

Assessing the Interplay of Shape and Physical Parameters by Nonlinear Dimensionality Reduction Methods

A. Serani,¹ D. D'Agostino,^{1,2} E. F. Campana,¹ and M. Diez,¹

(¹CNR-INM, Natl. Research Council–Institute of Marine Engineering, Italy,
²Dept. of Computer, Control, and Management Engineering “A. Ruberti”, Sapienza
University of Rome, Italy)

ABSTRACT

The paper presents an exploratory study on the application to ship hydrodynamics of nonlinear design-space dimensionality reduction methods, assessing the interaction of shape and physical parameters. Nonlinear extensions of the principal component analysis (PCA) are applied, namely local and kernel PCA. An artificial neural network approach, specifically a deep autoencoder (DAE) method is also applied and compared to PCA-based approaches. The data set under investigation is formed by the results of 9,000 potential flow simulations coming from an extensive exploration of a 27-dimensional design space, associated to a shape-optimization problem of the DTMB 5415 model in calm water at 18kn (Froude number, $Fr = 0.25$). Data include three heterogeneous distributed and suitably discretized parameters (shape modification vector, pressure distribution on the hull, and wave elevation pattern) and one lumped parameter (wave resistance coefficient), for a total of $5,101 \times 9,000$ elements. The reduced-dimensionality representation of shape and physical parameters is set to provide a normalized mean squared error smaller than 5%. The standard PCA meets the requirement using 19 principal components/parameters. Local and kernel PCA provide the most promising compression capability with 14 parameters required by the reduced-dimensionality parametrizations. DAE achieve the same error with 17 components, indicating significant nonlinear interactions in the data structure of shape and physical parameters. Although the focus of the current work is on design-space dimensionality reduction, the formulation goes beyond shape optimization and can be applied to large sets of heterogeneous physical data from simulations, experiments, and real operation measurements.

INTRODUCTION

The simulation-based design (SBD) analysis and optimization paradigm has demonstrated the capability of supporting the design decision process, not only provid-

ing large sets of design options but also exploring operational spaces by assessing the hydrodynamic and structural performance for a large number of operating and environmental conditions. The recent development of high performance computing (HPC) systems has driven the SBD towards integration with global optimization (GO) algorithms and uncertainty quantification (UQ) methods, moving the SBD paradigm to automatic deterministic and stochastic SBD optimization (SBDO, Diez et al. 2018a,b) possibly aiming at global solutions to the design problem. In shape design, SBDO consists of three main elements: (i) a deterministic and/or stochastic simulation tool (integrating physics-based solvers with UQ), (ii) an optimization algorithm, and (iii) a shape modification tool (see Figure 1, right box).

Even if efficient GO algorithms have been proposed (Jones et al., 1993; Campana et al., 2009; Serani and Diez, 2017a) and applied with success to SBDO (Campana et al., 2015; Serani et al., 2016; Serani and Diez, 2017b), finding a potentially global optimal solution within reasonable computational time/cost remains a critical issue and a technological challenge. Additionally, UQ of complex applications is computationally very demanding, especially if high-order statistical moments and/or quantiles need to be assessed as in robust and reliability-based design optimization. Both GO and UQ are affected by the *curse of dimensionality* as the algorithms' complexity and computational cost rapidly increase with the problem dimension. This is generally also true if metamodels are applied. Therefore, the assessment and breakdown of the design-space dimensionality are key elements for the efficiency and affordability of the SBDO (Diez et al., 2015).

In order to solve complex optimization problems, on-line linear design-space dimensionality reduction techniques have been developed, requiring the evaluation of the objective function or its gradient. Principal component analysis (PCA) or proper orthogonal decomposition (POD) methods have been applied for local reduced-dimensionality representations of feasible design regions (Raghavan et al., 2013). A PCA/POD-

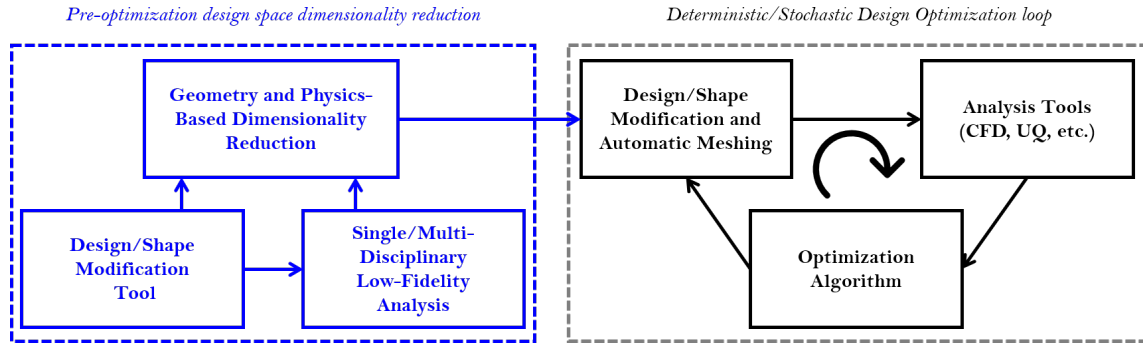


Figure 1: SBDO scheme, including pre-optimization design-space dimensionality reduction (Diez et al., 2016).

based approach is used in the active subspace method (Lukaczyk et al. 2014) to discover and exploit low-dimensional monotonic trends in the objective function, based on the evaluation of its gradient. This type of methods improve the optimization efficiency by basis rotation and/or dimensionality reduction. Nevertheless, they do not provide an assessment of the design space and associated shape parametrization before optimization is performed or objective function and/or gradient are evaluated. Moreover, if gradients are not directly provided (as in the case of black-box tools) they might be inaccurate due to noise and/or residuals affecting the approximation of derivatives by finite differences. Finally, these methods are local in nature and their extension to global optimization is not trivial nor straightforward.

Off-line linear models have been developed with focus on design-space variability and dimensionality reduction for efficient optimization procedures. A method based on the Karhunen-Loève expansion (KLE, equivalent to POD) has been formulated in (Diez et al., 2015) for the assessment of the shape modification variability and the definition of a reduced-dimensionality global model of the shape modification vector. No objective function evaluation nor gradient is required by the method. The KLE is applied to the continuous shape modification vector, requiring the solution of a Fredholm integral equation of the second kind. Once the equation is discretized, the problem reduces to the principal component analysis (PCA) of discrete geometrical data. These methods improve the shape optimization efficiency by reparametrization and dimensionality reduction, providing the assessment of the design space and the shape parametrization before optimization and/or performance analysis are carried out. The assessment is based on the geometric variability associated to the design space, making the method fully off-line and computationally very efficient and attractive (no simulations are required). Nevertheless, significant physical phenomena induced by small shape modifications (such as transitions, separations, etc.) may be overlooked as no phys-

ical information is processed by the method. Furthermore, linear methods such as PCA may not be efficient when a complex nonlinear relationship between design variables are involved.

An extension of the off-line design-space dimensionality reduction method to a combined shape and physics based formulation was introduced and developed in earlier research (Diez et al., 2016; Serani et al., 2017; Serani and Diez, 2018). The extension improved the effectiveness of the dimensionality reduction, bringing physics based information (provided by low-fidelity computations) into the variability breakdown analysis. The concept is summarized in Figure 1 (left box).

In order to address data with nonlinear structures, nonlinear dimensionality reduction methods have been developed and investigated. Among others, local PCA (LPCA) divides the initial design space in k clusters and the PCA is applied to each of them, assuming each cluster having approximately a linear structure. LPCA techniques (Kambhatla and Leen, 1997) may be differentiated on the basis of the clustering method which may follow for instance k -means (Lloyd, 1982) or spectral approaches (Von Luxburg, 2007). Kernel PCA (KPCA) solves the PCA in a new space (called feature space) using kernel methods (Schölkopf et al., 1998). Artificial neural networks have been also used to reduce data dimensionality (Hinton and Salakhutdinov, 2006), by performing both encoder and decoder tasks. Specifically, a deep autoencoder (DAE) is an architecture of a deep feedforward neural network that both compress (encoding) and reconstruct (decoding) the original data. Autoencoders or autoassociative neural networks have been studied and proposed as nonlinear extension of PCA by several researchers (Bourlard and Kamp, 1988; Kramer, 1991; Oja, 1991). They have been applied and found more effective than their linear counterparts for image compression applications (DeMers and Cottrell, 1993; Hinton and Salakhutdinov, 2006). Preliminary studies on the application of nonlinear design-space dimensionality reduction methods in ship hydrodynamics have been pre-

sented by D’Agostino et al. (2017, 2018a,b) for a USS Arleigh Burk-class destroyer, namely the DTMB 5415 model, an early and open public version of the DDG-51.

The objective of the present work is the application of nonlinear methods to design-space dimensionality reduction in shape optimization for ship hydrodynamics, based on the interaction of shape and physical parameters evaluated by low-fidelity solvers. Nonlinear methods include LPCA, KPCA, and DAE. The efficiency and effectiveness of these methods are assessed considering their compression capability and associated reconstruction error compared to standard (linear) PCA. Current formulations and methods go beyond design-space dimensionality reduction for shape optimization and can be extended to large sets of heterogeneous physical data from simulations, experiments, and real operation measurements.

The methods are demonstrated for the DTMB 5415 model in calm water at 18kn (Froude number, $Fr = 0.25$). The data set under investigation is formed by the results of 9,000 potential flow simulations exploring extensively a 27-dimensional design space. Data include three heterogeneous distributed and suitably discretized parameters (geometry modification vector, pressure distribution on the hull, and wave elevation pattern) and one lumped parameter (wave resistance coefficient), for a total of $5,101 \times 9,000$ elements. The reduced-dimensionality representation of shape and physical parameters is set to provide a mean squared error smaller than 5%, normalized with the overall data variance.

DIMENSIONALITY-REDUCTION FORMULATIONS AND METHODS

General definitions and assumptions are presented in the following, along with the shape-based formulation for design-space dimensionality reduction, its extension to combined shape- and physics-based formulation, and the solution of data reduction by principal component analysis (PCA), local and kernel PCA (LPCA and KPCA), and deep autoencoder (DAE).

Shape-Based Formulation

Consider a geometric domain \mathcal{G} (which identifies the initial or parent shape) and a set of coordinates $\mathbf{x} \in \mathcal{G} \subset \mathbb{R}^n$ with $n = 1, 2, 3$. Assume that $\mathbf{u} \in \mathcal{U} \subset \mathbb{R}^M$ is the design-variable vector, which defines a continuous shape modification vector $\delta(\mathbf{x}, \mathbf{u}) \in \mathbb{R}^m$ with $m = 1, 2, 3$ (with m not necessarily equal to n). Consider \mathbf{u} as a random variable, with associated probability density function $p(\mathbf{u})$.

The aim of the dimensionality reduction is to identify a reduced-dimensionality representation of the

shape modification vector $\hat{\delta}(\mathbf{x}, \alpha)$, for which its shape modification depends on a new reduced order design variable $\alpha \in \mathcal{A} \subset \mathbb{R}^N$ with $N < M$. The reconstruction vector $\hat{\delta}(\mathbf{x}, \alpha)$ is estimated during a process of encoding/decoding by the dimensionality reduction methods: the original shape modification vector $\delta(\mathbf{x}, \mathbf{u})$ is encoded in a low dimensional latent space defining the new design variable α . The decoding process reconstruct the original shape modification vector as $\hat{\delta}(\mathbf{x}, \alpha)$ for which its modification on the original domain depends on the vector $\alpha(\mathbf{u})$ learned by the dimensionality reduction methods. Figure 2 shows an example with $n = 1$ and $m = 2$.

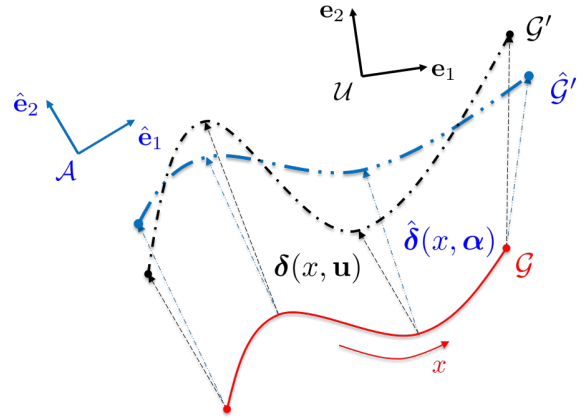


Figure 2: Scheme and notation for the current formulation, example for $n = 1$ and $m = 2$.

A convenient metric to evaluate the goodness of fit of $\hat{\delta}(\mathbf{x}, \alpha)$ respect to $\delta(\mathbf{x}, \mathbf{u})$ is the mean squared error (MSE) normalized to the total original geometric variance (σ^2) as

$$\begin{aligned} \text{NMSE} &= \frac{\text{MSE}}{\sigma^2} = \\ &= \frac{\iint_{\mathcal{U} \times \mathcal{A}, \mathcal{G}} \|\hat{\delta}(\mathbf{x}, \alpha) - \delta(\mathbf{x}, \mathbf{u})\|^2 p(\mathbf{u}, \alpha) dx du d\alpha}{\iint_{\mathcal{U}, \mathcal{G}} \|\delta(\mathbf{x}, \mathbf{u})\|^2 p(\mathbf{u}) dx du}, \end{aligned} \quad (1)$$

where $p(\mathbf{u}, \alpha)$ is an unknown joint probability distribution over the product space $\mathcal{U} \times \mathcal{A}$.

Discretizing \mathcal{G} by a number of Q elements of equal measure $\Delta\mathcal{G}$, assuming for the sake of simplicity and without loss of generality a centered data set (i.e., with zero mean value) with $\Delta\mathcal{G} = 1$, and sampling \mathcal{U} by a statistically convergent number of Monte Carlo realizations S , so that $\{\mathbf{u}_k\}_{k=1}^S \sim p(\mathbf{u})$, the spatial discretization $\mathbf{d}(\mathbf{u})_k$ of $\delta(\mathbf{x}, \mathbf{u}_k)$ are organized in a $[S \times L]$ data matrix as

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}(\mathbf{u})_{k=1}^\top \\ \vdots \\ \mathbf{d}(\mathbf{u})_{k=S}^\top \end{bmatrix}, \quad (2)$$

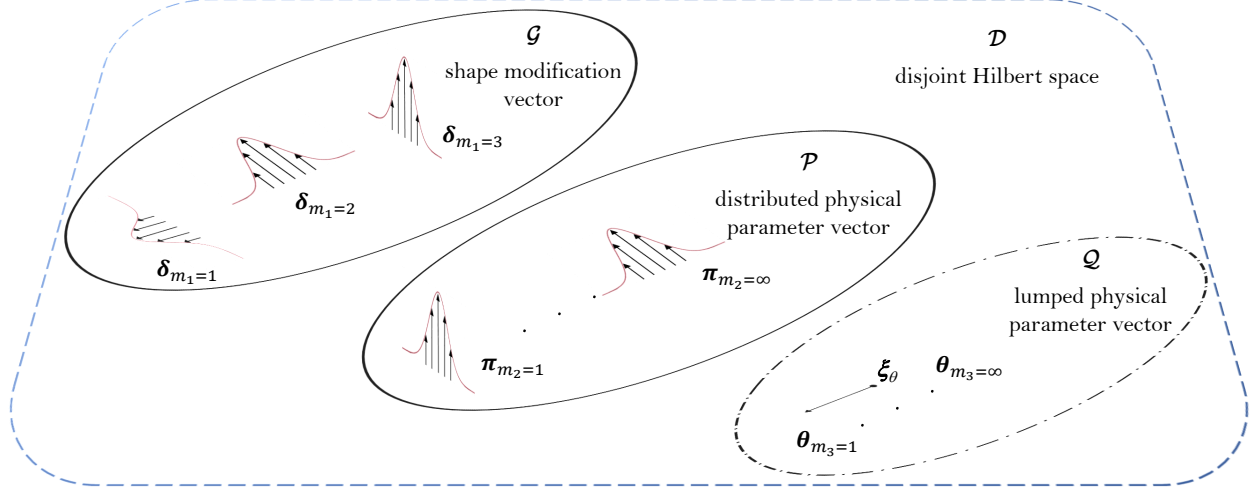


Figure 3: Domains for shape modification vector, distributed physical parameter vector, and lumped (or global) physical parameter vector.

where $L = mQ$. The expectation in Eq. 1 can be approximated by evaluating the MSE as

$$\text{MSE} = \frac{1}{S} \sum_{k=1}^S \|\hat{\mathbf{d}}(\boldsymbol{\alpha})_k - \mathbf{d}(\mathbf{u})_k\|^2 \quad (3)$$

and the total geometric variance σ^2 as

$$\sigma^2 = \frac{1}{S} \sum_{k=1}^S \|\mathbf{d}(\mathbf{u})_k\|^2. \quad (4)$$

Combining Eqs. 3 and 4, the discretized form of Eq. 1 can be expressed as

$$\text{NMSE} = \frac{\text{MSE}}{\sigma^2} = \frac{\sum_{k=1}^S \|\hat{\mathbf{d}}(\boldsymbol{\alpha})_k - \mathbf{d}(\mathbf{u})_k\|^2}{\sum_{k=1}^S \|\mathbf{d}(\mathbf{u})_k\|^2} \quad (5)$$

Details of formulation, equations, and numerical discretization can be found in Diez et al. (2015).

Extension to Combined Shape- and Physics-Based Formulation

Consider the shape modification vector $\boldsymbol{\delta} \in \mathbb{R}^{m_1}$, $m_1 = 1, \dots, 3$, along with a distributed physical parameter vector $\boldsymbol{\pi} \in \mathbb{R}^{m_2}$, $m_2 = 1, \dots, \infty$ (representing, e.g., velocity, pressure distribution, wave elevation, etc.), and a lumped (or global) physical parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{m_3}$, $m_3 = 1, \dots, \infty$ (representing, e.g., ship resistance, motion RMS, etc.).

For the sake of simplicity, consider one set of coordinates $\mathbf{x} \in \mathbb{R}^n$, and assume \mathcal{G} , \mathcal{P} , and \mathcal{Q} as the domain of $\boldsymbol{\delta}$, $\boldsymbol{\pi}$, and $\boldsymbol{\theta}$ respectively, as schematized in Figure 3. Note that \mathcal{Q} has a null measure and corresponds to

an arbitrary point \mathbf{x}_θ where the lumped physical parameter vector is virtually defined. Also note that generally $\mathcal{D} = \mathcal{G} \cup \mathcal{P} \cup \mathcal{Q}$ is not simply connected.

Consider a combined vector $\boldsymbol{\gamma} \in \mathbb{R}^m$ with $m = \max\{m_1, m_2, m_3\}$

$$\boldsymbol{\gamma}(\mathbf{x}, \mathbf{u}) \in \mathcal{D} = \begin{cases} \boldsymbol{\delta}(\mathbf{x}, \mathbf{u}) & \text{if } \mathbf{x} \in \mathcal{G} \\ \boldsymbol{\pi}(\mathbf{x}, \mathbf{u}) & \text{if } \mathbf{x} \in \mathcal{P} \\ \boldsymbol{\theta}(\mathbf{x}, \mathbf{u}) & \text{if } \mathbf{x} \in \mathcal{Q} \end{cases}, \quad (6)$$

The aim of the dimensionality reduction is to identify a reduced-dimensionality representation of the vector $\hat{\boldsymbol{\gamma}}(\mathbf{x}, \boldsymbol{\alpha})$, able to minimize the MSE:

$$\begin{aligned} \text{MSE} &= \quad (7) \\ &= \iint_{\mathcal{U} \times \mathcal{A}, \mathcal{D}} \|\hat{\boldsymbol{\gamma}}(\mathbf{x}, \boldsymbol{\alpha}) - \boldsymbol{\gamma}(\mathbf{x}, \mathbf{u})\|^2 p(\mathbf{u}, \boldsymbol{\alpha}) d\mathbf{x} d\mathbf{u} d\boldsymbol{\alpha}. \end{aligned}$$

The MSE and the data matrix are recast in the same form of Eqs. 3 and 2, provided that \mathbf{d}_k represent the discretized values of $\boldsymbol{\gamma}(\mathbf{x})$. Furthermore, the data matrix is assumed centered and normalized such that each domain (shape, distributed/lumped physical parameters) provides an associated data variance equal to one. Details of formulation, equations, and numerical discretization can be found in Diez et al. (2016) and Serani et al. (2017).

Principal Component Analysis

PCA allows to reduce the dimensionality of the data matrix by representation in a linear subspace defined by the eigenvectors \mathbf{z} of the $[L \times L]$ sample covariance matrix \mathbf{C} :

$$\mathbf{C} = \frac{1}{S} \mathbf{D}^\top \mathbf{D}, \quad (8)$$

Thus, PCA reduces to the solution of the eigenproblem

$$\mathbf{C}\mathbf{z} = \lambda\mathbf{z}. \quad (9)$$

The eigenvectors $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^S$ have the properties to maximize the variance (σ^2) associated to the linear representation and minimize the associated MSE (Hotelling, 1933; Pearson, 1901). Moreover, the eigenvalues λ represent the variance resolved along the eigenvectors. Given this property, a subset of N eigenvectors is used to define a reduced-dimensionality representation of the shape modification vector as

$$\hat{\mathbf{d}} = \hat{\mathbf{Z}}\boldsymbol{\alpha} \quad \text{with} \quad \boldsymbol{\alpha} = \mathbf{Z}^\top \mathbf{d} \quad (10)$$

where $\hat{\mathbf{Z}}$ has dimension $[L \times N]$ and is composed by the first N largest-variance eigenvectors.

For practical purposes, Eq. 9 may be solved by the singular value decomposition of the matrix \mathbf{D} .

Local Principal Component Analysis

LPCA is a nonlinear extension of PCA. LPCA performs a PCA for each disjoint region of the input space \mathcal{I} . If local regions are small enough the associated data manifold will not curve much over the extent of the region and the linear model is assumed to be a good fit (Kambhatla and Leen, 1997).

The first step in LPCA is clustering the data in k sets, such that $\mathbf{D} = \{\mathbf{D}_i\}_{i=1}^k$. Here, the k -means clustering algorithm (Lloyd, 1982) is used as described in Alg. 1.

Algorithm 1 k -means clustering algorithm

Require: A number of k random centroids as representative points of each cluster $\mathbf{D}_i \forall i = 1, \dots, k$.

- 1: **repeat**
 - 2: Assign each point \mathbf{d}_j to the nearest centroid $\boldsymbol{\mu}_i$ using the Euclidean distance as similarity measure.
 - 3: Update the centroids according to: $\boldsymbol{\mu}_i = \frac{1}{|\mathbf{D}_i|} \sum_{\mathbf{d}_j \in \mathbf{D}_i} \mathbf{d}_j$
 - 4: **until** $\boldsymbol{\mu}_i \forall i = 1, \dots, k$ remains unchanged
-

After k clusters are defined, k PCA eigenproblems are solved

$$\mathbf{C}_i \mathbf{z}_i = \lambda_i \mathbf{z}_i \quad \forall i = 1, \dots, k. \quad (11)$$

LPCA results are highly dependent on the clustering method and the number of clusters. The number of clusters needs to be defined carefully to avoid increasing the computational cost and data overfitting.

Kernel Principal Component Analysis

The KPCA method (Schölkopf et al., 1998) finds directions of maximum variance in a higher (possibly infinite) dimensional feature space \mathcal{F} , mapping data points

from the input space \mathcal{I} by a (possibly) nonlinear function $\Phi : \mathcal{I} \rightarrow \mathcal{F}$

$$\mathbf{d}_k \rightarrow \Phi(\mathbf{d}_k), \quad \forall k = 1, \dots, S. \quad (12)$$

The PCA is computed in the feature space \mathcal{F} . Assuming $\sum_k \Phi(\mathbf{d}_k) = 0$, the kernel principal component $\{\mathbf{z}_p\}_{p=1}^P$ can be found solving the eigenproblem

$$\boldsymbol{\Sigma}_\Phi \mathbf{z}_p = \lambda_p \mathbf{z}_p, \quad (13)$$

where $\boldsymbol{\Sigma}_\Phi$ is the $[P \times P]$ covariance matrix in the feature space \mathcal{F} :

$$\boldsymbol{\Sigma}_\Phi = \frac{1}{S} \sum_{k=1}^S \Phi(\mathbf{d}_k) \Phi(\mathbf{d}_k)^\top. \quad (14)$$

Defining $K(\mathbf{d}_i, \mathbf{d}_k) = \Phi(\mathbf{d}_i)^\top \Phi(\mathbf{d}_k)$ and \mathbf{z}_p as

$$\mathbf{z}_p = \sum_{k=1}^S c_{pk} \Phi(\mathbf{d}_k), \quad (15)$$

Eq. 13 can be rewritten as

$$\mathbf{K} \mathbf{c}_p = \lambda_p S \mathbf{c}_p, \quad (16)$$

where \mathbf{K} is the symmetric and positive-semidefinite $[S \times S]$ kernel matrix, with $\mathbf{K}_{ik} = K(\mathbf{d}_i, \mathbf{d}_k)$. The length of the S -component vector \mathbf{c}_p is chosen such that $\mathbf{z}_p^\top \mathbf{z}_p = \lambda_p S \mathbf{c}_p^\top \mathbf{c}_p = 1$. Once the eigenproblem of Eq. 16 is solved, the new parametrization can be found projecting $\Phi(\mathbf{d})$ on \mathbf{z}_p as

$$\begin{aligned} \boldsymbol{\alpha} = \Phi(\mathbf{d}) \mathbf{z}_p &= \sum_{k=1}^S c_{pk} \Phi(\mathbf{d})^\top \Phi(\mathbf{d}_k) \\ &= \sum_{k=1}^S c_{pk} K(\mathbf{d}, \mathbf{d}_k). \end{aligned} \quad (17)$$

The reconstruction of the original data from the feature space \mathcal{F} in KPCA is more problematic than PCA. Here, the approximate pre-images technique proposed by Bakır et al. (2004) is used.

Deep Autoencoders

An autoencoder is a feedforward artificial neural network that performs two main tasks: (i) an encoder function \mathcal{E} maps the input data \mathbf{d}_k into compressed data $\boldsymbol{\alpha}_k$; (ii) a decoder function \mathcal{D} maps from the compressed data $\boldsymbol{\alpha}_k$ back to $\hat{\mathbf{d}}_k$. The overall operation is performed setting the same number of neurons (L) in the input and output layer. The hidden layer is set to have $N < M$ neurons and is responsible for the data compression.

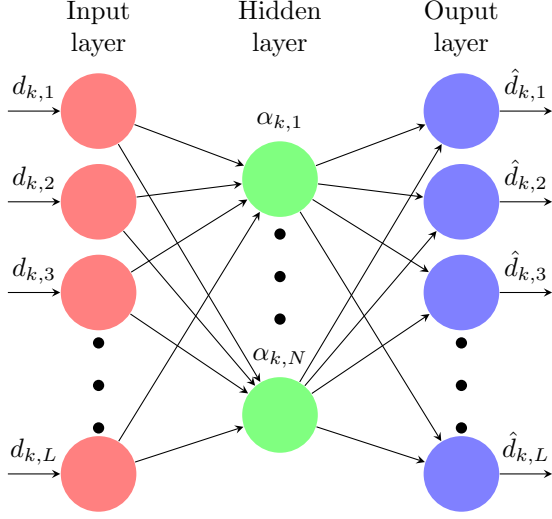


Figure 5: A single hidden layer autoencoder.

Consider a single hidden layer autoencoder (see Figure 5) and assume no bias vector. New design variables α_k can be expressed as

$$\alpha_k = \mathcal{E}(\mathbf{H}_{(1)} \mathbf{d}_k), \quad (18)$$

where \mathbf{H} is a weight matrix and subscript “(1)” indicates the encoding operation.

The reconstruction vector $\hat{\mathbf{d}}_k$ can be expressed as

$$\hat{\mathbf{d}}_k = \mathcal{D}(\mathbf{H}_{(2)} \alpha_k), \quad (19)$$

where subscript “(2)” indicates the decoding operation.

Finally, the network parameters $\mathcal{H} = \{\mathbf{H}_{(1)}, \mathbf{H}_{(2)}\}$, are evaluated by the (non trivial) mini-

mization of the MSE in the form:

$$\text{MSE}(\mathcal{H}) = \frac{1}{S} \sum_{k=1}^S \|\hat{\mathbf{d}}_k - \mathbf{d}_k\|^2 = \quad (20)$$

$$\frac{1}{S} \sum_{k=1}^S \|\mathcal{D}(\mathbf{H}_{(2)} \mathcal{E}(\mathbf{H}_{(1)} \mathbf{d}_k)) - \mathbf{d}_k\|^2. \quad (21)$$

Using nonlinear activation functions and multiple hidden layers, DAE provides a nonlinear generalization of the PCA. The intrinsic dimensionality of the data (the number of neurons N in the central hidden layer) is usually unknown and defined based on parametric minimization of the MSE, varying N .

APPLICATION

Figure 6 shows the geometry of the CNR-INSEAN 2340 model, a geosim replica of the DTMB 5415 model used for towing tank experiments, as seen in Stern et al. (2000). The main particulars of the model scale and tests conditions are summarized in Table 1. Since no rudder is considered here, the length between perpendiculars (L_{pp}) is calculated from the fore perpendicular to the transom bottom edge.



Figure 6: A 5.720 m length model of the DTMB 5415 (CNR-INSEAN model 2340).

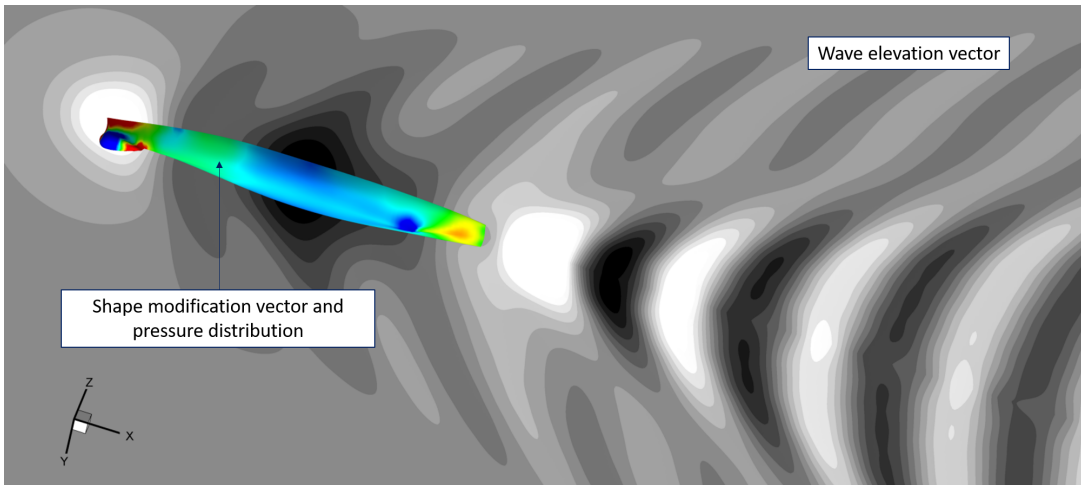


Figure 4: Distributed shape and physical parameters for current application.

Table 1: DTMB 5415 model scale main particulars and test conditions

Description	Unit	Value
Displacement	tonnes	0.549
Length between perpendiculars	m	5.720
Beam	m	0.760
Draft	m	0.248
Longitudinal center of gravity	m	2.884
Vertical center of gravity	m	0.056
Water density	kg/m ³	998.5
Kinematic viscosity	m ² /s	1.09E-06
Gravity acceleration	m/s ²	9.803
Froude number	–	0.250
Reynolds number	–	9.82E+06

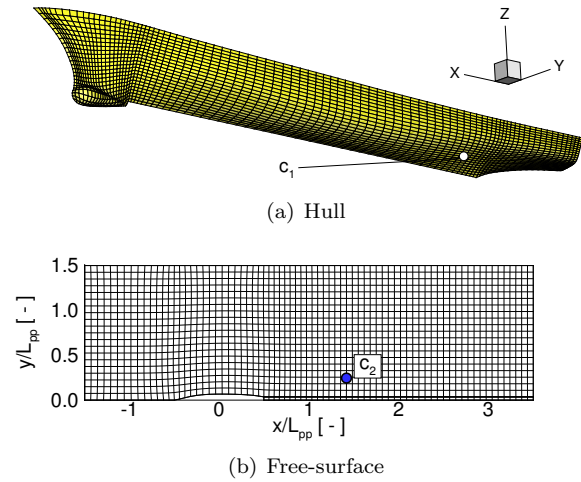


Figure 7: DTMB 5415 body and free-surface discretization used for both design-space dimensionality reduction and potential flow simulation.

The shape parameter vector used for design-space dimensionality reduction collects the y -component (δ_y) of the shape modification vector (δ). The heterogeneous/distributed physical parameter vector collects values of the pressure distribution (p) and the wave elevation pattern (η) (see Figure 4), whereas the lumped physical parameter vector includes the wave resistance coefficient (C_w). Physical parameters are based on a calm-water potential flow solution at $Fr = 0.25$.

Shape Modification Method

The shape modification is defined using a linear combination of $M = 27$ orthogonal basis functions (OBFs) over a hyper-rectangle embedding the demi hull (Serani et al., 2016)

$$\psi_i(\mathbf{x}) : \mathcal{V} = [0, L_{x_1}] \times [0, L_{x_2}] \times [0, L_{x_3}] \in \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad (22)$$

with $i = 1, \dots, M$, as

$$\delta(\mathbf{x}, \mathbf{u}) = \sum_{i=1}^M u_i \psi_i(\mathbf{x}), \quad (23)$$

where the coefficients $u_i \in \mathbb{R}$ ($i = 1, \dots, M$) are the (original) design variables,

$$\psi_i(\mathbf{x}) := \prod_{j=1}^3 \sin\left(\frac{a_{ij}\pi x_j}{L_{x_j}} + r_{ij}\right) \mathbf{e}_{q(i)}, \quad (24)$$

and the following orthogonality property is imposed:

$$\int_{\mathcal{V}} \psi_i(\mathbf{x}) \cdot \psi_k(\mathbf{x}) d\mathbf{x} = \delta_{ik}. \quad (25)$$

In Eq. 24, $\{a_{ij}\}_{j=1}^3 \in \mathbb{R}$ define the order of the function along j -th axis; $\{r_{ij}\}_{j=1}^3 \in \mathbb{R}$ are the corresponding spatial phases; $\{L_{x_j}\}_{j=1}^3$ are the hyper-rectangle edge lengths; $\mathbf{e}_{q(i)}$ is a unit vector. Modifications are applied along x_1 , x_2 , or x_3 , with $q(i) = 1, 2$, or 3 respectively. Details of equations and setting parameters may be found in Serani et al. (2016).

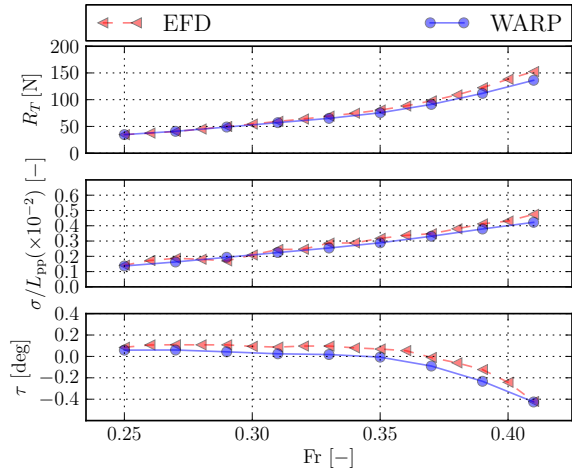


Figure 8: Calm-water model scale validation. The sinkage is positive if the center of gravity sinks; the trim is positive if the bow sinks.

Hydrodynamic Solvers and Computational Grids

Hydrodynamic simulations are conducted using the code WARP (Wave Resistance Program), developed at CNR-INSEAN. Wave resistance computations are based on linear potential flow theory using Dawson (double-model) linearization (Dawson, 1977). The frictional resistance is estimated using a flat-plate approximation, based on the local Reynolds number (Schlichting and Gersten, 2000). Details of equations, numerical implementations, and validation of the numerical solver are

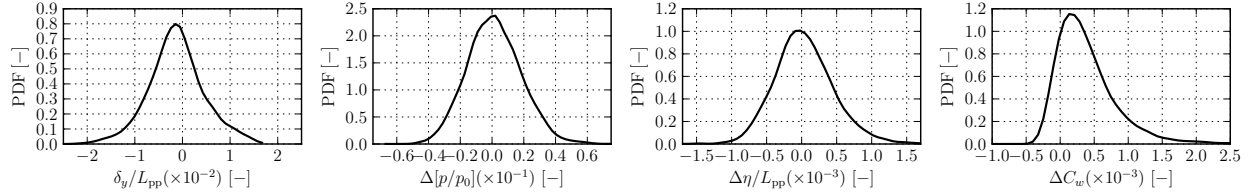


Figure 9: Probability density function of the heterogeneous parameters, using $S = 9,000$ random designs.

given in Bassanini et al. (1994). Grid sensitivity study can be found in Diez et al. (2015).

Simulations are performed for the right demi-hull, taking advantage of symmetry about the xz -plane. The computational domain for the free-surface is defined within $1L_{pp}$ upstream, $3L_{pp}$ downstream, and $1.5L_{pp}$ sideways, for a total of 75×20 grid nodes. The associated hull grid (90×25 nodes) is shown in Figure 7.

The solvers validation versus the experimental fluid dynamics (EFD) data collected at the INSEAN (Olivieri et al., 2001) is depicted in Fig. 8, showing a reasonable agreement.

Numerical Results

The original design space is sampled using a uniform random distribution of $S = 9,000$ hull-form designs. Figure 9 shows the geometry and physics based variability associated to the design space. The PDF is shown for normalized geometric variation δ_y/L_{pp} of a point on the hull (C_1 , see Fig. 7a), variation of the normalized pressure distribution $\Delta[p/p_0]$ of C_1 (where p_0 is the pressure reference), variation of the normalized wave elevation $\Delta\eta/L_{pp}$ of a point downstream (C_2 , see Fig. 7b), and variation of the wave resistance coefficient ΔC_w .

The reduced-dimensionality models are built using a training set $\mathbf{D}_{\text{train}}$ formed by the 80% of the original data set \mathbf{D} , whereas the remaining 20% is used as a test set \mathbf{D}_{test} to evaluate the normalized MSE (cross-validation procedure). The reduced-dimension N is set so as to achieve a normalized MSE equal to 0.05.

A number of cluster $k = 45$ is used for LPCA. A quadratic polynomial kernel is used for the KPCA. Three hidden layers are used for DAE (composed by $L-600-N-600-L$ neurons) with an exponential linear units (ELUs, Clevert et al. 2015) activation function for the first and the third hidden layer. A linear activation function is used for the output layer and for the central hidden layer. Adam optimization algorithm (Kingma and Ba, 2014) is used for train the autoencoder using a mini-batch size of 512 data point for gradient evaluation by the backpropagation algorithm (Rumelhart et al., 1988).

Table 2 summarizes the dimensionality-reduction results in terms of number of compo-

nents/parameters N required by the methods to reconstruct successfully the data set along with the associated normalized MSE. The number of components N also indicates the reduced-dimensionality parametrization of the shape modification vector for future SBDO.

Table 2: Methods' compression capability and NMSE (p -value < 0.05)

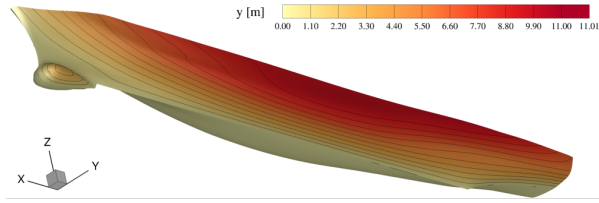
Method	N	NMSE% (training)	NMSE% (test)
PCA	19	4.5	4.6
LPCA	14	3.6	4.6
KPCA	14	4.1	4.6
DAE	17	4.3	4.5

The non-linear methods outperform linear PCA. Specifically, LPCA and KPCA are found the most effective methods for the current problem in terms of dimensionality reduction capability ($N = 14$). DAE and PCA show a sufficient compression capability, whereas the latter is found the least effective method requiring $N = 19$ principal components. This suggests the presence of significant nonlinear structures the data set. The breakdown per shape/physical variable of the normalized MSE is provided in Table 3.

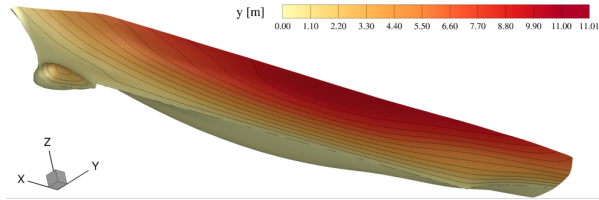
Table 3: Breakdown per variable of NMSE

Method	NMSE%			
	δ_y	p	η	C_w
PCA	9.42	9.58	2.37	5.26e-04
LPCA	10.7	8.24	2.29	1.01e-01
KPCA	10.0	8.84	2.39	8.86e-04
DAE	9.43	9.06	2.60	3.27e-04

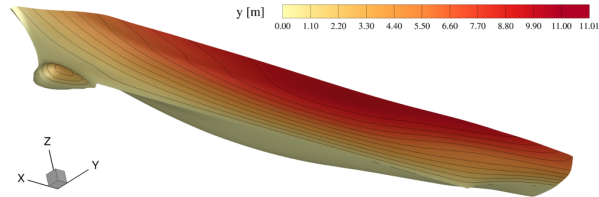
Figures 10, 11, and 12 show the reconstruction of the shape (δ_y) and the distributed physical (p and η) parameters vector for an example design in the test set. Longitudinal wave cut comparison at $y = 0$ is also provided in Fig. 13. A good agreement between the target and reconstructed data is achieved by all methods. Finally, Figure 14 shows the reconstruction of the C_w for the whole test set, showing a remarkable agreement.



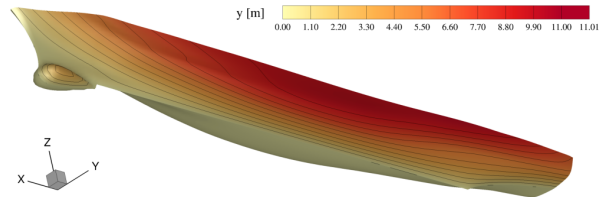
(a) Target geometry ($M = 27$)



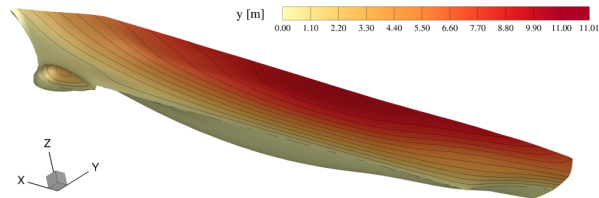
(b) PCA ($N = 19$)



(c) LPCA ($N = 14$)



(d) KPCA ($N = 14$)

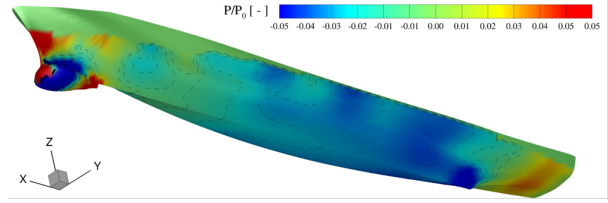


(e) DAE ($N = 17$)

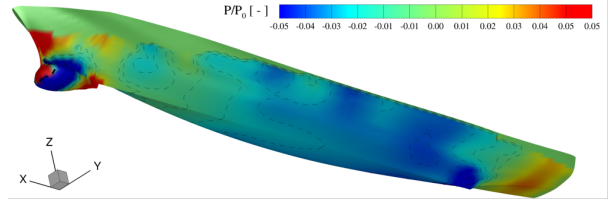
Figure 10: Reconstruction of hull form of the DTMB 5415: original versus reduced-dimensionality parametrization.

CONCLUSIONS AND FUTURE WORK

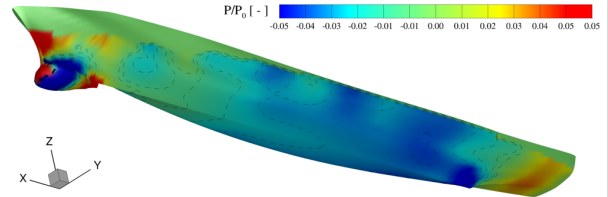
Nonlinear design-space dimensionality reduction methods have been applied to the DTMB 5415 in calm water at $Fr = 0.25$, based on the interaction of shape and heterogeneous physical parameters. Nonlinear extensions of PCA have been applied, namely local and kernel PCA.



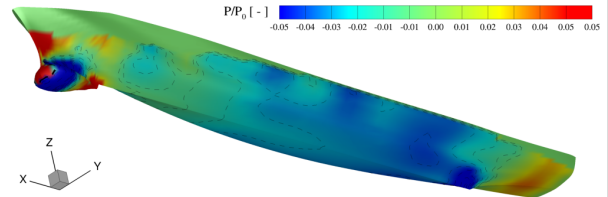
(a) Target pressure ($M = 27$)



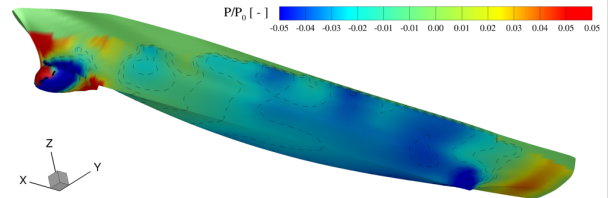
(b) PCA ($N = 19$)



(c) LPCA ($N = 14$)



(d) KPCA ($N = 14$)



(e) DAE ($N = 17$)

Figure 11: Reconstruction of pressure distribution for the DTMB 5415 at $Fr=0.25$: original versus reduced-dimensionality parametrization.

An artificial neural network approach, namely a DAE method, has been also applied and compared to PCA-based approaches.

The data matrix under investigation was formed by the results of 9,000 potential flow simulations coming from an extensive exploration of a 27-dimensional design space associated to a shape-optimization prob-

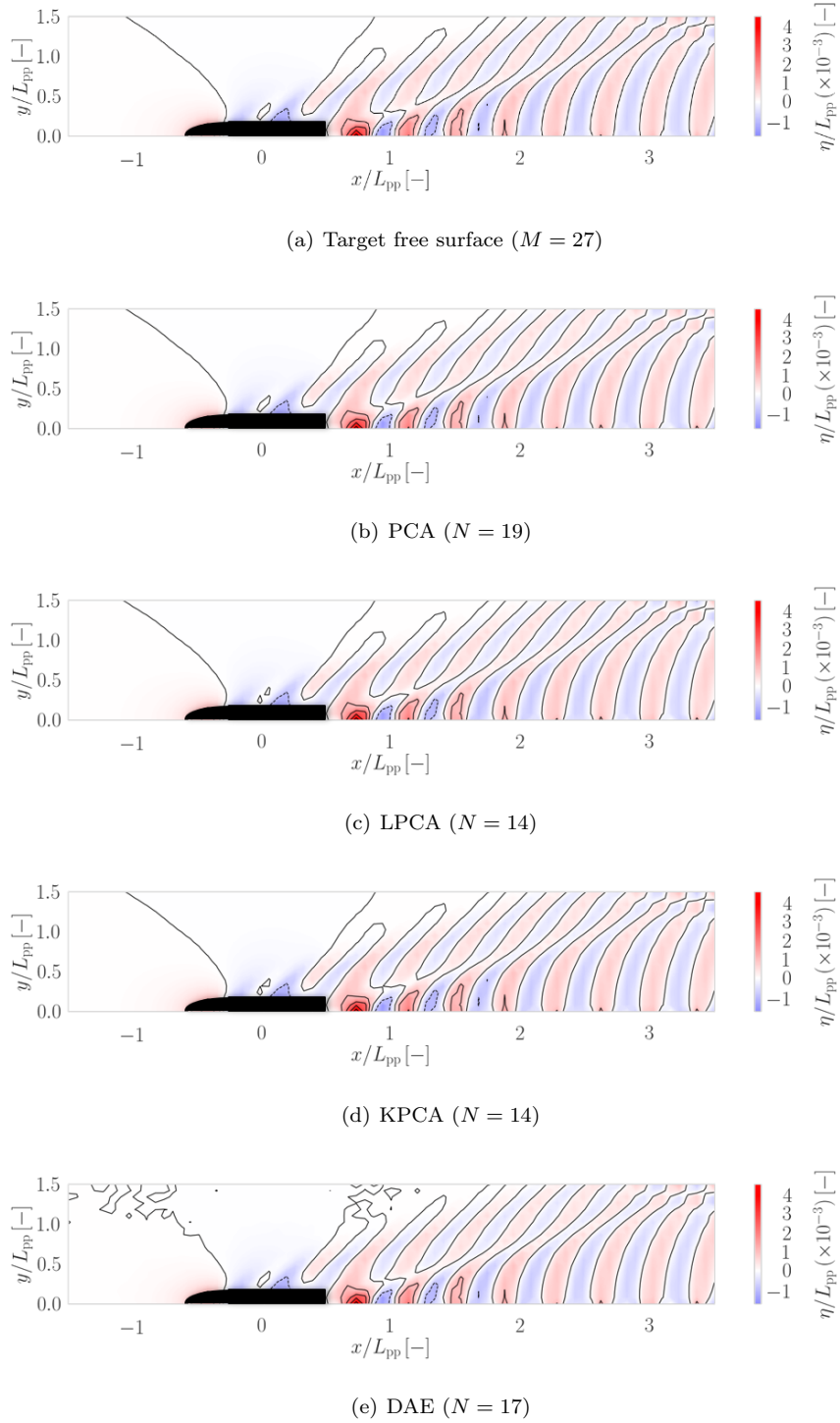


Figure 12: Reconstruction of the free surface for the DTMB 5415 at $Fr=0.25$: original versus reduced-dimensionality parametrization.

lem. Data include three heterogeneous distributed parameters (shape modification vector, pressure distribution on the hull, and wave elevation pattern) and one

lumped parameter (wave resistance coefficient), for a total of $5, 101 \times 9, 000$ elements.

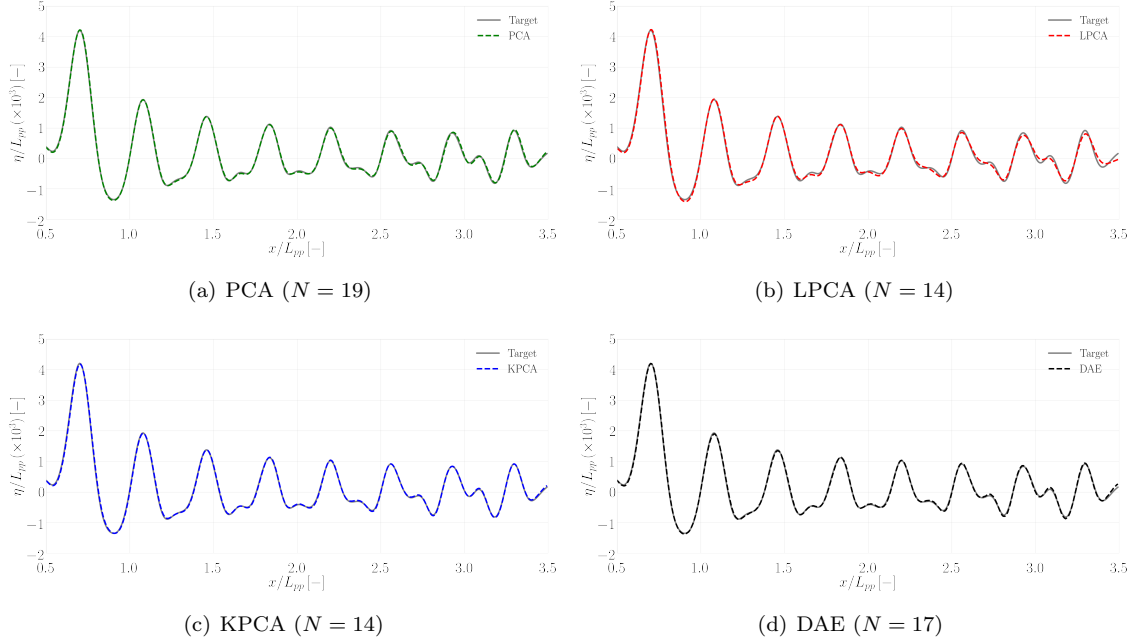


Figure 13: Reconstruction of longitudinal wave cut at $y = 0$ for the DTMB 5415 at $Fr=0.25$: original versus reduced-dimensionality parametrization.

The reduced-dimensionality representation of shape and physical variables was set to achieve an MSE smaller than 5% of the data variance. The standard (linear) PCA meets the requirement using 19 principal components/parameters. DAE achieve the worst compression capability among the nonlinear methods with 17 parameters. Finally, Kernel PCA and Local PCA provides the most promising compression capability with 14 components required by the reduced-dimensionality parametrization. Reconstructed data for shape, pressure, wave elevation, and wave resistance coefficients were presented, showing a remarkable agreement to target values. As a significant result, the number of parameters required by the methods also indicates the reduced dimensionality for shape parametrization in future SBDO studies by higher-fidelity solvers.

The current results are very promising, representing a first step towards data compression and reduced-order model prediction of complex physical phenomena. Current formulations go beyond shape optimization and can be applied to large sets of heterogeneous physical data from simulations, experiments, and real operation measurements.

Future work includes extensions to seakeeping and multiple operating conditions. The possibility of using higher-fidelity analysis such as Reynolds-averaged Navier-Stokes solver with metamodels will be addressed. In parallel, a similar approach is being applied to parti-

cle image velocimetry data of complex flows to assess the compression capability of nonlinear extensions of the POD technique (Clement et al., 2019).

ACKNOWLEDGEMENTS

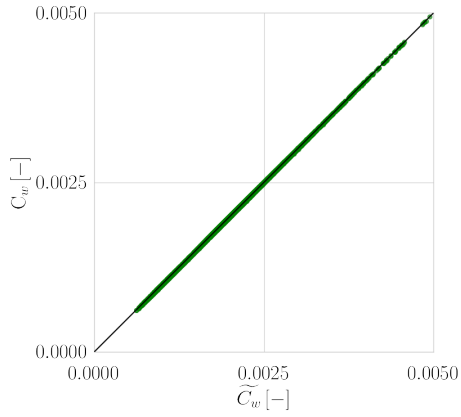
The work is supported by the US Office of Naval Research Global, NICOP grants N62909-15-1-2016 and N62909-18-1-2033, under the administration of Dr. Woei-Min Lin, Dr. Salahuddin Ahmed, and Dr. Ki-Han Kim, and by the Italian Flag-ship Project RITMARE. The research is performed in collaboration with NATO STO Task Group AVT-252 “Stochastic Design Optimization for Naval and Aero Military Vehicles.”

REFERENCES

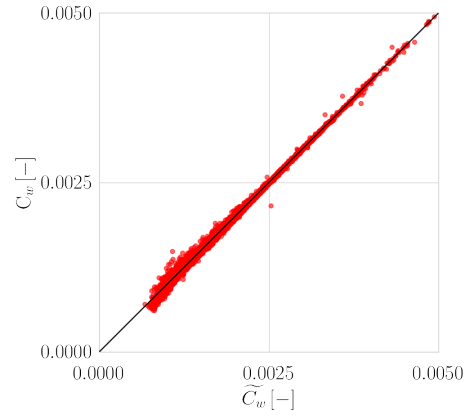
- Bakır, G. H., Weston, J., and Schölkopf, B., “Learning to find pre-images”, *Advances in neural information processing systems*, Vol. 16, 2004, pp. 449–456.
- Bassanini, P., Bulgarelli, U., Campana, E. F., and Lalli, F., “The wave resistance problem in a boundary integral formulation”, *Surveys on Mathematics for Industry*, Vol. 4, 1994, pp. 151–194.
- Bourlard, H. and Kamp, Y., “Auto-association by multilayer perceptrons and singular value decomposition”, *Biological Cybernetics* 59(4), 1988, pp. 291–294.

- Campana, E. F., Diez, M., Iemma, U., Liuzzi, G., Lucidi, S., Rinaldi, F., and Serani, A., “Derivative-free global ship design optimization using global/local hybridization of the DIRECT algorithm”, Optimization and Engineering 17(1), 2015, pp. 127–156.
- Campana, E. F., Liuzzi, G., Lucidi, S., Peri, D., Piccialli, V., and Pinto, A., “New global optimization methods for ship design problems”, Optimization and Engineering 10(4), 2009, pp. 533–555.
- Clement, S., Badra, J., Andre, M., Habukawa, M., Bardet, P., D’Agostino, D., Serani, A., Campana, E. F., and Diez, M., “Nonlinear Proper Orthogonal Decomposition by k-means Clustering and Local Principal Component Analysis of a Transient Buoyant Jet in a Stratified Environment”, 57th AIAA Aerospace Sciences Meeting, SciTech 2019, Manchester Grand Hyatt San Diego, San Diego, California, January 7-11, 2019.
- Clevert, D.-A., Unterthiner, T., and Hochreiter, S., “Fast and accurate deep network learning by exponential linear units (ELUs)”, arXiv preprint arXiv:1511.07289, 2015.
- D’Agostino, D., Serani, A., Campana, E. F., and Diez, M., “Nonlinear Methods for Design-Space Dimensionality Reduction in Shape Optimization”, 3rd International Conference on Machine Learning, Optimization, and Big Data, MOD 2017, Volterra, Italy, 2017.
- D’Agostino, D., Serani, A., Campana, E. F., and Diez, M., “Deep Autoencoder for Off-line Design-Space Dimensionality Reduction in Shape Optimization”, 56th AIAA Aerospace Sciences Meeting, SciTech 2018, Gaylord Palms, Kissimmee, Florida, USA, January 8-12, 2018a.
- D’Agostino, D., Serani, A., and Diez, M., “On the Combined Effect of Design-space Dimensionality Reduction and Optimization Methods on Shape Optimization Efficiency”, 19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (MA&O), AVIATION 2018, Atlanta, GA, USA, June 25-29, 2018b.
- Dawson, C. W., “A practical computer method for solving ship-wave problems”, Proceedings of the 2nd International Conference on Numerical Ship Hydrodynamics, Berkeley, pp. 30–38, 1977.
- DeMers, D. and Cottrell, G. W., “Non-linear dimensionality reduction”, Advances in Neural Information Processing Systems, pp. 580–587, 1993.
- Diez, M., Campana, E. F., and Stern, F., “Design-space dimensionality reduction in shape optimization by Karhunen–Loève expansion”, Computer Methods in Applied Mechanics and Engineering, Vol. 283, 2015, pp. 1525–1544.
- Diez, M., Serani, A., Campana, E. F., Goren, O., Sarioz, K., Danisman, D. B., Grigoropoulos, G., Aloniati, E., Visonneau, M., Queutey, P., and Stern, F., “Multi-objective Hydrodynamic Optimization of the DTMB 5415 for Resistance and Seakeeping”, Proceedings of the 13th International Conference on Fast Sea Transportation, FAST 2015, Washington, D.C., USA, 2015.
- Diez, M., Serani, A., Campana, E. F., and Stern, F. “Assessing the Ability to Optimize Hull Forms of Sea Vehicles for Best Performance in a Sea Environment, STO-TR-AVT-204, Chapter 3: INSEAN/UI Optimization Approach”. Technical report, 2018a, March, NATO.
- Diez, M., Serani, A., Campana, E. F., and Stern, F. “Stochastic Design Optimization for Naval and Aero Military Vehicles, STO-TR-AVT-252, Chapter 7: Reliability-Based Robust Hull-Form Optimization of a Naval Destroyer in Waves (TC5)”. Technical report, 2018b, NATO.
- Diez, M., Serani, A., Stern, F., and Campana, E. F., “Combined Geometry and Physics Based Method for Design-Space Dimensionality Reduction in Hydrodynamic Shape Optimization”, Proceedings of the 31st Symposium on Naval Hydrodynamics, Monterey, CA, USA, 2016.
- Hinton, G. E. and Salakhutdinov, R. R., “Reducing the dimensionality of data with neural networks”, Science 313(5786), 2006, pp. 504–507.
- Hotelling, H., “Analysis of a complex of statistical variables into principal components”, Journal of Educational Psychology 24(6), 1933, pp. 417.
- Jones, D., Perttunen, C., and Stuckman, B., “Lipschitzian optimization without the Lipschitz constant”, Journal of Optimization Theory and Applications 79(1), 1993, pp. 157–181.
- Kambhatla, N. and Leen, T. K., “Dimension reduction by local principal component analysis”, Neural Computation 9(7), 1997, pp. 1493–1516.
- Kingma, D. and Ba, J., “Adam: A method for stochastic optimization”, arXiv preprint arXiv:1412.6980, 2014.
- Kramer, M. A., “Nonlinear principal component analysis using autoassociative neural networks”, AIChE Journal 37(2), 1991, pp. 233–243.
- Lloyd, S., “Least squares quantization in PCM”, IEEE Transactions on Information Theory 28(2), 1982, pp. 129–137.

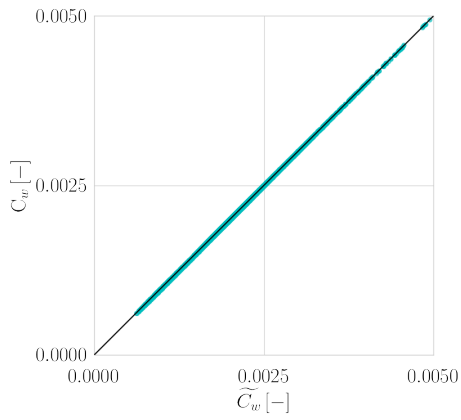
- Lukaczyk, T., Palacios, F., Alonso, J. J., and Constantine, P., “Active subspaces for shape optimization”, Proceedings of the 10th AIAA Multidisciplinary Design Optimization Specialist Conference, National Harbor, Maryland, USA, 13-17 January, 2014.
- Oja, E., “Data Compression, Feature Extraction, and Autoassociation in Feedforward Neural Networks”, T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas (Eds.), Artificial Neural Networks, Volume 1, pp. 737–745. Elsevier Science Publishers B.V., North-Holland, 1991.
- Olivieri, A., Pistani, F., Avanzini, A., Stern, F., and Penna, R. “Towing tank, sinkage and trim, boundary layer, wake, and free surface flow around a naval combatant INSEAN 2340 model”. Technical report, 2001, DTIC.
- Pearson, K., “LIII. On lines and planes of closest fit to systems of points in space”, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2(11), 1901, pp. 559–572.
- Raghavan, B., Breitkopf, P., Tourbier, Y., and Villon, P., “Towards a space reduction approach for efficient structural shape optimization”, Structural and Multidisciplinary Optimization, Vol. 48, 2013, pp. 987–1000.
- Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al., “Learning representations by back-propagating errors”, Cognitive Modeling 5(3), 1988, pp. 1.
- Schlichting, H. and Gersten, K. Boundary-Layer Theory. Springer-Verlag, Berlin, 2000.
- Schölkopf, B., Smola, A., and Müller, K.-R., “Nonlinear component analysis as a kernel eigenvalue problem”, Neural Computation 10(5), 1998, pp. 1299–1319.
- Serani, A., Campana, E. F., Diez, M., and Stern, F., “Towards Augmented Design-Space Exploration via Combined Geometry and Physics Based Karhunen-Loève Expansion”, 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (MA&O), AVIATION 2017, Denver, USA, June 5-9, 2017.
- Serani, A. and Diez, M. “Dolphin Pod Optimization”. Y. Tan, H. Takagi, and Y. Shi (Eds.), Advances in Swarm Intelligence. ICSI 2017, Lecture Notes in Computer Science, vol 10385, pp. 63–70. Cham: Springer, 2017a.
- Serani, A. and Diez, M., “Dolphin Pod Optimization: A Nature-Inspired Deterministic Algorithm for Simulation-Based Design”, 3rd International Conference on Machine Learning, Optimization, and Big Data, MOD 2017, Volterra, Italy, 2017b.
- Serani, A. and Diez, M., “Reliability-based Robust Design Optimization by Design-space Augmented Dimensionality Reduction”, 19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (MA&O), AVIATION 2018, Atlanta, GA, USA, June 25-29, 2018.
- Serani, A., Fasano, G., Liuzzi, G., Lucidi, S., Iemma, U., Campana, E. F., Stern, F., and Diez, M., “Ship hydrodynamic optimization by local hybridization of deterministic derivative-free global algorithms”, Applied Ocean Research, Vol. 59, 2016, pp. 115–128.
- Stern, F., Longo, J., Penna, R., Olivieri, A., Ratcliffe, T., and Coleman, H., “International Collaboration on Benchmark CFD Validation Data for Surface Combatant DTMB Model 5415”, Proceedings of the Twenty-Third Symposium on Naval Hydrodynamics, Val de Reuil, France, September 17-22, 2000.
- Von Luxburg, U., “A tutorial on spectral clustering”, Statistics and computing 17(4), 2007, pp. 395–416.



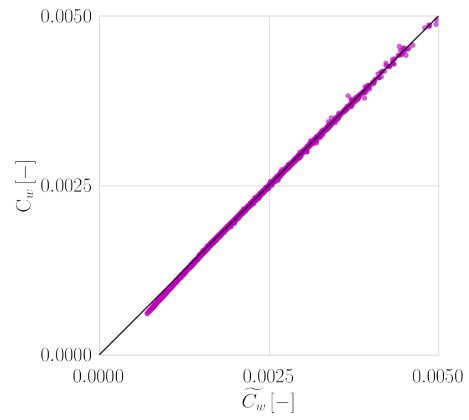
(a) PCA ($N = 19$)



(b) LPCA ($N = 14$)



(c) KPCA ($N = 14$)



(d) DAE ($N = 17$)

Figure 14: Reconstruction of wave resistance coefficient C_w for the DTMB 5415 at $Fr=0.25$: original versus reduced-dimensionality parametrization.