Real-time prediction by data-driven models applied to induction heating process

Khouloud Derouiche[1,2] · **Monzer Daoud**[2] · **Khalil Traidi**[3] · **Francisco Chinesta**[4]

Abstract
Data-driven modeling approach constitutes an appealing alternative to the finite element method for optimizing complex multiphysics parametrized problems. In this context, this paper aims at proposing a parametric solution for the temperature-time evolution during the multiphysics induction heating process by using a data-driven non-intrusive modeling approach. To achieve this goal, firstly, a set of synthetic solutions was collected, at some sparse sensors in the space domain and for properly selected process parameters, by solving the full-order finite element models using FORGE® software. Then, the gappy proper orthogonal decomposition method was used to complete the missing data. Next, the proper orthogonal decomposition method coupled with the nonlinear sparse proper generalized decomposition regression method was applied to find a low-dimensional space onto which the original solutions were projected and a model for the low-dimensional representations was, therefore, created. Hence, a real-time prediction of the temperature-time evolution and for any new process parameters could be efficiently computed at the predefined positions (sensors) in the space domain. Finally, spatial interpolation was carried out to extend the solutions everywhere in the spatial domain by applying a strategy based on the nonlinear dimensionality reduction by locally linear embedding method and the proper orthogonal decomposition method with radial basis functions interpolation. It was shown that the results are promising and the applied approaches provide good approximations in the low-data limit case.

Keywords Data-driven modeling · Induction heating process · Finite element method · Proper orthogonal decomposition · Gappy proper orthogonal decomposition · Nonlinear regression · Interpolation

Introduction

Most optimization problems in science and engineering require an abundant number of experiments to evaluate certain quantities of interest as a function of predefined process parameters. Besides, generating data is often regarded as a complex and a very expensive task that requires some expertise. In mechanics, for example, the optimization and the development of industrial processes are in continuous progress in order to enhance the fatigue behavior of many critically loaded mechanical parts.

In particular, induction hardening (IH) is one of the most appealing heat treatment processes widely employed in automotive and aerospace industries to provide optimal mechanical and microstructural material properties, corresponding to different predefined performance criteria [1, 2]. Basically, it consists of a rapid electromagnetic induction heating and a subsequent quenching phase. It has the advantage of providing a very short surface heat-up time, a precise

✉ Khouloud Derouiche
khouloud.derouiche@ensam.eu

Monzer Daoud
monzer.daoud@irt-m2p.fr

Khalil Traidi
khalil.traidi@safrangroup.com

Francisco Chinesta
francisco.chinesta@ensam.eu

[1] PIMM Laboratory, Arts et Métiers Institute of Technology, 151 boulevard de l'Hôpital, 75013 Paris, France

[2] French Technological Research Institute for Materials, Metallurgy and Processes (IRT- M2P), 4 rue Augustin Fresnel, 57070 Metz, France

[3] Safran Tech, Rue des Jeunes Bois, 78117 Châteaufort, France

[4] PIMM Laboratory, ESI Group Chair, Arts et Métiers Institute of Technology, CNRS, Cnam, HESAM Université, 151 boulevard de l'Hôpital, 75013 Paris, France

control of the treated zone, a good fatigue performance, a good reproducibility, and a more respectful operating mode to environmental requirements, compared to thermochemical treatments such as carbonitriding and carburizing [3]. However, the multidisciplinarity of the IH process in terms of its multiphysics property (electromagnetism, mechanical, metallurgical, and thermal fields) as well as the large number of involved process parameters, make it complex and difficult to handle.

Many works based on analytical and numerical methods have been reported in the literature in order to understand the involved physical mechanisms for better controlling of the influencing parameters and to ensure optimization in time and costs [4–7]. Among those methods, the finite element modeling (FEM) has shown a big success to deal with multiphysics-based parametrized problems thanks to the advanced numerical simulation codes [8]. However, the multi-query simulation approach becomes ineffective as far as the optimization procedure is concerned.

In order to overcome such issues, a data-driven parametric metamodeling approach coupled with a reduced-order modeling (ROM) technique for IH process constitutes an appealing alternative to the FEM approach. This is because this approach enables to compress the sampled data, by extracting the most "energetic" features that best describe the behavior of the system, and to achieve accurate parametric-based solutions at very low computational costs.

However, several problems may appear when dealing with data collected across space, time, or both of them and representing missing values. Such gaps can give rise to issues in terms of time series analysis, modal decomposition, or other data mining techniques requiring continuous data. Many researchers have been working on "gappy" data problems for many decades and many statistical approaches [9, 10] and non-statistical ones based on modal decomposition [11] have been used to provide an accurate reconstruction of the missing data.

Besides, another important issue when dealing with data-driven approaches is the quantity of sampling points needed to create the metamodel that lies the response to the input parameters. In fact, a large amount of data will provide a better approximation. However, the data generation is often expensive and consequently building a robust metamodel with a reduced amount of data is regarded as a challenging goal.

Furthermore, when a parametric solution is only known at few positions in a complex-shaped workpiece, extending the solution to more spatial points by interpolation is regarded as another key issue, especially when nonlinear structures are considered. To overcome such issue, dimensionality reduction approaches could be applied to the spatial data points in order to map their high-dimensional coordinates, in a global nonlinear structure, into a lower-dimensional space where a good interpolation of the solution could be obtained as a function of the new embedded coordinates. In the literature, different dimensionality reduction approaches have been developed which can be classified into linear (principal component analysis (PCA) [12] and linear discriminant analysis (LDA) [13]) and nonlinear (kernel PCA [14], locally linear embedding (LLE) [15], isometric mapping (ISOMAP) [16], etc.) techniques.

All of the above-mentioned difficulties were considered in the current work and appropriate approaches were applied to deal with them. This work aims at proposing a parametric metamodel that enables a real-time prediction of the temperature-time evolution in a spur gear of C45 steel during the induction heating process and under the effect of the process parameters. To achieve such a study, a data-driven non-intrusive ROM was applied.

Firstly, a set of synthetic solutions (called snapshots), representing the temperature-time evolution, was collected by using the multiphysics finite element (FE) simulations at some sparse points in the space domain and for different values of input process parameters associated with the latin hypercube sampling (LHS) design of experiments. Then, the gappy proper orthogonal decomposition (GPOD), which is an extension of the well-known proper orthogonal decomposition (POD) method, was applied to reconstruct an approximate solution for the missing data obtained from the FE simulations. Next, the POD with interpolation (PODI) method was applied as follows. The POD method was first applied on the repaired solutions to extract the reduced basis minimizing the error between the original snapshots and their orthogonal projections. The POD was then coupled with the nonlinear sparse proper generalized decomposition (sPGD) regression method, based on the separated representation approach, to fit the low-dimensional representation of the snapshot data, the so-called POD modal coefficients. Finally, extending the solution for more spatial data points in the heat-affected zones of the workpiece was carried out based on the nonlinear dimensionality reduction by LLE coupled with PODI for both surface and volume interpolation.

The applied methodology is summarized in Fig. 1, where it follows an offline–online strategy such that all the work was almost done at the offline phase while rapid computations were achieved on the online phase by using the constructed metamodels and the spatial interpolations.

The rest of the paper was organized as follows: Section "Dimensionality reduction and gappy data reconstruction" presents an overview on the applied methods starting with the POD and its extensions, then the nonlinear dimensionality reduction by LLE. Section "Problem statement: process and data" presents the studied induction heating process and the data generation. Section "Numerical results" illustrates the results of the created metamodel and the spatial interpolation for surface and volume cases, and section "Conclusions" addresses the main conclusions.

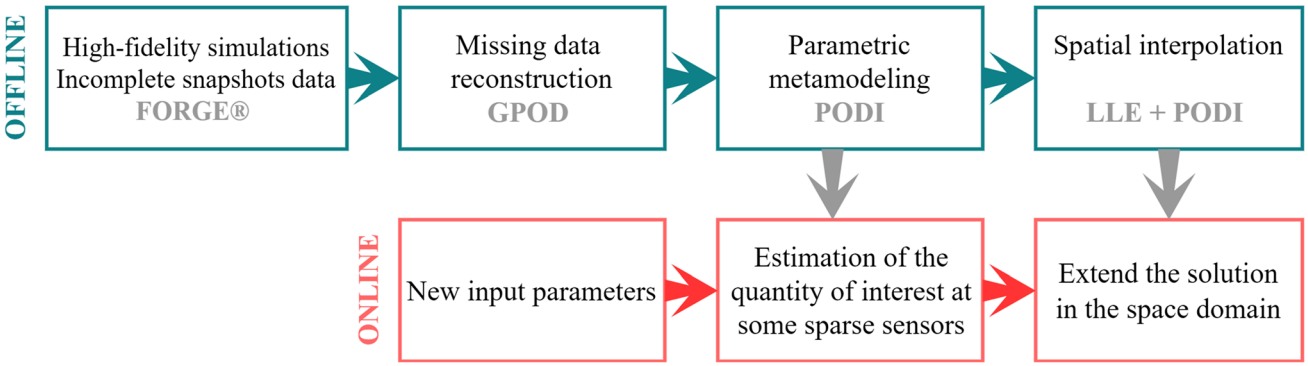


Fig. 1 Offline–online strategy for the temperature-time evolution during the induction heating process

Dimensionality reduction and gappy data reconstruction

Proper orthogonal decomposition (POD): fundamentals and extensions

POD

The POD is a method that enables to compute a reduced basis that provides a low-dimensional representation of a high-dimensional system state [17]. The reduced-order POD basis vectors are obtained thanks to the method of snapshots developed by Sirovich [18].

Consider a set of N field solutions or "snapshots", $\{\boldsymbol{T}_i\}_{i=1}^{N}$, where $\boldsymbol{T}_i \in \mathbb{R}^D$ is a vector containing the temperature field at D time steps of the discretized time intervals, and assuming that each snapshot is computed by solving the full-order model describing the problem for different values of the input parameters $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{is})$, where $\boldsymbol{\mu}_i$ is the vector containing the s parameters associated to the ith snapshot.

In order to find the reduced basis, minimizing the error between the original snapshots and their reconstructed solutions, the singular value decomposition (SVD) is applied on the snapshot matrix $\boldsymbol{T} \in \mathbb{R}^{D\times N}$, defined by $\boldsymbol{T} = [\boldsymbol{T}_1\ \boldsymbol{T}_2 \ldots\ \boldsymbol{T}_N]$, as follows:

$$\boldsymbol{T} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{1}$$

where $\boldsymbol{U} \in \mathbb{R}^{D\times D}$ and $\boldsymbol{V} \in \mathbb{R}^{N\times N}$ are unitary matrices containing the left and right singular vectors of $\boldsymbol{T}$, respectively. $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times N}$ is a rectangular diagonal matrix containing the singular values σ_i of $\boldsymbol{T}$, sorted in a decreasing order.

The reduced POD basis vectors, $\{\boldsymbol{\phi}_k\}_{k=1}^{R}$, is defined as the first R left singular vectors of $\boldsymbol{T}$ (i.e. first R columns of $\boldsymbol{U}$) corresponding to the R largest singular values. Therefore, the singular values provide a quantitative guidance for choosing the size of the POD basis, and a typical criterion to choose R is to find the cumulative energy captured by the POD basis vectors, also known as POD modes, that satisfies:

$$\frac{\sum_{i=1}^{R}\sigma_i^2}{\sum_{i=1}^{min(N,D)}\sigma_i^2} > \varepsilon \tag{2}$$

where ε is a certain threshold specified by the user, typically taken to be 99% or higher. Hence, an efficient representation of the snapshot data in a low-dimensional subspace of dimension R ($R \ll N$) can now be provided by a linear combination of the POD modes

$$\boldsymbol{T}_i = \boldsymbol{T}(t, \boldsymbol{\mu}_i) \approx \sum_{k=1}^{R} \alpha_{ki}\boldsymbol{\phi}_k(t) \quad \text{for} \quad i = 1, \ldots, N \tag{3}$$

where α_{ki} are called POD modal coefficients.

POD with interpolation (PODI)

While the POD-Galerkin model reduction, widely employed in different studies [19–22], intends to project the full-order problem, defined by partial differential equations (PDE), onto the POD reduced space to obtain a low-dimensional representation of the original operators, the PODI is a data-driven approach that relies only on data and does not require the original equations describing the system [23]. It enables a fast evaluation of the full-field solution for any value of the input parameters $\boldsymbol{\mu}$ by interpolating the reduced output representations defined by the POD coefficients with respect to $\boldsymbol{\mu}$.

This type of modeling is also referred as "Galerkin-free" reduced-order models [24], "Non-intrusive" POD [25, 26], or "surrogate" POD [27]. Different applications of PODI can be found in literature for automotive and aeronautic applications [28, 29] as well as for naval application [30].

The PODI proceeds in the same way as POD where the original snapshots are projected onto the subspace defined by the POD basis vectors in order to reduce their dimensionality then the POD modal coefficients α_{ki} are calculated using the inner product:

$$\alpha_{ki} = (\boldsymbol{\phi}_k, \boldsymbol{T}_i) \tag{4}$$

Then, an interpolation technique is applied to the set of POD modal coefficients $\{\alpha_{ki}\}_{i=1}^{N}$ and for each k, a surrogate model is constructed using a regression method such as sparse proper generalized decomposition (sPGD) [31–33], support vector regressions (SVRs) [34], gaussian processes (GPs) [35], etc.

Hence, the POD coefficients can be predicted for new values of $\boldsymbol{\mu}$ that are not included in the original set, and the prediction of the high-dimensional solution is achieved by reconstruction using Eq. 3.

In this work, the sPGD regression method was applied. For ease of explanation, a scalar function u that depends on a couple of parameters (μ_1, μ_2), with $u^i \equiv u(\mu_{1i}, \mu_{2i})$ known at N sampling points is considered in what follows, such that $u(\mu_1, \mu_2) : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$. The sPGD tries to find the approximate solution of u using a low-rank separated representation. By following the proper generalized decomposition (PGD) rationale [36], an approximated function of $u(\mu_1, \mu_2)$, denoted by $\tilde{u}^M(\mu_1, \mu_2)$, is expressed as a finite sum of M terms:

$$u(\mu_1, \mu_2) \approx \tilde{u}^M(\mu_1, \mu_2) = \sum_{j=1}^{M} P_j^1(\mu_1)\, P_j^2(\mu_2) \tag{5}$$

where P_j^1 and P_j^2 are one-dimensional functions depending on μ_1 and μ_2 respectively. The determination of the functions $P_j^1(\mu_1)$ and $P_j^2(\mu_2)$ is done by first approximating them using a set of basis functions as follows:

$$\begin{aligned} P_j^1(\mu_1) &= (\boldsymbol{N}_j^{\mu_1})^T \boldsymbol{a}_j \\ P_j^2(\mu_2) &= (\boldsymbol{N}_j^{\mu_2})^T \boldsymbol{b}_j \end{aligned} \tag{6}$$

where $\boldsymbol{N}_j^{\mu_1}$ represents the basis considered for approximating the jth mode depending on the μ_1 parameter, while $\boldsymbol{a}_j$ represents the associated weights, and similarly for the second direction (parameter μ_2). The choice of the set of basis functions, in which the one-dimensional functions is expressed, is made based on the studied problem. For example, a polynomial basis or a kriging basis can be selected.

Then, a minimization problem is considered to find $\tilde{u}^M$:

$$\tilde{u}^M = \arg\min_{u^*} \sum_{i=1}^{N} \left(u^i - u^*(\mu_{1i}, \mu_{2i})\right)^2 \tag{7}$$

where N is the number of sampling points for training models. The determination of the one-dimensional functions at each mode $j = 1,...,M$ is achieved by using a greedy algorithm such that the M th order term is calculated once the approximation up to order $M-1$ is known:

$$\tilde{u}^M(\mu_1, \mu_2) = \sum_{j=1}^{M-1} P_j^1(\mu_1)\, P_j^2(\mu_2) + P_M^1(\mu_1)\, P_M^2(\mu_2) \tag{8}$$

A nonlinear system of equations is derived, due to products of terms. Hence, an iterative scheme based on an alternating direction strategy is usually used to linearize the problem and to solve it [36].

In order to tackle more complex problems in the low-data limit case, where high oscillations may appear out of the training points because of the increased risk of overfitting, the sPGD uses the modal adaptivity strategy (MAS) [31]. The idea is to minimize spurious oscillations out of the training set by starting the PGD algorithm looking for functions with low degree. When it is observed that the residual decreases slowly or stagnates, higher order approximation functions are introduced. Therefore, oscillations are reduced, since a higher-order basis will try to capture what remains in the residual.

Gappy POD (GPOD)

The GPOD is an extension of POD to handle with incomplete data sets and to reconstruct an approximate solution for gappy data. It was first developed by Everson and Sirovich in the context of image reconstruction, such as human faces, from incomplete data [11] and was then applied successfully for reconstruction of PDE solutions in aerodynamics [23, 37, 38]. The GPOD has also used in the nonlinear model reduction methods of missing point estimation [39].

Consider the same set of snapshots $\{\boldsymbol{T}_i\}_{i=1}^{N}$ presenting the parametric temperature field and suppose that snapshots are not completely known and contain a certain percentage of gappiness, and consider their associated set of masks $\{\boldsymbol{n}_i\}_{i=1}^{N}$ defined as follows:

$$n_i^j = \begin{cases} 1 & \text{if } T_i^j \text{ is known} \\ 0 & \text{if } T_i^j \text{ is missing} \end{cases} \tag{9}$$

where T_i^j denotes the jth element of the vector $\boldsymbol{T}_i$. The core of the GPOD technique consists of an iterative implementation of POD, where the missing data for each snapshot are filled by an initial guess. The POD is then applied such that the guess for the data in the gaps is updated based on a POD approximation using a particular number of modes. The process is iterated until a convergence criterion is reached. The Everson-Sirovich procedure is described as follows:

– Step 1: fill in the missing elements of each snapshot $\boldsymbol{T}_i$, corresponding to $n_i^j = 0$, with an initial guess resulting in an initial complete field $\boldsymbol{T}_i^{(0)}$ such that:

$$T_i^{j(0)} = \begin{cases} T_i^j & \text{if } n_i^j = 1 \\ \bar{T}_i^j & \text{if } n_i^j = 0 \end{cases} \tag{10}$$

– Step 2: the POD analysis is performed on the snapshot set $\{\boldsymbol{T}_i^{(l)}\}_{i=1}^N$ ($l = 0$ for the first iteration) using P modes, $\{\boldsymbol{\phi}_k\}_{k=1}^P$, resulting in an intermediate repaired data for the current iteration l:

$$\tilde{\boldsymbol{T}}_i^{(l)} \approx \sum_{k=1}^{P} \beta_{ki}^{(l)} \boldsymbol{\phi}_k^{(l)} \quad \text{for} \quad i = 1, \ldots, N \tag{11}$$

such that the POD modal coefficients $\beta_{ki}^{(l)}$ are computed by minimizing the error between the original and repaired vectors using the gappy norm where only elements for which data are available in $\boldsymbol{T}_i$ are compared. The error is defined as:

$$E = \left|\left| \boldsymbol{T}_i - \tilde{\boldsymbol{T}}_i^{(l)} \right|\right|_{\boldsymbol{n}_i}^2 \tag{12}$$

This minimization leads to the linear system of algebraic equations:

$$\boldsymbol{K}^{(l)} \boldsymbol{\beta}^{(l)} = \boldsymbol{f}^{(l)} \tag{13}$$

where $K_{pq}^{(l)} = (\boldsymbol{\phi}_p^{(l)}, \boldsymbol{\phi}_q^{(l)})_{\boldsymbol{n}_i}$, $f_p^{(l)} = (\boldsymbol{T}_i, \boldsymbol{\phi}_p^{(l)})_{\boldsymbol{n}_i}$, and $(.,.)_{\boldsymbol{n}_i}$ is the gappy inner product.

– Step 3: each snapshot $\boldsymbol{T}_i$ is reconstructed by replacing its missing elements by the corresponding repaired elements in $\tilde{\boldsymbol{T}}_i^{(l)}$, and the previous guesses are overwritten for the next iteration as follows:

$$T_i^{j(l+1)} = \begin{cases} T_i^j & \text{if } n_i^j = 1 \\ \tilde{T}_i^{j(l)} & \text{if } n_i^j = 0 \end{cases} \tag{14}$$

– Step 4: repeat step 2 and 3 until convergence while considering the reconstructed fields in step 3 as the new initial guesses for the next iteration.

As described above, the GPOD method attempts to improve an initial guess at the missing data using P POD modes for the reconstructions. However, the optimum number of modes, for which the error is the smallest among all possible converged reconstructions, depends on the initial guess. In order to avoid this dependency, an extension of the Everson–Sirovich procedure was developed in [40] where the GPOD was applied iteratively with an increasing number of modes, improving the accuracy of the full-field reconstruction. The steps of the extended procedure are summarized as follows:

– Step 1: perform the standard Everson–Sirovich procedure with a low value of modes in the reconstruction ($P_0 = 1$ for example).
– Step 2: the converged result from the previous step is used as a new initial guess and the Everson–Sirovich procedure is then reapplied, but with $P_1 = P_0 + 1$ modes in the reconstruction.
– Step 3: proceed similarly until a convergence criterion is reached.

Therefore, the final approximate solution will only depend on the degree of gappiness and not on the initial guesses of the gappy regions. It is worth noting that if the percentage of missing data per snapshot is very high, the GPOD technique might fail to provide an accurate approximation.

A suitable convergence criterion for the selection of the optimum number of modes P is important. This is because the reconstruction error does not monotonically decrease with P. Among the proposed criteria listed in [40, 41], the one based on the total energy of the POD reconstruction is presented here, where the optimal number of modes is defined as:

$$\text{Choose } P \text{ such that: } \quad |E_P - E_0| \quad \text{is maximum, where} \quad E_P = \sum_{i=1}^{N} \sigma_{i,P}^2 \tag{15}$$

Here $\sigma_{i,P}^2$ is the ith POD eigenvalue of the P-modes reconstructed data, knowing that the eigenvalues are given by the squares of the singular values. E_P is the total energy for the converged reconstruction using P modes and E_0 is the total energy associated with a fixed reference value. It is worth noting that the sum of eigenvalues is calculated over all the modes, not just the P modes used for reconstruction.

Nonlinear manifold learning by LLE

LLE is a nonlinear dimensionality reduction method proposed by Roweis and Saul [15]. It is defined as an unsupervised manifold learning algorithm that computes a low-dimensional embedding manifold of high-dimensional data by mapping the inputs into a single global coordinate system of lower dimensionality while preserving their neighborhood. In other words, the close points in the high-dimensional space should also be close in the low-dimensional embedding space, and the same for the far points.

The basic idea of LLE is to recover global nonlinear structure from locally linear fits [42–44]. Hence, nonlinear dimensionality reduction can be simplified into local linear dimensionality reduction where each data in the high-dimensional space can be expressed by a linear combination of its nearest neighbors, and then it can be reconstructed on the low-dimensional embedding manifold by minimizing a cost function.

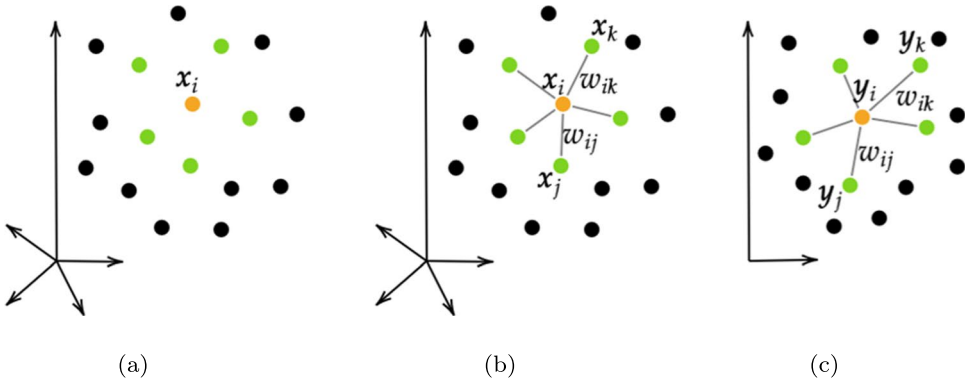


Fig. 2 Steps of locally linear embedding: (**a**) assign k nearest neighbors, (**b**) reconstruct with linear weights, and (**c**) map to embedded coordinates

Consider a set of N real-valued input points of D dimensional vectors, $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}$. It is assumed that sufficient well-sampled data are provided such that these data points lie on or near a smooth nonlinear manifold of lower dimensionality d ($d < D$). The LLE algorithm has three steps, as illustrated in Fig. 2:

- Step 1: assign k nearest neighbors to each data point $\boldsymbol{x}_i$ (Fig. 2(a)). The k nearest neighbors are identified using pairwise euclidean distance between the data points. When data points are widely separated, a large value of k is needed, and a small k is required when the density of the data is high.
- Step 2: compute the weights w_{ij} that provide the best linear reconstruction of each point $\boldsymbol{x}_i$ from its neighbors (Fig. 2(b)). The weights w_{ij} identify the contribution of the jth data point to the ith one. The optimal reconstruction weights are computed by minimizing reconstruction errors measured by the cost function:

$$\epsilon(\boldsymbol{W}) = \sum_{i=1}^{N} \left| \left| \boldsymbol{x}_i - \sum_{j=1}^{N} w_{ij} \boldsymbol{x}_j \right| \right|^2 \tag{16}$$

  where $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ is the weight matrix that includes the weights of linear reconstruction of all data points using their neighbors.

  The minimization problem should satisfy two constraints: (1) the weights of linear reconstruction for each point satisfy $\sum_j w_{ij} = 1$ for $i = 1,...,N$; (2) each data point $\boldsymbol{x}_i$ is reconstructed by its neighbors such that $w_{ij} = 0$ if $\boldsymbol{x}_j$ does not belong to $\boldsymbol{x}_i$ neighbors set.
- Step 3: compute the low-dimensional embedding vectors, $\boldsymbol{Y} = \{\boldsymbol{y}_i\}_{i=1}^{N}$, based on the idea that LLE preserves the local linearity from neighbors and the corresponding reconstruction weights (Fig. 2(c)). Therefore, the same weights computed in step 2 for the linear reconstruction of data points in the D dimensional space are used. The global internal coordinates on the embedding manifold are computed by minimizing the following embedding cost function:

$$\boldsymbol{\Phi}(\boldsymbol{Y}) = \sum_{i=1}^{N} \left| \left| \boldsymbol{y}_i - \sum_{j=1}^{N} w_{ij} \boldsymbol{y}_j \right| \right|^2 \tag{17}$$

  subject to two constraints: $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i \boldsymbol{y}_i^T = \boldsymbol{I}_{d \times d}$ and $\sum_{i=1}^{N} \boldsymbol{y}_i = 0$, where $\boldsymbol{y}_i \in \mathbb{R}^d$ ($d < D$) is the ith embedded data point.

For more details about the resolution of minimization problems, the reader can refer to [45]. Compared with other nonlinear dimensionality reduction methods, LLE has only one hyperparameter, which is the number of neighbors k. Hence, the algorithm implementation is quite simple.

Problem statement: process and data

In the present work, the induction heating simulation was carried out on a C45 steel spur gear of module 2.5 and 22 teeth. The gear of 10 mm width was surrounded by a copper coil having a 65 mm inner diameter and a rectangular section of 8 × 5 mm. During heating, electromagnetic and thermal fields in addition to the phase transformation are combined. In this study, two process parameters were considered as listed in Table 1.

Table 1 Input parameters and their lower and upper limits

Input Parameters	Lower Limit	Upper Limit
Frequency (kHz)	10	250
Power (kW)	50	600

Changing these input parameters could highly impact the process outputs. Hence, a parametric metamodel for the temperature-time evolution was constructed in order to evaluate this effect. To this end, a set of precomputed high-fidelity solutions was collected by solving the original full-order FE models for different values of input parameters using the commercial FE software FORGE®.

Heating simulation was carried out using two coupled solvers available in FORGE®. The first one is the electromagnetic solver to solve Maxwell's equations based on the electromagnetic period, which is divided into a series of 60-time increments, to give rise to a heating power. The second one is the thermal solver to compute the temperature field by solving the heat equation using the calculated heating power. The time step in this latter is managed by FORGE® such that a coarse time step could be found for simulations having low values of input parameters and a dense one for those with high values of input parameters. More details about the numerical simulation could be found in [46].

It is worth noting that only one half of the gear tooth was modeled, and two symmetry planes were imposed in order to improve the computational efficiency. The implemented material properties of C45 steel were taken from the literature [47–51]. According to the LHS design of experiments (DoE) [52], a total of 20 FE simulations were generated, as shown in Fig. 3. The temperature-time evolution was extracted at 14 specific points, obtained by implementing sensors on the studied geometry, as shown in Fig. 4, and a model for each sensor was then constructed.

All simulations were supposed to finish at the same time instant, initially chosen to be equal to 1 sec, which is sufficient for such a rapid heating process. However, some simulations cannot reach this time, generating hence incomplete data sets which can highly affect the data analysis.

To overcome this issue, reconstructing the missing data and modeling the temperature as a function of time were carried out because it seems to be more suitable and more general for spatial interpolation.

Numerical results

Modeling the temperature-time evolution

In fact, the missing data appeared at the last reached time instant such that the percentage of missingness, defined by the number of temporal data points that are missing with respect to the total number of time instants, was different from one simulation to another. Among the total number of simulations, 55% of them had a completely known data vectors because they reached the end of computation and the other ones had a gappiness percentage varying between 25% and 75%. In order to reconstruct an approximate solution for the incomplete data, the extended Everson-Sirovich procedure described in section "Gappy POD (GPOD)" was applied.

The number of modes was initialized to $P = 1$ and increased iteratively, enabling an improved full-field reconstruction accuracy, and the gappy values per snapshot were initially filled by a constant value equal to the last calculated temperature. An example of the initial temperature field evolutions (complete and gappy) and their reconstruction using GPOD are shown in Fig. 5 for

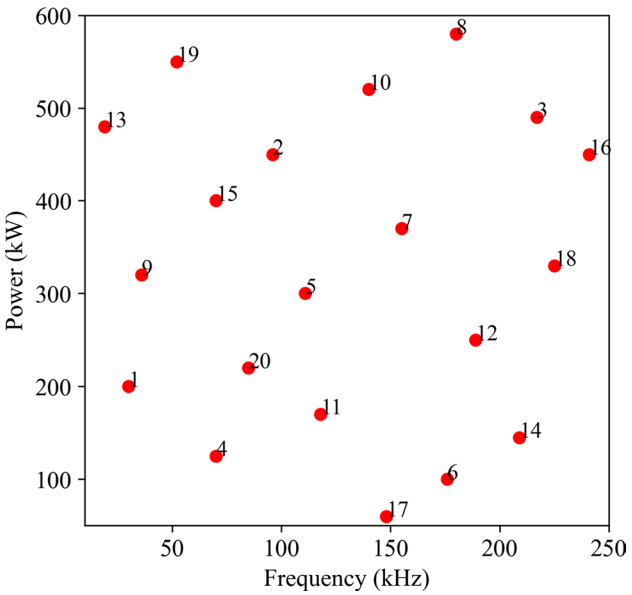


Fig. 3 LHS design of experiments

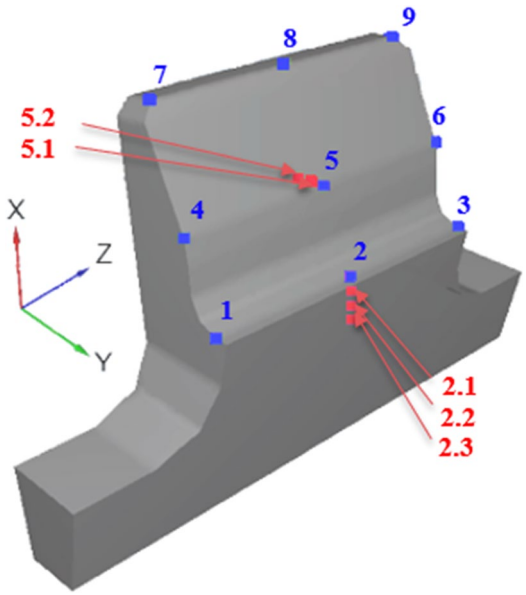
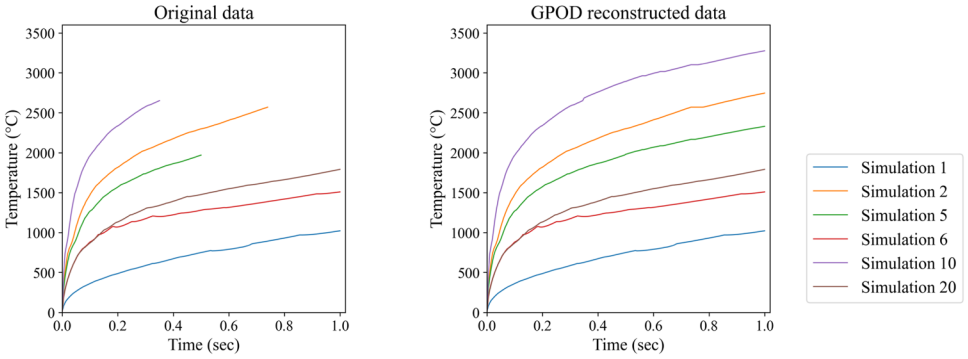
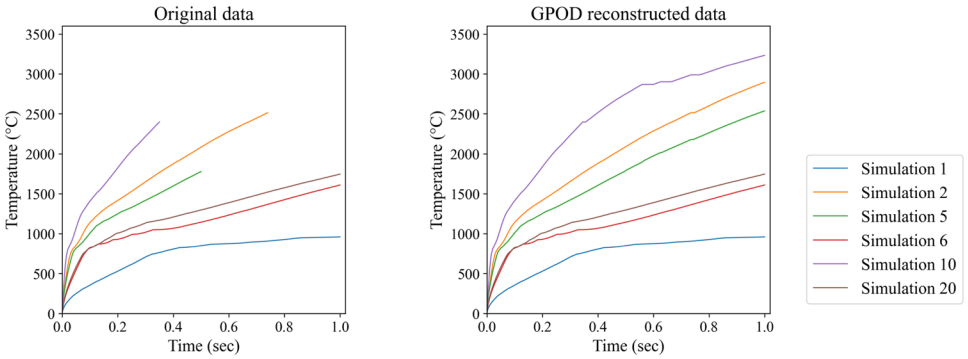


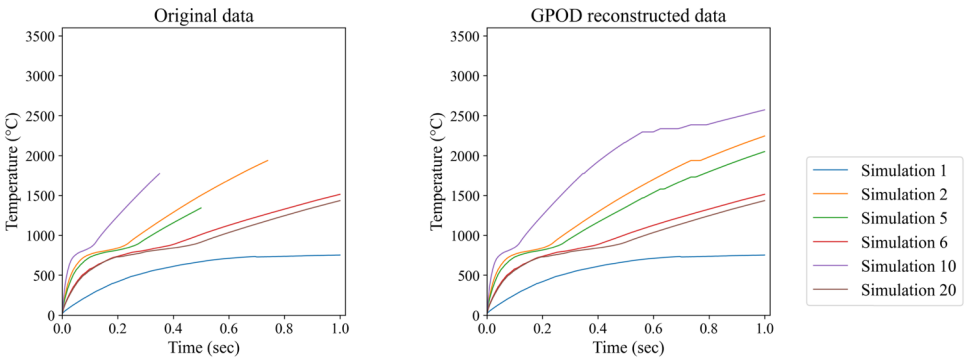
Fig. 4 Sensors positions. Blue points represent the surface sensors while the red ones represent those located at a certain depth



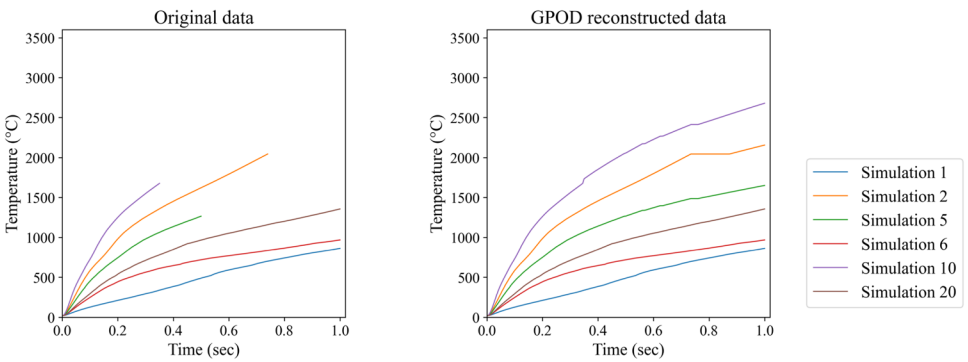
(a) Sensor #1



(b) Sensor #5



(c) Sensor #9



(d) Sensor #2.2

Fig. 5 Original and reconstructed gappy data at 4 sensors

6 different simulations. It is worth pointing out that only results for sensors (#1, #5, #9, and #2.2) in reference to Fig. 4 are shown for the sake of clarity.

Since the gappy fields are completely unknown, comparing the approximated gappy values with the "true" ones can not be achieved. It is worth mentioning that the reconstructed missing data of temperature are unreachable in reality since they exceed the melting temperature of C45 steel. However, having solutions with the same discretization in time was needed to get a well-posed problem for applying the POD. In this work, a time interval composed by 201-time steps was considered for the post-processing of the data. Since the temperature-time evolution is monotonic and doesn't show any local phenomena, a coarser discretization will not affect the predicted results.

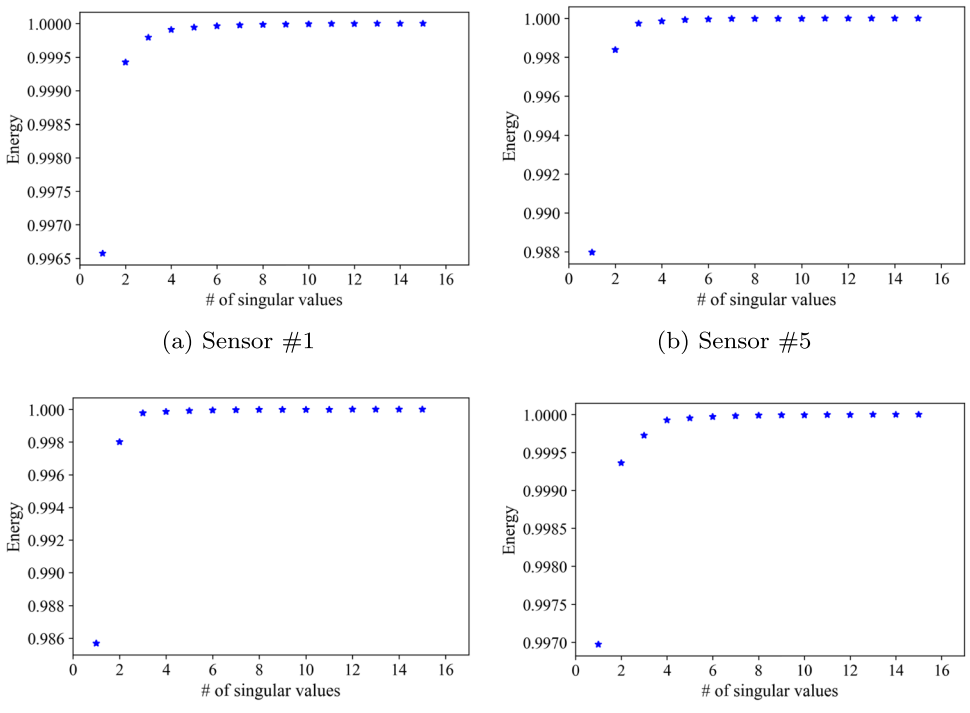
Once the gappy data reconstruction were established for each sensor, simulations were split into two subsets. Simulations 2, 9, 11, 12, and 16 inside the DoE were used as an error indicator. While the other 15 simulations were taken into account to construct the parametric solution. In other words, 75% of data were used to build the models and 25% to evaluate their prediction accuracy.

The PODI was then performed on the 15 simulations (snapshots) and for each sensor. The average of the temperature was first computed over the snapshot set and the mean value was subtracted from each snapshot in order to improve the numerical conditioning and avoid the magnitude of the first singular value numerically dominating all others. The Eq. 3 becomes:

$$\boldsymbol{T}_i = \boldsymbol{T}(t, \boldsymbol{\mu}_i) \approx \bar{\boldsymbol{T}} + \sum_{k=1}^{R} \alpha_{ki} \boldsymbol{\phi}_k(t) \quad \text{for} \quad i = 1, \ldots, N \tag{18}$$

By applying the POD, the left singular vectors of the snapshot matrices were truncated to the two or three first singular vectors, depending on the sensor reconstructed snapshots. This choice was made such that more than 99.9% of the cumulative energy was recovered, as shown in Fig. 6 at 4 different sensors.

Following the steps defined in section "POD with interpolation (PODI)", the original snapshots per sensor were projected on their corresponding reduced basis and the POD modal coefficients were computed. A surrogate model for the POD modal coefficients and for each sensor was then constructed by using the sPGD regression method, and by considering the frequency and power as input parameters for the model. Fig. 7 shows the temperature-time evolution obtained by the full-order FE model and the sPGD regression model (i.e. using Eq. 18 to reconstruct the approximate solution using the predicted values of the modal coefficients as a function of the frequency and power) for training and testing data, respectively. It can be seen that the predicted temperature-time profiles obtained by sPGD model (dashed red curves) were in good agreement with the FE ones (blue



(a) Sensor #1 (b) Sensor #5

(c) Sensor #9 (d) Sensor #2.2

Fig. 6 Cumulative energy computed from 15 snapshots of the thermal field at 4 different sensors

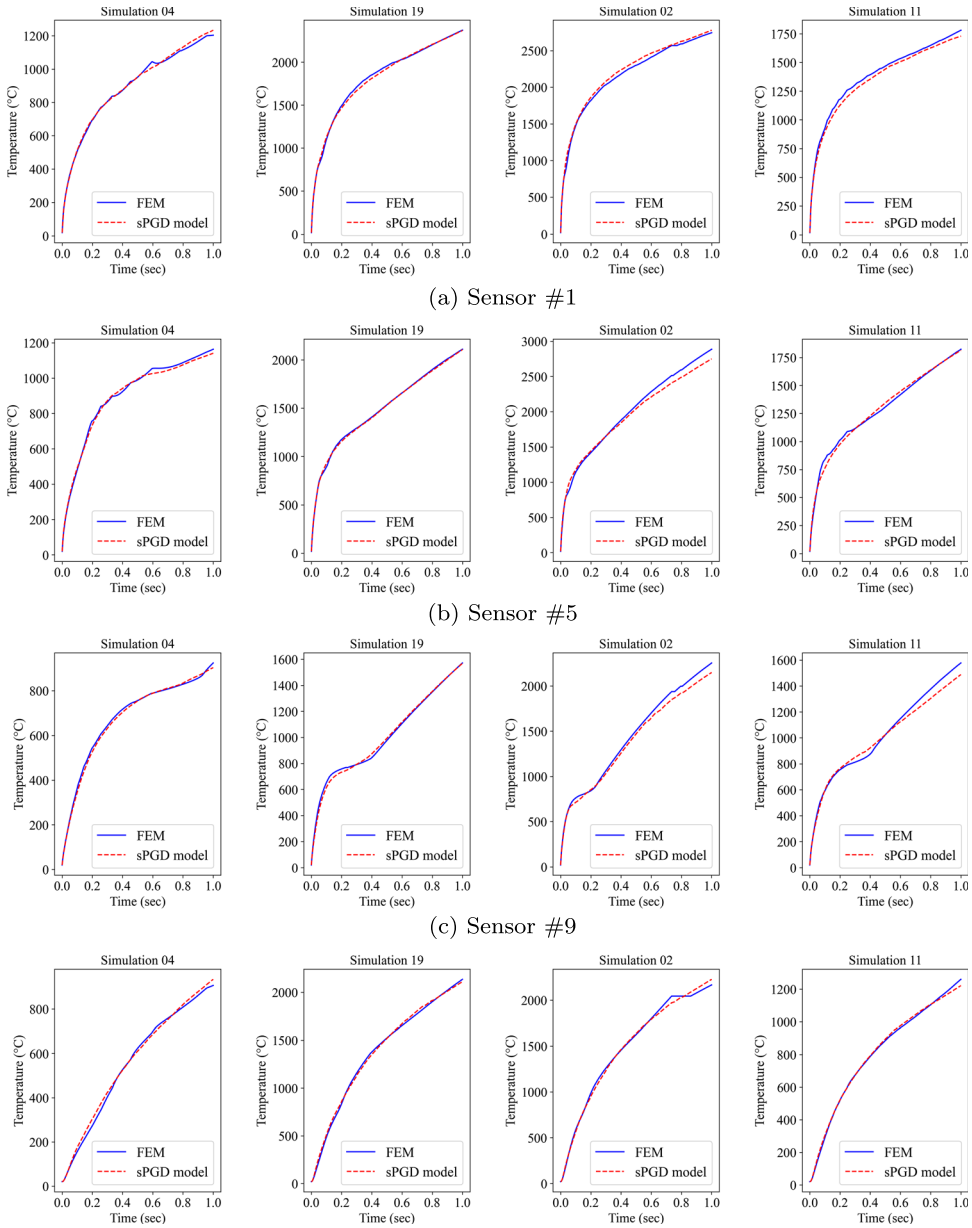


Fig. 7 Comparison between the reference FEM and the sPGD regression model of the temperature-time evolution at 4 sensors: simulations 4 and 19 were used to create the regression model and simulations 2 and 11 were used to verify the regression model accuracy

curves) even with different levels of temperature under the imposed values of input parameters. This shows that the applied methodology provides good parametric solutions even with a small amount of data.

The relative error to measure the prediction accuracy for training and testing data sets is presented in Table 2. The computed error is defined by:

$$\text{Error}(\%) = \frac{1}{n}\left(\sum_{i=1}^{n} \frac{\sqrt{\int (\boldsymbol{T}_i^{pred} - \boldsymbol{T}_i^{FEM})^2 dt}}{max(\boldsymbol{T}_i^{FEM})}\right) \times 100 \qquad (19)$$

Table 2 Relative errors of temperature-time evolution using sPGD model

Error (%)	sensor #1	sensor #5	sensor #9	sensor #2.2
Error_{train}	1.1%	0.9%	1.1%	1.2%
Error_{test}	2.9%	2.8%	3.0%	2.5%

where n, $\boldsymbol{T}_i^{FEM}$, $\boldsymbol{T}_i^{pred}$ are the number of data points (simulations), the vectors of the FEM solutions, and the predicted solutions, respectively. As shown in Table 2, the obtained errors do not exceed 3% for testing data at the 4 considered

sensors. These results indicate that the sPGD model gives a quite good prediction.

Spatial interpolation

A model for the temperature-time evolution was constructed for each sensor (illustrated in Fig. 4). Extending the approximate solutions for more spatial points in the heat-affected zones of the gear tooth was required. Interpolation at the extreme surface of the workpiece was first studied. Then, volume interpolation was carried out at a certain depth.

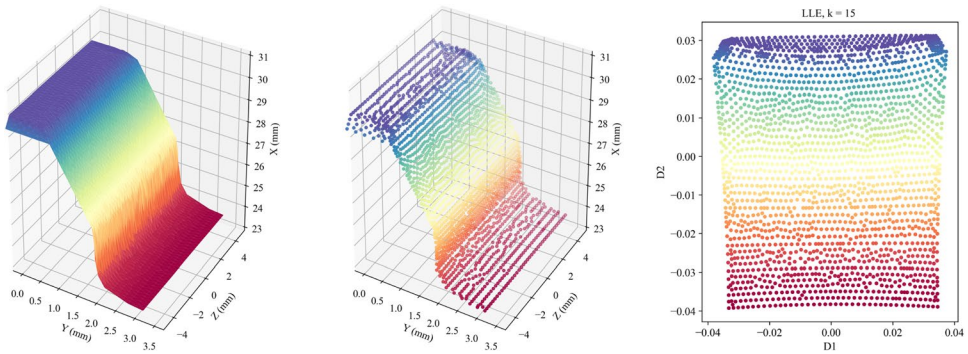
Surface interpolation

As shown in Fig. 4, only 9 sensors are considered at the gear tooth surface, and hence the approximate solution is only known at those sensors. The lack of data at other points in addition to the nonlinear geometry of the studied half-gear tooth make the interpolation difficult to achieve. To overcome such problem, a nonlinear dimensionality reduction by LLE was applied on the scattered data, representing the mesh of the half-tooth surface, in order to map the 3D spatial coordinates into a 2D embedding manifold while preserving the neighborhood. Therefore, the interpolation becomes much easier to handle.

The applied methodology for surface interpolation is defined as follows. The surface was first isolated from the rest of the workpiece as shown in Fig. 8(a). The mesh of the surface was then extracted as shown in Fig. 8(b) where the scattered points represent the mesh nodes (2092 points). LLE was next applied, as described in section "Nonlinear manifold learning by LLE", where the nearest neighbors ($k = 15$) were determined by euclidean distance. This number of neighbors k was chosen in accordance with the density of data points, and it was assumed to be enough to approximate each data point.

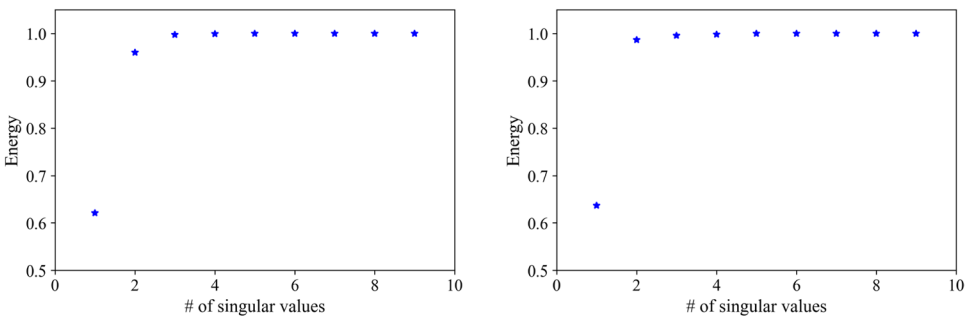
The result of LLE is illustrated in Fig. 8(c). It can be seen that the three-dimensional data were mapped into a single global coordinate system of two dimensions, where the color coding illustrates the neighborhood preserving mapping.

The next step was to apply the PODI, beginning by considering a snapshot matrix where each column represents the temperature-time evolution per sensor (the 9 available sensors). In fact, those snapshots should be obtained at the online step for new selected values of input parameters (frequency and power). However, two simulations from the existed DoE, presented in section "Problem statement: process and data', were considered for validation. All the results



(a) (b) (c)

Fig. 8 LLE approach: (**a**) isolated surface geometry (**b**) mesh nodes in 3D coordinates (**c**) embedded coordinates in the 2D manifold



(a) Simulation 1 (b) Simulation 14

Fig. 9 Cumulative energy computed from 9 snapshots of the thermal field for 2 different simulations

will be shown for simulation 1 and 14 where the medium and high frequency values were used.

By applying the POD on the snapshot matrices, reduced bases composed by three POD modes, for both simulations 1 and 14, were retained to recover the 99% of cumulative energy and to produce accurate reconstructions, as shown in Fig. 9.

Then, the POD modal coefficients were approximated by using interpolation with radial basis functions (RBF) which provide a general and flexible way of interpolation in multidimensional spaces, even for unstructured data where polynomial or spline interpolations are difficult to apply [53–55]. The RBF interpolation with a multiquadric basis function was applied to the POD modal coefficients as a function of the global internal coordinates in the low-dimensional space obtained by LLE.

The available sensor positions in the three-dimensional space and their corresponding coordinates in the embedded manifold, used for interpolation, are illustrated with blue points in Fig. 10(a) and (b), respectively. The validation of the applied methodology was provided at 4 new positions (orange points) as shown in Fig. 10.

A comparison between the FEM and the PODI approximation of the temperature-time evolution is shown in Fig. 11 for the test points illustrated in Fig. 10. It can be seen that the two curves were almost overlapped for all test points.

In addition, the relative error defined by Eq. 20 was used to determine the estimation accuracy for the test points and for the two simulations.

$$\text{Error}(\%) = \frac{\sqrt{\int (\boldsymbol{T}^{pred} - \boldsymbol{T}^{FEM})^2 dt}}{max(\boldsymbol{T}^{FEM})} \times 100 \tag{20}$$

As shown in Table 3, the obtained relative errors were less than 3% at the 4 test points. These results indicate that the applied methodology for surface interpolation was robust and provided accurate results.

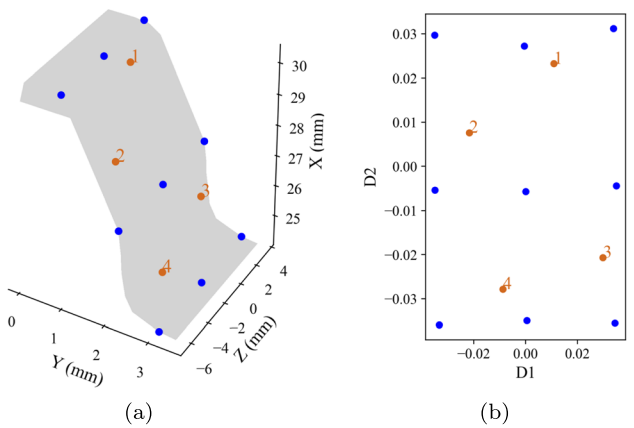


Fig. 10 Sensors in 3D surface (**a**) and their corresponding coordinates in the embedded manifold (**b**)

Table 3 Relative errors of the temperature-time evolution using POD with RBF interpolation - surface interpolation

Error (%)	point #1	point #2	point #3	point #4
Simulation 1	1.5%	2.0%	2.2%	2.4%
Simulation 14	1.0%	1.3%	3.0%	1.4%

This methodology can successfully be applied to other geometries. However, it is worth noting that the position of the considered sensors is very important and directly affect the interpolation accuracy and a good choice of the position is highly recommended.

Volume interpolation

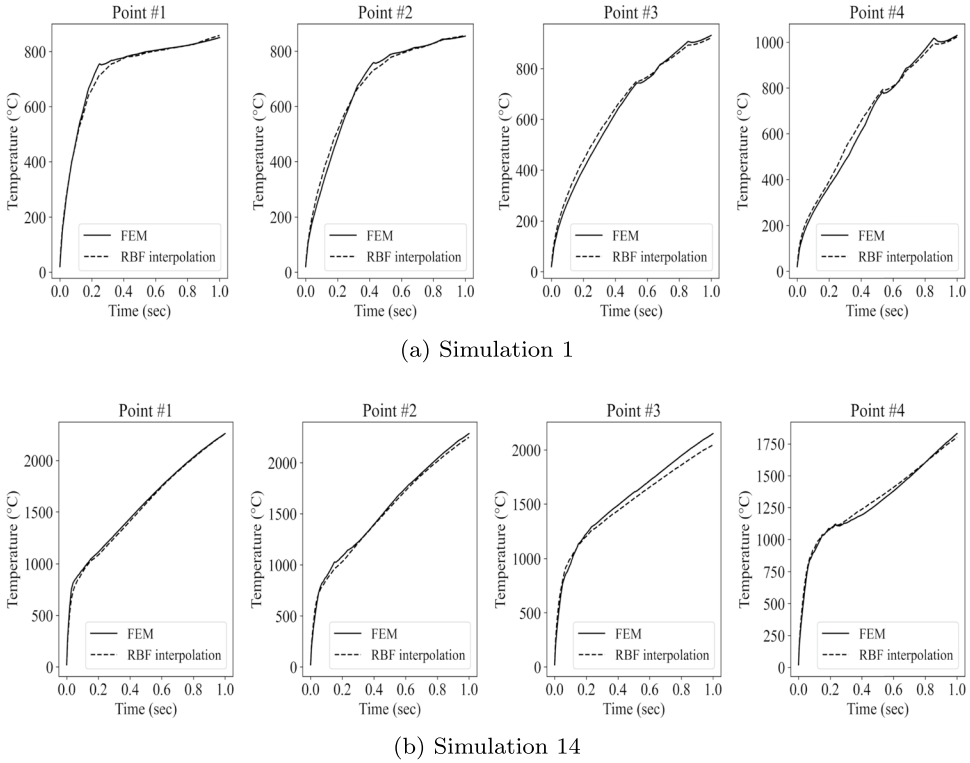
According to Fig. 4, only few sensors are located in volume, and all of them have the same z coordinate. Therefore, more sensor solutions should be provided in order to carry out the interpolation at different zones within the gear tooth. The applied methodology have several steps, beginning with collecting data in sensors by a post-processing on the simulation result files such that no new computations are required. The considered sensors are illustrated in Fig. 12 with red points.

Since the solution is currently known for all sensors in Fig. 12, an interpolation function for the temperature-time evolution over each surface sensor and its corresponding internal ones as a function of the distance from the surface point was created. A total of 9 interpolations were performed using the POD with linear interpolation. Hence, the temperature-time evolution can be computed by reconstruction (i.e. using the predicted POD modal coefficients) at any new point on the gray lines that pass over the red points (see Fig. 12).

Then, for any new point located in the domain limited by the existing sensors into which the prediction of temperature is required, multiple steps were applied. Therefore, to evaluate the performance of the methodology, 4 test points located at certain depths and at different Z coordinates were considered, as shown in Fig. 13(a). Figure 13(b) shows the projection of the test points onto the plane Z = 0 for a better visualization.

For each test point, its associated points in the 9 gray lines in Fig. 12 satisfying the equality of the minimum distance from the surface were determined. For the sake of clarity, an illustration of this step is shown in Fig. 14 at the 4 test points and at mid-width of the tooth because the X and Y coordinates of the three orange crosses, extracted at mid-width, remain the same for the other gray lines with other Z coordinates.

Basically, the real important information to be extracted from the test points is the shortest distance to the black curve, representing their orthogonal projection onto the gear tooth surface. Otherwise, for each test point, the orange



(a) Simulation 1

(b) Simulation 14

Fig. 11 Comparison between the FEM and the PODI approximation of the temperature-time evolution at 4 test positions and for simulations 1 and 14

crosses on the gray lines (Fig. 14) were determined such that the distance between those points and their orthogonal projections onto the tooth surface (green crosses) should be equal to the known distance between the test point and its orthogonal projection onto the same surface (i.e. equal dashed segments). Then, the prediction of the temperature-time evolution on the orange crosses were computed by using the previously created interpolation functions.

The coordinates of the orthogonal projection of all orange crosses in Fig. 14 onto the surface, represented by the green crosses, were extracted. A total of 9 points coordinates were provided for each test point, corresponding to the projection of each extracted point in gray lines of Fig. 12 onto the surface. Those 3D surface coordinates were then mapped into the 2D embedding manifold previously generated by LLE for surface interpolation, as shown in section "Surface interpolation". The representation of the mapped crosses for

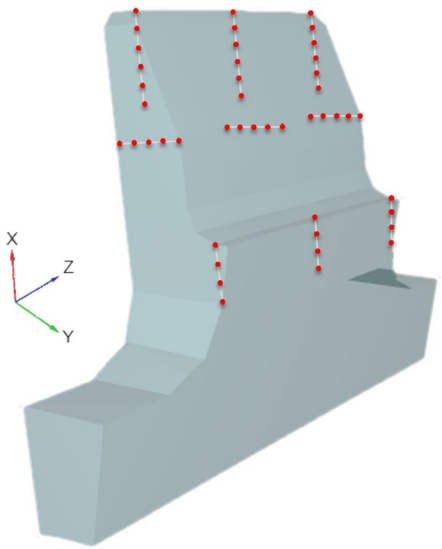
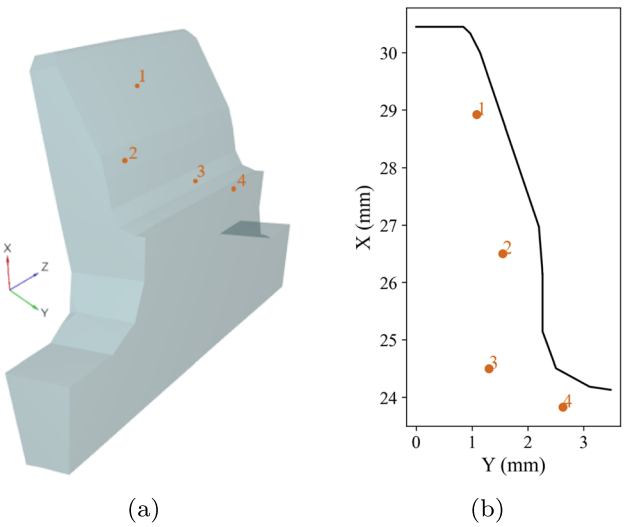


Fig. 12 Position of sensors for volume interpolation



(a) (b)

Fig. 13 Test points: (**a**) in the 3D coordinate system (**b**) projected onto the plane Z = 0

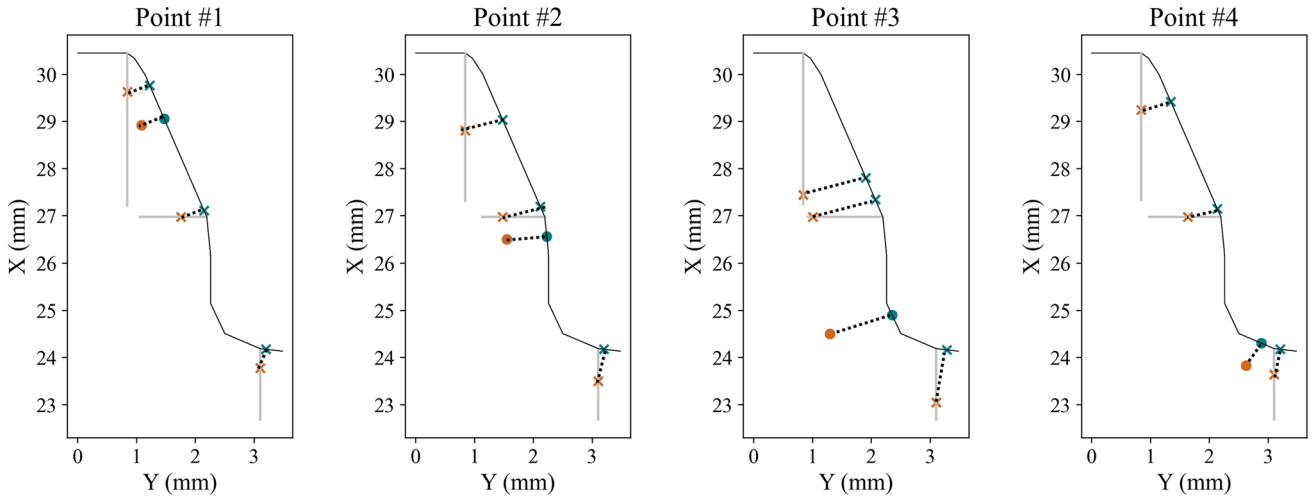


Fig. 14 Test points and their associated ones on the lines containing the available sensors

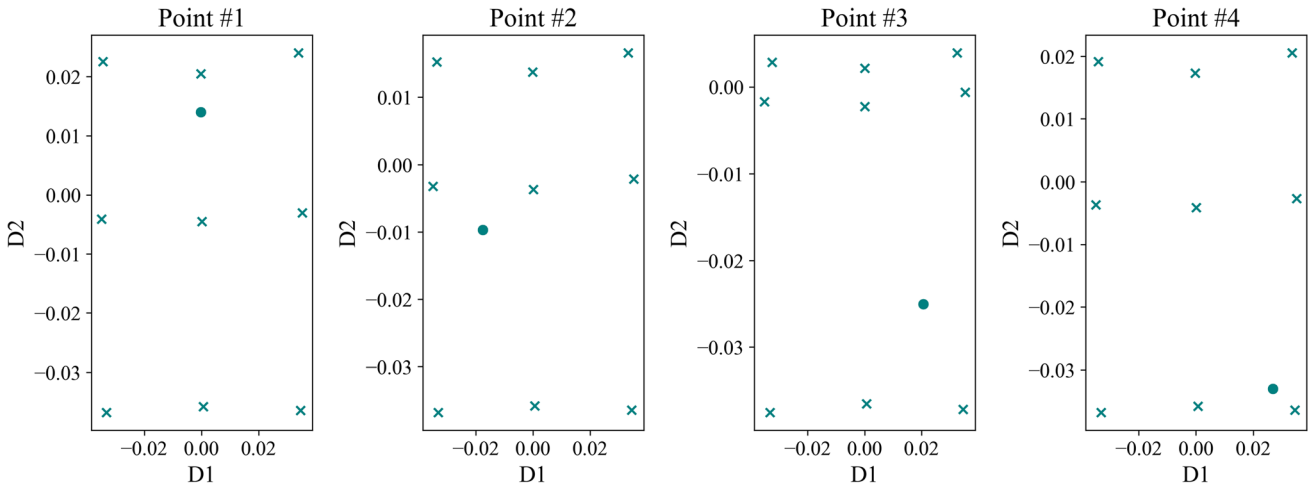


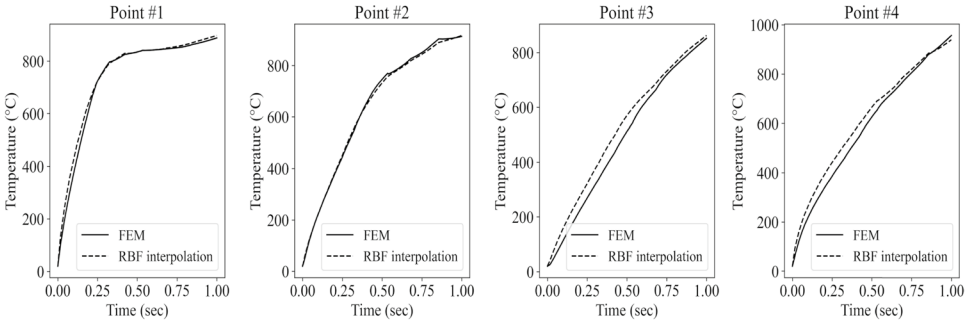
Fig. 15 Embedded coordinates in the 2D manifold

the 4 test points is illustrated in Fig. 15, where the projection of the test points on the surface were also mapped, and represented in the 2D space by green points.

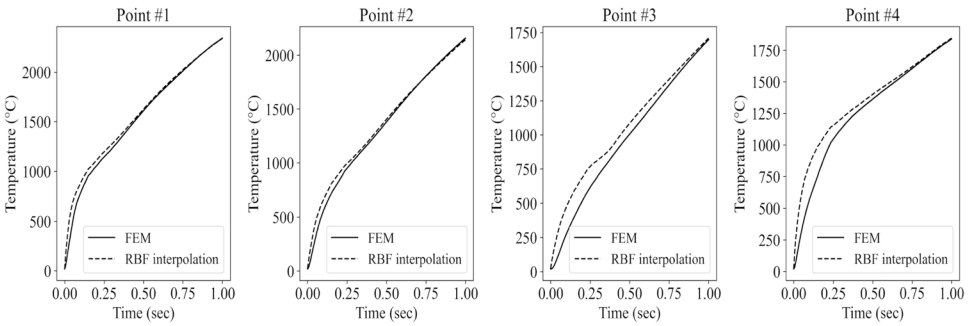
Now, following the same rationale as the surface interpolation, the PODI was applied on the snapshot matrices containing, as columns, the just predicted temperature-time evolution at the orange cross points in Fig. 14 for each test point, where the applied strategy was again tested for simulation 1 and 14 from the DoE. Similarly to the surface interpolation, the applied POD on the snapshot matrices resulted in a reduced basis having three POD modes, for both simulations and for each test point, to recover 99% of cumulative energy. Then, the POD modal coefficients were approximated by using RBF interpolation with a multiquadric basis function, as a function of the coordinates in the 2D space obtained by LLE.

After reconstruction, a comparison between the FEM and the PODI approximation of the temperature-time evolution is shown in Fig. 16, for simulation 1 (Fig. 16(a)) and simulation 14 (Fig. 16(b)) and at the 4 test points. It can be seen that all the curves show the same trend and a good correlation was obtained. However, a small gap between curves was observed for points #3 and #4. This could be explained by the position of the green points relative to the green crosses. However, such difference can be tolerated in the industrial application.

The relative error defined by Eq. 20 was also calculated to determine the estimation accuracy. The obtained errors, listed in Table 4, do not exceed 7% for simulation 14, and 5% for simulation 1. These results prove that the applied strategy for volume interpolation was very efficient and provided good approximations.



(a) Simulation 1



(b) Simulation 14

Fig. 16 Comparison between the FEM and the PODI approximation of the temperature-time evolution at 4 test points and for simulations 1 and 14

Table 4 Relative errors of the temperature-time evolution using POD with RBF interpolation - volume interpolation

Error (%)	point #1	point #2	point #3	point #4
Simulation 1	2.2%	0.8%	4.8%	3.8%
Simulation 14	2.6%	2.3%	6.2%	6.4%

Conclusions

In this work, the multiphysics parametrized induction heating process was studied and parametric metamodels were developed, for 14 sparse sensors located at different positions in the gear tooth, to predict the temperature-time evolution under different process parameters chosen from the predefined intervals.

To achieve this goal, firstly, a data-driven approach was applied on a set of synthetic data provided by the FE simulations for different values of frequency and power. Then, the GPOD method was applied in order to complete the missing values for certain simulations. Next, dimensionality reduction by POD coupled with sPGD regression method was applied to fit a model to the POD modal coefficients, providing the low dimensional representation of the original snapshots. The proposed approach was successfully applied, and a good approximation was provided using the sPGD regression model even with a small amount of data.

Since these models were built for some sparse points in the space domain, the solutions were extended to address the heat-affected zones in the space domain. Therefore, an interpolation over the gear tooth surface was first addressed by applying LLE method in order to map the three-dimensional coordinates into a two-dimensional manifold while preserving the neighborhood and hence recover the global nonlinear structure of the gear tooth from locally linear fits. The LLE was followed by POD with RBF interpolation on the snapshot matrix containing the temperature-time evolution on the available surface sensors, where interpolation was achieved with respect to the embedded coordinates in the lower-dimensional space. This approach was tested and approved for new points in the space domain.

Finally, internal spatial points were addressed for volume interpolation. Therefore, local interpolations with respect to the distance from the surface were first applied at different zones over a set of points by using POD coupled with linear interpolation. For any new point located within a certain depth at which the prediction of temperature is required, multiple steps were then applied including an intelligent search algorithm, a linear interpolation, an orthogonal projection onto the surface, a mapping by LLE, and a POD with RBF interpolation. This approach for volume interpolation resulted in good approximations as well.

The developed approach has the advantage of an almost real-time prediction that is highly suitable for active control

of the process parameters. In the light of the obtained results, the proposed approach can be extended to consider geometrical parameters such as dimensions of the workpiece and the inductor.

Acknowledgements This work was conducted with the help of the French Technological Research Institute for Materials, Metallurgy and Processes (IRT-M2P). The authors would like to acknowledge IRT-M2P and the partners of the project TRANSFUGE led by IRT-M2P.

Author contributions Conceptualization: F.C.; Methodology: K.D.; Formal analysis and investigation: K.D.; Writing - original draft preparation: K.D.; Writing - review and editing: K.D., M.D., F.C.; Supervision: M.D., K.T. All authors read and approved the final manuscript.

Funding This research was carried out as part of a project TRANSFUGE sponsored by the french institute of research and technology in materials, metallurgy and processes (IRT-M2P), France. The project benefited from funding from a consortium of industrial partners and the French PIA (Plan d'Investissement d'Avenir) granted the French National Agency for Research (ANR).

Declarations

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Rokicki P (2017) Induction hardening of tool steel for heavily loaded aircraft engine components. Arch Metall Mater 62:315–320
2. Candeo A, Ducassy C, Bocher P, Dughiero F (2011) Multiphysics modeling of induction hardening of ring gears for the aerospace industry. IEEE Trans Magn 47:918–921
3. Rudnev V, Loveless D, Cook R (2017) Handbook of induction heating. Second Edition, Boca Raton, FL : CRC Press, Taylor & Francis Group
4. Hömberg D (2004) A mathematical model for induction hardening including mechanical effects. Nonlinear Anal-Real World App 5:55–90
5. Spezzapria M, Dughiero F, Forzan M, Candeo A (2012) Multiphysics fem simulation of contour induction hardening process on aeronautical gears. J Iron Steel Res 19:95–98
6. Wang K, Chandrasekar S, Yang H (1992) Finite-element simulation of induction heat treatment. J Mater Eng Perform 1:97–112
7. Samiuddin M, Muzamil M (2016) Experimental investigation and optimization of process parameters for through induction hardening using factorial design of experiments. J Eng Res 5:174–185
8. Hutton D (2004) Fundamentals of finite element analysis, First Edition. McGraw-Hill Higher Education, Boston
9. Little RJA, Rubin DB (2019) Statistical analysis with missing data, Third Edition. Wiley, Hoboken, NJ, USA
10. Yates F (1933) The analysis of replicated experiments when the field results are incomplete. Am J Exp Agric 1:129–42
11. Everson R, Sirovich L (1995) Karhunen–loève procedure for gappy data. J Opt Soc Am A 12:1657–1664
12. Jolliffe IT (1986) Principal component analysis and factor analysis. In: Principal component analysis. Springer series in statistics. Springer, New York, pp 115–128
13. Ye J, Janardan R, Li Q (2004) Two-Dimensional Linear discriminant analysis. In: Advances in neural information processing systems (NIPS). MIT Press, vol 17, pp 1569–1576
14. Honeine P (2012) Online kernel principal component analysis: a Reduced-Order model. IEEE Trans Pattern Anal Mach Intell 34:1814–1826
15. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326
16. Geng X, Zhan D-C, Zhou Z-H (2005) Supervised nonlinear dimensionality reduction for visualization and classification. IEEE Trans Syst Man Cybern Part B (Cybernetics) 35:1098–1107
17. Chinesta F, Huerta A, Rozza G, Willcox K (2015) Model order reduction. In: Stein E, de Borst R, Hughes T (eds) The encyclo- pedia of computational mechanics. Second Edition. Wiley
18. Sirovich L (1987) Turbulence and the dynamics of coherent structures. I - Coherent structures. II - Symmetries and transformations. III - Dynamics and scaling. Quarterly of Applied Mathematics - Quart Appl Math 45:561–571
19. Hesthaven JS, Rozza G, Stamm B (2015) Certified reduced basis methods for parametrized partial differential equations. Springer International Publishing, SpringerBriefs in Mathematics, 978-3-319-22469-5
20. Girfoglio M, Quaini A, Rozza G (2021) A POD-galerkin reduced order model for a LES filtering approach. J Comput Phys 436:110260
21. Ballarin F, Manzoni A, Quarteroni A, Rozza G (2015) Supremizer stabilization of POD–galerkin approximation of parametrized steady incompressible Navier–Stokes equations. Int J Numer Methods Eng 102:1136–1161
22. Karatzas EN, Stabile G, Nouveau L, Scovazzi G, Rozza G (2019) A reduced basis approach for PDEs on parametrized geometries based on the shifted boundary finite element method and application to a Stokes flow. Comput Methods Appl Mech Eng 347:568–587
23. Bui-Thanh T, Damodaran M, Willcox K (2003) Proper orthogonal decomposition extensions for parametric applications in transonic aerodynamics
24. Shinde V, Longatte E, Baj F, Hoarau Y, Braza M (2015) A Galerkin-free model reduction approach for the Navier-Stokes equations. J Comput Phys 309:148–163
25. Guénot M, Lepot I, Sainvitu C, Goblet J, Coelho R (2013) Adaptive sampling strategies for non-intrusive POD-based surrogates. Eng Comput 30:521–547
26. Xiao D, Yang P, Fang F, Xiang J, Pain C, Navon IM, Chen M (2017) A non-intrusive reduced-order model for compressible fluid and fractured solid coupling and its application to blasting. J Comput Phys 330:221–244
27. Hamdaoui M, Le Quilliec G, Breitkopf P, Villon P (2013) Surrogate POD models for parametrized sheet metal forming applications. AIP Conf Proc 1532:607–615
28. Ripepi M, Verveld M, Karcher N, Franz T, Abu-Zurayk M, Görtz S, Kier T (2018) Reduced-order models for aerodynamic applications, loads and MDO. CEAS Aeronaut J 9:171–193
29. Dolci V, Arina R (2016) Proper orthogonal decomposition as surrogate model for aerodynamic optimization. Int J Aerosp Eng 2016:1–15
30. Demo N, Tezzele M, Mola A, Rozza G (2019) A complete data-driven framework for the efficient solution of parametric shape design and optimisation in naval engineering problems. In: Proceedings of MARINE 2019: VIII international conference on computational methods in marine engineering, pp 111–121

31. Ibáñez R, Abisset-Chavanne E, Ammar A, Gonzalez D, Cueto E, Huerta A, Duval JL, Chinesta F (2018) A multi-dimensional data-driven sparse identification technique: the sparse proper generalized decomposition. Complexity, Article ID 5608286:1–11
32. Limousin V, Delgerie X, Leroy E, Ibáñez R, Argerich C, Daim F, Duval J, Chinesta F (2019) Advanced model order reduction and artificial intelligence techniques empowering advanced structural mechanics simulations: application to crash test analyses. Mech Ind 20:804
33. Derouiche K, Garois S, Champaney V, Daoud M, Traidi K, Chinesta F (2021) Data-driven modeling for multiphysics parametrized problems-application to induction hardening process. Metals 11:738
34. Smola A J, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222
35. Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. Adaptive computation and machine learning. MIT Press, I-XVIII, 1–248
36. Chinesta F, Keunings R, Leygue A (2014) The proper generalized decomposition for advanced numerical simulations - a primer springerbriefs in applied sciences and technology. Springer, Cham
37. Bui-Thanh T, Damodaran M, Willcox K (2004) Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. AIAA J 42:1505–1516
38. Bui-Thanh T, Damodaran M, Willcox K (2003) Proper orthogonal decomposition extensions for parametric applications in compressible aerodynamics. In: Proceedings of the 21st AIAA applied aerodynamics conference, AIAA, pp 2003–4213
39. Astrid P, Weiland S, Willcox K, Backx T (2008) Missing point estimation in models described by proper orthogonal decomposition. IEEE Trans Autom Control 53:2237–2251
40. Venturi D, Karniadakis GEM (2004) Gappy data and reconstruction procedures for flow past a cylinder. J Fluid Mech 519:315–336
41. Gunes H, Sirisup S, Karniadakis G (2006) Gappy data: To Krig or not to Krig?. J Comput Phys 212:358–382
42. Saul LK, Roweis ST (2003) Think globally, fit locally: Unsupervised learning of low dimensional manifolds. J Mach Learn Res 4:119–155
43. Yotov K, Pingali K, Stodghill P (2005) Think globally, search locally. In: Proceedings of the 19th annual international conference on supercomputing, New York, NY, USA, pp 141–150
44. Wu H-T, Wu N (2018) Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding. Ann Stat 46:3805–3837
45. Saul L, Roweis S (2001) An introduction to locally linear embedding
46. Cardinaux D (2008) ÉTude et modélisation numérique 3D par éléments finis d'un procédé de traitement thermique de tôles emboutie après chauffage par induction : Application à un renfort de pied central automobile, Thesis, École Nationale Supérieure des Mines de, Paris, France
47. Magnabosco I, Ferro P, Tiziani A, Bonollo F (2006) Induction heat treatment of a ISO C45 steel bar: Experimental and numerical analysis. Comput Mater Sci 35:98–106
48. Ding H, Shin Y (2012) A metallo-thermomechanically coupled analysis of orthogonal cutting of AISI 1045 steel. J Manuf Sci Eng 134:051014
49. Geijselaers H (2003) Numerical simulation of stresses due to solid state transformations: The simulation of laser hardening. Thesis university of twente, Enschede, The Netherlands
50. Lee S-J, Pavlina E, Vantyne C (2010) Kinetics modeling of austenite decomposition for an end-quenched 1045 steel. Materials Science and Engineering A-structural Materials Properties Microstructure and Processing 527:3186–3194
51. Barba P (2018) Field models of induction heating for industrial applications. Przeglad Elektrotechniczny 1:3–7
52. Stein M (1987) Large sample properties of simulations using latin hypercube sampling. Technometrics 29:143–151
53. Jakobsson S, Andersson B, Edelvik F (2009) Rational radial basis function interpolation with applications to antenna design. J Comput Appl Math 233:889–904
54. Buhmann M (2000) Radial basis functions. Acta Numerica 9:1–38
55. Wendland H (2004) Scattered data approximation cambridge monographs on applied and computational mathematics. Cambridge University Press, Cambridge