reduced predicted calm-water resistance, with stability retained through the training data and displacement matched in a subsequent screening. An automated viscous free-surface CFD pipeline, validated against model-scale measurements for two benchmark container ships, is then applied to a generated variant at the displacement of the KRISO Container Ship (KCS). The total resistance of this variant is comparable to that of KCS at $F_n = 0.195$ but about 45 to 50 percent higher at $F_n = 0.250$ and 0.260 , an increase that the empirical screening did not capture. To examine the parametric representation within which the variants are generated, the pipeline is also applied to GenHull E, a hull refined by a naval architect toward KCS within the same representation. The refined hull has six to eight percent lower total resistance than the benchmark up to $F_n = 0.240$ and higher total resistance above $F_n = 0.250$, with the crossover occurring where a friction advantage of 8.07 N and a pressure penalty of 8.49 N nearly cancel. The two components are measured separately and have opposite speed dependence, so that an objective aggregated across the crossover can obscure the opposing performance trends. The evaluation also shows that the representation cannot express a faired bulbous bow, and it indicates limitations of the reward formulation, which contains neither viscous pressure nor a stability term. These findings demonstrate how high-fidelity evaluation can guide improvements to design representations and objectives, helping ensure that predicted gains are physically realizable.

*Corresponding author: hadi.keramati@ubc.ca

Email address: hadi.keramati@ubc.ca (Hadi Keramati)

1. Introduction

The optimization of ship design involves a multidisciplinary and tightly coupled workflow integrating geometric modeling, hydrodynamic analysis, and numerical optimization within broader naval-architectural requirements [1]. For hull-form development, a central objective is to reduce resistance while retaining the displacement, carrying capacity, and operational characteristics required of the vessel. Lower resistance reduces the effective power required for propulsion, but resistance reduction must be considered alongside hydrostatic, geometric, and operational requirements rather than pursued as an isolated objective. The resulting optimization problem therefore combines nonlinear performance objectives with constraints on vessel particulars, hydrostatic properties, geometric feasibility, and allowable design variations [2]. In practice, this search frequently begins from an established parent hull rather than an unrestricted design space [3], with selected features modified while the remainder of the vessel is retained. Han et al. [4], for example, optimized the forebody while preserving the aftbody geometry. Such localized optimization requires a parameterization sufficiently flexible to represent useful modifications while controlling their extent and maintaining smooth transitions to retained surfaces. Expert judgment therefore remains important in defining the design envelope and interpreting trade-offs, while automation supports systematic exploration within this engineering-defined space.

The bulbous bow and adjacent forebody are particularly relevant to localized resistance optimization because their geometry influences the bow-wave system and immersed-volume distribution. Their hydrodynamic benefit is strongly dependent on operating condition: a modification advantageous at one speed or draft may provide less benefit, or become detrimental, at another. Lu et al. [5] demonstrated this dependence through bulbous-bow optimization under multiple operating conditions. Candidate designs should therefore be assessed across the relevant operating envelope rather than at a single design point, while accounting for accompanying changes in displacement, draft, and other characteristics relative to the parent hull. Evaluating these alternatives requires balancing prediction fidelity against computational cost. Semi-empirical methods, such as the Holtrop-Mennen resistance and power prediction method, provide rapid estimates based on relationships calibrated from existing ship data [6], but may not resolve the hydrodynamic effects of localized geometric changes. Viscous free-surface CFD provides a higher-fidelity assessment of geometry-dependent wave, pressure, and viscous effects, but its repeated use in many candidates and operating conditions is computationally expensive. This motivates workflows that use inexpensive methods for design-space exploration while reserving high-fidelity CFD for selected candidates [2].

Surrogate-assisted optimization provides an established means of reducing this expense. By approximating the relationship between design variables, operating conditions, and hydrodynamic responses, a surrogate can support the screening of many candidates at a fraction of the cost of repeated CFD evaluations. Coppedè et al. [7] combined high-fidelity CFD with Gaussian-process response surfaces for KCS hull optimization and reassessed selected designs using refined CFD calculations. This separation between inexpensive exploration and subsequent numerical verification is important because a favorable surrogate prediction does not, by itself, establish a hydrodynamic improvement. The quality and coverage of the training data, the representation of relevant operating conditions,

and the distance of proposed designs from the evaluated region all influence the reliability of the search. Consequently, surrogate-based design should be understood as a means of identifying promising candidates for further assessment, not as a substitute for the verification of the final hull forms. Even with inexpensive evaluations, the optimization efficiency depends on the geometric representation and search strategy. Gradient-based and derivative-free hull optimization depends on response smoothness, sensitivities, dimensionality, and evaluation budget. Reducing the design space to physically meaningful shape variations can therefore be as important as selecting the optimization algorithm. Harries and Abt [8] examined shape variation and reduced design spaces to accelerate simulation-driven design. These considerations motivate the use of existing hull families to guide coordinated parameter changes beyond exploration within individual variable bounds.

Data-driven generative models provide a complementary approach to constructing design spaces by learning relationships among geometric variables and using them to generate new candidate forms. Khan et al. [9] developed ShipHullGAN as a learned parametric representation for ship-hull generation and design-space exploration. Such models can retain correlations among hull characteristics while exploring beyond individual training examples, although geometric plausibility, hydrodynamic performance, and vessel-specific requirements must still be addressed explicitly. Diffusion models generate new samples by learning to reverse a progressive noise-corruption process [10, 11]. For hull design, ShipGen demonstrated diffusion-based generation with feasibility and performance guidance [12], while C-ShipGen introduced conditional generation using prescribed principal dimensions and guidance from a total-resistance regression model [13]. For design optimization, however, learning the distribution of feasible geometries is only a starting point. The sampling process must also be directed toward candidates that improve a prescribed objective while satisfying relevant constraints. Krishnamoorthy et al. [14] demonstrated objective-guided conditional diffusion for offline black-box optimization. In hull design, such guidance can connect generative sampling to hydrodynamic performance, for example by rewarding lower predicted resistance while constraining unacceptable designs.

Performance information can be incorporated into diffusion-based design through several mechanisms. Differentiable surrogates enable gradient-based guidance, as demonstrated for resistance-directed hull generation in C-ShipGen [13] and coupled thermal-fluid heat-sink design in HeatGen [15]. Alternatively, reward-directed methods use performance values without requiring reward derivatives. Keramati et al. [16] combined reward-directed fine-tuning with value-based sampling for engineering design, including ship-hull applications. These approaches differ primarily in their requirements for differentiable feedback, model retraining, and sampling-time evaluation. Inference-time guidance is particularly useful when a trained generative model is reused for different performance objectives, since sampling can be redirected without retraining the diffusion model. Li et al. [17] demonstrated derivative-free value-based guidance without differentiating the reward or fine-tuning the generative model. Sequential Monte Carlo (SMC) provides a population-based alternative that propagates and resamples weighted particles [18], allowing reward-guided diffusion to preferentially sample promising candidate trajectories while retaining the learned denoising model as the proposal mechanism.

In this study, we propose a diffusion framework for localized optimization of the KRISO Container Ship (KCS), focusing on the bulbous bow and adjacent forebody. A diffusion model is trained on KCS-based geometries with controlled variations in this region, while parametric masking retains the remaining hull parameters at their baseline values. A

surrogate provides rapid performance evaluation, and reward-guided sequential Monte Carlo (SMC) sampling directs generation toward lower predicted resistance. Stability is retained through the training data and displacement is matched during subsequent screening, enabling data-driven exploration within an engineering-defined design envelope. The contribution is twofold. First, we introduce a derivative-free, reward-directed diffusion framework trained on a masked dataset for localized ship-hull optimization. Second, we provide a viscous hydrodynamic assessment of a sampled variant and of a hull refined by a naval architect within the same parametric representation. An automated CFD pipeline, with mesh and solver parameters determined systematically from hull geometry and Froude number, is validated against two benchmark container ships and applied to both hulls. The assessment reveals limitations in geometric representation, reward fidelity, operating-range aggregation, and preservation of stability margin that are not apparent from surrogate-based optimization alone.

The remainder of this paper is structured as follows. Section 2 presents the diffusion framework, including the masked design space, the denoising diffusion model, the reward-guided SMC sampling strategy, and the automated CFD pipeline together with its validation against the KCS and Duisburg Test Case benchmarks. Section 3 presents diagnostics of the diffusion model and sampler, summarizes the empirical resistance screening of the generated hull forms and the CFD assessment of one screened variant, and then reports the viscous assessment of the refined hull across seven Froude numbers, covering geometry, hydrostatics, resistance components, wave field and hull pressure. Section 4 discusses the speed dependence of the resistance comparisons and the representational, evaluative and constraint limitations they expose, and Section 5 concludes.

2. Methodology

This section presents the parametric diffusion framework for localized ship-hull optimization. The methodology enables targeted refinement of selected hull regions while preserving global hull-form continuity. It combines a denoising diffusion model with parametric masking and surrogate-guided sampling, with stability enforced through the training data and displacement matched during subsequent screening.

Fourteen design variables controlling the bow and bulbous-bow geometry are selected as the active subset, while the remaining variables are fixed at their baseline KCS values. Latin Hypercube Sampling (LHS) [19] is used to sample the active design space, producing 3,800 designs that meet the prescribed stability constraints. The full design vectors are used to train a hydrodynamic-resistance surrogate, with labels computed following Keramati et al. [16]: wave-making resistance is estimated using Michell’s integral and frictional resistance using the ITTC-57 correlation line applied to the wetted surface. The surrogate is a residual neural network (ResNet) with four residual blocks and 256 hidden units. Sampling is controlled through KL regularization and a temperature parameter.

The DDPM is trained on the fused full design vectors, while masking is imposed at the dataset level by varying only the active subset. This preserves a full-hull representation while concentrating geometric variability in the bow and bulbous-bow regions. After training, surrogate-guided sequential Monte Carlo sampling directs generation toward designs with lower predicted resistance. Figure 1 summarizes the workflow. In Stage 4, the CFD pipeline of Section 2.3 is applied to Model 1, one of the sampled variants (Section 3.2), and to GenHull E, a hull refined by a naval architect toward KCS within the

same parametric representation (Section 3.3). The DDPM and reward-guided sampling procedure are described next.

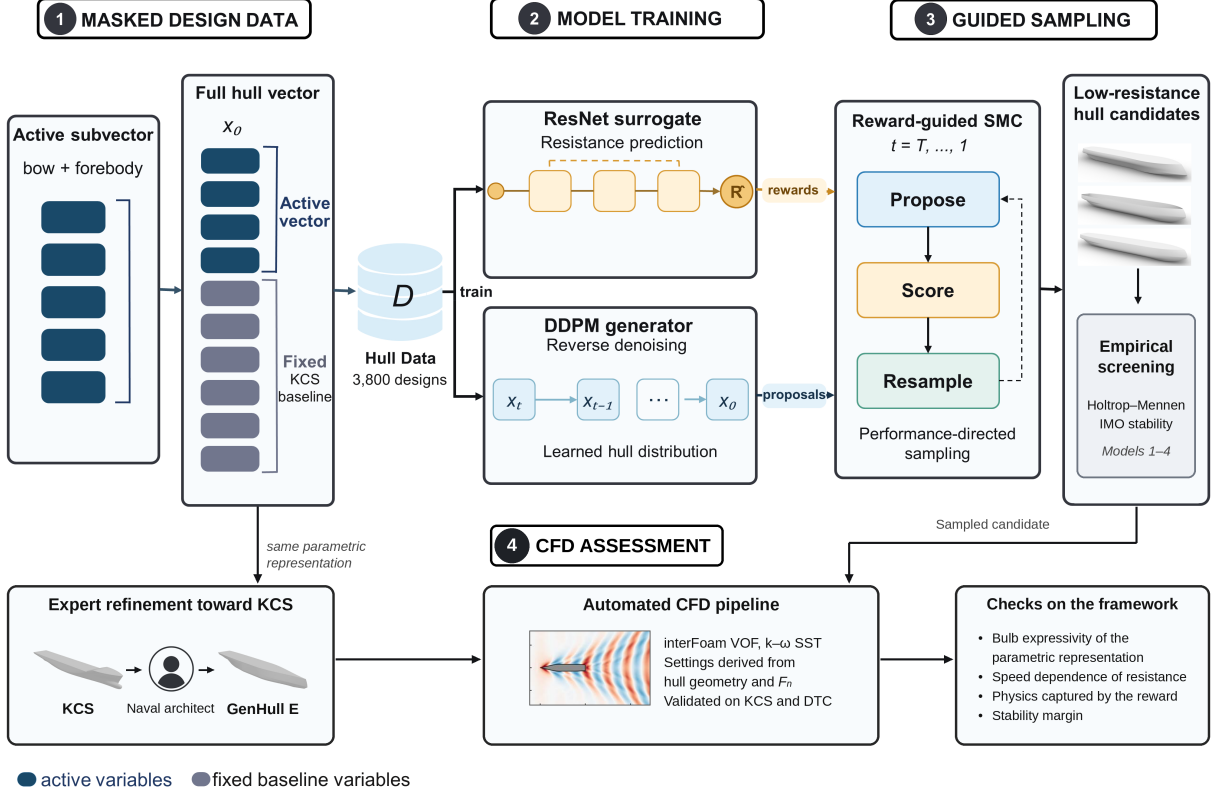


Figure 1: Schematic of the ship-hull optimization workflow with a partially variable design vector. The framework comprises data generation, training of a denoising diffusion probabilistic model (DDPM) and a deep residual surrogate model, reward-guided sampling via sequential Monte Carlo (SMC), and CFD assessment of a sampled variant and of a hull refined by a naval architect.

2.1. DDPM Formulation

Diffusion models are likelihood-based generative models that transform data into noise via a fixed forward Markov chain and learn the reverse denoising dynamics with a neural network [10, 11]. Given a data sample $x_0 \in \mathbb{R}^d$, DDPMs introduce latent variables $x_{1:T} = \{x_1, \dots, x_T\}$ over T diffusion steps and define the generative model

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t), \quad p(x_T) = \mathcal{N}(0, I), \quad (1)$$

where θ denotes the parameters of the neural network and I is an identity matrix.

Forward (noising) process. A fixed variance schedule $\{\beta_t\}_{t=1}^T$, with $\beta_t \in (0, 1)$, defines the forward Markov transitions

$$q(x_t | x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I\right), \quad (2)$$

where $q(\cdot)$ denotes the forward diffusion distribution.

Defining $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, the marginal distribution of x_t given x_0 admits the closed form

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \quad (3)$$

where $\epsilon \in \mathbb{R}^d$ is standard Gaussian noise.

Reverse (denoising) process. The reverse-time transitions are modeled as Gaussian conditionals

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I), \quad (4)$$

where $\mu_\theta(x_t, t)$ is a learned mean function and σ_t^2 denotes the variance at step t .

Using the noise-parameterization of [11], the reverse update can be written as

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z, \quad z \sim \mathcal{N}(0, I), \quad (5)$$

where $\epsilon_\theta(x_t, t)$ is a neural network predicting the injected noise and $z \in \mathbb{R}^d$ is standard Gaussian noise.

Training objective. Optimizing the evidence lower bound yields the standard DDPM loss

$$\begin{aligned} \mathcal{L}_{\text{simple}} &= \mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2], \\ x_t &= \sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t} \epsilon, \quad t \sim \text{Unif}\{1, \dots, T\}, \end{aligned} \quad (6)$$

where the expectation is taken over the data distribution, Gaussian noise ϵ , and uniformly sampled diffusion steps t . The diffusion formulation described above provides a generative mechanism for producing physically smooth and coherent hull-form variations within the prescribed parametric design space. However, in its standard form, the reverse diffusion process samples designs according to the learned data distribution without preferential emphasis on hydrodynamic performance. To transform the diffusion model from a passive generator into an active design optimization tool, a performance-aware guidance mechanism is required. In the present work, this guidance is introduced through a reward-based Sequential Monte Carlo (SMC) sampling strategy, which biases the reverse diffusion process toward hull geometries with lower predicted resistance while preserving diversity and geometric continuity. The details of this reward-guided sampling procedure are described next.

2.2. Sequential Monte Carlo sampling

After training, the surrogate model is used to evaluate candidate designs and assign rewards corresponding to their predicted hydrodynamic performance, with higher reward indicating lower total resistance. Sequential Monte Carlo sampling, following Algorithm 1, is then employed to bias the reverse diffusion process toward low-resistance hull geometries while remaining close to the learned data distribution. By balancing performance-driven selection with probabilistic resampling, the SMC procedure enables the diffusion model to function as an active optimization tool, promoting high-performing designs while preserving diversity and smooth geometric transitions.

In Algorithm 1, the function $\hat{v}_t(x_t)$ denotes the reward surrogate that evaluates the quality of a design state at diffusion step t . At each step, $\hat{v}_t(\cdot)$ is obtained by applying the ResNet surrogate model to the denoised design representation corresponding to x_t and returns a scalar reward proportional to the predicted drag value. In this work, the reward is defined such that higher values of $\hat{v}_t$ correspond to lower predicted drag. This reward applies the importance weights through an exponential tilting term $\exp(\hat{v}_t/\alpha)$, which biases the sampling process toward high-performing designs while preserving diversity through the temperature parameter α . The resampling threshold is set to 0.5 to mitigate weight degeneracy and prevent particle collapse, ensuring that the sampling process is not dominated by a single trajectory during the reverse diffusion steps. From

an engineering perspective, this procedure biases the generative process toward lower-resistance hull forms while retaining sufficient diversity to avoid premature convergence to a single design. For many generative settings, Algorithm 1 remains unchanged, with the hyperparameter α treated as a domain-dependent variable. Before presenting the results and discussion in the next section, we summarize the practical steps required for industrial practitioners and engineers to deploy generative models in practice. The first step is data curation in the structured format shown in Figure 1. It is generally advisable to begin with the full design vector, run and monitor the system performance, and validate feasibility. If successful, a component-wise model based on subvectors can then be constructed for localized optimization. Once sufficient data are generated and the empirical distribution is not severely skewed, the dataset can be used to train both predictive and generative models. These steps concern only the generation of data consistent with the model input representation. Subsequently, Algorithm 1 can be applied with a moderate temperature value to obtain improved designs without excessively biasing the search toward extreme high-performance regions. After robustness is confirmed, a systematic search over the temperature parameter can be conducted to identify the value that yields the best-performing designs.

We used Weights and Biases platform [20] to automate the search for parameters and hyperparameters. The final values of the parameters and hyperparameters of both the DDPM and ResNet models, including the corresponding loss functions, are reported in Table 1 and Table 2 to facilitate reproducibility.

Table 1: Parameters and hyperparameters for training DDPM

Parameter	Value
Input dimension (D)	44
Hidden dimension (H)	256
Activation	ReLU
Output dimension	44
Optimizer	AdamW
Loss function	Mean squared error (MSE)

Table 2: ResNet regression model parameters

Parameter	Specification
Hidden dimension	256 neurons
Residual block	Linear(256,256)
Optimizer	AdamW
Learning rate	5×10^{-4}
Weight decay	1×10^{-4}
Batch size	256
Loss function	Mean squared error (MSE)

Algorithm 1 SMC-Based Guidance

Require: Pre-trained diffusion models $\{p_t^{\text{pre}}\}_{t=T}^0$, proposal policies $\{q_t\}_{t=T}^0$, hyperparameter $\alpha \in \mathbb{R}$, Batch size N

1: **for** $t \in [T + 1, \dots, 1]$ **do**

2: **Importance sampling step:** Generate $i \in [1, \dots, N]$: $x_{t-1}^{[i]} \sim q_{t-1}(\cdot | x_t^{[i]})$

3: Update weights:

$$w^{[i]} \leftarrow \frac{p_{t-1}^{\text{pre}}(x_{t-1}^{[i]} | x_t^{[i]}) \exp(\hat{v}_{t-1}(x_{t-1}^{[i]}) / \alpha)}{q_{t-1}(x_{t-1}^{[i]} | x_t^{[i]}) \exp(\hat{v}_t(x_t^{[i]}) / \alpha)} w^{[i]}.$$

4: **Resampling step:** Calculate the effective sample size $\frac{(\sum_i w^{[i]})^2}{\sum_i (w^{[i]})^2}$. If it falls below $0.5N$, resample by selecting new indices with replacement:

$$\{x_{t-1}^{[i]}\}_{i=1}^N \leftarrow \{x_{t-1}^{s_{t-1}^{[i]}}\}_{i=1}^N, \quad \{s_{t-1}^{[i]}\}_{i=1}^N \sim \text{Cat}\left(\left\{\frac{w_{t-1}^{[i]}}{\sum_{j=1}^N w_{t-1}^{[j]}}\right\}_{i=1}^N\right).$$

5: **end for**

6: **Output:** $\{x_0^{[i]}\}_{i=1}^N$

2.3. CFD Evaluation of Resistance

The CFD results of the present study were obtained using an automated CFD pipeline developed for this purpose, in which every mesh and solver parameter is derived by rule from the hull geometry and the Froude number. The pipeline accepts a closed-hull STL file together with a set of ship particulars and produces a complete OpenFOAM case, and no parameter was adjusted manually for any of these computations. The principal benefit of this arrangement concerns comparability more than accuracy, since two hulls evaluated at the same Froude number are then treated identically by construction and not according to the judgment of the analyst preparing them. Differences between hulls may accordingly be interpreted as differences in geometry.

Computations use **interFoam** from OpenFOAM ESI v2412, solving the incompressible two-phase equations with the interface captured by the volume-of-fluid method [21] and turbulence closed by the $k-\omega$ SST model with wall functions [22]. A half domain with a centerplane symmetry condition extends $1.5 L_{pp}$ forward of the forward perpendicular, $2.0 L_{pp}$ to the side and $1.5 L_{pp}$ below the keel, with a Froude-scaled outlet. The background cell size is $0.0311 L_{pp}$, hull-surface refinement targets $y^+ = 56$, and the free-surface band targets 64 cells across the Kelvin wavelength $\lambda_K = 2\pi U^2/g$. The resulting meshes contain 3.52 and 9.92 million cells for KCS and DTC respectively. Because these targets are physical rather than integer, the same rules resolve to different refinement levels on different hulls. A single mesh, sized at the highest Froude number of a sweep, serves every speed in that sweep. Wave reflection from the domain boundaries is suppressed by a sponge zone relaxing velocity and turbulence quantities toward freestream values. Forces are averaged over at least thirty seconds of simulated time by the method of batch means, giving a 95 % confidence interval, and stationarity is verified by a half-to-half drift check. Figure 2 shows the mesh the rules produce on the KCS hull.

All computations are performed at fixed even keel and prescribed draft, whereas the

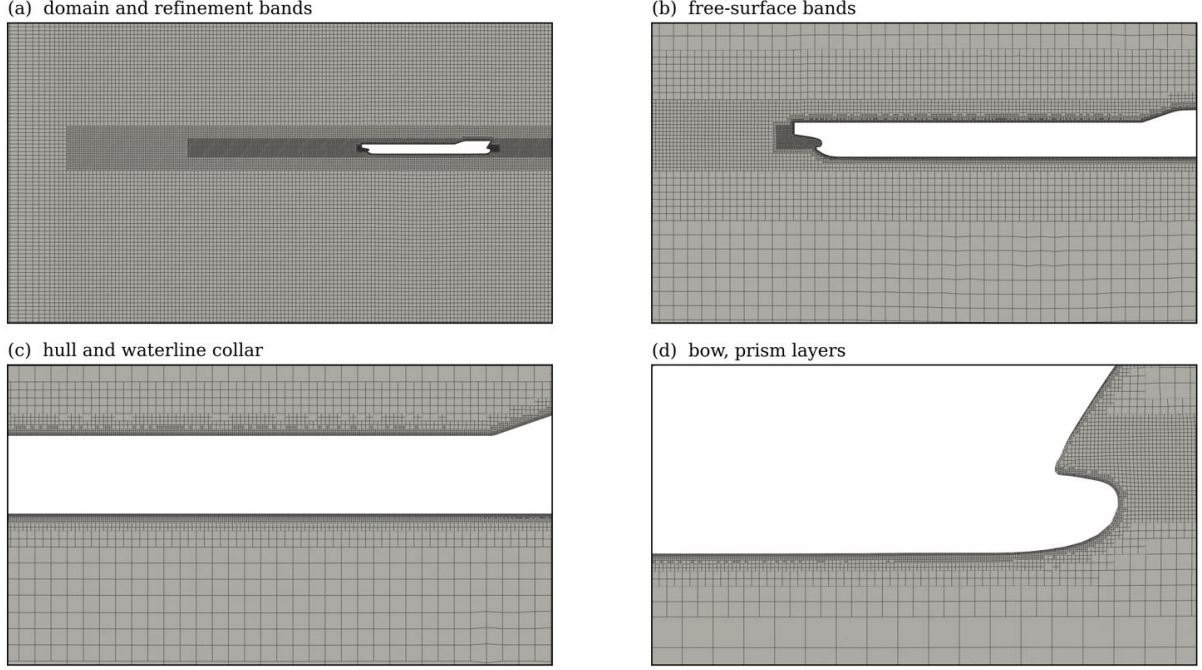


Figure 2: Computational mesh generated for the KCS hull, shown on the centerplane: (a) domain and refinement regions; (b) free-surface refinement bands; (c) hull surface and waterline refinement; (d) prism layers at the bow.

reference experiments are free to sink and trim. The magnitude of this difference was measured by repeating one computation on an identical mesh with the hull set at the experimentally measured attitude for $F_n = 0.260$, which increased total resistance by 0.88 % and skin friction by 1.9 %. Because sinkage scales approximately with the square of speed, a correction of this type cannot be applied uniformly across a speed range, and the raw comparison is therefore reported throughout. The same fixed condition is applied to every hull compared in Sections 3.2 and 3.3, so that each pair is compared at the same attitude.

2.3.1. Validation Against Benchmark Hulls

The pipeline was validated against model-scale resistance measurements for two benchmark container ships, the KRISO Container Ship and the Duisburg Test Case (DTC), at six Froude numbers each. The KCS results are given in Table 3 and Figure 3, and the DTC results in Table 4 and Figure 4. Experimental data for KCS are taken from the Tokyo 2015 workshop, Case 2.1 [23, 24], and for DTC from the SVA Potsdam measurements reported by el Moctar et al. [25], as tabulated by Feng et al. [26]. The KCS computations used a single mesh of 3.52×10^6 cells, sized at $F_n = 0.282$, for all six speeds. The tabulated y^+ is the average over the full hull patch, including dry faces above the waterline. On the wetted surface at $F_n = 0.260$, the distribution has a median of 98. The DTC computations used a mesh of 9 916 051 cells with 99.2 % boundary-layer coverage, generated by the identical rule set with no hull-specific intervention, and all six cases converged with a half-to-half drift below 0.80 %. The measured DTC resistance reaches a minimum at $F_n = 0.192$, and the computation reproduces a minimum one speed step higher, at $F_n = 0.200$.

The KCS points stay between +14.0 % and −8.9 % of the measurement, and between +6.7 % and −3.1 % over the range $F_n = 0.152$ to 0.227 targeted for the CFD-based

Table 3: Validation of the CFD pipeline against KCS model-scale measurements. ε_T is the error in total resistance coefficient relative to experiment.

F_n	y^+	$C_T \times 10^3$	95 % CI	EFD $\times 10^3$	ε_T [%]	C_{Fv} vs ITTC-57 [%]
0.108	25	4.326	± 0.08 %	3.796	+14.0	+6.3
0.152	32	3.886	± 0.40 %	3.641	+6.7	+2.9
0.195	39	3.455	± 0.90 %	3.475	-0.6	-2.1
0.227	44	3.361	± 0.24 %	3.467	-3.1	-4.5
0.260	48	3.382	± 0.46 %	3.711	-8.9	-6.3
0.282	60	4.148	± 0.55 %	4.501	-7.8	-7.8

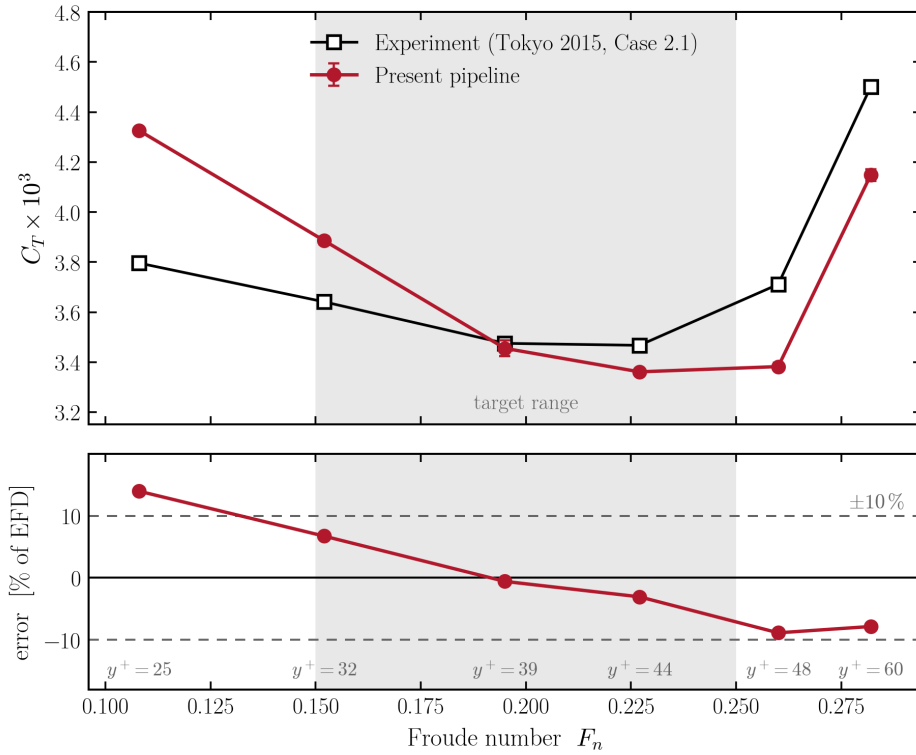


Figure 3: Computed and measured total resistance coefficients for KCS (top) and the corresponding comparison error (bottom). The shaded band indicates the Froude number range targeted for CFD-based training data.

Table 4: Validation of the CFD pipeline against DTC model-scale measurements, using the same rule set as for KCS.

F_n	Re	$C_T \times 10^3$	95 % CI	EFD $\times 10^3$	ε_T [%]	C_{Fv} vs ITTC-57 [%]
0.174	7.30×10^6	3.511	± 1.15 %	3.661	-4.1	-4.4
0.183	7.68×10^6	3.451	± 0.96 %	3.605	-4.3	-5.2
0.192	8.06×10^6	3.373	± 0.68 %	3.588	-6.0	-6.5
0.200	8.40×10^6	3.357	± 0.95 %	3.602	-6.8	-6.1
0.209	8.77×10^6	3.417	± 0.78 %	3.623	-5.7	-5.7
0.218	9.15×10^6	3.394	± 0.90 %	3.670	-7.5	-6.2

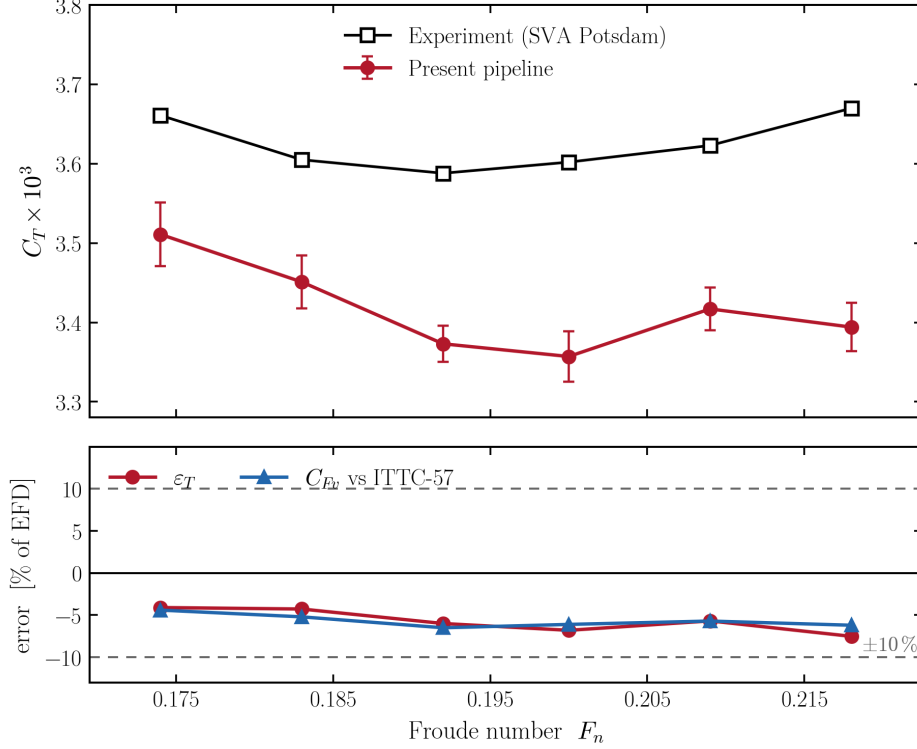


Figure 4: Computed and measured total resistance coefficients for DTC (top) and the corresponding comparison and skin-friction errors (bottom).

training data, while DTC points vary between -4.1% and -7.5% . Both hulls show skin friction trending with speed, at $+6.3\%$ to -7.8% on KCS and -4.4% to -6.2% on DTC.

The comparison error varies monotonically with Froude number, and this behavior follows from the meshing rules. Because a single mesh serves an entire sweep, the near-wall spacing remains fixed while the speed varies, so that the skin-friction error follows the wall resolution. The free-surface band is similarly sized for the highest speed of the sweep, so that the wave-resistance error follows the number of cells resolving the Kelvin wavelength, which falls from 64 at the top of the KCS range to approximately nine at the bottom. The comparison error therefore constitutes a bounded envelope with an identified cause, and may be taken into account when the comparisons presented below are interpreted. The lowest KCS speed, $F_n = 0.108$, lies outside the range used for any comparison in this paper and has been retained because it provides the coarse-resolution end of this explanation.

Two limitations of the validation are noted. A grid-convergence study was carried out for each hull on three geometrically similar grids, with the background cell size scaled by a factor of approximately $\sqrt{2}$, and the convergence ratio was evaluated separately for the pressure and viscous components of the resistance. The viscous component did not converge under refinement on either hull, with convergence ratios of 2.74 for KCS and 1.38 for DTC, whereas the pressure component converged monotonically with ratios of 0.68 and 0.52 respectively. The behavior of the total resistance under refinement follows that of the dominant component, and is divergent for KCS and convergent for DTC. A grid-convergence index for total resistance could therefore not be established and is not reported. The second limitation is that the comparison error at the KCS design speed exceeds the validation uncertainty, which indicates the presence of modeling

error according to the ITTC verification and validation procedure [27]. The residual is consistent with the behavior of the wall functions at the achieved y^+ , although this was not confirmed by a wall-resolved computation.

The pipeline is therefore suitable for comparative ranking of hulls at matched Froude number under a single rule set, in which the systematic error is common to both hulls and largely cancels in the difference, and not for absolute prediction of resistance. The degree of cancellation is bounded by the present validation data. At the closest matched speeds the comparison errors of the two benchmark hulls differ by four to five percentage points, so that cancellation cannot be assumed between dissimilar hull forms. For two hulls of matched length, draft and displacement, such as KCS and GenHull E (Section 3.3), the errors are expected to be more nearly equal and a greater degree of cancellation may be anticipated.

3. Results

This section evaluates denoising accuracy and examines SMC weight concentration, providing a diagnostic of diversity in particle contributions during sampling. It then summarizes the empirical resistance screening of the generated hull forms and reports a high-fidelity hydrodynamic assessment of one screened variant and of the parametric hull representation.

3.1. Diffusion Accuracy and Sampling Diversity

In order to assess the denoiser accuracy, we take a known, held-out hull design, add a known amount of random noise, and ask the model to predict that noise. This shows how accurately the model predicts the added noise in the forward process. Figure 5 shows the trained model’s accuracy on held-out designs across forward diffusion time. The left axis reports noise-prediction error, while the logarithmic right axis reports the conditional Gaussian reverse-posterior KL divergence using the reverse kernel with featurewise bounded denoised estimates. The KL compares the predicted reverse step distribution with the known reference hull and it shows how accurately does the model perform one reverse-diffusion step.

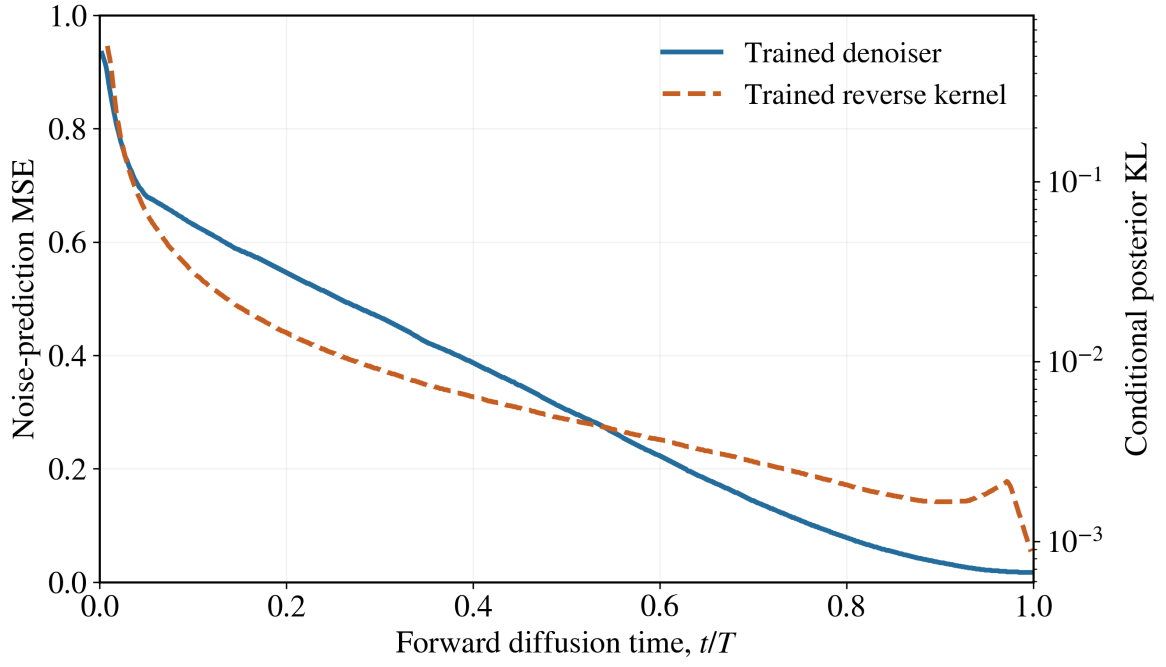


Figure 5: Denoising and reverse-step accuracy of the trained diffusion model on held-out hull designs. The left axis shows the mean squared error per dimension between the predicted and known forward-diffusion noise. The logarithmic right axis shows the conditional Gaussian posterior KL divergence, in nats per dimension, between the reference reverse-step posterior conditioned on the known clean hull and the trained reverse kernel using featurewise bounded denoised estimates. Both metrics are plotted against forward diffusion time; lower values indicate greater accuracy.

To assess SMC sampling performance across diffusion steps, Figure 6 examines the concentration of importance weights. The normalized effective sample size, ESS/N , is measured after each incremental weight update and before resampling. Intermediate resampling is triggered when $\text{ESS}/N < 0.5$.

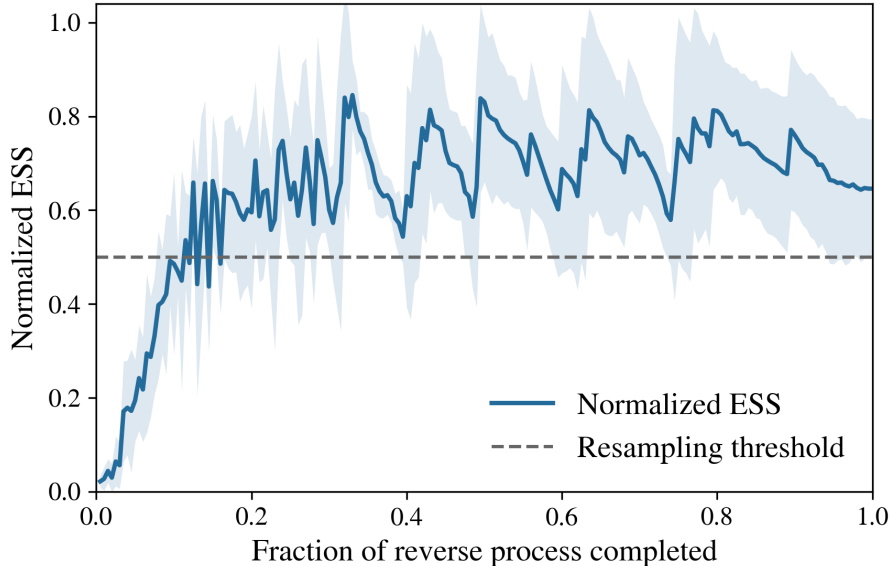


Figure 6: Normalized ESS during SMC sampling. The dashed line marks the resampling threshold of 0.5. The solid curve shows the mean and the shaded band shows an approximate pointwise 95% confidence interval across five sampler seeds, with the trained models held fixed. ESS is evaluated before resampling.

3.2. Screening and CFD Assessment of Generated Hull Forms

This subsection summarizes the empirical screening reported in the previous study on which the present work builds [28] and then assesses one of the screened variants with the CFD pipeline of Section 2.3. The diffusion framework outputs candidate hull geometries as triangulated surface meshes in STL format. Because STL is a faceted representation, the previous study applied a geometry-cleaning and surface-reconstruction step before CFD meshing. Each generated STL was imported into Rhinoceros and re-topologized using QuadRemesh to produce a quad-dominant surface mesh with improved element regularity, converted to a SubD representation and then to NURBS surfaces, from which a closed watertight polysurface was constructed. Geometric integrity was verified before export, covering watertightness, consistent surface normals and the absence of non-manifold edges and self-intersections. The exact preprocessing operations vary with the CAD and CFD toolchain; the requirement is a watertight, analysis-ready surface that can be robustly meshed. The generated STL of Model 1 is watertight and was meshed by the pipeline of Section 2.3 after scaling to model size, without surface reconstruction.

Four representative variants, denoted Models 1–4, were selected from the reward-guided sampling output and compared with the KCS baseline using the Holtrop–Mennen method over $F_n = 0.108$ to 0.45 at matched displacement. All four variants returned lower predicted total resistance coefficients than KCS over most of that range, and the largest reduction in aggregate predicted resistance was obtained for Model 3 [28]. Compared with KCS, all four variants have a reduced, effectively absent bulb and a finer bow entrance.

These results were obtained while satisfying preliminary hydrostatic and intact-stability screening checks, in which displacement is matched by adjusting the draft and

intact stability is evaluated through GM and GZ righting-arm curves against the IMO 2008 IS Code criteria. These checks provide early-stage feasibility screening and do not constitute class-society compliance verification. The more aggressive variants achieved their predicted margins in part by adopting finer forms at substantially greater draft for equal displacement. Models 3 and 4 require drafts of approximately 17.1 and 17.9 m, respectively, compared with 10.8 m for KCS. Such configurations would ordinarily be restricted by route and port draft limits. The geometric characteristics of the four variants and the influence of the parameter bounds are reported in full in [28] and are not repeated here. In that study, a CFD check of Model 1 at drafts of $0.25D$ and $0.67D$, where D is the hull depth, showed that the Holtrop–Mennen method underpredicted its resistance by up to 40 % at moderate Froude number.

To compare a sampled variant with KCS under consistent conditions, Model 1 was assessed with the automated pipeline of Section 2.3 at the displacement of KCS, using the same mesh rules and solver settings as the KCS computations. At this displacement, the draft of Model 1 is 18.6 % greater than that of KCS and its waterline beam is 9.1 % greater. Its block and prismatic coefficients are 0.50 and 0.73, compared with 0.65 and 0.66 for KCS. The total resistance of Model 1 differs from that of KCS by -1.2% at $F_n = 0.195$, whereas it is 44.9 % and 49.3 % higher at $F_n = 0.250$ and 0.260 , respectively. The screening therefore selected a variant whose resistance is comparable to that of KCS at the lower speed, but it did not capture the increase in resistance near the design speed. Two properties of the screening are relevant to this discrepancy. The procedure is an empirical regression evaluated on hulls whose form coefficients may lie outside the range over which the regression was fitted, and it aggregates over a Froude range wide enough to contain qualitatively different regimes. The reward that guides the sampler is also a low-fidelity estimate aggregated over a wide speed range, and neither property is visible in the values it returns.

3.3. CFD Assessment of a Refined Hull

The discrepancy between the empirical and viscous assessments of Model 1, first observed in the earlier study [28], led to an expert examination of the parametric representation of Section 2. Because the sampler can propose only hulls that this representation is able to express, its capability bounds what the framework can achieve. In this section, a hull refined by a naval architect toward KCS is evaluated with the same pipeline (Section 2.3) to examine what the representation produces near the parent hull and how the resistance of that hull varies across the speed range. The hull considered, designated GenHull E, is distinct from Models 1–4. It was constructed by the authors as an expert check of the representation and was fitted to KCS by constrained least squares, starting from a KCS-like parameter set based on the principal dimensions of KCS. In the first of two fitting stages, nine integral form coefficients were matched, with length, draft and displacement imposed as hard constraints and the remaining quantities, including beam, midship area, longitudinal center of buoyancy and waterplane coefficient, entering as soft residuals. In the second, the station half-breadths of KCS were matched together with waterline length, displacement and beam, with the aft stations weighted most heavily and the bulb parameters disabled. No resistance or stability term entered the fit at either stage. GenHull E is therefore the closest form to KCS that the representation was found to express, and its behavior characterizes the representation within which the generative framework operates, not the output of the sampler. GenHull A, an earlier stage of the same sequence that retained the parameterized bulb, is discussed in Section 4.2.

3.3.1. Geometry and Hydrostatics

Figure 7 shows GenHull E together with the two benchmark container ships used to validate the evaluation pipeline, and Table 5 lists the principal particulars and form coefficients of GenHull E and KCS. All values were computed from the as-meshed geometry at the design draft. The coefficients are referred to L_{pp} , which is the basis for the Froude and Reynolds numbers and is common to both hulls by construction. The waterline length is reported separately, because the KCS bulb protrudes forward of the forward perpendicular and the measured waterline consequently exceeds L_{pp} by 0.95 %. The longitudinal center of buoyancy is measured identically on both hulls as a percentage of waterline length aft of the waterline mid-point; the published KCS value of +1.48 % L_{pp} is referred to the forward perpendicular, which the bulb displaces.

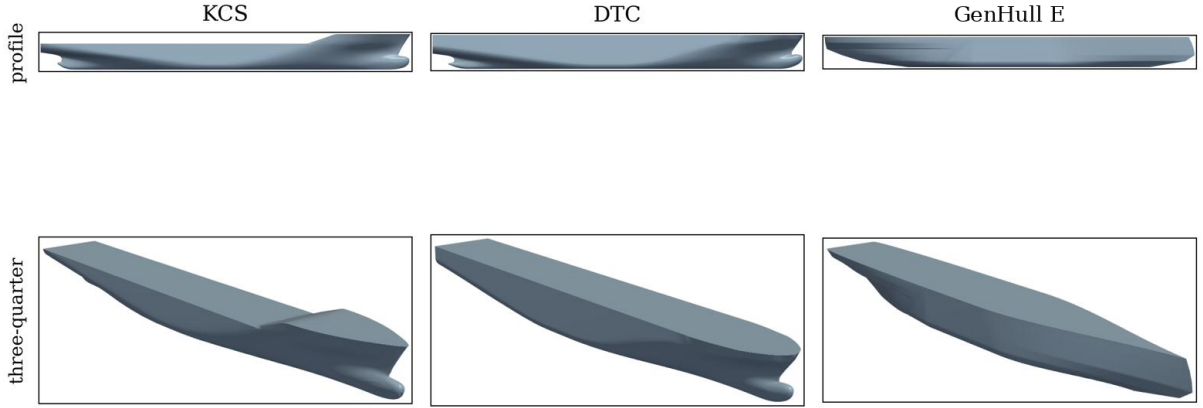


Figure 7: Profile (top) and three-quarter (bottom) views of the KCS, DTC and GenHull E hull forms.

Table 5: Principal particulars and form coefficients of KCS and GenHull E at full scale, computed from the as-meshed geometry at design draft.

Quantity	KCS	GenHull E	Difference
L_{pp} [m]	230.00	230.00	constrained
L_{wl} [m]	232.19	230.30	−0.81 %
B_{wl} [m]	32.22	34.46	+6.96 %
T [m]	10.80	10.80	−0.01 %
∇ [m ³]	51 765	52 042	+0.53 %
S [m ²]	9 462	8 802	−6.97 %
A_{wp} [m ²]	6 135	5 466	−10.91 %
A_m [m ²]	341.1	360.9	+5.80 %
C_B	0.6468	0.6080	−6.00 %
C_M	0.9802	0.9697	−1.08 %
C_P	0.6599	0.6270	−4.98 %
C_{WP}	0.8279	0.6896	−16.71 %
$S/\nabla^{2/3}$	6.812	6.315	−7.30 %
LCB [% L aft]	+1.00	+0.85	0.15 forward
BM_T [m]	9.00	7.84	−12.93 %
KM_T [m]	14.93	13.64	−8.62 %
Bulbous bow	present	absent	—

At matched length, draft and displacement, GenHull E has 6.97 % less wetted surface than KCS, 6.96 % greater beam, and a finer entry. These differences are consequences of the representation and not of any objective, since the fit contained no resistance term. With the bulb disabled and the hull fitted to the station half-breadths of KCS, the closest form that the representation produces at matched length, draft and displacement is beamier and finer at the ends than KCS and carries less wetted surface. GenHull E has no bulbous bow, whereas both benchmark hulls possess one, because the representation cannot express a faired bulb, as examined in Section 4.2.

The finer waterplane has a consequence for transverse stability that is not apparent from the form coefficients. Although the beam increases by 6.96 %, the waterplane inertia decreases by 12.46 %, because a fine entry places less waterplane area at large distance from the centerline. The metacentric radius BM_T consequently decreases by 12.93 % and the height of the transverse metacentre KM_T by 1.29 m. Figure 8 shows the cross curves of stability computed for both hulls at constant displacement, with both hulls truncated to a common depth of 19.0 m so that the freeboard treatment is identical. The two curves cross near 28° . GenHull E is less stiff than KCS near the upright condition but develops a larger maximum righting lever at a greater angle of heel, reaching $GZ_{max} = 1.39$ m at 37.5° against 1.11 m at 30° for KCS when both hulls are evaluated at $KG = 13.0$ m.

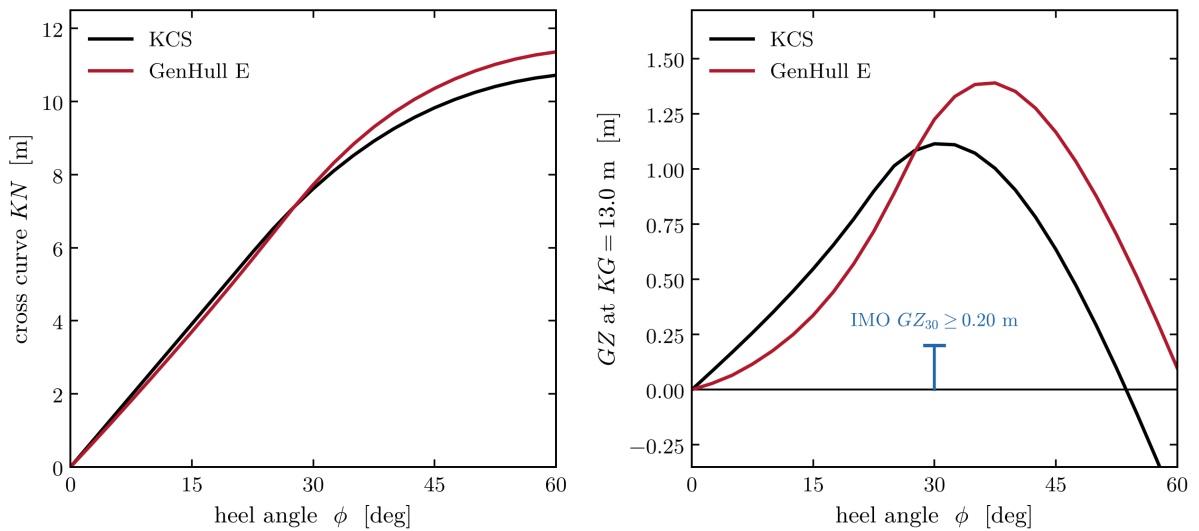


Figure 8: Cross curves of stability (left) and righting-arm curves at $KG = 13.0$ m (right) for KCS and GenHull E at design draft.

Expressed as the highest center of gravity satisfying the general intact-stability criteria of the 2008 IS Code, the limiting KG is 14.53 m for KCS and 13.49 m for GenHull E, a reduction of 1.04 m. The governing criterion differs between the two hulls. KCS is limited by the area under the righting-arm curve between 30° and 40° , whereas GenHull E is limited by the requirement for a minimum initial metacentric height. Where an operational metacentric height requirement applies equally to both hulls, as is usual for a container ship, the reduction corresponds to the full 1.29 m difference in KM_T . All stability quantities are reported at full scale, since the criteria of the IS Code are defined at that scale. The values were computed at upright trim on a bare hull with no downflooding angle defined, and the severe wind and rolling criterion was not evaluated. They therefore constitute a comparison of hull form and not an assessment of compliance.

3.3.2. Resistance Comparison with KCS

GenHull E was compared with KCS at seven Froude numbers, with length, draft and displacement as the matched basis and with beam and wetted surface free. Both hulls were computed on the same machine under the identical rule set, so that no difference in numerical setup enters the comparison, and the forces on both were averaged over matched windows of the force histories. The results are given in Table 6 and Figure 9. The final column of the table gives the resistance difference divided by the combined 95 % confidence half-width of the pair; it is a measure of separation and not a standard-deviation significance level, and it is not reported at the two lowest speeds, for which the confidence intervals were not recorded.

Table 6: Total resistance of GenHull E and KCS at model scale 1:31.6, with the pressure and friction components of the difference.

F_n	R_T KCS [N]	R_T GenHull E [N]	ΔR_T [%]	Δ pressure [N]	Δ friction [N]	separation
0.152	30.418	28.042	-7.81	+0.78	-3.15	—
0.195	44.596	40.843	-8.42	+2.06	-5.81	—
0.227	59.147	55.377	-6.37	+3.58	-7.35	2.1
0.240	65.366	60.661	-7.20	+3.18	-7.88	1.9
0.250	70.053	70.480	+0.61	+8.49	-8.07	0.2
0.260	77.716	83.114	+6.95	+14.04	-8.64	3.4
0.270	88.128	100.231	+13.73	+21.37	-9.27	8.4

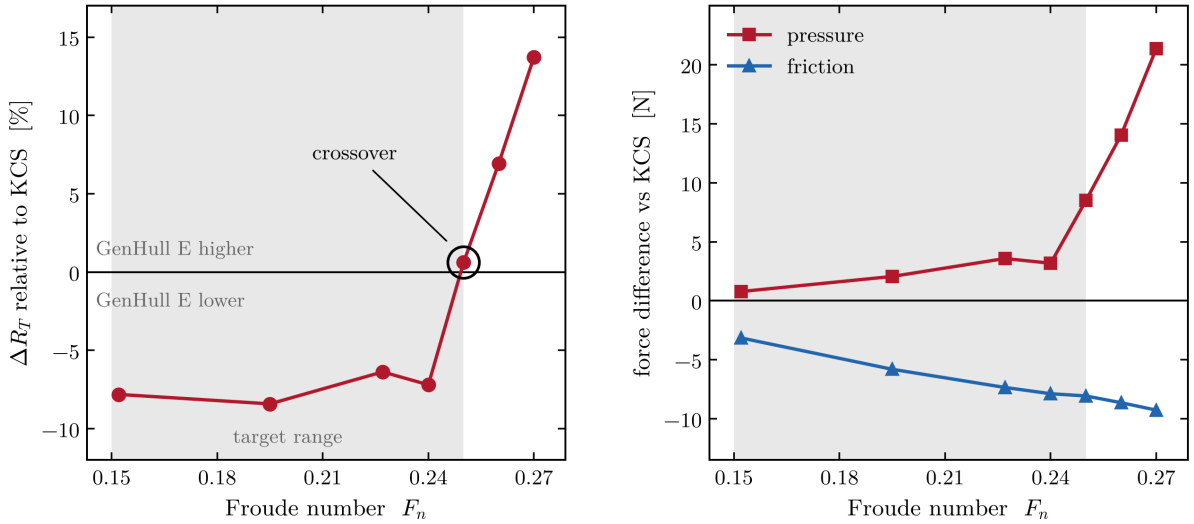


Figure 9: Total resistance of GenHull E relative to KCS (left) and the pressure and friction components of the difference (right). The shaded band indicates the Froude number range targeted for CFD-based training data.

GenHull E has six to eight percent lower total resistance than KCS from $F_n = 0.152$ to 0.240 and higher total resistance above that range. The crossover occurs at $F_n = 0.250$, where the difference is +0.61 %, or 0.43 N on 70 N. This difference is two tenths of the combined confidence half-width and is therefore indistinguishable from zero. When evaluated on four successive averaging windows as the run lengthened, the difference at this speed read +0.27, +0.02, +0.22 and +0.61 %, all within half a percent of zero.

The crossover is thus located at a computed speed and not interpolated between distant points, and the four window lengths place it at $F_n = 0.250$ to within the repeatability of the method.

The component columns of Table 6 identify the mechanism responsible. GenHull E has lower friction than KCS at every speed, and the difference increases smoothly with dynamic pressure from -3.15 N at $F_n = 0.152$ to -9.27 N at 0.270 . The magnitude of this difference exceeds the value that the reduced wetted surface alone would produce, since at the crossover the friction difference amounts to 14.3% of the KCS value against a 6.97% difference in wetted area. This excess is examined in Section 3.3.4. The pressure difference remains below 4 N up to $F_n = 0.240$ and then increases sharply, reaching 8.5 , 14.0 and 21.4 N at 0.250 , 0.260 and 0.270 respectively. Between $F_n = 0.240$ and 0.270 it increases nearly sevenfold, whereas the dynamic pressure increases by 27% . The rate of this increase identifies the pressure difference as wave-making resistance and not as viscous pressure resistance. The two components are measured separately and have opposite speed dependence, and the crossover occurs at the speed where the pressure penalty exceeds the friction advantage.

The matched-displacement basis has a further consequence for the manner in which the comparison is reported. Because the wetted surfaces differ by approximately 7% , the total resistance coefficient and the total resistance force do not rank the hulls by the same margin, and at $F_n = 0.270$ a coefficient difference of 22% corresponds to a force difference of 13.7% . Forces are therefore reported in this section, and coefficients are reserved for the comparison with measurement in Section 2.3.1.

3.3.3. Free-Surface and Hull Pressure Fields

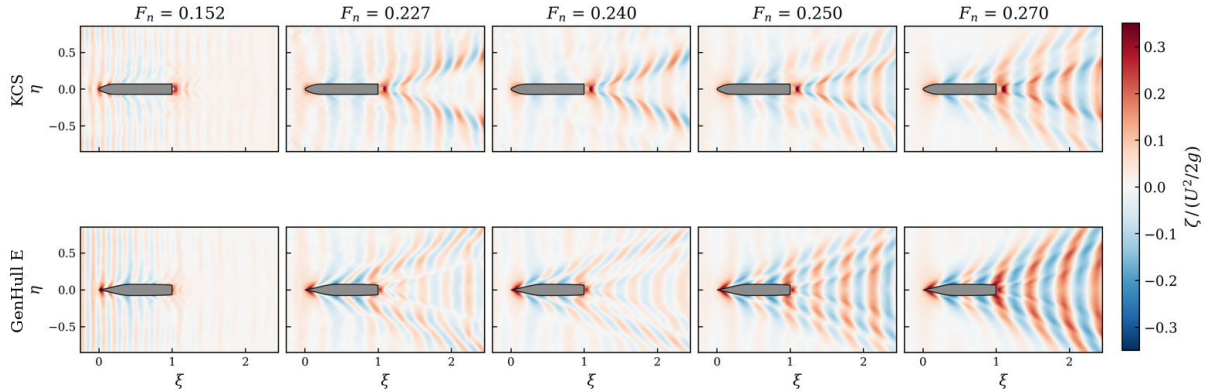


Figure 10: Free-surface elevation, normalized by $U^2/2g$, for KCS (top) and GenHull E (bottom) at five Froude numbers.

Figure 10 shows the computed free-surface elevation for both hulls at five Froude numbers and indicates the physical origin of the crossover. At $F_n = 0.152$ the Kelvin wavelength is $0.145 L_{pp}$ and the two wave systems are nearly indistinguishable, so that wave-making forms a small fraction of the total resistance and the difference in wetted surface determines the comparison. At $F_n = 0.270$ the wavelength has increased to $0.458 L_{pp}$, both wave systems are strong, and the system generated by GenHull E is visibly the stronger. The elevation is normalized by $U^2/2g$ because wave amplitude scales with the square of speed; on a common dimensional scale the low-speed panels would be nearly blank and differences between speeds would obscure differences between hulls.

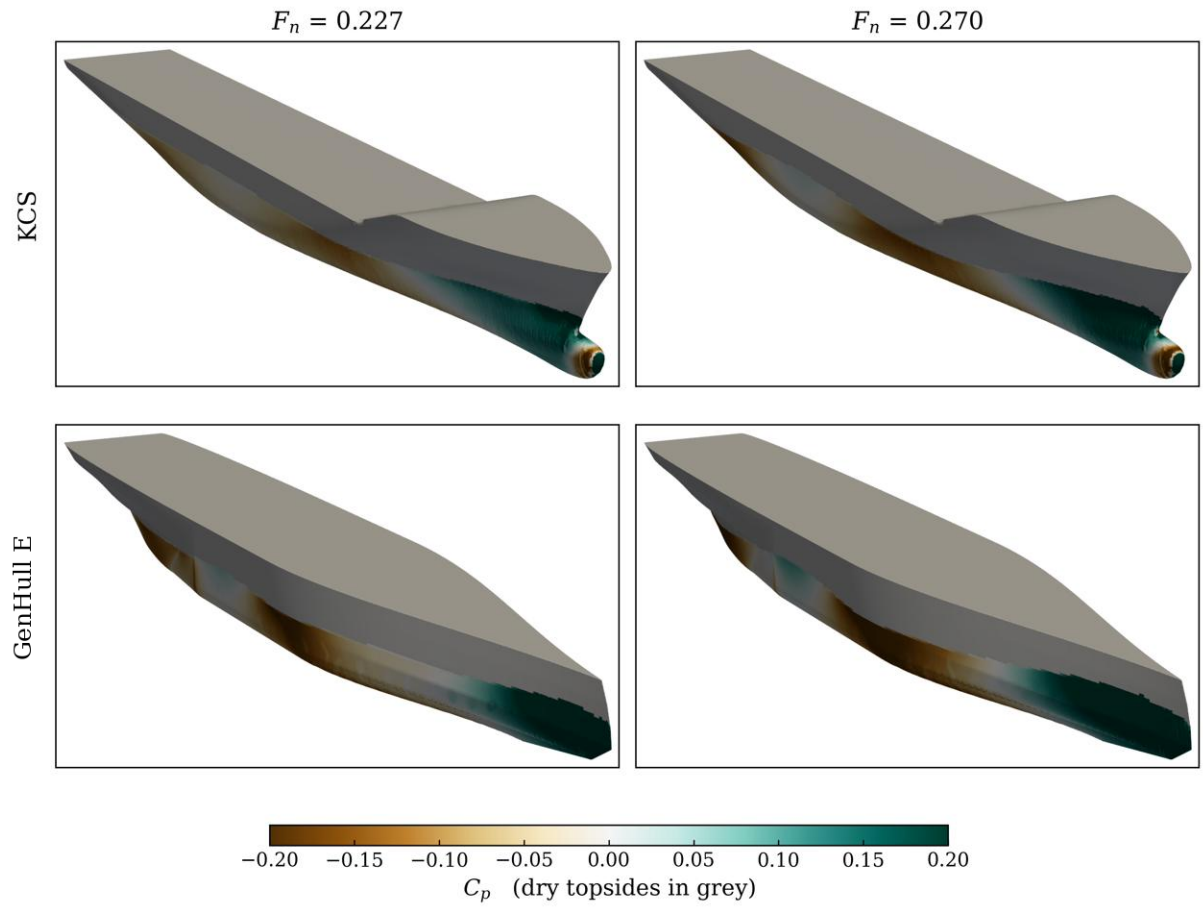


Figure 11: Pressure coefficient on the wetted forebody of KCS (top) and GenHull E (bottom) at $F_n = 0.227$ and 0.270 .

Figure 11 shows the pressure coefficient on the wetted forebody of both hulls at Froude numbers on either side of the crossover. KCS exhibits a concentrated stagnation region ahead of the stem at the bulb, which is the wave-cancellation mechanism that a bulbous bow provides near the design speed. GenHull E exhibits its stagnation region on the stem and develops stronger forebody suction at the higher speed, with no canceling wave system ahead of the hull.

3.3.4. Limitations of the Comparison

Two aspects of the computed viscous force qualify the comparison. The first concerns boundary-layer coverage, which lies between 87 and 92 % on GenHull E against 98 % on KCS. A direct measurement on the DTC hull, in which meshes achieving 75.8 % and 99.2 % coverage were compared at the same condition, established that 23.4 percentage points of coverage correspond to 3.7 % of skin friction. Scaling this result linearly, the present shortfall of approximately 11 points corresponds to 1.7 % of skin friction, or approximately 1.4 % of total resistance, in favor of GenHull E. A contribution of this size is attributable to discretization and should be discounted from the measured advantage.

The second concerns the remaining discrepancy, which is larger and has a different origin. At the crossover the viscous force on GenHull E lies 13.4 % below the ITTC-57 flat-plate value, against 6.1 % for KCS, a gap of 7.3 percentage points of which coverage accounts for fewer than two. A station-wise diagnostic located the balance of this gap at the bow and the forward shoulder, which is consistent with local separation on a forebody that has no bulb and carries excess beam at the shoulder. Separation of this kind reduces wall shear and increases pressure drag simultaneously, so that its effect is not confined to a reduction in friction, and the associated increase in pressure drag is included in the pressure column of Table 6. The present meshes do not permit the physical and numerical contributions to be distinguished, and no correction has been applied. When only the coverage contribution is discounted, the resistance advantage attributable to the hull form up to $F_n = 0.240$ is five to seven percent, against the six to eight percent measured.

No grid-convergence study was performed on GenHull E. The meshing rules applied to both hulls are identical to those used for the validation cases, so that the discretization error is largely common to the two members of each pair, although this argument is weakest where the flow physics differs most, which is above the crossover. Both hulls were computed at fixed even keel, consistent with the validation cases.

4. Discussion

4.1. Influence of Froude Number on the Resistance Comparison

The comparison of Section 3.3.2 shows that the sign of the resistance difference between GenHull E and KCS depends on the Froude number at which the comparison is made. At $F_n = 0.195$ GenHull E has 8.4 % lower total resistance than the benchmark, whereas at $F_n = 0.270$ its resistance is 13.7 % higher. Model 1, the sampled variant of Section 3.2, shows a similar dependence on speed with a larger penalty at the higher speeds. Its resistance is comparable to that of KCS at $F_n = 0.195$ but 45–50 % higher at $F_n = 0.250$ and 0.260, whereas the empirical screening predicted a lower resistance coefficient than that of KCS near $F_n = 0.26$ [28]. For both hulls, an assessment made at a single speed would therefore not hold across the speed range.

For GenHull E, the reversal arises from the two mechanisms identified in Section 3.3.2, in which the friction advantage of the smaller wetted surface grows smoothly with dynamic

pressure while the wave-making penalty of the bulbless forebody grows much faster above $F_n = 0.240$. An objective aggregated over a Froude range that spans such a reversal assigns a single value to two opposing regimes, and the performance at any individual speed may differ substantially from that value. Locating the speed at which an advantage ends requires evaluation at a sufficient number of conditions to bracket the crossover, and in the present case seven paired computations were required to establish it. An automated evaluation pipeline makes computations of this number affordable for a single hull comparison.

The result also bears on the selection of the evaluation range within the generative loop. The reward of Section 2, following [16], aggregates resistance from $F_n = 0.10$ to 0.45, a range that spans slow-steaming conditions through the resistance hump characteristic of container-ship forms. Because resistance grows rapidly with speed, an aggregate of this kind is dominated by the highest speeds, which here lie above the crossover. The operational range of a given vessel is usually narrower. For a vessel operating predominantly below $F_n = 0.2$, weighting the objective toward the low-speed end would favor forms resembling GenHull E. For a container ship operating between $F_n = 0.25$ and 0.30, the CFD results indicate that the same form would be disadvantageous, and that the transition between the two regimes is sufficiently sharp that an objective averaged across it would not represent either regime adequately. Selection of the Froude number range and its associated weighting is therefore an essential part of the problem definition.

4.2. Bulbous Bow Representation in the Parametric Hull Model

GenHull E has no bulbous bow, and this characteristic was found to originate in the parametric hull representation and not in the fit. A sweep of 108 parameter configurations was carried out over the bulb-related parameters of the representation, and every configuration produced a constant-width cylindrical forebody, since the bulb profile blends against a section family that cannot close into a rounded forward body. One of the nominal bulb parameters was found to have no effect on the generated geometry.

The consequence of this limitation was quantified at $F_n = 0.152$ under the same meshing rules. GenHull A, an earlier stage of the fit that retained the parameterized bulb while matching KCS on the nine integral form coefficients, returned 27.7% higher total resistance than KCS, whereas GenHull E, with the bulb disabled, returned 7.8% lower resistance at the same speed. The excess resistance of GenHull A was traced to viscous pressure drag at its cylindrical forebody.

This observation has a general implication for generative hull models. Matching integral form coefficients does not constrain the local geometry that governs resistance, and a parameterization may be incapable of representing a feature without any indication of this from the coefficients on which it was fitted. Verification of a generative geometry model against integral coefficients alone is therefore insufficient, and verification against local geometry or against a resolved resistance computation is required in addition.

4.3. Transverse Stability and the Objective Function

The stability result of Section 3.3.1 illustrates a related limitation. The generative framework screens candidate geometries against the general intact-stability criteria of the 2008 IS Code, and GenHull E satisfies those criteria for a vertical center of gravity up to 13.49 m. The fit that produced GenHull E constrained integral form coefficients and station shapes, none of which is a measure of stability, and the reward of the generative framework likewise contains no stability term. The resulting hull has a limiting vertical

center of gravity 1.04m lower than that of the benchmark, or 1.29m lower where an operational metacentric height requirement governs in place of the regulatory minimum, corresponding to seven to nine percent of a limiting KG of approximately 14.5 m.

The cause of this reduction is not apparent from the form coefficients. GenHull E is 6.96 % beamier than KCS, which would ordinarily increase the metacentric radius, but its waterplane coefficient is 16.7 % lower and the waterplane inertia is 12.5 % lower despite the 22 % increase in B^3 . A fine entry places less waterplane area at large distance from the centerline, and near the upright condition this distribution governs the initial stability. The same fineness that reduces the wetted surface therefore also reduces the second moment of the waterplane. The reward of the generative framework contains a friction term proportional to the wetted surface but no term for the waterplane inertia, so that a reduction of wetted surface obtained in this way would be credited by the reward while the accompanying loss of initial stability would not be registered.

This example shows that a feasibility screen and an objective term are not equivalent means of enforcing a design requirement. A screen establishes an admissible region, and an optimization that has no gradient with respect to stability may reduce the stability margin up to the limit of that region without the reduction being reflected in the objective value. The large-angle behavior of the two hulls further shows that their ranking depends on which stability measure is considered, since GenHull E is less stiff near upright yet attains a larger maximum righting lever at a greater angle of heel. The choice of which requirements are enforced as constraints, which are included in the objective, and which are screened after generation therefore has consequences that the objective value alone does not reveal.

4.4. Requirements for a CFD-Based Reward

The result of Section 4.2 might suggest that improved training data would enable the generative model to recover a satisfactory bulbous bow. However, two distinct limitations are involved: geometric expressivity and reward fidelity. The first is representational. The adopted parameterization cannot express a faired bulbous bow; consequently, no amount of training data can recover geometry outside the admissible design space. Addressing this limitation requires a richer parameterization or geometric representation. The second concerns the physics represented by the reward. The present reward combines ITTC-57 friction based on wetted surface with an inviscid wave-resistance estimate and therefore omits viscous-pressure effects. These effects contribute to the penalty of the parameterized bulb (Section 4.2) and the separated flow around GenHull E (Section 3.3.4). Addressing this limitation requires a higher-fidelity reward, for example one trained on viscous CFD data. These limitations are coupled. A richer parameterization with the present reward could generate new geometries but rank them incorrectly, whereas a higher-fidelity reward with the present parameterization cannot recover geometries outside its admissible family. Reliable generative optimization therefore requires geometric expressivity and reward fidelity to advance together.

Two further properties of the reward bound the confidence that can be placed in the empirical ranking. The frictional term uses the wetted surface computed by the hull-generation routines of [29], which integrate the waterline perimeter over height and therefore undercount any part of the hull surface that is not vertical. For the KCS-like parameter set from which GenHull E was derived, this routine returns 45 % less wetted surface at the design draft than the area of the hull surface. The weight of this error in the reward depends on the share of friction in the aggregated resistance, which is small

at the highest speeds that dominate the aggregate but large within the speed range of present interest. In addition, the reward contains no displacement term, and displacement is accounted for only in the matched-displacement screening of Section 3.2.

A resistance dataset generated with the pipeline of Section 2.3 would provide three properties that the present reward lacks. Such a dataset contains viscous pressure, so that bulb-like features can be scored; it is resolved by speed using a consistent method, so that a crossover of the kind reported in Section 3.3.2 is represented and not averaged away; and, being produced under a single rule set, it permits hull-to-hull differences to be interpreted to the extent established in Section 2.3.1. Absolute accuracy is not among these properties. The comparison envelope extends from +14 % to −9 % across the full speed range and is approximately ± 7 % within the range of present interest, and a surrogate trained on such data would inherit this bias. The acceptability of this bias rests on its being common-mode within a hull family, which the two-hull validation supports and bounds at four to five percentage points across dissimilar forms. The generation of such a dataset at the scale required by a surrogate, of the order of 200 hulls at six speeds, is the motivation for automating the evaluation pipeline and is the subject of continuing work.

5. Conclusion

This study combined a reward-directed diffusion framework for localized ship-hull optimization with a viscous assessment of a generated variant and of a hull refined by a naval architect toward KCS. Parametric masking and surrogate-guided sequential Monte Carlo sampling produced hull forms with lower empirically predicted calm-water resistance than KCS while restricting geometric variation to a designated region.

An automated CFD pipeline derives mesh and solver parameters from the hull geometry and Froude number using a consistent rule set. Validated against two benchmark container ships at six speeds each, it reproduces total resistance within +14.0 % to −8.9 % for KCS and −4.1 % to −7.5 % for DTC, with errors varying monotonically with speed. The pressure converged monotonically under grid refinement, whereas the skin friction did not converge on either hull, preventing a single grid-convergence index for total resistance. These limitations support comparative assessment at matched Froude number rather than precise absolute prediction.

Applied to one of the sampled variants (Model 1), the pipeline gave a total resistance comparable to that of KCS at $F_n = 0.195$ and 45–50 % higher at $F_n = 0.250$ and 0.260 , an increase that the empirical screening did not capture. For the refined GenHull E, CFD identified a speed-dependent performance reversal. Its total resistance is six to eight percent lower than that of KCS up to $F_n = 0.240$, equal within the confidence interval at $F_n = 0.250$, where an 8.07 N friction advantage nearly cancels an 8.49 N pressure penalty, and higher above that speed. The friction advantage grows smoothly with dynamic pressure, while the pressure penalty increases nearly sevenfold between $F_n = 0.240$ and 0.270 . Aggregating resistance across this crossover obscures the opposing trends.

The assessment identified limitations in geometry, reward evaluation, and stability treatment. None of the 108 bulb configurations tested produced a rounded bulbous bow; the parameterized bulb increased the resistance of GenHull A by 27.7 % relative to KCS at $F_n = 0.152$ by viscous pressure drag. The reward omits this mechanism and uses a wetted-surface routine that undercounts sloping surfaces. GenHull E also loses 1.04 m of permissible vertical center of gravity while satisfying intact-stability criteria, illustrating a margin reduction that the reward’s lack of a stability term leaves unaccounted for.

These findings show that the matching of integral form coefficients and passing feasibility checks are insufficient to establish hydrodynamic performance or preserve stability margins. Continuing work will generate a CFD resistance data set of approximately 200 hulls at six speeds to train a speed-resolved reward incorporating viscous pressure and extend the assessment to further hulls sampled from the framework, before and after expert refinement. A richer geometric parameterization is also required: improved geometry and reward fidelity must advance together to generate useful features and evaluate their effects reliably.

Data and code availability

The code for the CFD pipeline and diffusion framework, along with the dataset, will be made publicly available upon acceptance of this manuscript.

Acknowledgments

The present study is supported by the NSERC Alliance Grant with Elomatic and InnovMarine. This research was also supported in part through computational resources and services provided by Advanced Research Computing (ARC) at the University of British Columbia and Compute Canada.

References

- [1] H. M. Gaspar, D. H. Rhodes, A. M. Ross, S. O. Erikstad, Addressing complexity aspects in conceptual ship design: A systems engineering approach, *Journal of Ship Production and Design* 28 (4) (2012) 145–159.
- [2] B.-J. Zhang, S.-L. Zhang, *Research on Ship Design and Optimization Based on Simulation-Based Design (SBD) Technique*, Springer, Singapore, 2019.
- [3] F. Pérez, J. Clemente, Constrained design of simple ship hulls with b-spline surfaces, *Computer-Aided Design* 43 (12) (2011) 1829–1840.
- [4] S. Han, Y.-S. Lee, Y. B. Choi, Hydrodynamic hull form optimization using parametric models, *Journal of Marine Science and Technology* 17 (2012) 1–17.
- [5] Y. Lu, X. Chang, A.-k. Hu, A hydrodynamic optimization design methodology for a ship bulbous bow under multiple operating conditions, *Engineering Applications of Computational Fluid Mechanics* 10 (1) (2016) 330–345.
- [6] J. Holtrop, G. G. J. Mennen, An approximate power prediction method, *International Shipbuilding Progress* 29 (335) (1982) 166–170.
- [7] A. Coppedè, S. Gaggero, G. Vernengo, D. Villa, Hydrodynamic shape optimization by high fidelity CFD solver and Gaussian process based response surface method, *Applied Ocean Research* 90 (2019) 101841.
- [8] S. Harries, C. Abt, Faster turn-around times for the design and optimization of functional surfaces, *Ocean Engineering* 193 (2019) 106470.

- [9] S. Khan, K. Goucher-Lambert, K. Kostas, P. Kaklis, ShipHullGAN: A generic parametric modeller for ship hull design using deep convolutional generative model, *Computer Methods in Applied Mechanics and Engineering* 411 (2023) 116051.
- [10] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in: *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37 of *Proceedings of Machine Learning Research*, PMLR, 2015, pp. 2256–2265.
- [11] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: *Advances in Neural Information Processing Systems*, Vol. 33, 2020, pp. 6840–6851.
- [12] N. J. Bagazinski, F. Ahmed, ShipGen: A diffusion model for parametric ship hull generation with multiple objectives and constraints, *Journal of Marine Science and Engineering* 11 (12) (2023) 2215.
- [13] N. J. Bagazinski, F. Ahmed, C-ShipGen: a conditional guided diffusion model for parametric ship hull design, in: *International Marine Design Conference*, 2024.
- [14] S. Krishnamoorthy, S. M. Mashkaria, A. Grover, Diffusion models for black-box optimization, in: *Proceedings of the 40th International Conference on Machine Learning*, Vol. 202 of *Proceedings of Machine Learning Research*, PMLR, 2023, pp. 17842–17857.
- [15] H. Keramati, M. Sadeghi, R. K. Jaiman, HeatGen: A guided diffusion framework for multiphysics heat sink design optimization, *arXiv preprint arXiv:2511.09578* (2025).
- [16] H. Keramati, P. Kirchen, M. Hannan, R. K. Jaiman, A reward-directed diffusion framework for generative design, *Engineering Applications of Artificial Intelligence* 165 (2026) 113378.
- [17] X. Li, Y. Zhao, C. Wang, G. Scalia, G. Eraslan, S. Nair, T. Biancalani, S. Ji, A. Regev, S. Levine, M. Uehara, Derivative-free guidance in continuous and discrete diffusion models with soft value-based decoding, *arXiv preprint arXiv:2408.08252* (2024).
- [18] P. Del Moral, A. Doucet, A. Jasra, Sequential Monte Carlo samplers, *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68 (3) (2006) 411–436.
- [19] M. D. Shields, J. Zhang, The generalization of latin hypercube sampling, *Reliability Engineering & System Safety* 148 (2016) 96–108.
- [20] L. Biewald, Experiment tracking with weights and biases, software available from wandb.com (2020).
- [21] C. W. Hirt, B. D. Nichols, Volume of fluid (VOF) method for the dynamics of free boundaries, *Journal of Computational Physics* 39 (1) (1981) 201–225.
- [22] F. R. Menter, M. Kuntz, R. Langtry, Ten years of industrial experience with the SST turbulence model, *Turbulence, Heat and Mass Transfer* 4 (1) (2003) 625–632.
- [23] Tokyo 2015 Workshop, KRISO container ship (KCS): EFD data and format of computed results, case 2.1, Tokyo 2015: A Workshop on CFD in Ship Hydrodynamics, National Maritime Research Institute (2015).

- [24] W. J. Kim, S. H. Van, D. H. Kim, Measurement of flows around modern commercial ship models, *Experiments in Fluids* 31 (5) (2001) 567–578.
- [25] O. el Moctar, V. Shigunov, T. Zorn, Duisburg Test Case: Post-panamax container ship for benchmarking, *Ship Technology Research* 59 (3) (2012) 50–64.
- [26] Y. Feng, O. el Moctar, T. E. Schellin, Parametric hull form optimization of containerships for minimum resistance in calm water and in waves, *Journal of Marine Science and Application* 20 (4) (2021) 670–693.
- [27] International Towing Tank Conference, Uncertainty analysis in CFD verification and validation methodology and procedures, Tech. Rep. 7.5-03-01-01, ITTC (2021).
- [28] H. Keramati, M. A. Hannan, P. Kirchen, S. Tousi, J. Riekkinen, M. Barrett, J. Pyra, R. K. Jaiman, A parametric diffusion framework for localized generative optimization of ship hulls, in: *International Conference on Offshore Mechanics and Arctic Engineering*, Vol. 89541, American Society of Mechanical Engineers, 2026, p. V006T08A050.
- [29] N. J. Bagazinski, F. Ahmed, Ship-D: Ship hull dataset for design optimization using machine learning, in: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 87301, American Society of Mechanical Engineers, 2023, p. V03AT03A028.