

Comparing Multi-Index Stochastic Collocation and Radial Basis Function Surrogates for Ship Resistance Uncertainty Quantification

C. Piazzola¹, L. Tamellini¹, R. Pellegrini², R. Broglia², A. Serani², and M. Diez²

¹*Istituto di Matematica Applicata e Tecnologie Informatiche “E. Magenes” (CNR-IMATI), Pavia*

²*Istituto di Ingegneria del Mare (CNR-INM), Roma*

Abstract—This extended abstract is a summary of [4], where a comparison of two methods for the forward uncertainty quantification (UQ) of complex industrial problems is presented. Specifically, the performance of Multi-Index Stochastic Collocation (MISC) and adaptive multi-fidelity Stochastic Radial Basis Functions (SRBF) surrogates is assessed for the UQ of a roll-on/roll-off passengers ferry advancing in calm water and subject to two operational uncertainties, namely the ship speed and draught. The CFD simulations needed by both methods are performed by a multi-grid Reynolds Averaged Navier-Stokes solver.

Keywords—Forward uncertainty quantification, surrogate models, RoPax ferry, Navier-Stokes multi-grid solver.

I. INTRODUCTION

SHIP performance depends on design and operational/environmental parameters. The accurate prediction of significant design metrics (such as resistance and powering requirements; seakeeping, maneuverability, and dynamic stability; structural response and failure) requires prime-principles-based high-fidelity computational tools (e.g., computational fluid/structural dynamics, CFD/CSD), especially for innovative configurations and off-design conditions. These tools are generally computationally expensive, making the exploration of design and operational parameters a technological challenge. Assessing the impact of the uncertainty which affects the problem parameters on the output quantities is a typical uncertainty quantification (UQ) problem.

There is by now a large consensus in the UQ community on the fact that large-scale, industrially relevant UQ analyses can only be performed by leveraging on multi-fidelity methodologies, i.e., methodologies that explore the bulk of the variability of the quantities of interest of the simulation over coarse meshes (or more generally, computationally inexpensive models with e.g. simplified physics), and resort to querying high-fidelity models (e.g., refined meshes or full-physics models) only sparingly, to correct the initial guess produced with the low-fidelity models. Several approaches to this general framework can be conceived, depending on the kind of fidelity models considered and on the strategy used to sample the parameter space.

The objective of this work is to assess and compare the use of two methods from two methodological families for the forward UQ of complex industrial problems. Specifically, we consider the Multi-Index Stochastic Collocation method (MISC, [1], [2]) and an adaptive multi-fidelity Stochastic Radial Basis Functions method (SRBF [3]), belonging to the

family of multi-level/multi-index methods and kernel-based surrogates, respectively.

The performance of MISC and SRBF are compared on the UQ of a roll-on/roll-off passengers (RoPax) ferry sailing in calm water with two operational uncertainties, specifically ship speed and draught, the latter being directly linked to the payload. The estimation of the expected value of the (model-scale) resistance is presented and discussed. Further results on standard deviation and probability density function can be found in the extended version of the work [4]. Both methods need to repeatedly solve the free-surface Navier-Stokes equations (i.e. perform CFD simulations) for different configurations of the operational parameters. The solutions are obtained by the Reynolds Averaged Navier-Stokes (RANS) equations solver χ navis [5]–[7], developed at CNR-INM. Both MISC and SRBF use as fidelity levels the intermediate grids employed by the RANS solver (which is a multi-grid solver): these grids are obtained as isotropic derefinement of an initial coarse grid.

II. FORWARD UNCERTAINTY QUANTIFICATION

Let us consider a single-patch mesh of the computational domain with non-cubic hexahedral elements of the same size¹ and assume that the level of refinement of the mesh along each physical direction can be specified by prescribing some user-defined integer values $\alpha_1, \alpha_2, \alpha_3$; to fix ideas, one can think e.g. that the size of each element of the mesh scales as $2^{-\alpha_1} \times 2^{-\alpha_2} \times 2^{-\alpha_3}$, but this is not necessary. The three values of α_i are collected in a multi-index $\alpha = [\alpha_1, \alpha_2, \alpha_3]$; prescribing the multi-index α thus prescribes the computational mesh to be generated. If this flexibility is not allowed by the mesh-generator (or by the problem itself), it is possible to set $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$, i.e., controlling the mesh-generation by a single integer value α . The same philosophy applies also to multi-patch meshes, where in principle there could be up to three values α_i for each patch. The quantity of interest of the simulation computed over the mesh specified by α is denoted by G_α . For the Navier-Stokes equations, this could be either the full velocity field or a scalar quantity associated to it.

Next, let us assume that the simulation depends on the value of one or more random/uncertain parameters, say N parameters collected in the random vector $\mathbf{y} = [y_1, y_2, \dots, y_N]$, and denote by Γ the set of all possible values of \mathbf{y} , and by $\rho(\mathbf{y})$ the probability density function (PDF) of the random vector \mathbf{y} over Γ . Thus, the primary goal of the forward UQ

¹This assumption can be relaxed, but it is kept for simplicity of exposition.

analysis is to compute an approximation of $\mathbb{E}[G_\alpha]$, i.e., of the expected value of G_α . This quantity is typically computed by a sampling approach, i.e., the partial differential equation (PDE) at hand is solved for several possible values of \mathbf{y} , and the results are averaged with some weights:

$$\mathbb{E}[G_\alpha] \approx \sum_{j=1}^J G_\alpha(\mathbf{y}_j) \omega_j. \quad (1)$$

The simplest averaging scheme is Monte-Carlo, where the values \mathbf{y}_j are chosen at random over Γ (according to the PDF ρ) and $\omega_j = 1/J$.

A. Multi-Index Stochastic Collocation

Roughly speaking, MISC is based on using as quadrature points \mathbf{y}_j in Eq. (1) the union of several Cartesian grids over the domain Γ , that are obtained by tensorization of univariate quadrature rules (which should be chosen according to $\rho(\mathbf{y})$ for computational efficiency). In the RoPax ferry example considered in this work, y_1, y_2, \dots, y_N are uniform and independent random variables and the univariate Clenshaw–Curtis (CC) quadrature is employed. The quadrature points for the K -points univariate CC quadrature rule are

$$t_{j,K} = \cos\left(\frac{(j-1)\pi}{K-1}\right), \quad 1 \leq j \leq K,$$

and the corresponding quadrature weights can be efficiently computed by fast Fourier transform. Similarly to what done with the multi-index α for the physical domain, a multi-index $\beta \in \mathbb{N}^N$ is introduced, that specifies how many points \mathbf{y} will be used to generate any of the above-mentioned Cartesian grids. More specifically, after having introduced the auxiliary function

$$m(0) = 0, \quad m(1) = 1, \quad m(i) = 2^{i-1} + 1 \text{ for } i \geq 2,$$

$m(\beta_1)$ CC points are generated for y_1 , $m(\beta_2)$ CC points are generated for y_2 etc., and the grid obtained by taking the Cartesian product of the N sets of points thus generated is considered. Note that this choice of m guarantees that, given any two multi-indices β_1 and β_2 , the grid obtained using β_1 is contained (*nested*) in the one obtained using β_2 whenever all components of β_1 are smaller or equal than the corresponding components in β_2 . This is clearly useful in the context of adaptive schemes, like the version of MISC that we advocate in this work. The quadrature weight ω_j of each point of the Cartesian grid is immediately obtained by taking the product of the corresponding univariate weights.

The approximation of $\mathbb{E}[G_\alpha]$ computed over this grid with Eq. (1) is denoted as $\mathcal{Q}_{\alpha,\beta}$. Clearly, it would be beneficial to have both multi-indices α and β with large components, say $\alpha = \alpha^*$ and $\beta = \beta^*$, i.e., to average the values of many PDE solutions over a refined computational mesh. However, this is typically unfeasible due to computational costs. One possible remedy is to exploit the fact that a single, highly refined approximation $\mathcal{Q}_{\alpha^*,\beta^*}$ can often² be approximated as a linear combination of many coarser $\mathcal{Q}_{\alpha,\beta}$, where whenever the spatial discretization α is refined, the quadrature level β is kept to a minimum and viceversa (of course, the combined cost of computing the set of coarse discretizations should be

²Whenever $G(\mathbf{y})$ is a smooth function with respect to \mathbf{y} , i.e., roughly speaking, small changes in \mathbf{y} imply small changes in $G(\mathbf{y})$.

smaller than the cost of the highly refined one). This is in a nutshell the idea of MISC. In other words, MISC is a classical multi-level scheme, where most of the statistical variability of G is explored by solving many PDEs with coarse meshes (large³ $\|\beta\|$ with small $\|\alpha\|$) and then the result is corrected with a few PDE solutions with refined meshes (large $\|\alpha\|$ with small $\|\beta\|$). In formula,

$$\mathbb{E}[G_{\alpha^*}] \approx \mathcal{Q}_{\alpha^*,\beta^*} \approx \sum_{[\alpha,\beta] \in \mathcal{I}} c_{[\alpha,\beta]} \mathcal{Q}_{\alpha,\beta}, \quad (2)$$

where $c_{[\alpha,\beta]}$ are real numbers and \mathcal{I} is a collection of feasible discretizations, designed with the purpose just explained. For example, $\mathcal{I} = \{\alpha \in \mathbb{N}^3, \beta \in \mathbb{N}^N : \|\alpha\| + \|\beta\| \leq l\}$, for some integer value l . A suitable set \mathcal{I} can be designed either a-priori, or on-the-run by adaptive algorithms. In this contribution we focus on the latter option, see [4].

Note that the tensor quadrature operator $\mathcal{Q}_{\alpha,\beta}$ can be replaced by a tensor interpolation operator using global Lagrange polynomials collocated at the CC points and a formula analogous to Eq. (2) is obtained. The resulting linear combination of interpolants is used as surrogate model for G_α (i.e. response surface).

B. Stochastic Radial Basis Functions

An alternative methodology to MISC is based on the estimation of the quantities of interest by numerical quadrature applied to a multi-fidelity SRBF surrogate model. Given a training set $\mathcal{T} = \{\mathbf{y}_i, G(\mathbf{y}_i)\}_{i=1}^J$, the RBF prediction (i.e. response surface) is (here) based on a power function kernel and reads

$$f(\mathbf{y}, \tau) = \sum_{j=1}^{\mathcal{K}} w_j \|\mathbf{y} - \mathbf{c}_j\|^\tau, \quad (3)$$

where w_j are unknown coefficients, \mathbf{c}_j are \mathcal{K} points in Γ called RBF centers, and $\tau \sim \text{unif}[\tau_{\min}, \tau_{\max}]$ is a tuning parameter. The prediction $\tilde{G}(\mathbf{y})$ is computed as the expected value (approximated by Monte-Carlo) of f over τ [8]:

$$\tilde{G}(\mathbf{y}) = \mathbb{E}[f(\mathbf{y}, \tau)]_\tau \approx \frac{1}{\Theta} \sum_{i=1}^{\Theta} f(\mathbf{y}, \tau_i), \quad (4)$$

where Θ is the number of Monte-Carlo samples for τ .

The choice of \mathbf{c}_j and the computation of w_j depend on the number of centers \mathcal{K} chosen. If \mathcal{K} is set to be equal to the training set size, \mathbf{c}_j are chosen as the sampling points and w_j are determined by imposing exact interpolation at the training points ($f(\mathbf{y}_i, \tau) = G(\mathbf{y}_i)$), i.e., by solving $\mathbf{A}\mathbf{w} = \mathbf{f}$, with $\mathbf{w} = \{w_j\}$, $a_{ij} = \|\mathbf{y}_i - \mathbf{c}_j\|^\tau$ and $\mathbf{f} = \{G(\mathbf{y}_i)\}$. Otherwise, \mathcal{K} can be chosen to be smaller than the training set size: in this case, the centers \mathbf{c}_j are chosen by k-means clustering and the coefficients w_j are determined through a least-squares regression by solving $\mathbf{w} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{f}$. In this case, multiple \mathcal{K} can be tested, with the optimal \mathcal{K} chosen by minimizing a leave-one-out cross-validation metric. Using the latter approach is beneficial in case the training data are affected by noise.

The uncertainty $U_{\tilde{G}}(\mathbf{y})$ associated with the SRBF prediction is quantified by the 95%-confidence band of the cumulative density function of $f(\mathbf{y}, \tau)$, evaluated using a Monte-Carlo sampling over τ .

³ $\|\cdot\|$ denotes the Euclidean norm.

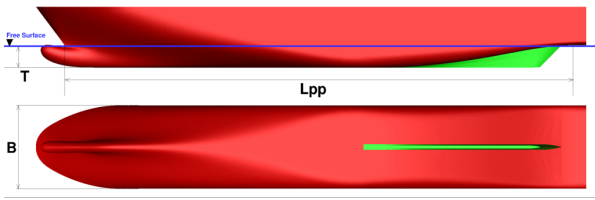


Fig. 1. RoPax ferry: hull form.

The multi-fidelity approximation is adaptively built as follows, see [4]. Extending the definition of the surrogate training set to an arbitrary number of M fidelity levels as $\{\mathcal{T}_\alpha\}_{\alpha=1}^M = \{\mathbf{y}_j, G_\alpha(\mathbf{y}_j)\}_{j=1}^{J_\alpha}$, the multi-fidelity approximation $\hat{G}_M(\mathbf{y})$ of $G_M(\mathbf{y})$ reads

$$\hat{G}_M(\mathbf{y}) \approx \tilde{G}_1(\mathbf{y}) + \sum_{i=1}^{M-1} \tilde{\varepsilon}_i(\mathbf{y}), \quad (5)$$

where $\tilde{\varepsilon}_i(\mathbf{y})$ is the inter-level error surrogate model with an associate training set $\mathcal{E}_i = \{(\mathbf{y}, G_{i+1}(\mathbf{y}) - \hat{G}_i(\mathbf{y})) \mid \mathbf{y} \in \mathcal{T}_{i+1} \cap \mathcal{T}_i\}$. Note that Eq. (5) does not require nested training sets.

The uncertainty of the final metamodel \hat{G}_M , $U_{\hat{G}_M}$, can be assessed with an ‘‘ANOVA-like’’ formula, inspired by the decomposition in Eq. (5). Upon having evaluated $U_{\hat{G}_M}$, the multi-fidelity surrogate is updated adding a new training point \mathbf{y}^* where the $U_{\hat{G}_M}$ is the largest (single-objective maximization problem). The fidelity to be used to this end is again chosen by a maximization criterion: more specifically, the uncertainties of each fidelity are evaluated at \mathbf{y}^* , and rescaled by the corresponding computational cost. The fidelity with the largest ratio is chosen, and the corresponding model is evaluated, see [9].

Finally, numerical quadrature is used on the SRBF surrogate model to estimate the expected value of the quantity of interest $\mathbb{E}[\hat{G}_M]$. More specifically, $\mathbb{E}[\hat{G}_M]$ is approximated using a multivariate midpoint rule, with a full-factorial sampling over the SRBF prediction with S samples, as

$$\mathbb{E}[G_M] \approx \mathbb{E}[\hat{G}_M] \approx \frac{1}{S} \sum_{j=1}^S \hat{G}_M(\mathbf{y}_j). \quad (6)$$

III. PROBLEM DESCRIPTION

The problem addressed in this work is the forward UQ analysis of the model-scale resistance (R_T) of a RoPax ferry in straight ahead advancement, subject to two operational uncertainties $\mathbf{y} = [U, T]$, namely the advancement speed (U) and the draught (T), uniformly distributed within the ranges in Tab. I. The RoPax ferry is characterized by a length between perpendicular at nominal draught (L_{PP}) of 162.85 m and a block coefficient $C_B = 0.5677$ (see Fig. 1). The parametric geometry of the RoPax is produced with the computer-aided design environment integrated in the CAESES[®] software, developed by FRIENDSHIP SYSTEMS AG, and made available in the framework of the H2020 EU Project Holiship. The analysis is performed at model scale with a scale factor equal to 27.14. The main dimensions and the operative conditions are summarized in Tab. I.

The hydrodynamic performance of the RoPax ferry is assessed for each $[U, T]$ configuration by the RANS code

χ navis developed at CNR-INM [5]–[7]. It is based on a finite volume scheme, with variables collocated at the cell centers. Turbulent stresses are taken into account by the Boussinesq hypothesis, with the Spalart-Allmaras turbulence model. Free-surface effects are taken into account by a single-phase level-set algorithm. Wall-functions are not adopted, therefore the wall distance $y^+ = 1$ is ensured on the wall. The numerical solutions are computed by means of a full multi-grid–full approximation scheme (FMG–FAS), with four grid levels, each obtained from the next finer grid with a coarsening ratio equal to 2, along each curvilinear direction. In the FMG–FAS approximation procedure, the solution is computed on the coarsest grid level first. Secondly, it is approximated on the next finer grid and the solution is iterated by exploiting all the coarser grid levels available with a V-Cycle. The process is repeated up to the finest grid level. For the present UQ problem all the four grid levels are used; to note, the number of grid volumes ranges from 5.5M for the finest grid, down to 11K for the coarsest one.

Based on the grid refinement ratio chosen, a normalized computational cost for the α -th grid level is estimated as:

$$\text{cost}(\alpha) = 8^{\alpha-1} \quad (7)$$

with $\alpha = 1, \dots, 4$. In the FMG-FAS scheme the computation on the α -th grid level involves computations on all the coarser meshes. However, with the estimation in Eq. (7), only the cost of the highest-fidelity level samples is taken into account, i.e. the computations on the coarser grids are considered negligible.

IV. NUMERICAL RESULTS

The performance of MISC and SRBF is assessed in terms of the estimation of the expected value of R_T . Fig. 2 shows the convergence of the expected value of R_T versus the computational cost. MISC and SRBF converge towards similar estimates. MISC achieves a good estimate already with low computational cost, whereas at later iterations, corresponding to computational cost above 1000, its behavior worsens; conversely SRBF gives a more stable estimate.

The response surfaces for the resistance obtained by the two methods are shown in Fig. 3. The response surface of MISC is very irregular and this can be attributed to the numerical noise which affects the CFD simulations (due to the fact that the solver is an iterative method which stops as soon as a prescribed tolerance is met), in particular the ones on the coarsest grid. As the overall idea of MISC is to solve most PDEs on the less expensive grids and building the

TABLE I
MAIN GEOMETRICAL DETAILS AND OPERATIVE CONDITIONS OF THE
ROPAX FERRY (MODEL SCALE 1 : 27.14).

Description	Full Scale	Model Scale
Length b. perp. (L_{PP})	162.85 m	6.0 m
Beam (B)	29.84 m	1.0993 m
Block coefficient (C_B)	0.5677	0.5677
Nominal disp. (∇)	19584.04 m ³	0.9996 m ³
Nominal draught (T_n)	7.10 m	0.261660 m
Draught range (T)	[7.812, 6.391] m	[0.236, 0.288] m
Speed range (U)	[6.173, 13.376] m/s	[1.185, 2.567] m/s
Froude range (Fr)	[0.154, 0.335]	[0.154, 0.335]
Reynolds range (Re)	[0.908, 1.968] · 10 ⁹	[0.642, 1.392] · 10 ⁷

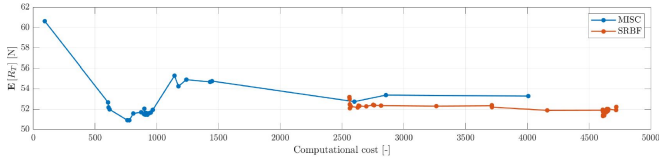


Fig. 2. Convergence of the expected value of R_T versus computational cost. For the sake of readability, we plot the results for MISC starting with the 4th iteration. The first three iterations correspond to computational cost 1, 12, and 14, and give poor results.

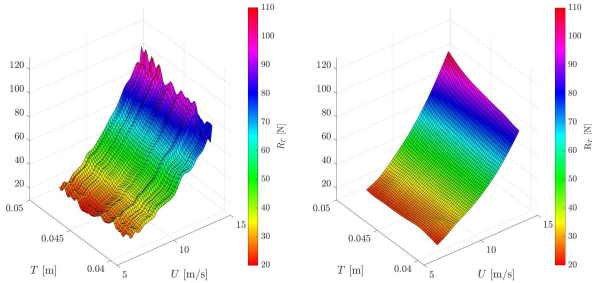


Fig. 3. Response surfaces at the final iteration: MISC (left), SRBF (right).

surrogate model with Lagrangian (hence exact) interpolation, the presence of numerical noise turns out to be problematic for this method. On the contrary, SRBF suffers the numerical noise of the CFD outputs until the exact interpolation is imposed (i.e. iteration 8 - figure not shown for brevity, see [4]), whereas for later iterations the use of regression improves the quality of the surrogates by smoothing the response surface and filtering out the numerical noise.

Fig. 4 displays the points in the parameter space selected by MISC and SRBF at the final iteration. At the beginning the MISC algorithm explores more in advancement speed direction (figure not shown for brevity, see [4]), suggesting a stronger dependence of the quantity of interest on the advancement speed rather than on the draught. Points exploring the direction of the draught are added only at later iterations. Until iteration 8 (exact interpolation only), the sampling performed by SRBF explores the domain extrema and starts to cluster samples in two zones among $U = [7, 8.8]$ m/s (figure not shown for brevity, see [4]), since the numerical noise negatively affects the prediction uncertainty of the interpolating SRBF. Switching to the least-squares approximation and filtering out the noise prevents an excessive clusterization of the samples at the successive iterations. Indeed, the SRBF samples are fairly spread over the domain.

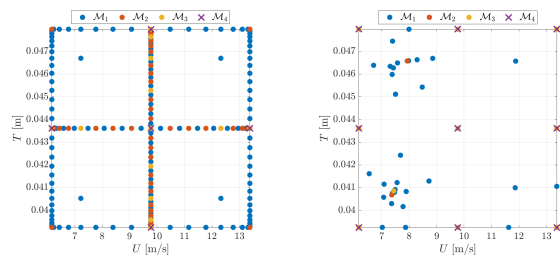


Fig. 4. Points in parameter space at the final iteration: MISC (left), SRBF (right). Note that every point required on grid \mathcal{M}_i is required also on all the grids with lower refinement level.

V. CONCLUSIONS

The results suggest that both methods give fairly accurate results at reasonable computational cost. MISC could be preferred when only limited data sets are available, whereas for larger data sets a slight preference may be cast for SRBF, due to its robustness to noise. Future research will address more complex test cases (larger number of uncertain parameters and more realistic conditions, such as regular/irregular waves) possibly validating the results against benchmark values.

ACKNOWLEDGMENT

CNR-INM is grateful to Dr. Woei-Min Lin, Dr. Elena McCarthy, and Dr. Salahuddin Ahmed of the Office of Naval Research and Office of Naval Research Global, for their support through NICOP grant N62909-18-1-2033. Dr. Riccardo Pellegrini is partially supported through CNR-INM project OPTIMAE. The HOLISHIP project (HOListic optimisation of SHIP design and operation for life cycle, www.holiship.eu) is also acknowledged, funded by the European Union's Horizon 2020 research and innovation program under grant agreement N. 689074. Lorenzo Tamellini and Chiara Piazzola have been supported by the PRIN 2017 project 201752HKH8 "Numerical Analysis for Full and Reduced Order Methods for the efficient and accurate solution of complex systems governed by Partial Differential Equations (NA-FROM-PDEs)". Lorenzo Tamellini also acknowledges the support of GNCS-INdAM (Gruppo Nazionale Calcolo Scientifico - Istituto Nazionale di Alta Matematica).

REFERENCES

- [1] Beck, J., Tamellini, L., and Tempone, R., "IGA-based multi-index stochastic collocation for random PDEs on arbitrary domains," *Comput. Methods Appl. Mech. Engrg.*, Vol. 351, 2019, pp. 330–350.
- [2] Jakeman, J. D., Eldred, M., Geraci, G., and Gorodetsky, A., "Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis," *Internat. J. Numer. Methods Engrg.*, Vol. 121, No. 6, 2020, pp. 1314–1343.
- [3] Wackers, J., Visonneau, M., Pellegrini, R., Ficini, S., Serani, A., and Diez, M., "Adaptive N-fidelity metamodels for noisy CFD Data," *AIAA AVIATION 2020 FORUM*.
- [4] Piazzola, C., Tamellini, L., Pellegrini, R., Broglio, R., Serani, A., and Diez, M., "Uncertainty quantification of ship resistance via multi-index stochastic collocation and radial basis function surrogates: a comparison," *AIAA AVIATION 2020 FORUM*.
- [5] Di Mascio, A., Broglio, R., and Muscarì, R., "On the application of the single-phase level set method to naval hydrodynamic flows," *Comput. & fluids*, Vol. 36, No. 5, 2007, pp. 868–886.
- [6] Di Mascio, A., Broglio, R., and Muscarì, R., "Prediction of hydrodynamic coefficients of ship hulls by high-order Godunov-type methods," *J. Mar. Sci. Technol.*, Vol. 14, No. 1, 2009, pp. 19–29.
- [7] Broglio, R., and Durante, D., "Accurate prediction of complex free surface flow around a high speed craft using a single-phase level set method," *Comput. Mech.*, Vol. 62, No. 3, 2018, pp. 421–437.
- [8] Volpi, S., Diez, M., Gaul, N. J., Song, H., Iemma, U., Choi, K. K., Campana, E. F., and Stern, F., "Development and validation of a dynamic metamodel based on stochastic radial basis functions and uncertainty quantification," *Struct. Multidiscip. Optim.*, Vol. 51, No. 2, 2015, pp. 347–368.
- [9] Serani, A., Pellegrini, R., Wackers, J., Jeanson, C.-J., Queutey, P., Visonneau, M., and Diez, M., "Adaptive multi-fidelity sampling for CFD-based optimization via radial basis functions metamodel," *Int. J. Comput. Fluid Dyn.*, Vol. 33, No. 6-7, 2019, pp. 237–255.