

Empowering optimal transport matching algorithm for the construction of surrogate parametric metamodel

Maurine Jacot^{1,2,*}, Victor Champaney¹, Sergio Torregrosa Jordan¹, Julien Cortial², and Francisco Chinesta^{1,3}

¹ PIMM Lab, Arts et Métiers Institute of Technology, 155 Boulevard de l'Hôpital, 75013 Paris, France

² Safran Tech, Department of Digital Sciences and Technologies, 1 rue des Jeunes Bois, 78117, Châteaufort, France

³ CNRS@CREATE LTD, 1 Create Way, # 08-01 CREATE Tower, 138602 Singapore

Received: 10 August 2023 / Accepted: 10 January 2024

Abstract. Resolving Partial Differential Equations (PDEs) through numerical discretization methods like the Finite Element Method presents persistent challenges associated with computational complexity, despite achieving a satisfactory solution approximation. To surmount these computational hurdles, interpolation techniques are employed to precompute models offline, facilitating rapid online solutions within a metamodel. Probability distribution frameworks play a crucial role in data modeling across various fields such as physics, statistics, and machine learning. Optimal Transport (OT) has emerged as a robust approach for probability distribution interpolation due to its ability to account for spatial dependencies and continuity. However, interpolating in high-dimensional spaces encounters challenges stemming from the curse of dimensionality. The article offers insights into the application of OT, addressing associated challenges and proposing a novel methodology. This approach utilizes the distinctive arrangement of an ANOVA-based sampling to interpolate between more than two distributions using a step-by-step matching algorithm. Subsequently, the ANOVA-PGD method is employed to construct the metamodel, providing a comprehensive solution to address the complexities inherent in distribution interpolation.

Keywords: Model order reduction / optimal transport / proper generalized decomposition / bayrcentric projection

1 Introduction

Many problems in science and engineering can be modelled mathematically by a system of Partial Differential Equations (PDEs). Generally, solving numerically these equations requires a numerical method e.g. the Finite Element (FE) method which approximates the problem solution from its approximation in each element of the mesh covering the domain [1]. This process may be time-consuming and computationally expensive, specifically when a new configuration associated with the problem parameters has to be evaluated.

One possible approach is to use interpolation to build offline a model that can rapidly calculate online a new solution of the problem at any given point in the parametric space, without requiring expensive computation resources. These techniques are already used in a wide range of scientific and engineering applications, such as parametrised Model Order Reduction techniques (pMOR) [2], because the ability to evaluate solutions in near real time leads to a more effective exploration of the parametric space.

However, interpolating in a high-dimensional space faces the so-called *curse of dimensionality* [3]. Many methodologies exploiting sparsity assumptions have been proposed to alleviate these difficulties [4–6] and succeed in constructing high-dimensional models from limited data.

Furthermore, the choice of a suitable technique for interpolation and the resulting performance is impacted by how data is represented and the purpose for which it is being used.

In the fields of physics, statistics, and machine learning, the framework of probability distributions is widely used for modelling and representing data by associating a probability measure to a collection of samples that form a solution [7]. This enables to describe and model the statistical characteristics of the data and to develop a probabilistic model for predicting and interpolating efficiently.

The continuous field description of a collection samples is viewed as a continuous probability density function assigning a probability density (also called weight) to each point of a regular grid that discretises the space (usually a Euclidean space). In the Finite Element method, the solution of a problem is typically represented by a continuous field (e.g. velocity, stress). The nodal values are defined

* e-mail: maurine.jacot@ensam.eu

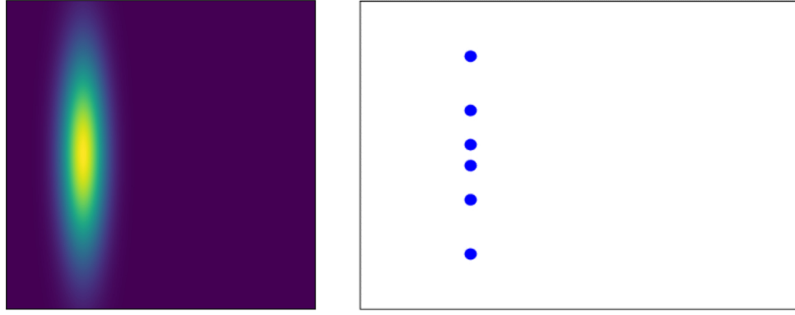


Fig. 1. (*left*) Continuous field description of data associated with a discretised probability density (also called weight) for each point of the regular grid. (*right*) Particle cloud description associated with a set of particles (point clouds).

at fixed points in space where the physical quantity is evaluated. The particle cloud description of data refers to describing the probability distribution by a set of particles characterised by a mass and a position. Figure 1 illustrates the two types of descriptions.

One of the most accurate and mathematically rigorous approaches to interpolate between distributions into the particle cloud description is to follow the movement of individual particles, within the Optimal Transport (OT) framework [8].

Optimal Transport, also known as the Wasserstein distance or transportation distance was formalised in 1781 by the French mathematician Gaspard Monge and is a mathematical framework that deals with the optimal way to transport mass from one probability distribution to another [9]. The goal is to find the most efficient way to transform one distribution into another while minimizing a certain cost or distance metric.

The basic idea is to consider the two probability distributions as piles of mass, and the transportation plan as a way to move mass from one pile to another. The cost of transporting mass from one point to another is typically given by a cost function, and the optimal transport problem involves finding the optimal way to move mass that minimizes the total cost [10,11].

Despite the recent advances in the field, such as the entropy regularised Optimal Transport and more precisely, the Sinkhorn-Knopp algorithm [12], solving an Optimal Transport problem can be computationally expensive and not well suited for an online approach [13]. First, if distributions are matched using the Optimal Transport theory in their continuous descriptions, it involves tackling an online transportation problem that requires solving partial differential equations. Thus, to decrease this complexity, distributions in their continuous form needs to be discretized into a collection of particles. However, the choice of the type of discretization can have a significant impact on the accuracy and efficiency.

A technique for discretization uses the Smoothed Particle Hydrodynamics (SPH) decomposition, as explained in [14]. In this method, all particles have the same mass, and their positions are established by solving an optimization problem. Contrastingly, an alternative method using Finite Element discretization eliminates the need for solving an optimization problem. In this setting, the positions

of particles align with the locations in the continuous field within the Finite Element discretization, and the mass assigned to each particle is directly proportional to the magnitude of the continuous field at the corresponding particle location. Ultimately, even if the distributions are discretized, performing online Optimal Transport between collections of particles presents challenges, as it entails solving a discrete linear programming problem, which can be both numerically costly and statistically inefficient in an offline-online approach [15].

In practice, to be able to interpolate between distributions, it is also common to consider more than two distributions simultaneously. The concept of interpolation between more than two distributions is relevant in various fields, including Machine Learning and Statistics. In these context, Euclidean distance is often used as a measure of dissimilarity or similarity between these distributions as is it applied in [16]. When interpolating between more than two distributions using the Wasserstein distance, the mathematical theory does not have a solid foundation. In fact, the goal is to define a matching between an arbitrary number of discrete densities where the distributions are discretized into collections of particles [17].

Some approaches have been developed to overcome this problem, in particular the Wasserstein Barycenters [18] which aim to find a representative that lies “in between” a set of given collections of particles. However, this method is often computationally intractable and cannot ensure a fast real time (online) solution for new data points due to the complexity of the problem and its formulation and, moreover, it is difficult to operate in highly multidimensional settings.

The aim of this paper is to provide a reliable approach for real-time evaluations of distributions of a parametric problem by using Optimal Transport by constructing a parametric metamodel. Once the metamodel is constructed, it can be deployed for real-time use. This means that the model is capable of making predictions or providing insights on new data as it becomes available, without significant delays.

The approach proceeds in two phases: an offline training in the parametric space and an online application.

In the offline phase, we introduce a structured sampling represented as a cross-centred sampling that will explore the influence of each dimension of the parametric space $\mathcal{D}$. Numerical simulations are performed for each

point of the sampling (problem parameters), solving a PDE defined in a domain and subjected to initial and boundary conditions. To manipulate efficiently the model and apply transport, the data has to be transformed into a particle cloud description, i.e. expressed in the form of particles with a mass and position. Performing SPH decomposition indeed involves solving an optimization problem, which can be computationally demanding. However, despite this computational cost, the structure of the SPH decomposition is often more efficient than alternative methods.

To avoid the computational cost of discretizing each distribution in their continuous form following the SPH decomposition, we introduce a new methodology where only the first distribution is discretized following the SPH decomposition and the optimal matching is performed using a step by step matching algorithm in the proposed sampling where the others distributions are discretized based on the Finite Element localization and the matching problem is solved based on the use of the notion of barycentric projection [19].

By using this approach, the dimensionality of the matching problem is reduced from matching the distributions in pairs, rather than matching all distributions simultaneously. Finally, the parametric displacement field between solutions associated to sampling points is learned using a Model Order Reduction technique, the ANOVA-PGD [6].

While the distinct arrangement of the sampling provides an opportunity to acquire knowledge of the parametric displacement field with reduced computational effort, it is possible that these sampling points may not cover the entire parametric space. This limitation can result in errors in the predictions made by the metamodel during online operation. To enhance the accuracy of the surrogate parametric metamodel, additional sampling points are generated using a Latin Hypercube Sampling (LHS) with associated solutions and then decomposed into particles. The difference between the projected positions of the particles and their actual positions will be used for computing a correction component. The surrogate parametric metamodel is then self-enriched by taking into account the correction. Testing points are then used to verify the accuracy of the corrected metamodel.

Finally, the self-enriched parametric surrogate parametric metamodel is used online to predict the particles' position and therefore the solution of the problem for any point of the parametric space.

The paper is organised as follows. [Section 2](#) first details in a nutshell the theory of Optimal Transport. In [Section 3](#), the construction of the parametric surrogate parametric metamodel using the ANOVA-PGD technique is detailed as well as its enrichment. The step by step matching algorithm using a structured sampling is presented. [Section 4](#) presents the global methodology and the error evaluation. Two physical problems are introduced in the [Section 5](#) and the proposed methodology is applied. Conclusion and remarks are finally discussed in [Section 6](#).

2 Theory of optimal transport

The objective of this section is to revisit the mathematical and numerical aspects of Optimal Transport. We restrict to present only the fundamentals that will be discussed later and used in the proposed methodology. To obtain further information regarding the theory, readers can refer to: [20], [21], [22] to cite a few.

2.1 Monge's problem

The Optimal Transport problem was first introduced in 1781 by the mathematician Gaspard Monge. His idea consists in finding the cheapest way to transport a pile of sand to a hole in the ground while minimising the distance that sand must travel.

We denote two independent and distributed discrete samples α and β on two compact metric spaces $(\mathcal{X}, \mathcal{Y})$ that correspond to the source and target distributions,

respectively $\alpha = \sum_{i=1}^n a_i \delta_{x_i}$ and $\beta = \sum_{j=1}^m b_j \delta_{y_j}$. In our case

$\mathcal{X}$ and $\mathcal{Y}$ are both subspaces of a common larger metric space and are endowed with the same metric. The Monge Problem (MP) searches for a map $T : \{x_1, x_2, \dots, x_n\} \rightarrow \{y_1, y_2, \dots, y_m\}$ that associates each point x_i to a single point y_j , while minimising the sum of the costs $c(x_i, y_j) = d(x_i, y_j)^2$. T must satisfy the mass conservation $\forall j \in [m], b_j = \sum_{i: T(x_i)=y_j} a_i$, where the push-forward

measure $T_{\#}\alpha$ is defined by $T_{\#}\alpha = \sum_{i=1}^n a_i \delta_{T(x_i)}$. The Monge

Problem can be formulated as a non-convex optimisation problem:

$$(MP) \quad M(\alpha, \beta) = \min_{T_{\#}\alpha = \beta} \sum_{i=1}^n c(x_i, T(x_i)). \quad (1)$$

2.2 Kantorovich relaxation and linear programming problem

The Monge formulation defines an ill-posed problem, since sometimes there is no transport map to solve the constraint $T_{\#}\alpha = \beta$. The main idea of Kantorovich is the relaxation of the deterministic transport map. At any point x_i , the mass can be split and transported from a source towards several targets. Thus, we adopt a probabilistic transportation. Numerically, the Kantorovich relaxation leads to a strictly convex problem that is easier to solve. Instead of a map T , a transportation plan $\gamma \in \Gamma(\alpha, \beta)$ is used, where $\Gamma(\alpha, \beta)$ is the set of measure couplings between two discrete probability measures α and β . Specifically, $\gamma : (\mathcal{X}, \mathcal{Y}) \mapsto \mathbb{R}_+$ is a function that satisfies $\int \gamma(x, y) dx dy = 1$, $\int \gamma(x, \bullet) dx = \beta$ and $\int \gamma(\bullet, y) dy = \alpha$. For the sake of simplicity, γ also refers to the discretised version of the transport plan, which belongs to $\mathbb{R}_+^{n \times m}$. Then, to define the optimal transport

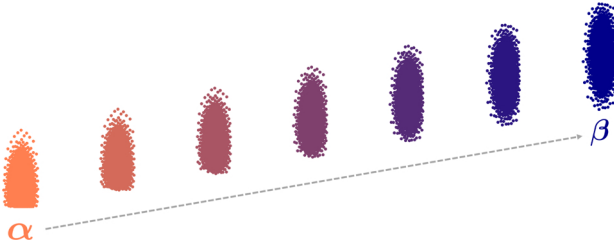


Fig. 2. Linear displacement interpolation between the source distribution α and the target distribution β .

map $\gamma^* \in \Gamma(\alpha, \beta)$, the Kantorovitch Problem (KP) reads:

$$(KP) \quad \gamma^* = \operatorname{argmin}_{\gamma \in \Gamma(\alpha, \beta)} \sum_{i=1}^n \sum_{j=1}^m c(x_i, y_j) \gamma_{i,j}. \quad (2)$$

where $c(x, y)d\gamma(x, y)$ is the transport cost associated with a transportation plan $\gamma \in \Gamma(\alpha, \beta)$ and can be interpreted as the work needed to move one unit of mass from x to y . The p-Wassertein distance for $p \in [1, \infty[$ measures similarity between distributions and is one natural way of defining the cost:

$$\mathcal{W}_p(\alpha, \beta) = \left(\min_{\gamma \in \Gamma(\alpha, \beta)} \sum_{i,j} \gamma_{i,j} d(x_i, y_j)^p \right)^{1/p} \quad (3)$$

The linear problem in equation (2) can be solved using the network simplex in $\mathcal{O}((n+m)nm \log(n+m))$ [23], such as proposed by the solver in open source Python Library POT [24] and will return the optimal transport plan γ^* .

2.3 Interpolation using optimal transport

OT defines a distance $\mathcal{W}_p$ between two discrete probability measures α and β . To interpolate between two distributions, the minimisation of the cost yields the map T which matches each point of the source α to one in the target distribution β . Then, the linear displacement interpolation is obtained by moving each mass along the line defined by $(1-t)x_i + tT(x_i)$ for $t \in [0, 1]$ as it is illustrated in the Figure 2. For $\mathcal{W}_2$, the interpolation is called displacement interpolation [25].

Once the displacement interpolation between two measures is defined, it can be generalised to an interpolation between more measures using the concept of Wasserstein barycenters as it is presented in [11].

2.4 Notion of barycentric projection

In the context of the Kantorovitch formulation, the map T is not defined, therefore an extra step is required to perform the interpolation, the so-called barycentric projection [19]. Once an optimal coupling γ^* has been computed, one needs to map x_i from $\mathcal{X}$ to $\mathcal{Y}$. For each source sample, the barycentric mapping is expressed as

follows:

$$x_i \in \mathcal{X} \mapsto \frac{1}{\sum_j \gamma_{i,j}^*} \sum_j \gamma_{i,j}^* y_j \in \mathcal{Y} \quad (4)$$

3 Offline learning approach using optimal transport

In the fields of statistical and machine learning, Optimal Transport has gained significant attention due to its ability to compare, align and model probability distributions while underlying geometry and structure of data. It is therefore an appealing route to create parametric models when more traditional methods fail due to the presence of localised features in the data.

When dealing with many parameters, in order to obtain a precise approximation of the system, a sufficient number of sampling points generated from a $N_{\mathcal{D}}$ -dimensional sampling should be used. However, this leads to solving optimal transport with many distributions, and therefore to high-dimensional matching problems. Learning transportation plans in this context is a challenge due to the *curse of dimensionality*. One effective approach to address this issue involves utilizing the concept of Wasserstein barycenter that aims to find a representative distribution that lies “in between” a set of given probability distributions. It can be thought of as a way to compute a central or average distribution that captures the essential characteristics of the input distributions. However, the computational complexity of this approach can be high, especially when dealing with high multidimensional settings and cannot ensure a fast real time (online) solution for new data points. Therefore, finding an offline representation of the solution to the OT problem and learning its parametric dependencies is essential to evaluate online a new configuration of the system. In addition, finding a strategy to efficiently compute transportation plans is crucial since, as the number of parameters and therefore of data samples increases, the traditional optimization methods become computationally expensive or even infeasible.

Motivated by such an idea, our offline learning approach follows a step by step matching algorithm based on the structure of an adequate sampling. Point clouds are matched successively by pairs based on the notion of barycentric projection. The distinct arrangement of the sampling provides a way to learn the displacement field between numerous sets of point clouds. Finally, we introduce a non-linear regression technique, the ANOVA-PGD, to learn the parametric model of the transportation plan. This technique is presented in this section.

3.1 Definition of a structured sampling

Computing parametric solutions based on surrogates (metamodels or response surfaces) requires to define a design of experiments (also called sampling) to work with. We consider the parametric space $\mathcal{D}$ where $N_{\mathcal{D}}$ parameters (also known as dimensions) are introduced, $\mathbf{p} = (p^1, \dots, p^{N_{\mathcal{D}}})$. The chosen sampling is a centred cross

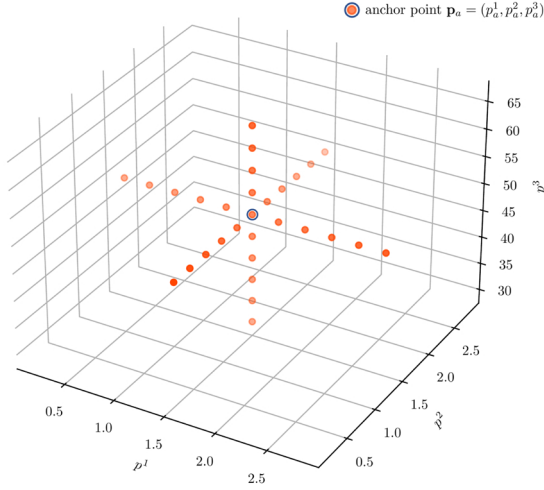


Fig. 3. Cross-centred sampling (DoE) considered as example with anchor point $\mathbf{p}_a$.

that explores the influence of each dimension individually [6] and is composed of B sampling points, $b \in \llbracket B \rrbracket$ around the so-called anchor point $\mathbf{p}_a$. Figure 3 shows an example of sampling with $N_D = 3$ used to demonstrate our proposed methodology.

3.2 Step by step matching algorithm based on barycentric projection

Our approach relies on a step-by-step algorithm designed to address the challenge of estimating transportation plans in a high-dimensional space. While the Optimal Transport framework, as explained in Section 2, is centered around matching two distributions, extending this matching to more than two distributions introduces increased complexity. Additionally, if the distributions are represented by a continuous field, discretizing them into a set of particles can be computationally expensive, especially when using Smoothed Particle Hydrodynamics (SPH) decomposition.

To mitigate these challenges, we operate within the defined parametric space $\mathcal{D}$ and leverage a distinctive arrangement of sampling. The distributions at the anchor point are discretized following the SPH decomposition, while the other distributions are decomposed using the Finite Element discretization. Subsequently, each pair of point clouds is matched sequentially by solving a classical Optimal Transport problem. The results are then projected using a barycentric projection, facilitating interpolation between the two distributions. This principle is described in Figure 4.

We first provide some essential notation. Let B be a set of points defining a cross-centred sampling as shown in the Figure 3. B is decomposed into N_c overlapping subsets, $n_c \in \llbracket N_c \rrbracket$ which correspond to the number of branches of the cross. Then, each branch contains the associated N_b sampling points with the same anchor point, $n_b \in \llbracket N_b \rrbracket$ and an associated solution of a high-fidelity problem represented by a continuous field, $f_{n_b}^{n_c} : \Omega_N \subset \mathbb{R}^2 \rightarrow \mathbb{R}$

with f_0 , the solution at the anchor point. Each dimension of the problem contains $N_c = 2$ branches. Figure 5 shows the different parts that compose the sampling presented in the Figure 3 with the associated high-fidelity solutions.

For every $n_c \in \llbracket N_c \rrbracket$, our step by step matching algorithm follows a four-steps procedure:

1. **High-fidelity solutions:** High-fidelity solutions are generated based on the resolution of PDEs with the FE method, for instance, on the points of the cross-centred sampling presented in Figure 5 which is expected to express the individual effect of each dimension. Solutions of the problem are denoted $f_{n_b}^{n_c} : \Omega_N \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ and correspond to the solution of each node $j \in \llbracket N \rrbracket$ of the mesh. The mesh is the same for all the solutions and the spatial coordinates of each node are denoted (x_j, y_j) and stored in a vector $\mathbf{X}$ as it is shown in the Figure 6. Then, solutions are normalised in the domain Ω , denoted as a distribution $\varphi_{n_b}^{n_c}$ and correspond to the integral of the solution at each node of the mesh.
2. **Discretisation into point cloud:** The distribution at the anchor point φ_0 is discretised based on the SPH decomposition, into a 2D point cloud and denoted $\bar{\varphi}_0$. For the others distributions, we consider that particles are positioned on the FE points and theirs masses are proportional to the high-fidelity solution at each point.

For that, we consider a numerical solution f as $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+$. The solution is first normalised over the domain to transform features on a similar scale such that:

$$\varphi = \frac{f}{I} \text{ with } I = \int_{\Omega} f d\Omega. \quad (5)$$

Then, following the Smoothed-Particle Hydrodynamics method (SPH), the continuous function is discretised in a set of points referred to as particles [17].

Indeed, the idea is to decompose φ into a sum of P Gaussian functions, $i \in \llbracket P \rrbracket$, centred on the coordinates of each particle, located at μ_x and μ_y , and stored in the vector $\boldsymbol{\mu} \in \mathbb{R}^{(P \times 2)}$. The sum of all the Gaussian functions should represent the approximated distribution φ such as:

$$\varphi \approx \bar{\varphi} = \sum_{i=1}^P G_{\boldsymbol{\mu}_i, \sigma}(\mathbf{x}), \quad (6)$$

where $G_{\boldsymbol{\mu}_i, \sigma}$ is a 2D-Gaussian with the standard deviation σ (a model hyperparameter) and the mean $\boldsymbol{\mu}_i$. This decomposition is obtained by minimizing $\|\bar{\varphi} - \varphi\|^2$.

3. **Transport plan:** An optimal transport plan γ^{*1, n_c} is computed by resolving the Kantorovich problem in the equation (2) between the particle's positions $\boldsymbol{\mu}_0^{n_c}$ and $\varphi_1^{n_c}$ as it is presented in Figure 7.
4. **Projection:** Particles are projected from $\boldsymbol{\mu}_0$ using the notion of barycentric projection following the

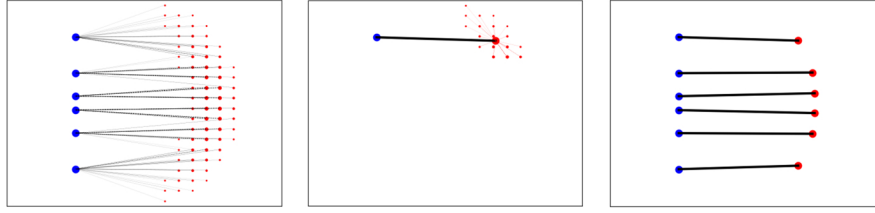


Fig. 4. (left) Optimal matching between two distributions represented by point clouds. (middle) Barycentric projection focusing in one matched particle on the first point cloud. (right) Particles optimal matching following the barycentric projection that allow to interpolate between these two collections of particles.

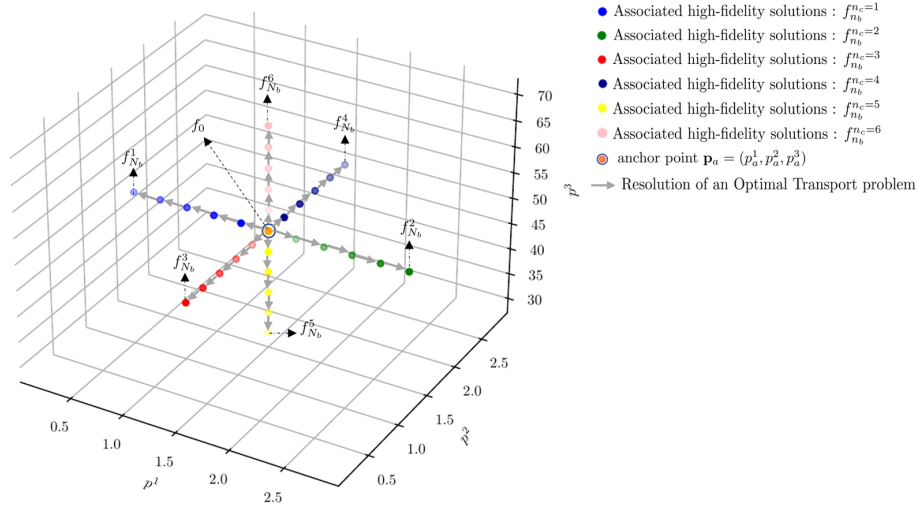


Fig. 5. Different parts composing the initial cross-centred sampling introduced in the Figure 3.

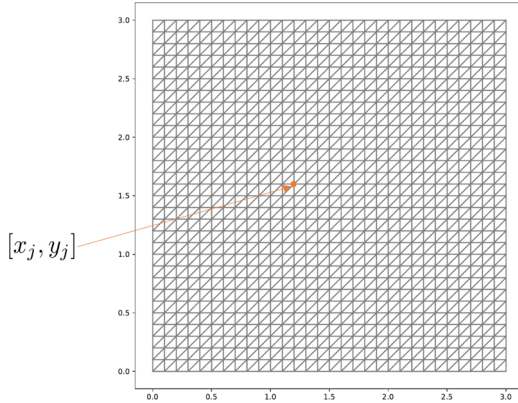


Fig. 6. Example of a element mesh with associated spatial coordinates.

methodology presented in Figure 4. The projection is performed between each pair of discretized distributions ($\mu_{n_b}^{n_c}, \mu_{n_b+1}^{n_c}$) by computing the transport plan between $\mu_{n_b}^{n_c}$ and $\varphi_{n_b+1}^{n_c}$ iteratively.

Each point cloud representation of a distribution $\varphi_{n_b}^{n_c}$ is characterised by μ_x and μ_y of each particle stored in the vector $\mu_{n_b}^{n_c}$. A natural way to estimate the particles' position $\mu_{n_b+1}^{n_c}$, would be to approximate it using the optimal transport plan γ^{*n_c, n_b} between the source represented by

a point cloud $\mu_{n_b}^{n_c}$ and the target $\varphi_{n_b+1}^{n_c}$ distributions represented by a continuous field. By relying on the notion of barycentric projection presented in the Section 2.4, the mapping of the i -th component of $\mu_{n_b+1}^{n_c}$ which allows to reconstruct to the distribution $\varphi_{n_b+1}^{n_c}$ is approached as below and presented in Figure 8:

$$(\mu_{n_b+1}^{n_c})_i = \frac{1}{\sum_{j=1}^m \gamma_{ij}^{*n_c, n_b}} \sum_{j=1}^m \gamma_{ij}^{*n_c, n_b} (\mathbf{X})_j, \quad (7)$$

where $\mathbf{X}$ is the vector that contains the spatial coordinates of the mesh (x_j, y_j) as shown in Figure 6 and γ^{*n_c, n_b} is the transportation plan between $\mu_{n_b}^{n_c}$ and $\varphi_{n_b+1}^{n_c}$.

Then, the four-step procedure is repeated for each part n_c of the sampling. At the end, we obtain the particles' positions μ_B from all the B distributions corresponding to the sampling points presented in the Figure 3. The particles' positions allow to recover the distribution φ_B using equation (6). The number of particles P and the standard deviation of each particle σ are fixed from the decomposition of the first distribution φ_0 and are the same for all the distributions.

3.3 Construction of the surrogate parametric metamodel

Real-time applications require fast and accurate computations, often with limited resources. Hence, the parametric

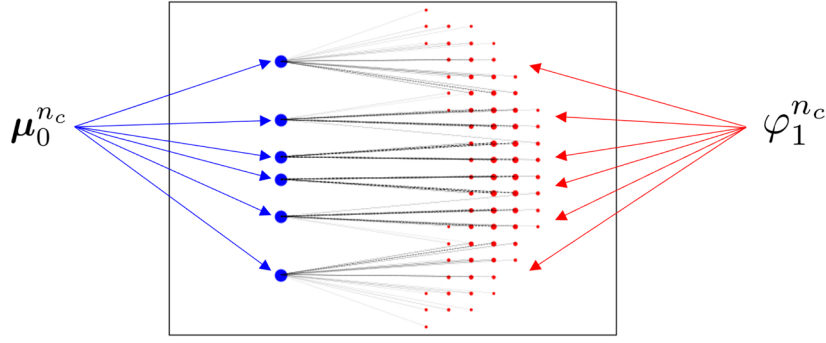


Fig. 7. Optimal matching between two point clouds (collections of particles.)

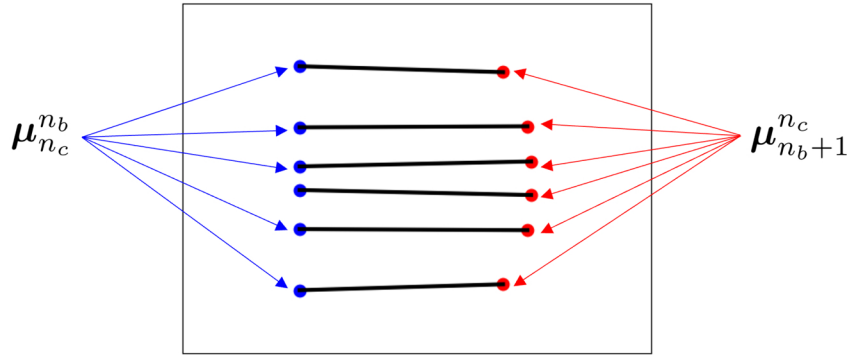


Fig. 8. Projection of particles $\mu_{n_b+1}^{n_c}$ from $\mu_{n_c}^{n_b}$ using the notion of barycentric projection proposed in equation (7).

displacement field between points in the sampling is learned using non-intrusive Model Order Reduction techniques. We ensure to find a sparse approximation that will allow estimating the particles' parametric displacement, to be able to represent the solution for any parameter choice without having to compute the associated solution.

3.3.1 Model Order Reduction using ANOVA-Proper Generalized Decomposition (ANOVA-PGD)

The ANOVA-PGD consists in combining the Analysis of Variance and the sparse Proper Generalized Decomposition (sPGD), to approximate solutions which depend on multiple parameters.

ANOVA-PGD works by decomposing the solution as a sum of functions defined with respect to the anchor point $\mathbf{p}_a$ [26] and then using interpolation to approximate the terms depending on a single dimension (parameter), defining N_D one-dimensional approximation problem. The difference between this approximation and the data, referred to as the residual, is then learned using a sparse PGD (sPGD).

We apply the ANOVA-PGD to learn the particles' positions stored in a vector $\boldsymbol{\mu}$ which will allow to reconstruct distributions $\bar{\varphi}$. Then, following the use of a sampling, it is possible to define a parametric vector $\boldsymbol{\mu}(\mathbf{p})$ where $\mathbf{p}$ are the parameters of the problem in the parametric space $\mathcal{D}$. The ANOVA decomposition of a function $\boldsymbol{\mu}(\mathbf{p})$ is an orthogonal decomposition based on the analysis of

variance and written as a sum of orthogonal functions [6]:

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{p}) &= \boldsymbol{\mu}(p^1, \dots, p^{N_D}) = \boldsymbol{\mu}_0 + \sum_{i=1}^{N_D} \boldsymbol{\mu}_i(p^i) \\ &+ \sum_{1 \leq i_1, i_2 \leq N_D} \boldsymbol{\mu}_{i_1, i_2}(p^{i_1}, p^{i_2}) + \dots + \boldsymbol{\mu}_{1, \dots, N_D}(p^1, \dots, p^{N_D}) \end{aligned} \quad (8)$$

The anchored-ANOVA allows computing all the one-dimensional functions from the DoE introduced in the Section 3.1 and presented in Figure 3. Then, the remaining terms of the ANOVA decomposition are grouped in the residual, function of all the variables, defined as:

$$\boldsymbol{\mu}_{RES}(\mathbf{p}) = \boldsymbol{\mu}(\mathbf{p}) - \boldsymbol{\mu}_{ANOVA}(\mathbf{p}), \quad (9)$$

where $\boldsymbol{\mu}_{RES}$ is approximated with the Sparse Proper Generalized Decomposition [27] and $\boldsymbol{\mu}_{ANOVA}(\mathbf{p})$ contains the first terms of the sum such as:

$$\boldsymbol{\mu}_{ANOVA}(\mathbf{p}) = \boldsymbol{\mu}_0 + \sum_{i=1}^{N_D} \boldsymbol{\mu}_i(p^i), \quad (10)$$

where the constant term $\boldsymbol{\mu}_0$ is the point cloud at the anchor point $\mathbf{p}_a$.

Once the model is built and $\boldsymbol{\mu}$ can be obtained for any choice of the parameter, φ can be recovered from $\boldsymbol{\mu}$ using the formula in (6). To recover f from φ , the total mass $I(\mathbf{p})$ is learned with a different surrogate parametric meta-model following the same methodology (ANOVA-PGD based regression).

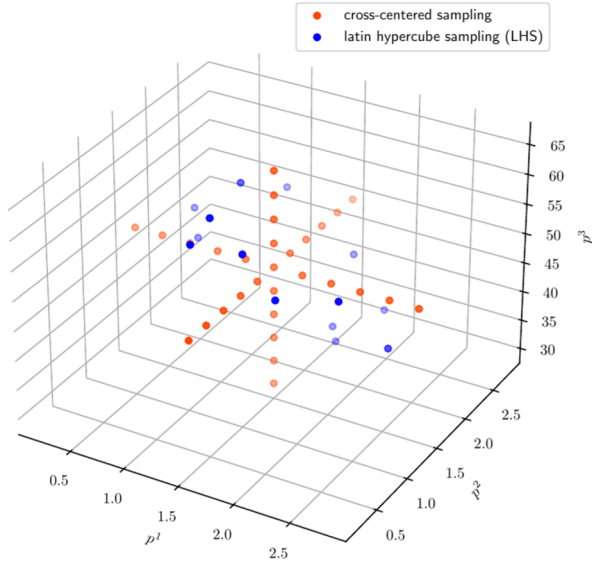


Fig. 9. Example of enrichment of the cross-centred sampling points introduced in the Figure 3. The blue points are the additional sampling points from a LHS.

3.3.2 Enrichment and correction of the parametric surrogate parametric metamodel

The question that arises now is, how to ensure that the surrogate parametric metamodel predicts particles' position μ in new points of the parametric space which fall outside of the cross and how to correct it?

One possibility is to consider additional parameter configurations based on the use of a Latin Hypercube Sampling (LHS) with the associated high-fidelity solution to self-enrich the parametric surrogate parametric metamodel. One example of LHS is depicted in Figure 9 and represented by the blue points. The sampling points are split into two sets: the training set B_{train} , $b_{train} \in \llbracket B_{train} \rrbracket$ that will allow to self-enrich the parametric model and the testing set B_{test} , $b_{test} \in \llbracket B_{test} \rrbracket$ in order to evaluate the approximation.

For each new sampling point, the parametric reduced order model trained with the cross-centred sampling predicts the particles' position $\hat{\mu}_{b_{train}}$ which allow reconstructing the distribution $\hat{\varphi}_{b_{train}}$. The transport plan between the predicted distribution and the reference one resulting from the normalised high-fidelity model distribution $\varphi_{b_{train}}$ is calculated. The displacement related to this transport plan represents a correction to add to the predicted particles' position $\hat{\mu}_{b_{train}}$. This approach is performed for every new point in the sampling. This correction is then learned as the residual of the ANOVA-PGD presented in 3.3.1. A larger learning sampling points leads to a better enrichment and therefore more precise predictions from the model for new unseen configurations.

To sum-up, for each new parameter configuration $b_{train} \in \llbracket B_{train} \rrbracket$, the model is corrected following a four-step process:

1. **Enrichment:** New high-fidelity solutions are generated associated to the additional sampling points

and normalised $\varphi_{b_{train}}$ on the domain Ω . Normalised solutions correspond to the weight at each node of the mesh $\mathbf{X} = (x_j, y_j)$.

2. **Prediction:** The surrogate parametric metamodel introduced in the Section 3.3.1 (Eq. (10)) is used to predict $\hat{\varphi}_{b_{train}}$ based on the particles' position $\hat{\mu}_{b_{train}}$ of the Gaussian functions for each new simulation b_{train} .
3. **Transport plan:** An Optimal Transport plan is computed between the prediction of each distribution $\hat{\varphi}_{b_{train}}$ and the normalised high-fidelity model distribution $\varphi_{b_{train}}$ by solving the Kantorovitch problem (Eq. (2)).
4. **Correction and training:** Using the barycentric projection presented in equation (7), the prediction of the particles' positions $\hat{\mu}_{b_{train}}$ is corrected and denoted $\hat{\mu}_{b_{train}}^c$. The parametric surrogate parametric metamodel is then self-enriched with $\hat{\mu}_{b_{train}}^c$ from the LHS sampling.

In detail, this approach corrects the prediction of the initial surrogate parametric metamodel by adding an additional term known as a correction term to each particle's displacement and denoted $(\delta_{b_{train}})_i$ for the i -th particle and are obtained from the matching between the output of the initial ROM and the new high-fidelity simulations. Enrichment is necessary to include the effect of the interaction between the parameters to the particles' positions. The corrected particles' position are obtained such as:

$$(\hat{\mu}_{b_{train}}^c)_i = (\hat{\mu}_{b_{train}})_i + (\delta_{b_{train}})_i, \quad (11)$$

as illustrated in Figure 10.

4 Model overview and online evaluation

In this section, we first summarise the methodology presented so far to build the offline phase. Then, the online model is evaluated for the testing configurations and the errors are calculated.

4.1 Model overview

The construction of the parametric model is composed of two phases: an offline phase of the surrogate construction and its online evaluation. In the offline phase, the aim is to build a surrogate parametric metamodel by combining Optimal Transport and the ANOVA-PGD technique to be able to predict online the solution for any set of parameters in the parametric space $\mathcal{D}$.

The model is built by the following steps:

- **High-fidelity simulations:** A database of high-fidelity solutions f_b from the parametric space $\mathcal{D}$ is obtained using a cross-centred sampling (DoE). Continuous solutions are normalised and denoted as distributions φ_b .
- **SPH decomposition and matching:** The distribution at the anchor point φ_0 is decomposed into P particles and, by dividing the sampling into subsets,

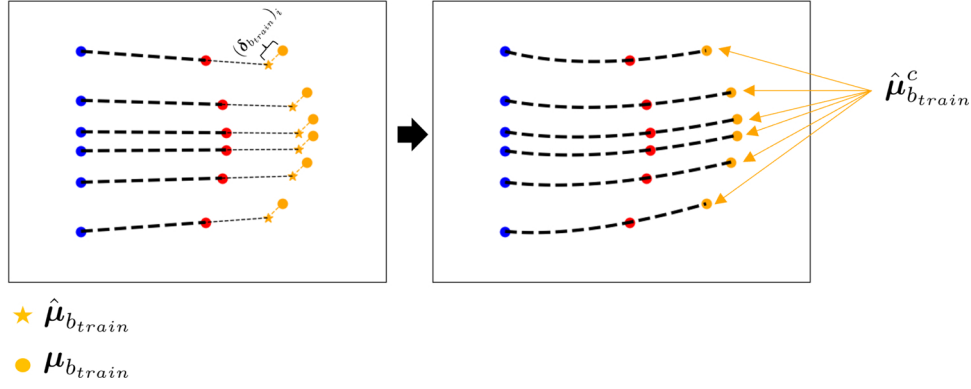


Fig. 10. Methodology of the enrichment and correction of the parametric surrogate parametric metamodel for the b_{train} new sampling point.

a step by step matching is performed using Optimal Transport and barycentric projection. Then, the particles' positions in each distribution $b \in \llbracket B \rrbracket$ are obtained and stored in the vector μ_b .

- **Surrogate parametric metamodel training:** The ANOVA-PGD technique is used to construct a surrogate parametric metamodel of the particles' position. The inputs are the parameters values of the problem $\mathbf{p}$ (cross-centred sampling points) and the outputs are the predicted particles' positions μ_B . The I_b norm used for the normalization of f_b is also learned with another metamodel.
- **Surrogate parametric metamodel enrichment and correction:** B_{train} additional sampling points are generated from a sparse sampling technique such as LHS. These new sampling points are used to run high-fidelity simulations. They are then used with the surrogate parametric metamodel to predict the particles' positions $\mu_{B_{train}}$. The surrogate parametric metamodel is corrected by adding an additional term to these predictions, coming from the matching performed between the predictions and the normalised high-fidelity solutions. The ANOVA-PGD is self-enriched with the predicted-corrected particles' positions denoted $\hat{\mu}_{B_{train}}^c$. New sampling points B_{test} from the Latin Hypercube sampling proposed in Figure 9 are used to check and quantify the accuracy of the model.
- **Online evaluation:** The surrogate parametric metamodel can be used for any set of unknown parameters and will predict the distribution $\hat{\varphi}$ using the particles' position as presented in the equation (6). To recover $\hat{f}$, $\hat{\varphi}$ is multiplied by the predicted norm $\hat{I}$.

4.2 Error evaluation

To evaluate the accuracy of the surrogate parametric metamodel B_{test} , $b_{test} \in \llbracket B_{test} \rrbracket$ sampling points from the LHS in Figure 9 are used. Then, predicted solutions $\hat{f}_{b_{test}}$ are compared with the reference solutions $f_{b_{test}}$ obtained from high-fidelity solutions. To recover the denormalised prediction $\hat{f}_{b_{test}}$, we first reconstruct $\hat{\varphi}_{b_{test}}$ using the predicted particles' positions $\hat{\mu}_{b_{test}}$ following equation (6) and

Table 1. Parametric space

Parameters	Minimum	Maximum
Coordinated of the Gaussian heat source Sx (cm)	0.2	2.8
Coordinated of the Gaussian heat source Sy (cm)	0.2	2.8
Initial temperature T_0 ($^{\circ}\text{C}$)	30	70

then multiply the result by the predicted mass $\hat{I}_{b_{test}}$. The relative error is calculated as follows:

$$E_{rr} = \frac{\|f_{b_{test}} - \hat{f}_{b_{test}}\|^2}{\|f_{b_{test}}\|^2}. \quad (12)$$

5 Numerical applications

In order to illustrate and analyse the proposed methodology, two different 2D heat problems are studied. In both cases, we consider a rectangular plate of dimensions $[0, 3] \times [0, 3]$ made of a conductive material (isotropic). The 2D heat equation is solved using the Finite Element (FE) method implemented in FEniCS [28] in the domain Ω . We consider the parametric space $\mathcal{D}$ were three parameters (also known as dimensions) are introduced, $\mathbf{p} = (p^1, p^2, p^3)$ which correspond respectively to Sx , Sy , the coordinates of the Gaussian heat source and the initial temperature T_0 . The sampling is the one presented in the Figure 9 referred to the parameter intervals in Table 1. We also referred to the element mesh in Figure 6.

5.1 Parametric 2D heat equation problem with one Gaussian heat source

First, we consider one unique heat source with the parameters $\mathbf{p} = (Sx, Sy, T_0)$. Then, the 2D problem reads:

$$\begin{cases} k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) = \rho C_P \frac{\partial T}{\partial t} & \text{in } \Omega \times [0, t_f], \\ \nabla T \cdot \mathbf{n} = 0 & \text{on } \partial\Omega \\ T(x, y, t = 0) = T_0 \frac{1}{2\pi\sigma^2} \exp \left(-\frac{(x-Sx)^2 + (y-Sy)^2}{2\sigma^2} \right) & \text{in } \Omega, \end{cases} \quad (13)$$

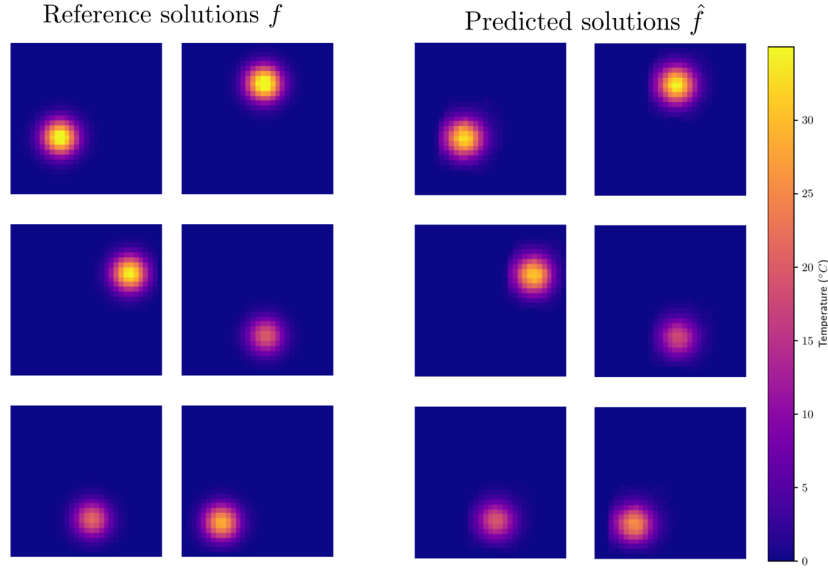


Fig. 11. Comparison between the reference solutions and the predicted solutions for $B_{test} = 6$.

Table 2. Relative errors obtained between the predicted solutions and the real solutions for $B_{test} = 6$.

Testing points $\mathbf{p}_{B_{test}}$	Relative error
$\mathbf{p}_1 = (Sx = 0.96, Sy = 1.10, T_0 = 57.94)$	0.1012
$\mathbf{p}_2 = (Sx = 1.62, Sy = 2.2, T_0 = 59.16)$	0.07888
$\mathbf{p}_3 = (Sx = 2.40, Sy = 2.10, T_0 = 53.81)$	0.0930
$\mathbf{p}_4 = (Sx = 1.66, Sy = 0.75, T_0 = 31.04)$	0.0891
$\mathbf{p}_5 = (Sx = 1.61, Sy = 0.72, T_0 = 33.08)$	0.07828
$\mathbf{p}_6 = (Sx = 0.76, Sy = 0.659, T_0 = 44.7)$	0.09723

In this case, and for each parameter configuration $\mathbf{p}_m$, $m \in \llbracket M \rrbracket$, where $M = B + B_{train}$ is the number of sampling points of the structured sampling plus the number of sampling points of the LHS as it shown in the Figure 9, $T(x, y, t = t_f)_m$ is a high-fidelity solution. After the normalisation, they are denoted φ_i . Then, based on the methodology described in Section 4, a surrogate parametric metamodel is created and returns a predicted particles' position $\hat{\boldsymbol{\mu}}_M$. The number of particles and the standard deviation are fixed with $P = 1500$ particles and $\sigma = 0.035$. In the sPGD part, $M_{modes} = 100$ modes were used with a classical polynomial basis. To evaluate the relative error of the surrogate parametric metamodel, equation (12) is used with $B_{test} = 6$ distributions. Results are presented in the Table 2.

The Figure 11 shows a comparison between reference solutions f from the high-fidelity model and reconstructed predicted solutions $\hat{f}$ from the surrogate parametric metamodel for $B_{test} = 6$.

5.2 Parametric 2D heat equation problem with two Gaussian heat sources

In the previous physical problem, the solution can be easily approximated by a Model Order Reduction technique because of the localised solution due to one Gaussian heat

source. To evaluate our approach, a more complex numerical example is formulated in this section while taking into consideration two Gaussian heat sources. Then, the 2D heat problem with the parameters $\mathbf{p} = (Sx, Sy, T_0)$ sampled as shown in Figure 9 reads:

$$\begin{cases} k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) = \rho C_P \frac{\partial T}{\partial t} & \text{in } \Omega \times [0, t_f], \\ \nabla T \cdot \mathbf{n} = 0 & \text{on } \partial\Omega \\ T(x, y, t = 0) = T_0 \frac{1}{2\pi\sigma^2} \exp^{-\frac{(x-Sx)^2 + (y-Sy)^2}{2\sigma^2}} \\ \quad + 1.2 \cdot T_0 \frac{1}{2\pi\sigma^2} \exp^{-\frac{(x-Sx)^2 + (y-Sy)^2}{2\sigma^2}} & \text{in } \Omega, \end{cases} \quad (14)$$

As presented in the first case, and for each parameter configuration $\mathbf{p}_m$, $m \in \llbracket M \rrbracket$, where $M = B + B_{train}$ is the number of sampling points of the structured sampling plus the number of sampling points of the LHS as it shown in the Figure 9, $T(x, y, t = t_f)_m$ is a high-fidelity solution. After the normalisation, they are denoted φ_i . Then, based on the methodology presented Section 4, a surrogate parametric metamodel is created and returns predicted particles' position $\hat{\boldsymbol{\mu}}_M$. The number of particles and the standard deviation are fixed with $P = 1500$ particles and $\sigma = 0.05$. In the sPGD part, $M_{modes} = 100$ modes were used with a classical polynomial basis. To evaluate the error of the surrogate parametric metamodel, equation (12) is used with $B_{test} = 6$ distributions. Results are presented in the Table 3.

Then, Figure 12 shows a comparison between reference solutions f from the high-fidelity model and reconstructed predicted solutions $\hat{f}$ from the surrogate parametric metamodel for $B_{test} = 6$.

6 Conclusion

In this paper, with the aim of speeding up the offline stage and increasing the accuracy of the former methodology,

Table 3. Relative errors obtained between the predicted solutions and the real solutions for $B_{test} = 6$.

Testing points $\mathbf{p}_{B_{test}}$	Relative error
$\mathbf{p}_1 = (Sx = 0.49, Sy = 1.10, T_0 = 57.94)$	0.09412
$\mathbf{p}_2 = (Sx = 1.62, Sy = 2.2, T_0 = 59.16)$	0.08015
$\mathbf{p}_3 = (Sx = 2.40, Sy = 2.10, T_0 = 53.81)$	0.09407
$\mathbf{p}_4 = (Sx = 1.66, Sy = 0.75, T_0 = 31.04)$	0.09044
$\mathbf{p}_5 = (Sx = 1.61, Sy = 0.72, T_0 = 33.08)$	0.08164
$\mathbf{p}_6 = (Sx = 0.76, Sy = 0.659, T_0 = 44.7)$	0.10078

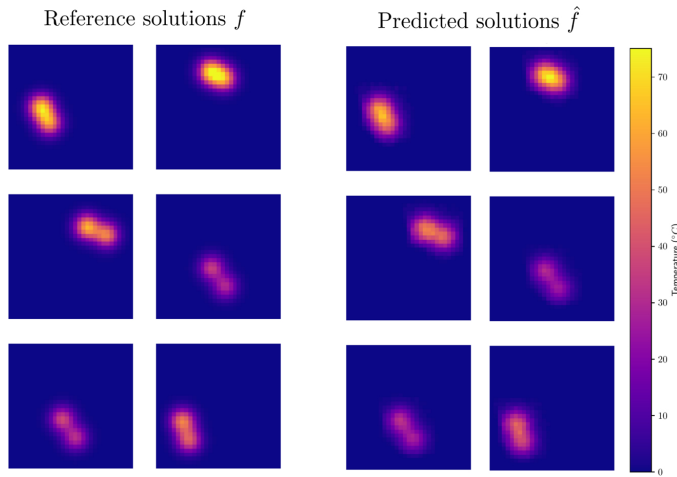


Fig. 12. Comparison between reference solutions and predicted solutions for $B_{test} = 6$.

we address Optimal Transport as a step by step matching algorithm, matching solutions by pairs rather than all together.

Relying on the ANOVA-PGD and its corresponding sampling strategy, the dimensionality of the complex matching problem is drastically reduced leading to an easily solved optimization problem. Moreover, based on the notion of barycentric projection, we no longer need to decompose all the solutions into Gaussian functions but just the one for the ANOVA anchor point. Also, the use of the ANOVA-PGD methodology ensures that the quantity of offline calculations remains reasonable as a result of the linear increase in sampling.

Finally, the ANOVA-PGD based surrogate is self-enriched with a classical LHS, relying again on the barycentric projection approach. This method enables the use of particle cloud representation of data, which may not typically be employed for this particular problem. To evaluate the efficiency, the work is in progress at extending our approach in a given complex industrial problem.

Funding

The work described in this paper was performed in the frame of the MORPHO project which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 101006854. This research was supported by the French National Association for Research and Technology (ANRT) through the research grant CIFRE 2020/0310.

Conflict of interest

All authors do not have conflict of interest to this submission.

References

- [1] O. Cecil Zienkiewicz, P.B. Morice, The finite element method in engineering science, Vol. 1977, McGraw-Hill London, 1971
- [2] F. Chinesta, P. Ladeveze, E. Cueto, A short review on model order reduction based on proper generalized Decomposition, Arch. Comput. Methods Eng. **18**, 395 (2011)
- [3] M. Verleysen, D. François, The curse of dimensionality in data mining and time series prediction, in Computational Intelligence and Bioinspired Systems: 8th International Work-Conference on Artificial Neural Networks, IWANN 2005, Vilanova i la Geltrú, Barcelona, Spain, June 8–10, 2005, Proceedings 8. Springer, 2005, pp. 758–770
- [4] A. Belloni, V. Chernozhukov, L1-penalized quantile regression in high-dimensional sparse models, 2011
- [5] G. Yi, J.Q. Shi, T. Choi, Penalized Gaussian process regression and classification for high-dimensional nonlinear data, Biometrics **67**, 1285–1294 (2011)
- [6] A. Sancarlos, V. Champany, J.-L. Duval, E. Cueto, F. Chinesta, Pgd-based advanced nonlinear multiparametric regressions for constructing metamodells at the scarce-data limit, [arXiv:2103.05358], 2021
- [7] U. Wilensky, Paradox, programming, and learning probability: a case study in a connected mathematics framework, J. Math. Behav. **14**, 253–280 (1995)
- [8] C. Villani et al., Optimal transport: old and new, Vol. 338, Springer, 2009
- [9] Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, Int. J. Comput. Vis. **40**, 99 (2000)
- [10] M. Perrot, N. Courty, R. Flamary, A. Habrard, Mapping estimation for discrete optimal transport, Adv. Neural Inform. Process. Syst. **29**, 4197–4205 (2016)
- [11] M. Cuturi, A. Doucet, Fast computation of wasserstein barycenters, in: International Conference on Machine Learning, PMLR, 2014, 685–693
- [12] P.A Knight, The sinkhorn-knopp algorithm: convergence and applications, SIAM J. Matrix Anal. Appl. **30**, 261–275 (2008)
- [13] P. Dvurechensky, A. Gasnikov, A. Kroshnin, Computational optimal transport: complexity by accelerated gradient descent is better than by sinkhorn's algorithm, in: International Conference on Machine Learning, PMLR, 2018, 1367–1376
- [14] X. Bian, Z. Li, G.E. Karniadakis, Multi-resolution flow simulations by smoothed particle hydrodynamics via domain decomposition, J. Comput. Phys. **297**, 132–155 (2015)
- [15] J. Weed, F. Bach, Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance, 2019
- [16] M. Jacot, V. Champany, F. Chinesta, J. Cortial, Parametric damage mechanics empowering structural health monitoring of 3D woven composites, Sensors **23**, 1946 (2023)
- [17] S. Torregrosa, V. Champany, A. Ammar, V. Herbert, F. Chinesta, Surrogate parametric metamodel based on

- optimal transport, *Math. Comput. Simul.* **194**, 36–63 (2022)
- [18] J. Rabin, G. Peyré, J. Delon, M. Bernot, Wasserstein barycenter and its application to texture mixing, in: *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*. Springer, 2012, pp. 435–446
- [19] N. Deb, P. Ghosal, B. Sen, Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections, *Adv. Neural Inform. Process. Syst.* **34**, 29736–29753 (2021)
- [20] G. Peyré, M. Cuturi, et al., Computational optimal transport, *Center for Research in Economics and Statistics Working Papers*, 2017-86, 2017
- [21] G. Peyré, M. Cuturi, et al., Computational optimal transport: with applications to data science, *Found. Trends Mach. Learn.* **11**, 355–607 (2019)
- [22] C. Villani, *Topics in optimal transportation*, Vol. 58, American Mathematical Society, 2021
- [23] N. Bonneel, M.V. De Panne, S. Paris, W. Heidrich, Displacement interpolation using lagrangian mass transport, in: *Proceedings of the 2011 SIGGRAPH Asia Conference*, 2011, pp. 1–12
- [24] R. Flamary, N. Courty, A. Gramfort, M.Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, et al., Pot: Python optimal transport, *J. Mach. Learn. Res.* **22**, 3571–3578 (2021)
- [25] R.J. McCann, A convexity principle for interacting gases, *Adv. Math.* **128**, 153–179 (1997)
- [26] M. El Fallaki Idrissi, F. Praud, V. Champaney, F. Chinesta, F. Meraghni, Multiparametric modeling of composite materials based on non-intrusive pgd informed by multi-scale analyses: application for real-time stiffness prediction of woven composites, *Composite Struct.* **302**, 116228 (2022).
- [27] R. Ibáñez, E. Abisset-Chavanne, A. Ammar, D. González, E. Cueto, A. Huerta, J.L. Duval, F. Chinesta, et al., A multidimensional data-driven sparse identification technique: the sparse proper generalized decomposition, *Complexity* **2018** (2018).
- [28] I. Torre, Diffusive solver: a diffusion-equations solver based on fenics, [arXiv:2011.04351], 2020

Cite this article as: M. Jacot, V. Champaney, S. Torregrosa Jordan, J. Cortial, F. Chinesta, Empowering optimal transport matching algorithm for the construction of surrogate parametric metamodel, *Mechanics & Industry* **25**, 9 (2024)