Model order reduction and parametric metamodeling of the multiphysics induction heating process

Khouloud Derouiche

▶ **To cite this version:**

Khouloud Derouiche. Model order reduction and parametric metamodeling of the multiphysics induction heating process. Modeling and Simulation. HESAM Université, 2023. English. NNT : 2023HESAE027 . tel-04468582

HAL Id: tel-04468582
https://pastel.hal.science/tel-04468582

Submitted on 20 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HESAM UNIVERSITÉ

Arts et Métiers Sciences et Technologies

ÉCOLE DOCTORALE SCIENCES ET MÉTIERS DE L'INGÉNIEUR

Laboratoire Procédés et Ingénierie en Mécanique et Matériaux - PIMM
Campus de Paris

THÈSE

présentée par : **Khouloud DEROUICHE**
soutenue le : **23 mars 2023**

pour obtenir le grade de : **Docteur d'HESAM Université**
préparée à : **École Nationale Supérieure d'Arts et Métiers**
Spécialité : **Génie mécanique et matériaux**

Réduction de modèle et métamodélisation paramétrique du procédé multiphysique de chauffage par induction

THÈSE dirigée par :
M. Francisco CHINESTA

et co-encadrée par :
M. Monzer DAOUD & M. Khalil TRAIDI

Jury

M. Nazih MECHBAL	Professeur, Arts et Métiers ParisTech	Président
M. Elías CUETO	Professeur, Université de Saragosse	Rapporteur
M. Pierre JOYOT	Maître de conférences HDR, ESTIA	Rapporteur
M. Amine AMMAR	Professeur, Arts et Métiers ParisTech	Examinateur
M. Mohamed JEBAHI	Maître de conférences, Arts et Métiers ParisTech	Examinateur
Mme Marianne BERINGHIER	Maître de conférences, ISAE-ENSMA	Examinatrice
M. Francisco CHINESTA	Professeur, Arts et Métiers ParisTech	Examinateur
M. Monzer DAOUD	Expert en modélisation, IRT-M2P	Examinateur
M. Khalil TRAIDI	Ingénieur chercheur, SAFRAN	Invité
M. Philippe BRISTIEL	Expert en procédé de fabrication et en simulation de durabilité, STELLANTIS	Invité

T H È S E

Remerciements

Le bon accomplissement d'un travail de thèse dépend de l'ambiance dans laquelle celui-ci se déroule. En cela, je suis très reconnaissante à l'équipe d'encadrement et à la direction du projet. L'ensemble de ce travail n'aurait jamais été possible sans leur soutien. Je tiens à exprimer ma profonde gratitude à :

- M. Francisco CHINESTA pour m'avoir accordé la chance de réaliser cette aventure, pour son implication dans l'orientation de mes travaux, et pour toutes les connaissances qu'il a su m'apporter. Je mesure et je réalise la chance que j'ai eue en le côtoyant durant ces trois années.

- M. Monzer DAOUD pour son investissement total, son implication dans le projet, sa réactivité, ses précieux conseils, et son soutien continu tout au long de la thèse. Je le remercie également pour son aide précieux et sa rigueur à la correction de mes articles et du manuscrit. Je n'ai sûrement pas les mots suffisants pour lui témoigner ma reconnaissance.

- M. Khalil TRAIDI pour son implication et pour son apport scientifique et humain. J'ai eu un plaisir de travailler avec lui pendant mon stage de master et ma thèse !!

Je tiens à remercier les membres du jury pour le temps consacré à la lecture et à l'analyse de ce travail.

Je remercie tous les partenaires industriels du projet TRANSFUGE pour leur contribution au bon déroulement de la thèse et leur participation active aux réunions d'avancement.

Un grand merci à tous les membres du laboratoire PIMM; chercheurs, techniciens, et personnel administratives pour les moments de convivialité, leur accueil chaleureux, leur bonne humeur, et leur professionnalisme. Je tiens plus particulièrement à remercier Jacques GROS LA FAIGE pour son aide et ses conseils précieux.

Je tiens aussi à remercier TRANSVALOR de m'avoir accordé un contrat de 2 mois afin de finaliser mes travaux et la rédaction du manuscrit dans les meilleures conditions.

Je remercie infiniment SAFRAN de m'avoir accordé le temps nécessaire pour finir ma thèse et rejoindre le pôle matériaux et procédé de Safran Tech en temps qu'ingénieur modélisation des procédés d'usinage et de grenaillage.

Je saisis cette occasion pour remercier profondément mes amis du PIMM pour tous les beaux moments qu'on a passé ensemble, pour tous les afterworks que j'attendais avec impatience pour bien rigoler. Je les remercie également pour leur soutien moral et leurs encouragements. Une pensée particulière à mon cher ami Sebastian pour son aide et son soutien continu.

Mes chaleureuses pensées pour toute ma famille de m'avoir soutenue et encouragée tout au long de ce parcours. Merci à mes parents Hedi DEROUICHE et Samia MAMI pour leur soutien quotidien et leur affection. Un merci particulier à mes soeurs Khaoula et Sabrine d'être toujours présentes et à mes frères Khaled et Fradj. Et un merci spécial à mes nièces Salma et Hedi d'être des créatures si mignons et source de bonheur.

Ces trois années étaient une expérience unique tant au niveau humain qu'au niveau professionnel. A tous ceux qui ont participé de près ou de loin à la réalisation de ce travail, merci beaucoup !

Résumé

La modélisation par éléments finis (MEF) représente aujourd'hui l'outil de calcul le plus attrayant pour prédire et optimiser de nombreux problèmes industriels. Cependant, la MEF devient inefficace en ce qui concerne les problèmes complexes multiphysiques paramétrés, tels que le traitement de chauffage par induction, en raison de son coût de calcul élevé. L'objectif de cette thèse est de définir une méthodologie de réduction de modèles multi-physiques adaptée au procédé de chauffage par induction et de proposer une solution paramétrique pour quelques quantités physiques d'intérêt, notamment l'évolution temporelle de la température et du taux d'austénite sur un pignon droit en acier C45, en utilisant une approche de modélisation non intrusive basée sur les données comme alternative à la MEF pour une prédiction en temps réel.

Pour ce faire, un ensemble de solutions synthétiques a été collecté, au niveau de certains capteurs dans la pièce et pour différentes valeurs de paramètres d'entrée (fréquence et puissance), en se basant sur des données de la simulation numérique obtenues via le logiciel de calcul par éléments finis FORGE®. En effet, une étude de faisabilité et de convergence a d'abord été effectuée afin de figer une configuration qui converge et qui suit les bonnes tendances. Les résultats obtenus par simulation selon un échantillonnage de type hypercube latin ont ensuite été traités.

Pour le modèle de température, une réduction dimensionnelle par la méthode "proper orthogonal decomposition" (POD) couplée avec trois méthodes de régression non linéaire (sPGD, SVR, et GB) ont été appliquées pour construire une base réduite et créer un modèle représentatif de la solution dans l'espace de faible dimension. Pour le taux d'austénite, deux métamodèles ont été développés pour différents instants qui caractérisent la transformation austénitique. Les résultats ont montré que les méthodes sPGD et SVR donnent comparativement une meilleure prédiction.

Par conséquent, une prédiction en temps réel de l'évolution temporelle de la température et du taux d'austénite peut être calculée pour un nouvel ensemble des paramètres d'entrée et au niveau des capteurs considérés. Ensuite, une interpolation spatiale a été réalisée pour étendre la solution partout dans la zone affectée thermiquement. Pour la température, deux approches ont été proposées; la première

est basée sur la réduction de dimensionnalité non linéaire par la méthode “locally linear embedding” et la méthode “POD” avec interpolation par fonction de base radiale, tandis que la deuxième est basée sur la “gappy POD”. Les deux approches génèrent de bonnes approximations malgré leurs différences. Pour le taux d’austénite, une généralisation de l’approche proposée précédemment a été effectuée en prenant en considération des paramètres géométriques. Une comparaison des trois méthodes de régression a été menée.

Enfin, une étude de l’effet d’un changement dimensionnel du pignon sur l’évolution de la température a été effectuée, ceci sans passer par un nouveau plan d’expérience, mais en s’appuyant sur les résultats de la géométrie de référence. Pour ce faire, deux approches ont été proposées pour prédire l’évolution de la température dans des nouvelles géométries. La première approche est basée sur le réseau de neurone en utilisant comme paramètres d’entrée quelques incréments initiaux des courbes temporelles de la température. La deuxième approche est basée sur la “POD” et la régression par sPGD en utilisant la puissance de chauffe comme quantité intermédiaire. Il a été montré que les résultats sont prometteurs, cependant, il est difficile d’approximer des phénomènes non-linéaires dépendant du temps à partir des données partielles extraites au début du procédé.

Mots clés: Modélisation par éléments finis; Modélisation non intrusive; Chauffage par induction; Réduction de dimensionnalité; Métamodèle; Acier C45; régression non linéaires; Interpolation

Abstract

Finite element (FE) modeling has recently become the most attractive computational tool to predict and optimize many industrial problems. However, it becomes ineffective as far as complex multiphysics parameterized problems, such as the induction heating process, are concerned because of the high computational cost. This thesis aims at defining a multiphysics model order reduction methodology for the induction heating process and proposing a parametric-based solution for some physical quantities of interest, namely the temporal evolution of temperature and austenite phase rate within a C45 steel spur-gear, using a non-intrusive data-driven modeling approach as an alternative to the FE modeling for a real-time prediction.

To achieve this goal, a set of synthetic solutions was collected, at some sparse sensors in the workpiece and for different values of input parameters (frequency and power), from numerical simulation via FORGE® software. Indeed, a convergence study was first conducted to choose the best numerical configuration that converges and follows the right trends. Next, according to the Latin hypercube sampling design of experiments (DoE), FE results were obtained and then treated.

For the temperature modeling, a dimensionality reduction by the proper orthogonal decomposition (POD) method coupled with three nonlinear regression methods (sPGD, SVR, and GB) was applied to build a reduced basis and create models for the low-dimensional representation of the initial snapshots. For the austenite rate, two metamodels were developed for the time instants t_Ac1 and t_Ac3 characterizing the beginning and the complete austenitic transformation. It was shown that better predictions were obtained with the sPGD and SVR methods, comparatively.

Therefore, a real-time prediction of the temperature and austenite phase evolution could be calculated for a new set of process parameters and for the considered sensors. Then, spatial interpolation was performed to extend the solution everywhere in the heat-affected zone. For the temperature field, two approaches were proposed; the first one is based on nonlinear dimensionality reduction by locally linear embedding and POD with radial basis function interpolation, while the second one is based on gappy POD. Both approaches generate good approximations despite their differences. For the

austenite rate, a generalization of the previously proposed approach was carried out by taking into consideration geometrical parameters. A comparison between the results of the three regression methods was conducted.

Finally, the effect of the gear geometrical change on the temperature-time evolution was analyzed, by using the results of the reference geometry considered so far and without using a new DoE. Two approaches were proposed to predict the temperature-time evolution in new geometries. The first approach is based on the artificial neural network by considering the beginning of the temperature curves, known at few time steps, as input parameters. The second approach is based on POD and sPGD regression by using the internal heat source as an intermediate quantity. The obtained results were promising, however, it remains difficult to approximate nonlinear time-dependent phenomena from partial data extracted at the beginning of the process.

Keywords: Finite element modeling; Non-intrusive modeling; Induction heating; Dimensionality reduction; Metamodel; C45 steel; Nonlinear regression; Interpolation

Contents

List of Tables **xv**

List of Figures **xx**

List of Acronyms **xxiii**

1 Introduction **1**

1.1 Industrial and scientific context 2

1.1.1 Surface induction hardening of steels 2

1.1.2 Induction heating 3

1.1.2.1 IH principle 3

1.1.2.2 Skin effect 4

1.1.2.3 Edge and end effects 5

1.1.3 Problematics and challenges 6

1.2 Objectives 13

1.3 Thesis outline 13

1.4 Scientific contributions 15

1.4.1 Papers in international journals 15

1.4.2 Paper in french magazine 15

1.4.3 Conference proceedings 15

1.4.4 International conferences 15

1.4.5 Global plenary of TRANSFUGE project 16

1.4.6 Presentation in a thematic day 16

2 Overview of model order reduction techniques and regression methods 17

2.1 Introduction . . . 18

2.2 From the FEM to the MOR . . . 19

2.2.1 POD: fundamentals and extensions . . . 19

2.2.1.1 POD . . . 19

2.2.1.2 POD with interpolation (PODI) . . . 23

2.2.1.3 Gappy POD (GPOD) . . . 24

2.2.2 PGD and its counterparts . . . 28

2.2.2.1 Standard intrusive PGD . . . 28

2.2.2.2 Non-intrusive sparse subspace learning . . . 32

2.2.2.3 Non-intrusive sparse PGD . . . 34

2.3 Learning physics from data with machine learning and interpolation algorithms . . . 37

2.3.1 Polynomial regression . . . 38

2.3.2 Support vector regression . . . 40

2.3.2.1 Concepts and mathematical formulation of linear SVR . . . 40

2.3.2.2 Nonlinear SVR and kernel trick . . . 43

2.3.3 Boosted regression trees or gradient boosting . . . 44

2.3.4 Artificial neural network . . . 46

2.3.5 Radial basis functions interpolation . . . 48

2.4 Dimensionality reduction for manifold learning . . . 49

2.4.1 Locally linear embedding . . . 51

2.5 Summary and technical choices . . . 53

2.5.1 Intrusive versus non-intrusive MOR . . . 53

2.5.2 Metamodel or surrogate model . . . 54

2.5.2.1 Design of experiments . . . 55

2.5.2.2 Data generation . . . 56

2.5.2.3 Data pre-processing . . . 56

2.5.2.4 Metamodels construction and validation . . . 56

2.6 Conclusion . . . 57

3 Physical modeling and numerical simulation of induction heating process 58

3.1 Introduction . . . 59

3.2 Mathematical formulation of physical phenomena . . . 59

3.2.1 Modeling of electromagnetic phenomena . . . 60

3.2.1.1 Maxwell's equations . . . 60

3.2.1.2 Constitutive relations . . . 61

3.2.2 Modeling of thermal phenomena . . . 64

3.2.2.1 Heat transfer by conduction . . . 65

3.2.2.2 Heat transfer by convection . . . 66

3.2.2.3 Heat transfer by radiation . . . 66

3.2.3 Modeling of metallurgical phenomena . . . 66

3.2.3.1 Introduction to the phase transformation diagram . . . 66

3.2.3.2 Transformation kinetics of austenetic phase during heating . . . 67

3.2.4 Modeling of mechanical phenomena . . . 69

3.2.4.1 Elastic strain rate $\dot{\boldsymbol{\varepsilon}}^{el}$ . . . 70

3.2.4.2 Inelastic strain rate $\dot{\boldsymbol{\varepsilon}}^{in}$ . . . 70

3.2.4.3 Thermal strain rate $\dot{\boldsymbol{\varepsilon}}^{th}$ . . . 71

3.2.4.4 Phase change strain rate $\dot{\boldsymbol{\varepsilon}}^{tr}$ . . . 71

3.3 Coupling between physics . . . 71

3.4 Numerical simulation of induction heating process . . . 73

3.4.1 Principle of the IH simulation . . . 73

3.4.2 Identification of material properties . . . 76

3.4.2.1 Chemical Composition . . . 76

3.4.2.2 Electromagnetic properties . . . 77

3.4.2.3 Thermal properties . . . 77

3.4.2.4 Metallurgical properties . . . 77

3.4.2.5 Mechanical properties . . . 78
3.4.3 Model development in FORGE® software . . . 78
3.4.3.1 Electromagnetic solver . . . 79
3.4.3.2 Thermo-metallo-mechanical solver . . . 83
3.4.4 Numerical results . . . 87
3.4.4.1 Temperature distribution . . . 87
3.4.4.2 Austenite phase distribution . . . 88
3.5 Conclusion . . . 88

4 Local parametric metamodels for induction heating process 89

4.1 Introduction . . . 90
4.2 Numerical design of experiments - DoE . . . 90
4.3 Data generation and sensors positions . . . 91
4.4 Modeling the temperature field evolution . . . 93
4.4.1 First proposed approach . . . 94
4.4.1.1 Methodology . . . 94
4.4.1.2 Results and discussion . . . 96
4.4.2 Second proposed approach . . . 101
4.4.2.1 Methodology . . . 101
4.4.2.2 Results and discussion . . . 102
4.5 Modeling the austenite-time evolution . . . 108
4.5.1 Methodology . . . 108
4.5.2 Results and discussion . . . 110
4.6 Conclusion . . . 114

5 Full-field approximation in the space domain 116

5.1 Introduction . . . 117
5.2 Spatial approximation for the temperature field . . . 117
5.2.1 First approach: manifold learning and PODI for full-field spatial reconstruction 117

5.2.1.1 Surface interpolation . . . 117
5.2.1.2 Volume interpolation . . . 123
5.2.2 Second approach: GPOD for full-field spatial reconstruction . . . 130
5.2.2.1 Data extraction . . . 130
5.2.2.2 GPOD application . . . 131
5.2.2.3 Results and discussion . . . 132
5.3 Spatial approximation for the austenite phase . . . 136
5.4 Conclusion . . . 141

6 Geometrical analysis **144**
6.1 Introduction . . . 145
6.2 Motivation . . . 145
6.3 ANN-based approach . . . 145
6.3.1 Methodology . . . 145
6.3.2 Data generation and ANN modeling . . . 146
6.3.3 Results and discussion . . . 147
6.4 Intermediate quantity-based approach . . . 155
6.4.1 Methodology . . . 155
6.4.2 Results and discussion . . . 157
6.4.2.1 Modeling the area of the internal heat source evolution . . . 157
6.4.2.2 Modeling the temperature field evolution . . . 160
6.5 Conclusion . . . 165

7 Conclusions and perspectives **167**
7.1 Outcomes . . . 167
7.2 Perspectives . . . 171

Résumé étendu **173**

Bibliography **193**

Appendix **212**

A Identification of the electromagnetic material properties **212**

A.1 Electric resistivity 212

A.2 Relative magnetic permeability 213

B Identification of the thermal material properties **216**

B.1 Density 216

B.2 Thermal conductivity 216

B.3 Specific heat 217

C Identification of the mechanical material properties **219**

C.1 Young's modulus and Poisson ratio 219

C.2 Thermal expansion coefficient 220

C.3 Parameters of the constitutive law 220

C.4 Volume change due to phase transformation 221

List of Tables

2.1 Usual kernel functions 44

2.2 Usual activation functions 47

2.3 Common radial basis functions 48

3.1 Chemical composition of the employed C45 steel (FPDBase, 2019) 77

3.2 Gear and inductor dimensions 79

3.3 Effect of the electromagnetic mesh size on computation time 81

3.4 Electromagnetic properties of the inductor and surrounded air 83

3.5 Effect of the mesh size on computation time 84

3.6 Process parameters 87

4.1 Input parameters and their lower and upper limits 90

4.2 Relative errors of time evolution using 3 regression methods 101

4.3 Relative errors of temperature-time evolution using 3 regression methods 108

4.4 Error for t_Ac1 using the three regression methods 112

4.5 Error for t_Ac3 using the three regression methods 114

5.1 Relative errors of the temperature-time evolution using POD with RBF interpolation - surface interpolation 122

5.2 Relative errors of the temperature-time evolution using POD with linear RBF interpolation 126

5.3 Relative errors of the temperature-time evolution using POD with RBF interpolation - volume interpolation 130

5.4 Relative errors of the temperature-time evolution using GPOD approximation - surface points . . . 134
5.5 Relative errors of the temperature-time evolution using GPOD approximation - volume points . . . 135
5.6 Error for t_Ac1 and t_Ac3 using three regression methods . . . 138
5.7 Error for t_Ac1 and t_Ac3 using three regression methods - surface test points . . . 139
5.8 Error for t_Ac1 and t_Ac3 using three regression methods - volume test points . . . 141

6.1 ANN architecture and hyperparameters . . . 147
6.2 Process and geometrical parameters of the new geometries . . . 147
6.3 Relative errors (%) of the temperature-time evolution for the reference and the new geometries . . . 148
6.4 Relative errors (%) of the temperature-time evolution for the 4 new geometries and at the 4 test points . . . 155
6.5 RMSPE (%) of the area of $\dot{Q}$ for the reference and the new geometries . . . 159
6.6 Relative errors (%) of the temperature-time evolution for the reference and the new geometries . . . 162
6.7 Relative errors (%) of the temperature-time evolution for the 4 new geometries and at the 4 test points using the FE values of $\dot{Q}$ area . . . 162
6.8 Relative errors (%) of the temperature-time evolution for the 4 new geometries and at the 4 test points using the predicted values of $\dot{Q}$ area . . . 165

A.1 Parameters implemented in the FORGE® model to describe the electrical resistivity evolution . . . 212
A.2 Parameters used in the FORGE® model to describe the relative permeability . . . 214

C.1 Constitutive law parameters for austenite . . . 220
C.2 Constitutive law parameters for ferrite . . . 221
C.3 Constitutive law parameters for pearlite . . . 221

List of Figures

1.1 Induction heating principle (DUOLIN) 4

1.2 Representation of the skin depth 5

1.3 Field lines and power distribution along a cylinder immersed in a magnetic field created by a solenoid inductor (Rudnev et al., 2017). 6

1.4 Effect of frequency in IH process (adapted from (Wolfgang and Hans-Joachim, 2002)) 7

2.1 Principle of the alternate direction fixed point algorithm 31

2.2 Illustration of underfitting (a), accurate polynomial regression model (b), and overfitting (c) (adapted from (Scikit-learn)) 39

2.3 Illustration of the linear SVR with slack variables and linear loss function 42

2.4 Illustration of the kernel trick (Zararsiz et al., 2012) 44

2.5 Illustration of a MLP architecture 46

2.6 Steps of locally linear embedding: (a) assign k nearest neighbors, (b) reconstruct with linear weights, and (c) map to embedded coordinates 53

2.7 Steps of metamodeling approach 55

3.1 Iron-Carbon equilibrium diagram (Geijselaers, 2003) 67

3.2 Isothermal TTA diagram of a low-alloy steel of the type 35CrMo4 (0.34 %C - 1.07 %Cr - 0.17 %Mo) (Orlich et al., 1973) 69

3.3 Multiphysics-based coupling 72

3.4 Global principle of induction heating process 74

3.5 Discretization of an electromagnetic period (TRANSVALOR, 2020) 74

3.6 Chaining algorithm between the two solvers 76

3.7 CHT diagram of C45 (Orlich et al., 1973) 78
3.8 Geometry of the electromagnetic model 79
3.9 Electromagnetic solver mesh: (a) box method, (b) global mesh, and (c) cross-sectional mesh view 80
3.10 Evolution of the in-depth temperature - Mesh convergence in the electromagnetic solver 82
3.11 Thermo-metallo-mechanical model: (a) geometry with symmetry planes (b) gear mesh 84
3.12 Evolution of the in-depth temperature - Mesh convergence in the thermo-metallo-mechanical solver 85
3.13 Manipulator to define the boundary conditions 86
3.14 Isovalues of temperature distribution at times 0.1, 0.3, and 0.5 sec 87
3.15 Isovalues of austenite phase distribution at times 0.1, 0.3, and 0.5 sec 88

4.1 LHS design of experiments 91
4.2 Sensors positions. Blue points represent the surface sensors while the green and orange ones represent those located at a certain depth 92
4.3 Illustration of the time evolution of the number of time steps and temperature for simulation 1 and simulation 18 at sensor #1 93
4.4 Illustration of the temperature field evolution 94
4.5 Data alignment 95
4.6 Illustration of the first proposed approach 96
4.7 Cumulative energy computed from 15 snapshots of the time at 4 different sensors . . . 97
4.8 Comparison between the reference FE and the regression models of the time-temperature evolution at 4 sensors and for 4 training simulations 99
4.9 Comparison between the reference FE and the regression models of the time-temperature evolution at 4 sensors and for 4 testing simulations 100
4.10 Illustration of the second proposed approach 102
4.11 Original and reconstructed data for surface and internal sensors 103
4.12 Cumulative energy computed from 15 snapshots of the thermal field at 4 different sensors 104

4.13 Comparison between the reference FE and the regression models of the temperature-time evolution at 4 sensors and for 4 training simulations 106

4.14 Comparison between the reference FE and the regression models of the temperature-time evolution at 4 sensors and for 4 testing simulations 107

4.15 Three possible states of austenite phase evolution 109

4.16 Reference FE versus predicted values of t_Ac1 for 4 sensors 111

4.17 Reference FE versus predicted values of t_Ac3 for 4 sensors 113

5.1 Methodology for surface interpolation 118

5.2 LLE approach: (a) isolated surface geometry (b) mesh nodes in 3D coordinates (c) embedded coordinates in the 2D manifold 119

5.3 Cumulative energy computed from 9 snapshots of the thermal field for 2 different simulations 119

5.4 3D surface sensors (a) and their corresponding coordinates in the embedded manifold (b) 120

5.5 Comparison between the FEM and the PODI approximation of the surface temperature distribution at t = 0.5 sec 121

5.6 Comparison between the FEM and the PODI approximation of the temperature-time evolution at 4 test positions and for simulations 1 and 14 122

5.7 Methodology for volume interpolation 123

5.8 Position of sensors for volume interpolation 124

5.9 Test points for local interpolation at mid-width (Z = 0) 124

5.10 Cumulative energy computed from 4 snapshots of the thermal field for 2 different simulations 125

5.11 Comparison between the FEM and the linear RBF interpolation of the temperature-time evolution at 2 test positions and for simulations 1 and 14 126

5.12 Test points (a) in the 3D coordinate system (b) projected onto the plane Z = 0 . . . 127

5.13 Illustration of the searching method 127

5.14 Test points and their associated ones on the lines containing the available sensors . . 128

5.15 Embedded coordinates in the 2D manifold 129

5.16 Comparison between the FEM and the PODI approximation of the temperature-time evolution at 4 test points and for simulations 1 and 14 130

5.17 Cumulative energy of the thermal field computed from snapshots at different input parameters and time steps 132

5.18 Comparison between the FEM and the GPOD approximation of the surface temperature distribution at t = 0.5 sec 133

5.19 Comparison between the FEM and the GPOD approximation of the temperature-time evolution at 4 test positions and for simulations 1 and 14 - surface points 134

5.20 Comparison between the FEM and the GPOD approximation of the temperature-time evolution at 4 test positions and for simulations 1 and 14 - volume points 135

5.21 Reference FE versus predicted values of t_Ac1 137

5.22 Reference FE versus predicted values of t_Ac3 137

5.23 Comparison between the FE and the predicted values of t_Ac1 and t_Ac3 at 4 surface test points and for simulations 1 and 14 138

5.24 Volume test points 140

5.25 Comparison between the FE and the predicted values of t_Ac1 and t_Ac3 at 4 volume test points and for simulations 1 and 14 141

6.1 Methodology for full-field approximation using ANN 146

6.2 Comparison between the FEM and the ANN approximations of the surface temperature distribution for the new geometries at t = 0.25 sec 150

6.3 Comparison between the FEM and the ANN approximations of the surface temperature distribution for the new geometries at t = 0.5 sec 152

6.4 Test points at the new gear geometries 153

6.5 Comparison between the reference FEM and the ANN approximations of the temperature-time evolution for the 4 new gear geometries and at the 4 test points 154

6.6 Methodology of the intermediate quantity-based approach 156

6.7 Real and normalized Cartesian coordinates 157

6.8 Reference FE versus predicted values of the area of $\dot{Q}$ over 5 different time intervals for the reference and the new geometries 159

6.9 Cumulative energy of the thermal field computed from snapshots at different input parameters and spatial nodes 160

6.10 Reference FE versus predicted values of the low-dimensional representation of temperature for the reference and the new geometries 161

6.11 Comparison between the FEM and the sPGD approximations of the temperature-time evolution for the 4 new geometries and at the 4 test points using the FE values of $\dot{Q}$ area 163

6.12 Comparison between the FEM and the sPGD approximations of the temperature-time evolution for the 4 new geometries and at the 4 test points using the predicted values of $\dot{Q}$ area 165

1 Principe du chauffage par induction (DUOLIN) 175

2 Représentation de la profondeur de peau 176

3 Lignes de champ et distribution de puissance le long d'un cylindre immergé dans un champ magnétique créé par un inducteur solénoïde (Rudnev et al., 2017). 177

4 Effet de la fréquence sur le procédé de chauffage par induction (adapté de (Wolfgang and Hans-Joachim, 2002)) 179

1 Electrical resistivity of C45 steel according to FORGE® model 213

2 Model of B for ferromagnetic steels 214

3 Magnetic induction B of C45 steel 215

1 Density evolution per phase of C45 steel 216

2 Evolution of the thermal conductivity per phase of C45 steel 217

3 Evolution of the specific heat per phase of C45 steel 218

1 Young's modulus evolution with temperature 219

2 Evolution of the thermal expansion coefficient 220

3 Volume change due to phase transformation 222

List of Acronyms

ANN	artificial neural networks
ANOVA	analysis of variance
BEM	boundary element method
BRT	boosted regression trees
CHT	continuous heating transformation
DoE	design of experiments
ELUs	exponential linear units
FDM	finite difference method
FVM	finite volume method
FEM	finite element method
GB	gradient boosting
GPOD	Gappy proper orthogonal decomposition
HAZ	heat-affected zone
IH	induction heating
IID	independent and identically distributed
IRT	institute of technological research
KKT	Karush-Kush-Tucker
KLT	Karhunen-Loève transform
LHS	latin hypercube sampling
LLE	Localy Linear Embedding
MAS	modal adaptivity strategy
MLP	multi-layer perceptron
ML	machine learning
MOR	model order reduction
MSE	mean square error
MAE	mean absolute error
MAPE	mean absolute percentage error
MASE	mean absolute scaled error
MAAPE	mean arctangent absolute percentage error
M2P	materials metallurgy processes
NRMSE	normalized root mean square error
PCA	principal component analysis
PDE	partial differential equation
POD	proper orthogonal decomposition
PODI	proper orthogonal decomposition with interpolation
PGD	proper generalized decomposition
QoI	quantity of interest

ReLUs	rectified linear units
RBF	Radial basis functions
RSM	response surface method
RMSE	root square error
sPGD	sparse proper generalized decomposition
SSL	sparse subspace learning
SVD	singular value decomposition
SVR	support vector regression
SMAPE	symmetric mean absolute percentage error
TTA	temperature time austenitization
TRANSFUGE	TRANSmission FUture GÉnération

Chapter 1

Introduction

Outline

1.1 Industrial and scientific context 2
1.1.1 Surface induction hardening of steels 2
1.1.2 Induction heating 3
1.1.3 Problematics and challenges 6
1.2 Objectives 13
1.3 Thesis outline 13
1.4 Scientific contributions 15
1.4.1 Papers in international journals 15
1.4.2 Paper in french magazine 15
1.4.3 Conference proceedings 15
1.4.4 International conferences 15
1.4.5 Global plenary of TRANSFUGE project 16
1.4.6 Presentation in a thematic day 16

1.1 Industrial and scientific context

The aeronautical and automotive industries would eventually like to optimize the fabrication process of the mechanical structures in order to meet the industrial requirements in terms of costs and relative environmental concerns while maintaining optimal mechanical properties and performances. The production of optimal industrial components with interesting mechanical properties requires a judicious choice of material, however, the most suitable material is often expensive. Therefore, in order to ensure a good compromise between material quality and cost, heat treatments represent an excellent option and a valuable solution to give the material its optimal microstructures and mechanical properties corresponding to the various predefined performance criteria (Biasutti et al., 2012; Rudnev et al., 2017). However, for many applications, only superficial layer material properties play an important role. In this context, surface heat treatments of industrial components by mechanical, thermal, or thermochemical means are particularly suitable to improve the fatigue strength and the resistance to the imposed external mechanical loads by changing the properties of the critical zones (Rudnev et al., 2017).

In the increasingly competitive market, French industrial suppliers want to replace conventional thermochemical treatments such as carburizing and nitriding (Czerwinski, 2012) by more economical and environmentally friendly solutions with less variability. An interesting alternative to thermochemical treatments is induction heat treatments. This latter is generally considered the last step in the production chain and it is one of the most appealing heat treatments widely employed in automotive and aerospace industries, especially for the hardening process (Barrere, 1992; Pascal, 2003; Candeo et al., 2011; Rokicki, 2017).

1.1.1 Surface induction hardening of steels

The surface induction hardening consists of a rapid electromagnetic induction heating (IH) to a temperature within or above the phase transformation range (austenitization) and subsequent rapid cooling (quenching). The austenitization enables the initial microstructure of the steel to progressively transform into a solid solution of carbon called gamma iron (γ-Fe) or austenite. During quenching, the fast cooling of the workpiece, brought to a high temperature, by shower or immersion in an agitated fluid has the effect of trapping the carbon in the iron crystal lattice, therefore forming martensite. Therefore, the main beneficial effect of the induction hardening is to introduce a fine-grain martensite

phase as well as a compressive residual stress field, induced under the action of non-uniform plastic deformation induced by thermal gradient and localized volume variation resulted from phase transformation (Durban, 1997; Denis, 1997; Grum, 2007; Rudnev et al., 2017), within the superficial layer without affecting the metallurgy of the bulk material. Indeed, the presence of the former enhances hardness, wear resistance, and contact fatigue strength while the presence of the latter acts to inhibit crack initiation development. Indeed, the induction hardening process has the advantage of providing a very short surface heat-up time, precise control of the treated zone, good aptitude for integration in the production line, extensive possibilities with regard to the dimensions of the parts to be treated, good fatigue performance, good reproducibility, and operating mode compatible with severe environmental requirements, compared to the thermochemical treatments (Chatterjee, 2003).

Although both heating and cooling steps are important and have to be well understood, the IH step is of major interest and represents the most critical part. This is because heating a workpiece on a limited depth and getting the desired temperature field is a delicate task because of the thermal conduction phenomena which could modify the metallurgical properties of the material under the hardened layer.

In this manuscript, the IH step was carried out in order to optimize it and provide a better understanding of this latter. Indeed, there is a growing interest in IH technology in industrial (Lozinskii, 1969), medical (Stauffer et al., 1984) and domestic applications (Moreland, 1973) thanks to its good performances and its advantages compared with other classical heating techniques. More details about its application, developments, and challenges are provided in (Lucía et al., 2013).

1.1.2 Induction heating

1.1.2.1 IH principle

The IH is a direct application of two fundamental physical laws: Lenz's law and the Joule effect. Indeed, by supplying an inductor with an alternating electric current at a given frequency (often between 50 Hz and a few hundred kHz), a magnetic field is created around the inductor. When an electrically conductive material is immersed in these fields, it is crossed by a magnetic flux whose variations induce, according to Lenz's law, an electromotive force that gives rise to eddy currents (Gié Hubert, 1982). These induced currents produce heat by the Joule effect, and the workpiece heats up. The principle of IH is presented in Figure 1. This mechanism is the only source of heat production

for non-magnetic or paramagnetic materials (e.g., aluminum, copper, and steels with a temperature above their Curie temperature). For ferromagnetic materials, heat is also produced by the presence of a hysteresis effect in the magnetization curve (magnetic induction B as a function of the magnetic field H). This second mechanism is generally of a much weaker effect than the first; the ratio remains difficult to quantify because it is very dependent on the material and the process parameters.

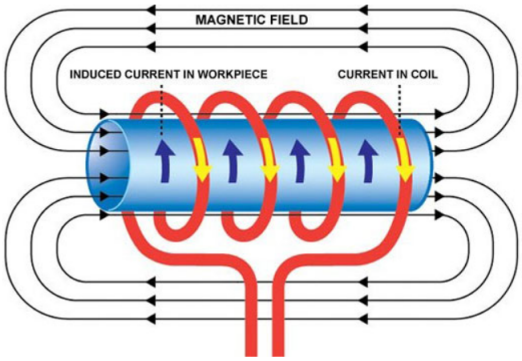


Figure 1.1: Induction heating principle (DUOLIN)

1.1.2.2 Skin effect

When a workpiece of conductive material is crossed by an alternating current, the current distribution on the crossed section is not uniform. The maximum value of the current density is always located at the surface of the workpiece and decreases inside it. This phenomenon of non-uniformity of the current distribution through a section of the conductor is called the skin effect. The heat generation area is concentrated in a thin layer under the surface of the part. The penetration depth called δ is the quantity that allows quantifying the importance of this phenomenon, it is defined in a usual way as the depth where the current density is e-1 times its value on the surface, Figure 2. The theoretical formula in Eq.(1) presents the order of magnitude of the skin thickness (Lamb, 1883).

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}} \tag{1.1}$$

where f is the current frequency, ρ is the electrical resistivity of the material, and μ is its magnetic permeability. It is easily interpretable that the higher the frequency, the smaller will be the skin depth. Indeed, this quantity is of primary importance in the induction heating, especially when applied to ferromagnetic materials. In this case, the penetration depth is not constant during heating but tends to vary due to the magnetic field and temperature-dependent electromagnetic properties of the material.

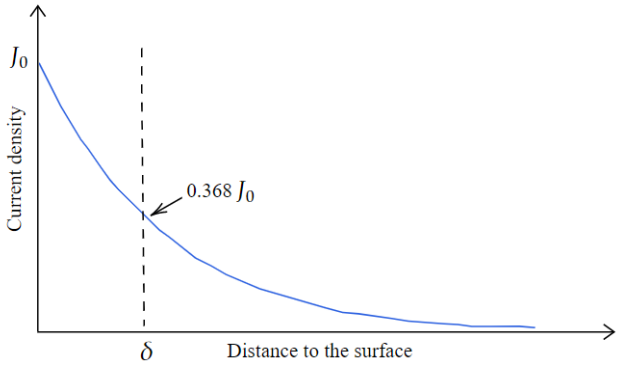


Figure 1.2: Representation of the skin depth

1.1.2.3 Edge and end effects

The temperature heterogeneity in an inductively heated part is due to the skin effect for the gradients between surface and core, and to the edge and tip effects for the surface gradients. If we consider a cylindrical part immersed in a homogeneous field (as inside a long solenoid inductor), the presence of the conducting part induces a distortion of the electromagnetic field (Figure 3). These field distortions that lead to variations in the induced current density and power distributions are referred to as edge effects. In real industrial cases, solenoid inductors can rarely be considered infinitely long. Under these conditions, there is a gap in the field lines at each end: the end effect. The end and edge effects are very similar and often influence each other. They are essentially governed by four variables: the skin effect, the depth to which the workpiece goes into the inductor (if it does not pass completely through the inductor), the ratio of the inner diameter of the inductor to the diameter of the workpiece, and the turn spacing factor. This last factor represents the ratio between the space between two turns and the width of a turn in the inductor. In the case of a single-turn inductor, the factor is equal to 1. An incorrect combination of these four factors can lead to overheating or underheating the workpiece's end part. It is therefore necessary to control them in order to observe the required temperature field distribution.

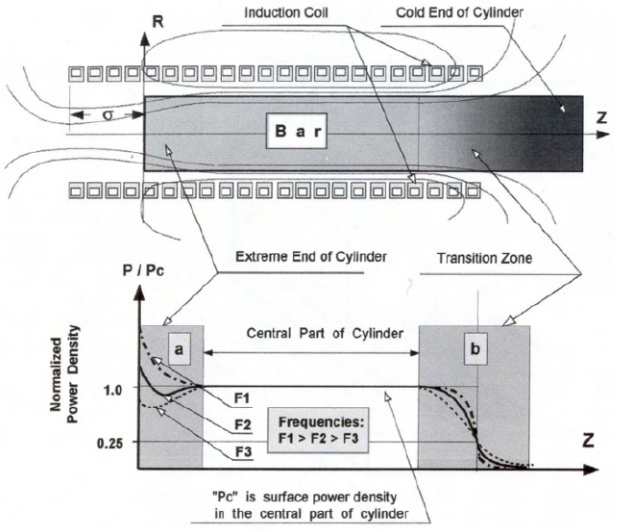


Figure 1.3: Field lines and power distribution along a cylinder immersed in a magnetic field created by a solenoid inductor (Rudnev et al., 2017).

1.1.3 Problematics and challenges

Optimizing the IH process provides better heating characteristics and consequently better mechanical properties after quenching. However, the main difficulty behind the IH process control and optimization is related to its multidisciplinarity. This is because it involves several fields of physics (electromagnetism, heat transfer, metallurgy, and mechanics). This multiphysics nature combined with very high heating rates (above 2000°C/s for this study) show high complexity for obtaining parts with desired mechanical properties.

The literature review illustrates that experimental, analytical, and numerical methods have been carried out to study the IH process. The objectives and the challenges behind the development of such methods are:

- reducing development time;
- reducing costs by “Do It Right The First Time (DIRFT)”, reduction of the NRC “Non-Recurring Costs” such as the costs of tools and prototypes and the RC “Recurring Cost” such as production costs;
- understanding the involved physical mechanisms for better controlling the influencing parameters and therefore being able to better predict the physical quantities that could be controlled;
- anticipating the possible defects and non-quality risks.

It is worth pointing out that we are interested, in the framework of our project, in power transmission components, gears in particular, which are considered as a complex geometry. They are commonly used for more than three thousand years in almost all sectors of activities; automobile, aeronautics, naval, watchmaking, and others. However, depending on their field of use, the control of the thermal dissipation inside the gear and the choice of the most suitable heat strategy should be achieved. Therefore, several parameters must be taken into consideration:

- the relative position of the inductor with respect to the workpiece (air gap, respective lengths, etc.);
- the current delivered to the inductor;
- the supply frequency and the skin effect which characterize the distribution of the induced currents in the workpiece: the more the frequency increases, the more the induced currents are concentrated on the surface. Typically, inductors are powered by alternating currents with a frequency varying from a few tens of Hertz to several hundred or thousand Hertz. Indeed, dual frequency generators (Esteve et al., 2006; Zgraja, 2019) have also been developed to feed the inductor, sequentially or simultaneously, with two different frequencies to achieve different penetration depths and to allow the heating of the contour. This approach is used for hardening workpieces with irregular surface geometry like gears (Rivat, 2002) such that the medium frequency range (3 to 10 kHz) favors the heating of the root of teeth, while the high-frequency range ($<$ 400 kHz) favors the heating of the tip and the flank of the gear (see Figure 4);

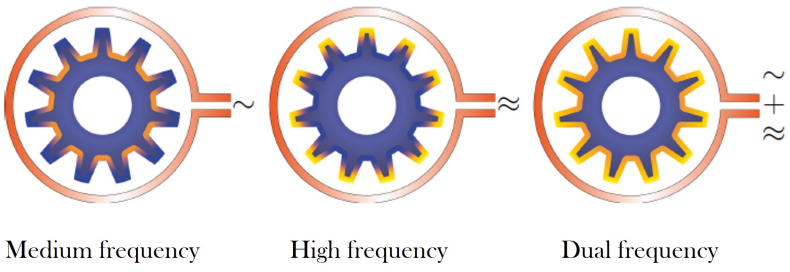


Figure 1.4: Effect of frequency in IH process (adapted from (Wolfgang and Hans-Joachim, 2002))

- the time process;
- the type of the inductor (geometry, size, technology). The geometry of the inductors can be very varied, ranging from a single coil to complex multi-coil inductors;

- the geometry of the gear. Indeed, gear wheels can have a diameter ranging from a few tens of millimeters in automotive to a few meters in naval for example, and with a modulus ranging from 1.5 to 12 mm.;

- the material and the initial microstructure of the gear to be heated.

In fact, the development of the process traditionally relies on industrial expertise based on a trial and error approach. Indeed, many studies for sensibility analysis, optimal process parameters selection, and optimization for already selected goals (maximizing, minimizing, or predicting a physical quantity of interest (QoI)) have been carried out for the induction hardening process (IH followed by quenching). In Misra et al. (2014), the effective case depth (the distance inward from the part's surface to a specific hardness) and hardness values have been analyzed for different conditions within fixed ranges of medium frequency power, feed rate, quench pressure, and temperature. A regression model has been developed to predict these physical qualities based on the response surface methodology coupled with the central composite design. In Onan et al. (2015), optimization studies have been conducted to evaluate the effect of the power supply, the scan rate, and the gap between the workpiece and coil on the induced hardening contour. Taguchi's L27 orthogonal array has been adopted for the experimental trials. The obtained data have been analyzed with analysis of variance (ANOVA). Further studies for optimizing the process by evaluating the effect of the IH process parameters have been developed, see for instance (Kohli and Singh, 2011).

However, the experimental approach is not only time consuming but requires significant experimental tests for a restricted validation range, given the various conditions and parameters controlling the process (e.g. frequency, power supply, heating time, shape, dimension, and material of the workpiece and inductor). Even if the physical phenomena encountered in the process are relatively well known, it is difficult to determine their effects on complex-shaped industrial components or inductor geometries. This is because the heat distribution can vary from one zone to another resulting in metallurgical and mechanical alterations which complicate the process optimization.

This made way for analytical investigations of the IH process. Many advances in analytical models related to the IH process have been developed during the last century. The first advanced analytical models for the calculation of one-dimensional and even two-dimensional electromagnetic and thermal fields in IH systems have been developed by Vologdin (1939), followed by Rodigin, Curtis, Stansel,

and others (Curtis, 1944; Stansel, 1949; Rodigin, 1950; Brunst, 1957; Simpson, 1960; Slukhotskii and Ryskin, 1974). The analytical solutions were based on many concepts such as series expansion, Fourier transformation, separation of variables, Bessel and other special functions. Analytical models include complex mathematical analyses which are less expensive but they are based on simplifications and assumptions which affect their accuracies when applied to complex geometries and nonlinear material properties. More promising approach to solve linear and nonlinear physical models related to coupled electromagnetic, thermal, and other fields involved in the IH process, is provided by numerical techniques. These techniques could be applied to any complex geometry with linear or nonlinear material properties. Among those methods, we can cite the finite difference method (FDM) to solve the heat transfer problems (Özişik et al., 2017), the boundary element method (BEM) to compute the 3D high-frequency electromagnetic fields during IH in (Kagami and Fukai, 1984; Muhlbauer et al., 1993), the finite volume method (FVM) to solve electromagnetic problems in (Beckstein et al., 2017) and the well-known and largely experienced finite element method (FEM) (Hutton, 2003) to solve all included physics for IH (Pascal, 2003; Cardinaux, 2008; Aliferov et al., 2010; Spezzapria et al., 2012; Jin, 2015).

Accordingly, the development of these methods in addition to the improvement of the computer capabilities led to the development of many advanced numerical simulation tools able to model the IH process, solve the physical equations behind it, and provide a precise thermal field, both spatially and temporally, and consequently a better understanding of the physical phenomena and their interactions. Several works on the simulation of the IH process have been carried out in the literature. Starting with simple part geometries with a single frequency, as in (Jacot et al., 1996) for axially symmetric parts using a mixed FEM and BEM to solve the electromagnetic problem. In Bay et al. (2003), a model combining electromagnetic, thermal, and mechanical phenomena for an axially symmetric part has been developed. Then, by the multifrequency approach supplied to one common inductor, the contour of complex-shaped workpieces such as gears has been treated (Petzold, 2014). Furthermore, FLUX 2D, FLUX 3D, and FLUX expert software have been used in (Delaleau et al., 2000) to solve the electromagnetic and thermal equations for a 3D gear model, but without taking into account metallurgical transformations and mechanical deformations. In Mingardi (2013), a 3D model of a spur-gear in carbon steel (AISI 4340) has been developed to perform IH treatment using FLUX 3D software. The effect of machine parameters (medium frequency for preheating, coil dimensions, and the presence of a flux concentrator over and below the gear) has been evaluated on the final temperature distribution in order to obtain a homogeneous hardening contour and to prevent the austenitization of the entire

tooth. It was shown that flux concentrators have a key role in reducing edge effects of IH and that the medium frequency in the preheating phase (1 to 10 kHz, for 1 to 5 sec) leads to better treated contour at the end of the high-frequency phase (200 kHz heating phase, for 0.2 sec). In Hömberg et al. (2016), the simulation of multifrequency induction hardening including phase transformation and mechanical effects has been carried out for a 3D gear model.

However, despite the improvement of computer capabilities, accurate and realistic modeling in 3D is still a challenging task, since the process involves mutual coupling of the aforementioned physical effects on complex geometric domains. Although the different physical fields related to the IH process can be predicted, 3D numerical modeling still suffers from some drawbacks. High computational cost, supporting the use of multiple frequencies, and data exchange between solvers, are to name a few (Li et al., 2013; Sumithra and Thiripurasundari, 2017).

In addition, sensibility analysis, QoI prediction, and optimal process parameters selection have been carried out in (Barglik et al., 2014; Senhaji, 2017) using numerical simulations. The employed design of experiments (DoE) such as Taguchi, or factorial DoE is generally constructed by defining first the input parameters and their levels (2 or 3 levels in general) and constructing then the so-called matrix of experiments by considering different possible combinations of the imposed levels of parameters. Those trials are then achieved by simulation using numerical tools such as FORGE®, ANSYS, COMSOL multiphysics, and FLUX software. Then, ANOVA and RSM are performed on the obtained results for optimization purposes. Those techniques are effective when an objective function is addressed, they enable to decrease the number of experiments required for optimization and also to create a response surface using low-order polynomials. However, depending on the number of parameters, their levels, and the possible interactions between them, the number of simulations to be run can be very high. Therefore, the multi-query simulation approach becomes ineffective as far as the optimization procedure is concerned.

A possible solution, investigated in this Ph.D. thesis, consists in employing model order reduction (MOR) techniques, that can express the solution of a given problem (expressed by partial differential equation (PDE), for instance) into a reduced basis with a strong physical or mathematical content. This basis is often extracted from experimental or numerical solutions of the problem at hand. This MOR can be done, for instance, by the proper orthogonal decomposition method (POD) or by the

reduced basis method (Chinesta et al., 2017). Such techniques allow a significant reduction in the resolution complexity such that the solution complexity scales with the size of the reduced basis instead of the size of the initial model, solved with FEM in most cases.

Even if the use of a reduced basis implies a certain loss of information, it enables impressive computation time savings by providing fast predictions. Besides, the predicted solution for a given problem is accurate enough as soon as the considered case continues living in the space spanned by the reduced basis, however, poor accuracy is expected when the required solution can not be accurately approximated within the space spanned by the reduced basis. To improve generality while ensuring accuracy, an appealing technique, the so-called proper generalized decomposition (PGD), enables the construction of the reduced basis with an interesting dimension reduction and the resolution of the problem simultaneously (Chinesta et al., 2013). However, despite its effectiveness, this method is very intrusive and requires the modification of the FE source code representing the physical system. This intrusiveness is even more pronounced for PGD than the methods based on the use of reduced bases.

To avoid intrusiveness, non-intrusive techniques have been proposed. They proceed by constructing the solution of the considered problem (a parametric problem in general) by using several high-fidelity solutions generated for different values of the model parameters according to a DoE. Among these techniques, we can mention the POD with interpolation (PODI) (Bui-Thanh et al., 2003a), where usual regression or interpolation methods expressing the dependence of the POD modal coefficients on the considered input parameters are employed. Moreover, regressions, part of supervised machine learning (ML), are widely employed in engineering applications to develop models (Brunton and Kutz, 2019; Chinesta et al., 2020; Sancarlos et al., 2021; Sun et al., 2022). Otherwise, sparse subspace learning (SSL) (Borzacchiello et al., 2017) interpolates the precomputed solutions by considering a hierarchical approximation basis, according to a hierarchically structured sampling over the whole parametric space. The SSL technique can provide a good interpolated parametric solution, however, the volume of data, which increases exponentially with the number of input parameters and the level of the hierarchical approximation, remains an issue. On the other hand, a sparsely sampled MOR technique, the so-called sparse PGD (sPGD) has been proposed in (Ibañez et al., 2018).
These data-driven non-intrusive model-based techniques, determined without a prior knowledge of physiques, are called surrogate models or metamodels. They enable, as any MOR method, real-time simulation, optimization, inverse analysis, and simulation-based control, in the online stage. Therefore,

from the engineering and artificial intelligence viewpoint, data could provide knowledge in different ways:

- producing data-based models, relating the outputs of interest to uncorrelated inputs;
- enriching the existing models based on well-established physics;
- extracting patterns with high information contents;
- classifying tendencies and variables;
- reducing dimensionality, especially when a high-dimensional problem is addressed;
- extracting the existing correlations and keeping only the valuable, sufficient, and explicative information.

However, multiple questions and challenges arise from those developments:

- the choice of the sampling strategy to get the more informative models;
- the variability of data;
- the choice of the interpolation or regression techniques for a better estimation of the response for new parameter values (not used during training). There are many possibilities (polynomial approximations, kriging, ML methods, etc.), however, there is no guarantee that the chosen method performs well on the studied data or not;
- the quantity of data to be used, given that collecting and treating data are often very expensive. Hence, the smart data paradigm should replace the big data-based procedures for many cases of applications. However, the main issue with applying regression in the low-data limit concerns nonlinear behaviors in multi-parametric settings. This last factor leads to the so-called "curse of dimensionality"; the exponential growth in the number of degrees of freedom and equivalently the number of necessary sampling points (or training points in ML terminology) in the parametric spaces;
- the imputation of missing values. For some modeling procedures, missing data have to be completed to offer a global map or to deduce measures in regions where measures cannot be directly performed. Many researchers have been working on "gappy" data problems for many decades and many statistical approaches (Yates, 1933; Little and Rubin, 2019) and non-statistical

ones based on modal decomposition (Everson and Sirovich, 1995) have been used to provide an accurate enough reconstruction of the missing data.

- etc.

From the above lists, it reveals that the use of data, today and in the future, drastically differs from its use in the past. The new developments require competence in data collection, data mining, and data analysis. This is why most of the just mentioned challenges are in continuous development.

1.2 Objectives

This thesis is part of a research project launched in 2019 by the Institute of Technological Research (IRT) Materials Metallurgy Processes (M2P) which is entitled: TRANSmission FUture GÉnération (TRANSFUGE). The overall objective of this research work, which brings together industrial partners[1] and relies on the expertise of the PIMM laboratory at Arts et Métiers ParisTech, is to optimize the IH process by developing a methodology based on MOR techniques to generate a parametric-based solution of the required physical QoI while providing a much lower computational cost with sufficient results accuracy and very fast predictions over the parameter space in almost real-time.
In particular, the specific objectives of the proposed research are to:

- develop a numerical model for the IH process using FORGE® FEM code.
- develop efficient local metamodels, for the temperature and austenite phase evolution, at some sparse sensors on the gear tooth from synthetic data collected via numerical simulation for different process parameters.
- propose a methodology to extend the parametric solution known at some sensor positions to address all the heat-affected zone (HAZ).
- study the effect of the spur-gear geometrical change on the temperature-time evolution during the IH process.

1.3 Thesis outline

The thesis is divided into seven chapters:

[1] Safran, Stellantis, Renault, Naval Group, NTN-SNR, Valeo, Ascometal

- Chapter 2 shows a review of intrusive and non-intrusive MOR techniques for complex problems with a focus on non-intrusive methods adapted for the IH process. Different common tools of MOR were presented, such as POD and its extensions, PGD and its counterparts, various ML methods for regression purposes, etc. Then a review of some manifold learning methods was shown. Finally, a summary of the technical choices was provided.

- Chapter 3 shows the physical modelisation and numerical simulation of the IH process. The first part of this chapter presents the mathematical formulations of the included physics and the assumptions related to the IH process. The second part illustrates the mutual coupling between physics, and the last part is devoted to the FE numerical simulation on FORGE® software: definition of the algorithm of resolution, identification of material properties, model development (geometries, mesh generation, boundary and initial conditions, process parameters, etc.), and finally some numerical results were illustrated.

- Chapter 4 is devoted to the development of local metamodels for temperature and austenite phase evolution during the IH. This chapter details the steps of the proposed methodologies: DoE, sensor locations, data pre-processing, MOR method, and machine learning methods. The results were shown and discussed.

- Chapter 5 shows different methodologies to address the spatial space for fast prediction of temperature field and austenite phase in all the HAZ. For the temperature field, two methodologies were proposed. A methodology based on manifold learning by locally linear embedding coupled with PODI was first proposed for surface and volume interpolation. Then, a second methodology based on the gappy POD method was applied. All the steps and obtained results were shown. For austenite, a generalization of the proposed approach in chapter 4 was achieved by applying regressions with adding parameters related to the geometry. Different results were shown as well.

- Chapter 6 details the two proposed methodologies to study the effect of the gear geometrical change on the temperature-time evolution, by using the results of the reference geometry considered in chapters 4 and 5 and without using a new DoE. The developed metamodels and results were illustrated.

- The thesis conclusion drawn from the current research work and recommendations for future work are provided at the end of this manuscript.

1.4 Scientific contributions

The thesis works have been published in several scientific reviews and presented at international conferences:

1.4.1 Papers in international journals

- Derouiche, K., Daoud, M., Traidi, K., Chinesta, F. (2022). Real-time prediction by data-driven models applied to induction heating process. International Journal of Material Forming, 15(4), 1-17.

- 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(5), 738.

1.4.2 Paper in french magazine

An article was published in the magazine "Traitements & Matériaux" dedicated to topics related to "Transformation of metallic materials"

- Derouiche, K., Garois, S., Daoud, M., Traidi, K., Chinesta, F. (2021). Approches basées sur la métamodélisation et l'intelligence artificielle pour le traitement thermique par induction. Traitements et matériaux. Dossier Matériaux, surfaces & revêtements, 473, 14 décembre 2021.

1.4.3 Conference proceedings

A short paper published for the ESAFORM conference that took place in April 2021. Derouiche, K., Daoud, M., Traidi, K., Chinesta, F. (2021). A non-intrusive model order reduction approach for multi-physics parametrized problems-Application to induction heating process.

1.4.4 International conferences

During my Ph.D., I participated in two international conferences:

- ESAFORM in April 2021, Liège, Belgium

- COMPLAS in September 2021, Barcelona, Spain

1.4.5 Global plenary of TRANSFUGE project

Five presentations were achieved in the global plenary of IRT-M2P in front of the industrial partners of the TRANSFUGE project.

1.4.6 Presentation in a thematic day

A thematic day entitled “Advanced methods in numerical simulation” that took place in April 2022 at the Laboratoire de Mécanique de Paris-Saclay (LMPS), which was created from the fusion of the LMT and the MSSMAT laboratory. This day was part of the Year of Mechanics.

Chapter 2

Overview of model order reduction techniques and regression methods

Outline

2.1 Introduction . . . 18
2.2 From the FEM to the MOR . . . 19
2.2.1 POD: fundamentals and extensions . . . 19
2.2.2 PGD and its counterparts . . . 28
2.3 Learning physics from data with machine learning and interpolation algorithms . . . 37
2.3.1 Polynomial regression . . . 38
2.3.2 Support vector regression . . . 40
2.3.3 Boosted regression trees or gradient boosting . . . 44
2.3.4 Artificial neural network . . . 46
2.3.5 Radial basis functions interpolation . . . 48
2.4 Dimensionality reduction for manifold learning . . . 49
2.4.1 Locally linear embedding . . . 51
2.5 Summary and technical choices . . . 53
2.5.1 Intrusive versus non-intrusive MOR . . . 53
2.5.2 Metamodel or surrogate model . . . 54
2.6 Conclusion . . . 57

2.1 Introduction

For years, the FEM shows a big success for studying the components of complex systems. Indeed, it is employed by most commercial software codes because of its high accuracy approximations. However, this accuracy is at the expense of computation time and ease of integration. Taking the example of multidimensional models (complex fluid dynamics (Bird et al., 1987), quantum chemistry (Cancès et al., 2003), etc.) where classical mesh-based approaches can no longer be applied because of the exponential increase in the number of degrees of freedom of the problem.

To overcome such issue, an alternative approach based on the MOR techniques exists today. It offers an interesting compromise in terms of computational cost, speed of execution, and result accuracy. This approach does not proceed to simplify the model, the models continue to be well established with a valid description of the physics. However, it does rely on an adequate approximation of the solution which allows for a simplified solution procedure without compromising the accuracy of the model solution. MOR techniques are numerous and have become very popular in recent years.

In the first part of this chapter, an overview of some reduction techniques is shown. Among those methods, we have the well-known proper orthogonal decomposition (POD) and its extensions which have been applied for different applications and continue to be developed (Chinesta et al., 2017). We have also the method based on the separated representation known as the proper generalized decomposition (PGD) (Chinesta et al., 2013), which has been widely employed in the last decades and has proven its ability to overcome certain difficulties. The latter give rise to other new methods, such as sparse proper generalized decomposition (sPGD) (Ibañez et al., 2018) and sparse subspace learning (SSL) (Borzacchiello et al., 2017) which are principally based on data.

The presented methods can be classified into different categories according to (1) their ease of integration (2) their ability to consider parametric and multidimensional problems (3) their ability to deal with nonlinear problems. Indeed, these classifications allow us to justify the choice of the methods retained for our case of study. In the second part, we focus on some interpolation and machine learning algorithms that allow to learn physics from data and to provide approximate solutions for complex problems. In the third part of this chapter, we focus on dimensionality reduction but for manifold learning and topology preservation. Finally, the choices carried out in terms of techniques

and software were clearly illustrated.

For notational convenience, capital boldface characters are used for matrices and tensors, lowercase boldface characters are employed for vectors, and non-bold characters for scalars and numbers.

2.2 From the FEM to the MOR

In this section, a transient problem described by a PDE is addressed for the explanation of the MOR methods. Indeed, the scalar function $u(\boldsymbol{x}, t)$, representing the solution of the problem, is to be approximated such that $(\boldsymbol{x}, t) \in \Omega \times \mathcal{T}$, $\Omega \subset \mathbb{R}^n$ ($n = 1, 2$ or 3) and $\mathcal{T} \subset \mathbb{R}$, are the space coordinates and time, respectively. The standard FEM considers the approximation of the solution as:

$$u(\boldsymbol{x}, t) = \sum_{i=1}^{N} U_i(t) N_i(\boldsymbol{x}) \tag{2.1}$$

where

- N: the number of mesh nodes in the domain Ω;
- $N_i(\boldsymbol{x})$: the interpolation function or shape function satisfying $N_i(\boldsymbol{x}_j) = \delta_{ij}\ \forall i, j$
- U_i: the value of the unknown field at node i.

In order to solve the problem using the FEM, it is first necessary to discretize the entire domain into a certain number of elements, i.e. to decompose the Ω domain into Ω_e subdomains. The basic idea consists in calculating the approximate solution over each subdomain rather than the entire domain. This approximation results in an algebraic problem of size N in the linear case, or many of them in the general transient and nonlinear cases, which is computationally expensive. In order to reduce the computational cost, model order reduction methods have been proposed and are nowadays widely employed. The next sections provide an overview of some techniques and their extensions.

2.2.1 POD: fundamentals and extensions

2.2.1.1 POD

The POD is an efficient technique for dimensionality reduction based on spectral decomposition for high-dimensional datasets (Astrid, 2004; Chinesta et al., 2017). The POD term was invented by Lumley (1967) for the extraction of coherent structures in the framework of fluid dynamics. The POD technique can appear in different names such as the principal component analysis (PCA) (Pearson,

1901; Jolliffe, 2005), or Karhunen-Loève Transform (KLT) (Karhunen, 1946; Kosambi, 1943; Loeve, 1948), depending on the field of application. The connection and the equivalence of these methods have been discussed in the literature (Liang et al., 2002; Wu et al., 2003).
Indeed, a wide range of POD applications can be found in the literature, such as data compression (Andrews et al., 1967), human face characterization (Kirby and Sirovich, 1990), and optimal control (Lombardi et al., 2009). Fundamentally, the POD method enables to compute a reduced basis that provides a low-dimensional representation of a high-dimensional system state and yields a set of empirical modes, which describes the dominant behavior of a given problem. Despite being a linear MOR technique, there are no assumptions about the linearity of the system, which explain its popularity in data analysis and model reduction. In order to efficiently determine the POD basis vectors (modes) for large problems, Sirovich (1987) introduced the method of snapshots. This technique can be used for a variety of applications, including several heat transfer problems (Newman, 1999), a derivation of reduced-order dynamical models (Holmes et al., 2012), a steady-state temperature distribution of flow in a square cavity by changing the Rayleigh number (Ly and Tran, 2001), and a multiphysics IH process (Derouiche et al., 2022).

Basically, a set of instantaneous or parametric solutions called "snapshots" have to be collected directly from experiments or FE simulations. The POD method then computes the set of basis vectors from these snapshots in such a way that for any given basis size, the energy is maximized or the error between the original and reconstructed data is minimized. Very often a reduced number of modes R ($R \ll N$, where N is the actual size of the FE solution) are sufficient to approximate the solution of the problem in question or problems similar to the one that served to extract the modes by slightly modifying the material properties, the boundary conditions or the geometrical parameters of the physical model with respect to the initial ones. In other words, while FE shape functions are general and can be employed in any problem, the reduced-order basis is restricted to the domain where the learned process is accomplished. Therefore, it cannot accurately describe the solution of models, different from the one from which the reduced basis has been extracted.
The solution $u(\boldsymbol{x}, t)$ is expressed in the reduced basis $\{\phi_k(\boldsymbol{x})\}_{k=1}^{R}$ as:

$$u(\boldsymbol{x}, t) \approx \sum_{k=1}^{R} a_k(t)\phi_k(\boldsymbol{x}) \tag{2.2}$$

where a_k is the scalar function of t representing the reduced coordinates associated with the kth POD mode $\phi_k(\boldsymbol{x})$ such that the resulting problem is of size R, instead of N.

Indeed, two equivalent representations have been proposed in the literature to express the POD; the first one is expressed in the form of a minimization problem:

$$\min_{\phi_1,\dots,\phi_R} \langle \|u(\boldsymbol{x},t) - \sum_{k=1}^{R} a_k(t)\phi_k(\boldsymbol{x})\|_2^2 \rangle \tag{2.3}$$

$$\text{subject to } (\phi_i, \phi_j) = \delta_{i,j} \ \forall i,j \in [1,R]^2 \tag{2.4}$$

where $\|\cdot\|$ defines the norm corresponding to the inner product $(\cdot,\cdot)$, $\langle\cdot\rangle$ denotes the appropriate average for the problem under consideration, and $\delta_{i,j}$ is the Kronecker delta symbol equal to 1 if $i=j$ and 0 otherwise. The aim is to find the orthonormal basis functions $\{\phi_k(\boldsymbol{x})\}_{k=1}^R$ such that the projection of u into the subspace generated by the R basis functions provides the best approximation. The second way of expressing POD aims at maximizing the mean projection of u into the orthonormal basis functions $\{\phi_k(\boldsymbol{x})\}_{k=1}^R$, leading to the following maximization problem:

$$\max_{\phi_1,\dots,\phi_R} \sum_{k=1}^{R} \langle |(u(\boldsymbol{x},t), \phi_k(\boldsymbol{x})|^2 \rangle \tag{2.5}$$

$$\text{subject to } (\phi_i, \phi_j) = \delta_{i,j} \ \forall i,j \in [1,R]^2 \tag{2.6}$$

In this work, only data coming from FE simulations were considered. These data are typically in discrete form because of the applied spatio-temporal discretization to evaluate the PDEs. Therefore, the discrete version of POD is presented hereafter in order to address the practical computation of the POD.

To proceed with the construction of the optimal POD modes, consider a set of P snapshots, $\{\boldsymbol{u}_i\}_{i=1}^P$, where $\boldsymbol{u}_i \in \mathbb{R}^N$ is a high-dimensional vector containing the real-valued full-field QoI at N nodes, and assuming that snapshots are computed by solving the full-order model describing the problem at some discrete time instants t_i such that $\boldsymbol{t} = [t_1,\dots,t_P] \in \mathbb{R}^P$. The snapshot matrix $\boldsymbol{M} \in \mathbb{R}^{N\times P}$, containing the snapshots as columns, is defined by $\boldsymbol{M} = [\boldsymbol{u}_1\boldsymbol{u}_2\cdots\boldsymbol{u}_P]$. When the snapshot matrix is tall and skinny (the number of the mesh nodes N is much greater than the number of snapshots P), the method of snapshots introduced by Sirovich represents an efficient way to compute the POD basis. It shows that the spatial basis vectors $\{\boldsymbol{\phi}_k\}_{k=1}^R$ can be expressed as a linear combination of the snapshots:

$$\boldsymbol{\phi}_k = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{P} \alpha_{ki}\boldsymbol{u}_i \quad \text{for } k = 1,\dots,R \tag{2.7}$$

such that λ_k and α_{ki} satisfy the eigen-problem:

$$\boldsymbol{C}\boldsymbol{A} = \boldsymbol{\Lambda}\boldsymbol{A} \tag{2.8}$$

where $\boldsymbol{A} = [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_{\boldsymbol{P}}]^T$, $\boldsymbol{\alpha}_i \in \mathbb{R}^R$ and $\boldsymbol{C} \in \mathbb{R}^{P \times P}$ is known as the snapshot correlation matrix

$$\boldsymbol{C} = \boldsymbol{M}^T \boldsymbol{M} \tag{2.9}$$

The eigenvectors of $\boldsymbol{C}$, associated to the highest R eigenvalues $\{\lambda_k\}_{k=1}^R$, enable to construct the POD basis vectors verifying the orthonormalization property by using Eq.(2.7). The eigenvalues of $\boldsymbol{C}$ show the importance of the basis vectors. Equivalently, the POD modes can be found by applying the singular value decomposition (SVD) on the snapshot matrix $\boldsymbol{M}$ as follows:

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{2.10}$$

where $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ and $\boldsymbol{V} \in \mathbb{R}^{P \times P}$ are unitary matrices containing the left and right singular vectors of $\boldsymbol{M}$, respectively. $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times P}$ is a rectangular diagonal matrix containing the singular values σ_i of $\boldsymbol{M}$, sorted in a decreasing order. The reduced POD basis vectors, $\{\boldsymbol{\phi}_i\}_{i=1}^R$, is defined as the first R left singular vectors of $\boldsymbol{M}$ (i.e. first R columns of $\boldsymbol{U}$) corresponding to the R largest singular values. Indeed, the SVD is closely related to the eigen-problem such that the right singular vectors of $\boldsymbol{M}$ (columns of $\boldsymbol{V}$) are the eigenvectors of $\boldsymbol{C}$ and the singular values of $\boldsymbol{M}$ are the positive square roots of the eigenvalues of $\boldsymbol{C}$. Therefore, the singular values or eigenvalues provide quantitative guidance for choosing the size of the POD basis. A typical criterion to choose R is to find the cumulative energy captured by the POD basis vectors, that satisfies:

$$\frac{\sum_{i=1}^{R} \sigma_i^2}{\sum_{i=1}^{min(N,P)} \sigma_i^2} > \varepsilon \tag{2.11}$$

where ε is a certain threshold specified by the user, typically taken to be 99% or higher (Jones et al., 1998; Cordier, 2008).

Hence, an efficient representation of the initial snapshot data in the low-dimensional subspace of dimension R ($R \ll N$) can now be provided by a linear combination of the POD modes

$$\boldsymbol{u}_i \approx \sum_{k=1}^{R} a_{ki} \boldsymbol{\phi}_k \quad \text{for} \quad i = 1, \ldots, P \tag{2.12}$$

where a_{ki} are the low-dimensional representation of the original solution, function of time in this case, and known by the so-called POD modal coefficients. The reduced-order model can then be derived by projecting the original model governed by PDEs onto the reduced space spanned by the POD modes or by interpolating the reduced representation of data. The reduced-order model is less complex and can be solved efficiently with interesting time savings.

2.2.1.2 POD with interpolation (PODI)

Typically, the POD-Galerkin method, widely employed in different studies (Liberge, 2008; Ballarin et al., 2015; Hesthaven et al., 2016; Karatzas et al., 2019; Girfoglio et al., 2021), intends to project the full-order problem, defined by PDEs, onto the POD reduced space to obtain a low-dimensional representation of the original operators. The main issue of this method is its intrusiveness: the construction of the reduced operators requires access to the full-order operators and thus the knowledge of the solvers discretization and solution algorithm. Therefore, the non-intrusive MOR technique, the so-called PODI, has been proposed. It is a data-driven approach that relies only on data and does not require the knowledge of the physical systems nor the modification of the source FE simulation code (Bui-Thanh et al., 2003a, 2004; Cao et al., 2016). The POD procedure outlined above considers a series of snapshots at different time instants. The procedure can also be applied with parameter space by providing snapshots with different values of parameters such as process parameters (frequency, power, etc.), geometrical parameters (module, number of teeth, etc.), or material properties (conductivity, permeability, etc.) for the IH process, for instance. Indeed, PODI is applicable for parametric problems and many works for parametric computational fluid dynamics problems have been carried out (Bui-Thanh et al., 2003a; Van den Eynde et al., 2020). In Van den Eynde et al. (2020), a comparison between solutions with PODI and POD Galerkin projection has been shown on the convection-diffusion equation while considering the thermal diffusivity and a uniform time-independent velocity field as parameters. The results show a good correlation.

This type of modeling has experienced great development in recent years and it has also been referred to as “Galerkin-free” reduced-order models (Shinde et al., 2016), “Non-intrusive” POD applied for fluid mechanics by solving the Navier–Stokes equations (Guénot et al., 2013; Xiao et al., 2015, 2016, 2017), “surrogate” POD in (Hamdaoui et al., 2013) for parametrized sheet metal forming applications, or Physics-based reduced-order approximation (Audouze et al., 2009). Different applications of PODI can be found in the literature for automotive and aerodynamic applications (Dolci and Arina, 2016; Verveld et al., 2016; Ripepi et al., 2018), naval application (Demo et al., 2019), as well as for geotechnical and geological application (Zhao, 2021). In relation to our study, many research works based on PODI have been achieved for nonlinear magnetostatic problems (Henneron et al., 2019), parameterized electromagnetic scattering problems using MOR with cubic spline interpolation (Li et al., 2021), and parametrized heat transfer problems (Chakir et al., 2019). However, most studies do not consider multiphysics problems which is a key point of interest in our present work.

Consider now a set of P snapshots, $\{\boldsymbol{u}_i\}_{i=1}^{P}$, corresponding to the set of parameter values $\{\boldsymbol{\mu}_i\}_{i=1}^{P}$ such that $\boldsymbol{\mu_i} = [\mu_{i1}, \ldots, \mu_{is}]$ is the vector containing the s parameters associated to the ith snapshot. The PODI enables a fast prediction of the full-field solution for any value of the input parameters $\boldsymbol{\mu}$ by interpolations or regressions of the reduced representation of the reference solution defined by the POD coefficients with respect to $\boldsymbol{\mu}$. PODI proceeds in the same way as POD where the original snapshots are projected onto the subspace defined by the POD basis vectors to reduce their dimensionality. Then the POD modal coefficients $\{a_{ki}\}_{i=1}^{P}$ for $k = 1, \cdots, R$ are calculated using the inner product

$$a_{ki} = (\boldsymbol{\phi}_k, \boldsymbol{u}_i) \tag{2.13}$$

Then, an interpolation or regression technique is applied to the set of POD modal coefficients $\{a_{ki}\}_{i=1}^{P}$ and for each k. Hence, the solution $\boldsymbol{u}$ for a new value of $\boldsymbol{\mu}$, not included in the original set, can be approximated by:

$$\boldsymbol{u} \approx \sum_{k=1}^{R} a_k \boldsymbol{\phi}_k \tag{2.14}$$

Among the different techniques used to build the parametric solution, we can mention:

- the novel sPGD method (Ibañez et al., 2018) applied to the IH process (Derouiche et al., 2021) and crash tests(Limousin et al., 2019);
- the support vector regression (SVR) (Smola and Schölkopf, 2004) which has been applied to approximate the displacement and stress field in geotechnical and geological engineering contexts in (Zhao, 2021);
- the radial basis functions (RBF) interpolation for contact problems (Nguyen and Kim, 2022) and nonlinear magnetostatic problem (Henneron et al., 2019);
- the artificial neural network (ANN) for supercritical flow problems (Sun et al., 2022);
- etc.

Those methods will be more detailed in the next sections.

2.2.1.3 Gappy POD (GPOD)

The GPOD is an extension of POD to handle incomplete datasets 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 (Everson and Sirovich, 1995) and was

then applied successfully for the reconstruction of PDE solutions in aerodynamics (Bui-Thanh et al., 2003a,b, 2004) and has also been used in the nonlinear model reduction methods for missing point estimation (Astrid et al., 2008). In this section, two cases are illustrated: GPOD for missing data reconstruction using a complete snapshot set and GPOD for an incomplete snapshot set.

GPOD for the reconstruction of missing data from completely known snapshots

We start by defining a mask vector, which describes for a particular state vector the position of the available and the missing values. For example, if we consider the solution vector $\boldsymbol{u}$, its corresponding mask vector $\boldsymbol{n}$ is defined as:

$$n^j = \begin{cases} 1 & \text{if} \quad u^j \quad \text{is known} \\ 0 & \text{if} \quad u^j \quad \text{is missing} \end{cases} \tag{2.15}$$

where u^j and n^j denote the jth element of the vectors $\boldsymbol{u}$ and $\boldsymbol{n}$, respectively. For ease of explanation, the pointwise multiplication is defined as $(\boldsymbol{n}, \boldsymbol{u})_i = n_i u_i$, the gappy inner product is defined as $(\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{n}} = ((\boldsymbol{n}, \boldsymbol{u}), (\boldsymbol{n}, \boldsymbol{v}))$, and the induced norm is $(\|\boldsymbol{v}\|_{\boldsymbol{n}})^2 = (\boldsymbol{v}, \boldsymbol{v})_{\boldsymbol{n}}$.

Consider $\{\boldsymbol{\phi}_k\}_{k=1}^R$ the POD basis for the snapshot set $\{\boldsymbol{u}_i\}_{i=1}^P$, where all snapshots are completely known. The POD modes $\{\boldsymbol{\phi}_k\}_{k=1}^R$ are computed as described in section 2.2.1.1. Consider $\boldsymbol{h}$ another solution vector that has some missing elements, with corresponding mask vector $\boldsymbol{n}$. The reconstruction of the full or "repaired" vector from the incomplete vector $\boldsymbol{h}$ is required. Such application might arise, for example, when $\boldsymbol{h}$ represents sparse measurements of the system state at some sensor locations (e.g., temperature measurements at some sensors for example), and there is a desire to reconstruct the corresponding full state information.

We assume that the vector $\boldsymbol{h}$ represents a solution whose behavior can be characterized with the existing snapshot set. The repaired vector $\tilde{\boldsymbol{h}}$ is then defined as a linear combination of the R POD basis vectors as follows:

$$\tilde{\boldsymbol{h}} \approx \sum_{k=1}^{R} b_k \boldsymbol{\phi}_k \tag{2.16}$$

To compute the POD modal coefficients b_k, the error E between the original and the repaired vectors, computed over the existing data elements in $\boldsymbol{h}$, must be minimized:

$$E = \|\tilde{\boldsymbol{h}} - \boldsymbol{h}\|_{\boldsymbol{n}}^2 \tag{2.17}$$

The coefficients b_k that minimize the error E can be found by differentiating Eq.(2.17) with respect to each b_k in turn and by using the Eq.(2.16). Therefore, a linear system of equations takes place

$$\boldsymbol{K}\boldsymbol{b} = \boldsymbol{f} \tag{2.18}$$

where $K_{pq} = (\boldsymbol{\phi}_p, \boldsymbol{\phi}_q)_{\boldsymbol{n}}$, $f_p = (\boldsymbol{h}, \boldsymbol{\phi}_p)_{\boldsymbol{n}}$. By solving Eq.(2.18) to find $\boldsymbol{b}$ and using Eq.(2.16), the intermediate repaired vector $\tilde{\boldsymbol{h}}$ is obtained. Finally, the full state vector $\boldsymbol{h}$ is reconstructed by replacing the missing elements in $\boldsymbol{h}$ by the corresponding repaired elements in $\tilde{\boldsymbol{h}}$ such that $h^j = \tilde{h}^j$ if $n^j = 0$.

GPOD for incomplete snapshot set

Consider the same set of snapshots $\{\boldsymbol{u}_i\}_{i=1}^P$ representing the solution field at P time instants and suppose that snapshots are not completely known and contain a certain percentage of gappiness, and consider their associated set of masks $\{\boldsymbol{n}_i\}_{i=1}^P$ defined as follows:

$$n_i^j = \begin{cases} 1 & \text{if} \quad u_i^j \quad \text{is known} \\ 0 & \text{if} \quad u_i^j \quad \text{is missing} \end{cases} \tag{2.19}$$

where u_i^j denotes the jth element of the vector $\boldsymbol{u}_i$. 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 $\boldsymbol{u}_i$, corresponding to $n_i^j = 0$, with an initial guess resulting in an initial complete field $\boldsymbol{u}_i^{(0)}$ such that:

$$u_i^{j(0)} = \begin{cases} u_i^j & \text{if } n_i^j = 1 \\ \bar{u}_i^j & \text{if } n_i^j = 0 \end{cases} \tag{2.20}$$

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

$$\tilde{\boldsymbol{u}}_i^{(l)} \approx \sum_{k=1}^{R} b_{ki}^{(l)} \boldsymbol{\phi}_k^{(l)} \quad \text{for} \quad i = 1, \ldots, P \tag{2.21}$$

  such that the POD modal coefficients $b_{ki}^{(l)}$ 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 $\boldsymbol{u}_i$ are compared. The error is defined as:

$$E = \left|\left| \boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i^{(l)} \right|\right|_{\boldsymbol{n}_i}^2 \tag{2.22}$$

  This minimization leads to the linear system of algebraic equations

$$\boldsymbol{K}^{(l)} \boldsymbol{b}^{(l)} = \boldsymbol{f}^{(l)} \tag{2.23}$$

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

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

$$u_i^{j(l+1)} = \begin{cases} u_i^j & \text{if } n_i^j = 1 \\ \tilde{u}_i^{j(l)} & \text{if } n_i^j = 0 \end{cases} \tag{2.24}$$

- step 4: repeat steps 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 R 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 (Venturi and Karniadakis, 2004) 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 ($R_0 = 1$ 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 $R_1 = R_0 + 1$ 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 R is important. This is because the reconstruction error does not monotonically decrease with R. Among the proposed criteria listed in (Venturi and Karniadakis, 2004; Gunes et al., 2006), the one based on the total energy of the POD reconstruction is presented here, where the optimal number of modes is defined as:

$$\text{Choose } R \text{ such that:} \quad |E_R - E_0| \quad \text{is maximum, where} \quad E_R = \sum_{i=1}^{P} \sigma_{i,R}^2 \tag{2.25}$$

Here $\sigma^2_{i,R}$ is the ith POD eigenvalue of the R-modes reconstructed data, knowing that the eigenvalues are given by the squares of the singular values. E_R is the total energy for the converged reconstruction using R 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 R modes used for reconstruction.

To end up, the reading of (Chinesta et al., 2017; Brunton and Kutz, 2019) is strongly recommended for further information about advances and development concerning the POD method.

2.2.2 PGD and its counterparts

2.2.2.1 Standard intrusive PGD

Equations (2.1) and (2.2) provide an approximate solution for the problem as a finite sum of the product of time-dependent coefficients (unknown to be computed) and space functions. The space functions are usually given by the well-known shape functions for the FEM or a series of empirical modes extracted by the POD. Indeed, POD is considered an "a posteriori" method since it allows the construction of the reduced-order model through snapshots from experiments or numerical simulations.
However, one would ideally like to be able to construct "a priori" a reduced representation of the solution, i.e. without relying on the knowledge of an existing approximation of the problem solution as in POD. In this case, one can assume that the space functions are also unknown. This yields to compute both space and time functions simultaneously (Ladevèze, 1999). Thus, an approximate solution can be written as a finite sum of unknown function product, which represents the basic idea behind the PGD method (Chinesta et al., 2013):

$$u(\boldsymbol{x}, t) \approx \sum_{k=1}^{M} X_k(\boldsymbol{x}) T_k(t) \tag{2.26}$$

where M is the number of terms in the approximation (also called modes in the PGD terminology) such that $M < N$ and $M \sim R$. The PGD consists of constructing a numerical approximation for the unknown field in a separated representation form involving "a priori" unknown functions of variables based on a successive enrichment strategy.
Regarding the nonlinear nature of the separated representation form (product of functions), a linearization procedure is required. This procedure has been studied in different works such as (Chinesta and Cueto, 2014; Cueto et al., 2016) and the references therein. This separated representation approach demonstrates its effectiveness in the context of high-dimensional problems to overcome the

so-called "curse of dimensionality" already highlighted in the literature (Ammar et al., 2006, 2007).

Moreover, for some geometries (shells, beams, plates, layered domains such as composite materials), the spatial domain Ω of u, supposed to be three-dimensional $u(x, y, z)$, is well-suited for a space domain separation (Bognet et al., 2012, 2014; Chinesta et al., 2014; Bordeu et al., 2015). We can have:

- the fully separable spatial domain Ω, this case gives rise to a sequence of one-dimensional problems instead of the typical three-dimensional complexity. The separated representation is written as:

$$u(x, y, z) \approx \sum_{k=1}^{M} X_k(x) Y_k(y) Z_k(z) \tag{2.27}$$

- the partially separable spatial domain Ω, where the obtained complexity of the problem is roughly the typical of a two-dimensional problem, i.e., the calculation of in-plane functions $X_k(x, y)$.

$$u(x, y, z) \approx \sum_{k=1}^{M} X_k(x, y) Z_k(z) \tag{2.28}$$

Equivalently, this approach can also be applied to parametric problems. If we consider the unknown field u as a function of space, time and a set of parameters $\boldsymbol{\mu} = \mu_1, \ldots, \mu_s$, the subsequent separated representation could be established as:

$$u(\boldsymbol{x}, t, \boldsymbol{\mu}) \approx \sum_{k=1}^{M} X_k(\boldsymbol{x}) T_k(t) \prod_{i=1}^{s} P_k^i(\mu_i) \tag{2.29}$$

The parameters $\boldsymbol{\mu}$ can be of different natures and considered as extra-coordinates to the problem just like the space and time physical variables. In this framework, a so-called computational vademecum (also known as abacus, virtual charts, etc.) can be developed in order to provide a sort of response surface for the considered problem. This approach has been successfully employed in different problems. Optimization, simulation-based control, inverse analysis, and uncertainty propagation are to name a few. Several studies have been carried out on this parametric approach, we can cite for instance (Ammar et al., 2010; Gonzalez et al., 2010; Ammar et al., 2014; Cueto et al., 2016). Once the response has been constructed at the offline stage, it provides results under real-time constraints, in the order of milliseconds, at the online stage by just invoking this response surface instead of simulating the whole problem for new possible scenarios.

In terms of performance, the efficiency of this approach is quite good. Indeed, using the standard incremental strategy to solve a transient problem, defined in a three-dimensional space with P time steps

(P can go up to a few million for industrial applications), requires the solution of P three-dimensional problems. However, using the PGD constructor, this yields to solve $M.m$ three-dimensional problems for the computation of the spatial functions $X_i(\boldsymbol{x})$ and $M.m$ one-dimensional problems for the computation of the time functions $T_i(t)$ where m is the number of nonlinear iterations needed to compute each term of the sum.

In order to illustrate the procedure for constructing the separate PGD representation, we consider the generic model governing the evolution of the field under study $u(\boldsymbol{x}, t)$, and we assume that the model could be affected by one or some parameters. However, a single parameter μ is considered here for the sake of simplicity

$$\mathcal{R}(u(\boldsymbol{x}, t); \mu) \equiv \mathcal{U}(u(\boldsymbol{x}, t); \mu) - F(\boldsymbol{x}, t; \mu) = 0, \quad \text{where } (\boldsymbol{x}, t, \mu) \in \Omega \times \mathcal{T} \times \mathcal{P} \tag{2.30}$$

where $\mathcal{R}(\bullet)$ represents a residual term, $\mathcal{U}(\bullet)$ is a generic linear or nonlinear differential operator, and $F(\bullet)$ the so-called forcing term. It is expected that the solution depends on the considered value of the parameter μ. Therefore, in the PGD framework and for the parametric setting, μ is considered as an extra-coordinate and the parametric solution approximation $u^M(\boldsymbol{x}, t, \mu)$ in the separated form is given by:

$$u^M(\boldsymbol{x}, t, \mu) = \sum_{k=1}^{M} X_k(\boldsymbol{x}) T_k(t) P_k(\mu) \tag{2.31}$$

It is worth noting that we kept unseparated spatial coordinates $\boldsymbol{x}$ for the simplicity of illustration. Then, to compute the different unknown functions involved in the Eq.(2.31), the usual weighted residual form is expressed for all suitable test functions $u^*(\boldsymbol{x}, t, \mu)$ as follows:

$$\int_{\Omega \times \mathcal{T} \times \mathcal{P}} u^*(\boldsymbol{x}, t, \mu) \mathcal{R}(u(\boldsymbol{x}, t, \mu)) \; d\boldsymbol{x} \; dt \; d\mu = 0 \tag{2.32}$$

A greedy algorithm is adopted such that the terms (product of functions) of the sum are computed one at a time, thus enriching the PGD approximation until an appropriate convergence criterion is satisfied.

When looking for the mth term in Eq.(2.31) ($m < M$), the first $m-1$ terms of the PGD approximation, $u^{m-1}(\boldsymbol{x}, t, \mu)$, are assumed to be known and the trial and test functions to be employed within the integral form in Eq.(2.32) are defined by:

$$u^m(\boldsymbol{x}, t, \mu) = \sum_{k=1}^{m-1} X_k(\boldsymbol{x}) T_k(t) P_k(\mu) + X_m(\boldsymbol{x}) T_m(t) P_m(\mu) \tag{2.33}$$

$$= u^{m-1}(\boldsymbol{x}, t, \mu) + X_m(\boldsymbol{x}) T_m(t) P_m(\mu) \tag{2.34}$$

and

$$u^*(\boldsymbol{x},t,\mu) = X^*(\boldsymbol{x})T_m(t)P_m(\mu) + X_m(x)T^*(t)P_m(\mu) + X_m(x)T_m(t)P^*(\mu) \tag{2.35}$$

The separated representation constructor deeply described in (Chinesta et al., 2013) proceeds by using an alternate direction fixed-point algorithm that computes the unknown functions at the enrichment step m as shown in Figure 2.1 such that the PGD approximation at the enrichment step m and iteration q is given by:

$$u^{m,q}(\boldsymbol{x},t,\mu) = u^{m-1}(\boldsymbol{x},t,\mu) + X_m^q(\boldsymbol{x})T_m^q(t)P_m^q(\mu) \tag{2.36}$$

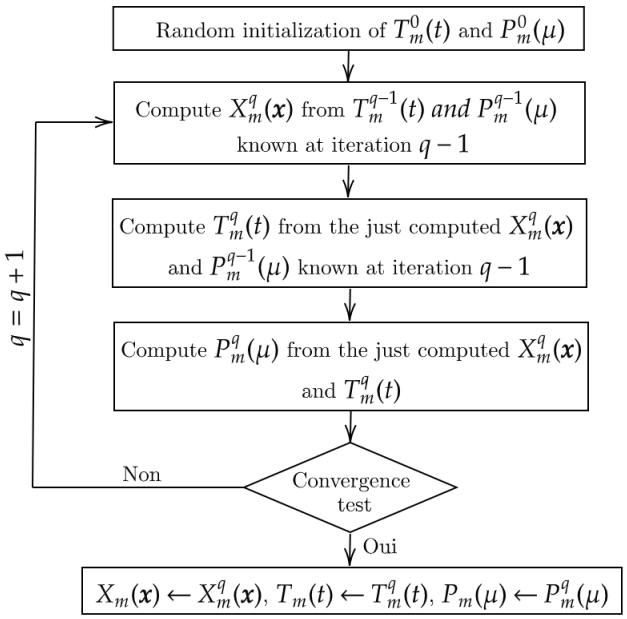


Figure 2.1: Principle of the alternate direction fixed point algorithm

More details about the function computations at each iteration could be found in (Chinesta et al., 2013). Furthermore, as illustrated in Figure 2.1, this nonlinear iterative process continues until the fixed point is reached such that:

$$\|X_m^q(\boldsymbol{x})T_m^q(t)P_m^q(\mu) - X_m^{q-1}(\boldsymbol{x})T_m^{q-1}(t)P_m^{q-1}(\mu)\| < \epsilon \tag{2.37}$$

where $\|\,\boldsymbol{.}\,\|$ is an appropriate norm and ϵ is a user-defined tolerance. Once the Eq.(2.37) is satisfied, the product of the $(m+1)$th term is considered.

The enrichment procedure also stops when a measure of error $E(m)$ becomes small enough such that $E(m) < \mathcal{E}$. Different stopping criteria can be applied and are reported in several references (Ladeveze

and Chamoin, 2011; Chinesta et al., 2013; Nadal et al., 2015). As an example, one can use:

$$E(m) = \frac{\|X_m(\boldsymbol{x})\ T_m(t)\ P_m(\mu)\|}{\|X_1(\boldsymbol{x})\ T_1(t)\ P_1(\mu)\|} \tag{2.38}$$

where $\|\ \boldsymbol{.}\ \|$ is an appropriate norm and it can be the L^2-norm for instance.

In the context of PGD, the model is solved only once in the offline stage allowing the construction of the solution. Then, its use in the online stage, by specifying the value of the parameter we want to evaluate, provides a rapid prediction in almost real time. Therefore, real-time simulation, process optimization, inverse identification, sensitivity analysis, and control become feasible for complex problems.
Finally, we would like to pay attention to two aspects of the standard PGD. The first one is its intrusiveness, which is clear from the building approach-based solution presented in this section. As shown above, the modification of the solver code by acting on the model resolution is required. However, such modification is not always obvious especially when access to the solver code is not possible. The second point is the offline resolution procedure which can be quite cumbersome, especially when dealing with multi-parametric problems.

To overcome the intrusiveness of standard PGD algorithms to be able to construct parametric solutions by using commercial simulation software, two recent methods have been proposed in the following two sections which have shown promising performances on several applications.

2.2.2.2 Non-intrusive sparse subspace learning

The SSL method is a new technique that has proven its efficiency in some application cases (Borzacchiello et al., 2017). Thanks to its non-intrusive character, the SSL can produce a parametric solution to a problem based only on the result of a deterministic solver. The SSL can be coupled with any simulation software and the treatment of non-linearity is no longer a problem because it is entirely managed by the direct solver through the high-fidelity simulation. The specificity of this technique is its well-structured aspect such that the parametric domain is explored in a hierarchical manner. Indeed, the numerical data from simulations are obtained by manipulating the input parameters such that sampling points are collocated at the Gauss-Lobato-Chebyshev integration points.
Therefore, the SSL method allows to provide a hierarchical approximation. This is by using a hierarchical basis enabling to improve the solution accuracy by adding hierarchical levels without disturbing

the previous ones. For a better understanding of the method, a demonstration of a simple case is illustrated in the following.

To simplify the illustration, we consider again a transient problem with a single parameter denoted by μ such that μ varies in the interval $[\mu_{min}, \mu_{max}]$. We seek to calculate the approximate parametric solution $u(\boldsymbol{x}, t, p)$. Firstly, the SSL consists in choosing a hierarchical basis of the parametric domain such that we associate to each Gauss-Lobato-Chebyshev collocation point, μ_i^m, a function $f_i^m(\mu_)$, where the indexes i and m correspond to the number of collocation point and the approximation level, respectively.

Let's consider the first level of approximation $m = 0$ with only one parameter. Here, the transient problem should be solved by running a direct solver code for two trials, corresponding to the minimum and maximum value of the parameter μ, that define the parametric domain, such that $\mu_1^0 = \mu_{min}$ and $\mu_2^0 = \mu_{max}$. We get the reference solutions defined by:

$$u(\boldsymbol{x}, t, \mu_1^0) = u_1^0(\boldsymbol{x}, t) \tag{2.39}$$

$$u(\boldsymbol{x}, t, \mu_2^0) = u_2^0(\boldsymbol{x}, t) \tag{2.40}$$

Thus, the approximate solution at level $m = 0$ is given by:

$$u^0(\boldsymbol{x}, t, \mu) = u_1^0(\boldsymbol{x}, t) f_1^0(\mu) + u_2^0(\boldsymbol{x}, t) f_2^0(\mu) \tag{2.41}$$

This solution shows the separated representation with a standard linear approximation where the basic interpolation functions $f_1^0(\mu)$ and $f_2^0(\mu)$ are:

$$f_1^0(\mu) = \frac{\mu - \mu_2^0}{\mu_1^0 - \mu_2^0} \tag{2.42}$$

$$f_2^0(\mu) = \frac{\mu - \mu_1^0}{\mu_2^0 - \mu_1^0} \tag{2.43}$$

At the approximation level $m = 1$, we add a single point denoted by μ_1^1, placed in the middle of the parametric domain ($\mu_1^1 = \frac{\mu_{min} + \mu_{max}}{2}$) and associated with the interpolation function $f_1^1(\mu)$. After simulation we obtain the solution $u(\boldsymbol{x}, t, \mu_1^1) = u_1^1(\boldsymbol{x}, t)$. In order to obtain the approximation at level $m = 1$, we enrich the one at level $m = 0$ as follows:

$$u^1(\boldsymbol{x}, t, \mu) = u^0(\boldsymbol{x}, t, \mu) + \bar{u}_1^1(\boldsymbol{x}, t) f_1^1(\mu) \tag{2.44}$$

The fonction $f_1^1(\mu)$ is now parabolic, and the function $\bar{u}_1^1(\boldsymbol{x}, t)$ is called surplus and defined by:

$$\bar{u}_1^1(\boldsymbol{x}, t) = u_1^1(\boldsymbol{x}, t) - u^0(\boldsymbol{x}, t, \mu_1^1) \tag{2.45}$$

The process continues adding surpluses as far as the hierarchical approximation level increases. It can be noticed that the SSL solution could contain too many terms. In that case, a post-compression takes place by looking for a more compact separated representation, more details could be found in (Borzacchiello et al., 2017).
The generalization to several parameters is simple and is done in the same way as the presented case while increasing the dimension of the hierarchical bases.

To sum up, SSL is a method that requires a well-structured and hierarchically processed dataset. It has been shown to be effective in several application cases. However, this technique shows a severe limitation for high-dimensional problems. This is because the SSL requires 2^s high-fidelity simulations just to reach the first level of approximation with s parameters. Thus, a high-dimensional model requires a very large number of computational points to get a good approximate solution. To overcome this issue the sPGD and its regularized variants were proposed.

2.2.2.3 Non-intrusive sparse PGD

Unlike the SSL, this method relies on a sparse unstructured dataset and allows building the parametric solution with a reduced number of full-order solutions compared to SSL. Indeed, the sPGD has shown promising results for academic examples (Ibañez et al., 2018) and industrial applications such as the crash test application (Limousin et al., 2019). It consists in defining an approximate solution to problems by using offline-data obtained via a commercial software (in most cases) with considerably reducing the number of simulations to run, compared to the SSL technique, especially when multi-parametric problems are addressed. This allows consequently to overcome the "curse of dimensionality". In other words, the sPGD can be viewed as a nonlinear regression method that employs the separation of variables, in the same way as the PGD, which explains its use in multi-parametric settings. Besides, as for the SSL, the nonlinearity of the physical models is entirely managed by the FEM implemented in the simulation codes.

To explain the method, the same transient problem with a single parameter μ is considered for ease of explanation such that:

$$\begin{aligned} u : \Omega \times \mathcal{T} \times \mathcal{P} \subset \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} &\to \mathbb{R} \\ (\boldsymbol{x}, t, \mu) &\to u(\boldsymbol{x}, t, \mu) \end{aligned} \tag{2.46}$$

with $u(\boldsymbol{x}, t, \mu)$ a scalar function known at P sampling points such that $u_i(\boldsymbol{x}, t) \equiv u(\boldsymbol{x}, t, \mu_i)$ $(i = 1, \cdots, P)$. The sPGD tries to find the approximate solution of u from sparse data using a low-rank separated representation. To achieve this goal, we first consider the Galerkin projection

$$\int_{\Omega \times \mathcal{T} \times \mathcal{P}} w^*(\boldsymbol{x}, t, \mu)(\tilde{u}(\boldsymbol{x}, t, \mu) - u(\boldsymbol{x}, t, \mu))\ d\boldsymbol{x}\ dt\ d\mu = 0 \tag{2.47}$$

where $w^*(\boldsymbol{x}, t, \mu)$ is an arbitrary test function and $u^M(\boldsymbol{x}, t, \mu)$ is the approximation of $u(\boldsymbol{x}, t, \mu)$, expressed as a finite sum of M terms by following the PGD rationale (Chinesta et al., 2013).

$$u(\boldsymbol{x}, t, \mu) \approx u^M(\boldsymbol{x}, t, \mu) = \sum_{k=1}^{M} X_k(\boldsymbol{x}) T_k(t) P_k(\mu) \tag{2.48}$$

where $X_k(\boldsymbol{x})$, $T_k(t)$ and $P_k(\mu)$ are the separated functions depending on $\boldsymbol{x}$, t and μ respectively. The determination of these functions is done by first approximating them using a set of basis functions as follows:

$$X_k(\boldsymbol{x}) = \sum_{j=1}^{D} N^{\boldsymbol{x}}_{j,k}(\boldsymbol{x}) a_{j,k} = (\boldsymbol{N}^{\boldsymbol{x}}_k)^T\ \boldsymbol{a}_k \tag{2.49}$$

$$T_k(t) = \sum_{j=1}^{D} N^{t}_{j,k}(t) b_{j,k} = (\boldsymbol{N}^{t}_k)^T\ \boldsymbol{b}_k \tag{2.50}$$

$$P_k(\mu) = \sum_{j=1}^{D} N^{\mu}_{j,k}(\mu) c_{j,k} = (\boldsymbol{N}^{\mu}_k)^T\ \boldsymbol{c}_k \tag{2.51}$$

where D is the number of degrees of freedom of the chosen approximation, $\boldsymbol{N}^{\boldsymbol{x}}_k$ represents the basis vectors considered for approximating the kth mode depending on the variable $\boldsymbol{x}$, while $\boldsymbol{a}_k$ represents the associated weights, and similarly for the other variables (t and μ). The choice of the set of basis functions, in which the functions are expressed, is made based on the studied problem. For example, a polynomial basis or a kriging basis can be selected (Ibañez et al., 2018). As described in the PGD section, the determination of the separated functions at each mode $k = 1, \ldots, M$ is achieved by using a greedy algorithm such that the mth order term is calculated once the approximation up to order $m - 1$ is known:

$$u^m(\boldsymbol{x}, t, \mu) = \sum_{k=1}^{m-1} X_k(\boldsymbol{x}) T_k(t) P_k(\mu) + X_m(\boldsymbol{x}) T_m(t) P_m(\mu) \tag{2.52}$$

A minimization problem (often using the L2-norm) is addressed to find u^m such that:

$$u^m(\boldsymbol{x}, t, \mu) = \arg\min_{\tilde{u}} \sum_{i=1}^{P} \|u_i - \tilde{u}(\boldsymbol{x}, t, \mu_i)\|_2^2 \tag{2.53}$$

where $\|u_i - \tilde{u}(\boldsymbol{x}, t, \mu_i)\|_2^2 = (u_i - \tilde{u}(\boldsymbol{x}, t, \mu_i))^2$.

It is worth noting that the product of the test function $w^*(\boldsymbol{x}, t, \mu)$ with the solution function $u(\boldsymbol{x}, t, \mu)$

is only evaluated at few locations, corresponding to the available sampled data. Thus, the test function is expressed as a set of Dirac delta functions collocated at the sampling data (denoted by P), instead of using the test function in a finite element context, such that:

$$w^*(\boldsymbol{x},t,\mu) = (X^*(\boldsymbol{x})T_m(t)P_m(\mu) + X_m(\boldsymbol{x})T^*(t)P_m(\mu) + X_m(\boldsymbol{x})T_m(t)P^*(\mu))\sum_{i=1}^{P}\delta(\boldsymbol{x},t,\mu_i) \quad (2.54)$$

A nonlinear system of equations is derived, due to the products of functions. Hence, an iterative scheme based on an alternating direction strategy is addressed to linearize the problem and to find the vectors $\boldsymbol{a}_k$, $\boldsymbol{b}_k$ and $\boldsymbol{c}_k$ of the Eq.(2.51), as shown in Figure 2.1.

Since one of the advantages of the sPGD method is its unstructured aspect of the dataset, a very important point to address is the definition of the numerical DoE according to a sampling method using the predefined input parameters and their interval of variation. The DoE should be fixed in a way to extract the maximum information from data to build the parametric solution. Among the existing sampling strategies, we can mention the Latin hypercube sampling (LHS) (Stein, 1987) which guarantees good coverage and a reasonable representation of the parametric domain.

Besides, when combining high-degree approximations (induced by the separated representations) with not sufficiently rich databases (as is always the case when operating in highly multi-parametric settings), the risk of overfitting may increase. The notion of overfitting is explained and illustrated in section 2.3.1. To overcome overfitting, the sPGD uses the modal adaptivity strategy (MAS) proposed in (Ibañez et al., 2018) that consists in adapting the approximation bases such that its degree is kept to a minimum in the first PGD modes. This degree is then increased progressively for the calculation of higher-level modes.

However, one of the limitations of the method is the use of the L2-norm for minimizing the residual. Indeed, to deal with sparsity in order to employ extremely rich approximations while exploiting parsimony, the use of L1-norm instead of L2-norm can provide more representative approximations with a good compromise between sparsity enforcement and computational efficiency. Recent works based on combining separated representation constructor with Lasso regularization or Lasso and Ridge regularization (the so-called Elastic-net regularization), that make use of the L1 and (L1 + L2) norms respectively, have been proposed in (Sancarlos et al., 2021). Except the just mentioned techniques, there is another recent technique, the so-called ANOVA-PGD, which aims at allying orthogonal hi-

erarchical bases with a more favorable scaling (with respect to the SSL) of the amount of data with the approximation richness. For that purpose, separated representations and sparse approximations (eventually regularized) will be combined for addressing multiple correlation terms (Sancarlos et al., 2021).

2.3 Learning physics from data with machine learning and interpolation algorithms

The term "machine learning" (ML) refers to a class of algorithms capable of extracting relevant information from various data without being explicitly programmed. It can be defined as an artificial intelligence technology. These algorithms are built from different fields such as statistics, computer science, and information. They require data to be analyzed and used to train models. In recent years, the increase in large datasets and the accessibility of the algorithms explain the growing popularity of ML and data science in general. Many applications that use ML can be found in everyday life, such as disease forecasting, detection of spam emails, or facial recognition. Most ML methods are usually divided into three categories:

- supervised learning: the goal is to find the relationship between some input parameters and a labeled output. A function is inferred from the training data forming pairs of inputs and outputs. If the output data are continuous, the supervised learning method solves a regression task. On the contrary, if discrete outputs are considered, the classification problem has to be solved;

- unsupervised learning: the algorithm extracts hidden patterns from unlabeled data. The training set is no more associated with some outputs. Unsupervised learning can be applied to several tasks, for instance, clustering, dimensionality reduction, or generative modeling;

- reinforcement learning: an agent is trained to interact with an environment by maximizing a numerical reward signal. Therefore, this agent must discover the best action to generate the best reward using "trial and error". Direct instructions are not given to the agent. Moreover, a specific action may affect the reward over an extended period, thus complicating the task of the agent. Typical examples of reinforcement learning include robot control, telecommunication, or games.

In particular, ML tools are starting to be widely used in engineering problems to fit models to some physical QoI. In this Ph.D. thesis, a coupling between MOR and ML techniques was adopted to opti-

mize the computational costs of the IH process.

This section shows some ML methods currently used for regression purposes and some interpolation methods as well. For notational convenience, we consider P observations (training data to build the model) such that $\{y_i\}_{i=1}^P$ is the physical quantity known at P training points and used for building an estimator $\tilde{y}$ of the objective function y, and $\{\boldsymbol{x}_i\}_{i=1}^P$ is the set of input parameters (predictors or features) such that $\boldsymbol{x} = [x_1, \cdots, x_s]^T \in \mathbb{R}^s$ with s is the number of input parameters, and $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_P]^T \in \mathbb{R}^{P\times s}$ is the matrix of input parameters.

2.3.1 Polynomial regression

The polynomial regression is usually considered one of the simplest regression methods. To simplify the mathematical illustration, the formalism of one-dimensional polynomial regression is presented. Therefore, each training sample is one dimensional such that $s = 1$ and $\boldsymbol{X} \in \mathbb{R}^{P\times 1}$. The generalization for $s > 1$ can be found in (Fang et al., 2005) for instance.
The classical polynomial regression $\tilde{y}$ of order m is defined by:

$$\tilde{y}(x) = w_0 + w_1 x + \cdots + w_m x^m = \sum_{i=0}^{m} w_i x^i, \ \forall\ x \in \mathbb{R} \tag{2.55}$$

such that the polynomial coefficients or weight vector $\boldsymbol{w} = [w_0, w_1, \cdots, w_m]^T \in \mathbb{R}^m$ is to be determined. The weights are computed by fitting the polynomial to the training data such that a cost function $\mathcal{F}$, introduced in Eq.(2.56), should be minimized using the method of least squares.

$$\mathcal{F}(\boldsymbol{w}) = \frac{1}{2P}\sum_{i=1}^{P}(y(x_i, \boldsymbol{w}) - \tilde{y}(x_i, \boldsymbol{w}))^2 = \|\boldsymbol{y} - \boldsymbol{V}\boldsymbol{w}\| \tag{2.56}$$

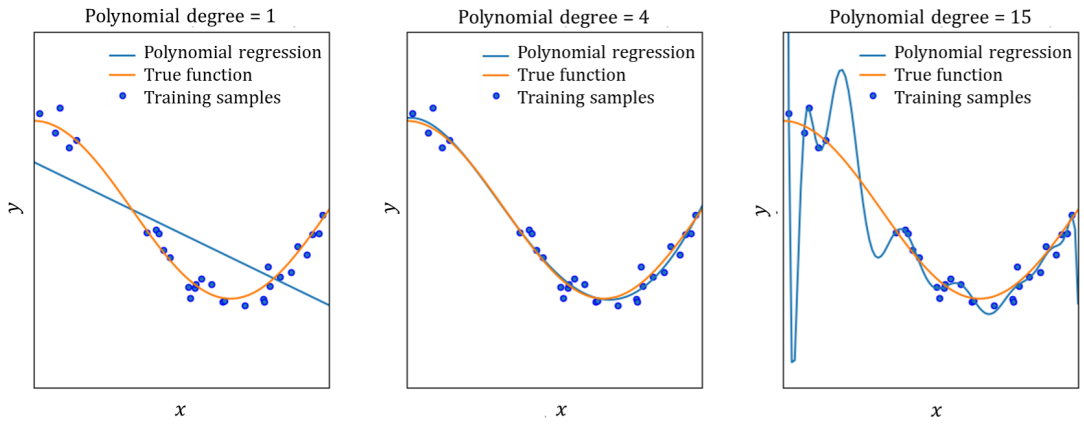
where $\boldsymbol{y}$ is the vector of true solutions at the P samples and $\boldsymbol{V}$ is the Vandermonde matrix defined as (Sobester et al., 2008):

$$\boldsymbol{V} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_P & x_P^2 & \cdots & x_P^m \end{bmatrix} \in \mathbb{R}^{P\times m} \tag{2.57}$$

The expression of the cost function $\mathcal{F}$ represents the mean of the sum of the squared differences between the true values and their associated predictions for each training point x_i. The cost function $\mathcal{F}$ should be minimized to obtain the optimal values of polynomial coefficients such that:

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \mathcal{F}(\boldsymbol{w}) \tag{2.58}$$

This minimization problem can be solved analytically by setting the derivative of $\mathcal{F}$ to zero or numerically by using the gradient descent algorithm (Brunton and Kutz, 2019), for instance. However, the order of the regression m remains a key issue, and it is usually fixed by the user. For low-order polynomials ($m = 1$, 2, or 3), regressions can also be referred to as response surface methods (RSM) (Myers et al., 2016) if they are built in some relatively small regions of the parameter space. Indeed, the choice of the polynomial degree m should be done judiciously, because prediction errors are highly impacted by this parameter, which can be interpreted as an example of the bias-variance tradeoff (Papachristoudis, 2019). For low polynomial orders, smooth and simple models are built with high bias and low variance that do not accurately capture the relationships between inputs and the response y, this is known as underfitting. While for high polynomial orders, complex models with high variance are generated and higher model error can be obtained for unseen parameters. This phenomenon is known as overfitting (Lawrence et al., 1997), which demonstrates the non-proportionality between model complexity and model error. Figure 2.2 shows an example of underfitting (left figure) and overfitting (right figure) where the polynomial regression differs strongly from the true solution function. Those phenomena of underfitting and overfitting are general and can be encountered in other regression methods.



(a) (b) (c)

Figure 2.2: Illustration of underfitting (a), accurate polynomial regression model (b), and overfitting (c) (adapted from (Scikit-learn))

To overcome overfitting, a classical solution consists of adding a penalty term in the cost function $\mathcal{F}$. This latter method is called regularization. Among the widely used regularization methods, we can mention Ridge, Lasso, and Elastic-net regularizations (Tian and Zhang, 2022).

The generalization of the polynomial regression to multivariate cases increases drastically the number of polynomial basis in the expression. For input parameter $\boldsymbol{x}$ of size s, a second-order polynomial model is expressed as follows:

$$\tilde{y}(\boldsymbol{x}) = w_0 + \sum_{i=1}^{s} w_i x_i + \sum_{i=1}^{s}\sum_{j=1}^{s} w_{ij} x_i x_j, \ \forall \boldsymbol{x} \in \mathbb{R}^s \tag{2.59}$$

Increasing the order of the multivariate polynomial regression increases the complexity of the model and may lead to numerical issues. For this reason, it is preferred to use an orthogonal basis instead of the canonical basis. In addition, the data size required to fully define the system increases exponentially with the order m and the size of the input parameters s.
In conclusion, despite its easy interpretation, implementation, and mathematical manipulation, polynomial regression has some major drawbacks, namely its incapacity of fitting high-dimensional problems.

2.3.2 Support vector regression

The SVR has been proposed by Drucker et al. (1996) as a supervised learning algorithm. It has been proven to be an effective method for regression and time series prediction (Smola and Schölkopf, 2004). One of its main advantages is that its computational complexity doesn't depend on the dimensionality of the input space. The SVR tries to find an estimated function $\tilde{y}$ by using the ε-insensitive approach. This latter consists in introducing a flexible ε-insensitive region around the estimated function, the so-called ε-tube. Basically, an optimization problem is first formulated by defining a convex ε-insensitive loss function to be minimized and to find the flattest tube that best approximates the function y and contains most of the training instances, while balancing model complexity and prediction error. Hence, a function is constructed from the loss function and the geometrical properties of the tube. The optimization, which has a unique solution, is then solved, using appropriate numerical optimization algorithms. It is worth noting that for SVR, the training and testing data are assumed to be independent and identically distributed (IID) in a supervised learning context.

2.3.2.1 Concepts and mathematical formulation of linear SVR

The general form of the scaler approximated function $\tilde{y}$ is given by the Eq.(2.60):

$$\tilde{y}(\boldsymbol{x}) = b + \langle \boldsymbol{w}, \boldsymbol{x} \rangle = b + \boldsymbol{w}^T \boldsymbol{x} \tag{2.60}$$

with $b \in \mathbb{R}$ is a bias, $\boldsymbol{w} = [w_1, \cdots, w_s]^T \in \mathbb{R}^s$ is the weight vector and $\langle \cdot, \cdot \rangle$ denotes the dot product. Like the polynomial regression, these two quantities (b and $\boldsymbol{w}$) are unknown and have to be determined. Then, a hyperplane is built by Eq.(2.60) and fitted to the training data with a given tolerance margin ε. Indeed, the concept of flatness, in this case, means that one seeks a small $\boldsymbol{w}$. The formulation of the linear SVR can be seen as a minimization problem of the norm of the weight vector $\boldsymbol{w}$ constrained by the training error contained in the ε-tube (Smola and Schölkopf, 2004) such that ε can be tuned to gain the desired accuracy of the model. In other words, we do not care about errors as far as they are less than ε, but will not accept any deviation larger than this. The new objective function and constraints are as follows:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 \\ &\text{subject to} \quad \begin{cases} y_i - b - \boldsymbol{w}^T \boldsymbol{x}_i \leq \varepsilon \\ b + \boldsymbol{w}^T \boldsymbol{x}_i - y_i \leq \varepsilon \end{cases} \end{aligned} \tag{2.61}$$

However, if there is no function $\tilde{y}$ that satisfies these constraints for all points. For example, if ε is very small, there is no solution to the problem. Hence, the existence of the solution is not guaranteed. To deal with otherwise infeasible constraints, a new optimization problem is created by introducing slack variables ξ_i and ξ_i^* for each training point, allowing the violation of some constraints. The concept of slack variables is simple: for any value that falls outside ε-tube, we denote its deviation from the margin as ξ. These deviations can potentially exist, but we would still like to minimize them as much as possible. Therefore, we can add these deviations to the objective function to get the formulation stated in (Vapnik, 1995).

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{P}(\xi_i + \xi_i^*) \\ &\text{subject to} \quad \begin{cases} y_i - b - \boldsymbol{w}^T \boldsymbol{x}_i \leq \varepsilon + \xi_i \quad \text{for } i = 1, \cdots, P \\ b + \boldsymbol{w}^T \boldsymbol{x}_i - y_i \leq \varepsilon + \xi_i^* \quad \text{for } i = 1, \cdots, P \\ \xi_i, \xi_i^* \leq 0 \quad \text{for } i = 1, \cdots, P \end{cases} \end{aligned} \tag{2.62}$$

where C is a positive constant, introduced to determine a trade-off between the flatness of the model and the tolerance of model error greater than ε and to prevent overfitting (regularization). This corresponds to dealing with a so-called ε-insensitive loss function that penalizes predictions farther than ε. Several loss functions can be adopted such as the linear, quadratic, and Huber functions (Smola and Schölkopf, 2004). The choice of the loss function is influenced by a priori information about the noise distribution affecting the training data, the model sparsity, and the training computational complexity. Figure 2.3 graphically depicts the situation, where the points outside the ε-tube are

penalized using a linear loss function:

$$L_\varepsilon(y,\tilde{y}) \begin{cases} 0 \text{ if } |y-\tilde{y}| \leq \varepsilon \\ |y-\tilde{y}| - \varepsilon \quad \text{otherwise} \end{cases} \tag{2.63}$$

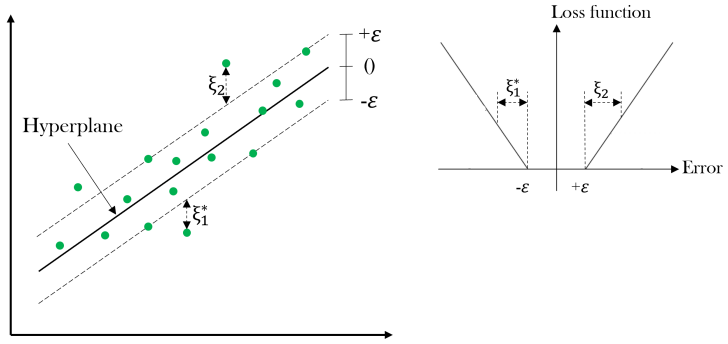


Figure 2.3: Illustration of the linear SVR with slack variables and linear loss function

It turns out that in most cases the optimization problem 2.62 can be solved more easily in its dual form. The key idea is to construct a Lagrange function $\mathcal{L}$ from the objective function and the corresponding constraints, by introducing Lagrange multipliers, or dual variables α, α^*, λ and λ^*. More details could be found in (McCormick, 1983; Mangasarian, 1994; Vanderbei, 1999). The Lagrangian function is given by:

$$\begin{aligned} \mathcal{L}(\boldsymbol{w},\xi,\xi^*,\alpha,\alpha^*,\lambda,\lambda^*) = &\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{P}(\xi_i+\xi_i^*) + \sum_{i=1}^{P}\alpha_i(y_i - b - \boldsymbol{w}^T\boldsymbol{x}_i - \varepsilon - \xi_i) + \\ &\sum_{i=1}^{P}\alpha_i^*(b + \boldsymbol{w}^T\boldsymbol{x}_i - y_i - \varepsilon - \xi_i^*) - \sum_{i=1}^{P}(\lambda_i\xi_i + \lambda_i^*\xi_i^*) \end{aligned} \tag{2.64}$$

such that dual variables should satisfy the positivity constraints. By differentiating $\mathcal{L}$ with respect to the primal variables $(\boldsymbol{w}, b, \xi, \xi^*)$, setting them to zero for optimality and substituting the result from the derivatives into Eq.(2.64), the dual problem formulation is as follows:

$$\begin{aligned} \text{maximize} \quad & -\varepsilon\sum_{i=1}^{P}(\alpha_i+\alpha_i^*) + \sum_{i=1}^{P}y_i(\alpha_i-\alpha_i^*) - \frac{1}{2}\sum_{i,j=1}^{P}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\boldsymbol{x}_i^T\boldsymbol{x}_j \\ \text{subject to} \quad & \begin{cases} \sum_{i=1}^{P}(\alpha_i-\alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \tag{2.65}$$

It can be noticed that the dual variables λ and λ^* are eliminated by expressing them as a function of C, α, and α^* using the derivatives of $\mathcal{L}$ with respect to ξ and ξ^*. Hence, by using derivative of $\mathcal{L}$ with

respect to $\boldsymbol{w}$, the final expression of the SVR is given by:

$$\tilde{y}(\boldsymbol{x}) = b + \sum_{i=1}^{P}(\alpha_i - \alpha_i^*)\boldsymbol{x}_i^T\boldsymbol{x} \tag{2.66}$$

where $\boldsymbol{w} = \sum_{i=1}^{P}(\alpha_i - \alpha_i^*)\boldsymbol{x}_i$. This is the so-called Support Vector expansion, such that $\boldsymbol{w}$ can be completely described by a linear combination of the training patterns $\boldsymbol{x}_i$. Hence, the complexity of a function's representation is independent of the dimensionality of the input space $\mathbb{R}^s$. The dual problem is solved by the quadratic program algorithm. Details about the algorithm can be found in (Smola and Schölkopf, 2004) and references therein. The constants C and ε are considered hyperparameters of the model and have to be well-chosen as they directly influence the shape of the model.

2.3.2.2 Nonlinear SVR and kernel trick

So far, the linear SVR has been explained, the next step is to address the nonlinear SVR for more complicated cases. This could be achieved by simply preprocessing the training input data $\boldsymbol{x}_i$ by a map $\phi : \mathbb{R}^s \to \mathcal{G}$ into some feature space $\mathcal{G}$, and then applying the standard SVR algorithm. This concept has been addressed in (Aizerman, 1964; Nils J., 1965). The dimension of $\mathcal{G}$ is usually higher than the dimension of the input space $\mathbb{R}^s$. However, since the SVR approximation function is provided by dot products between $\boldsymbol{x}_i$, expensive computation cost could be generated with dot product in high dimension. To overcome this issue, the dot product in Eq.(2.66) is replaced by a function k, the so-called kernel function, such that $k(\boldsymbol{x}, \boldsymbol{x}') := \langle\phi(\boldsymbol{x}), \phi(\boldsymbol{x}')\rangle = \phi(\boldsymbol{x})^T \cdot \phi(\boldsymbol{x}')$. This function k must satisfy several properties, explained in (Smola and Schölkopf, 2004). Hence, the optimization problem and the approximation function become, respectively

$$\begin{aligned} &\text{maximize} \quad -\varepsilon\sum_{i=1}^{P}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{P} y_i(\alpha_i - \alpha_i^*) - \frac{1}{2}\sum_{i,j=1}^{P}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &\text{subject to} \quad \begin{cases} \sum_{i=1}^{P}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \tag{2.67}$$

$$\tilde{y}(\boldsymbol{x}) = b + \sum_{i=1}^{P}(\alpha_i - \alpha_i^*)k(\boldsymbol{x}_i, \boldsymbol{x}) \tag{2.68}$$

In the nonlinear setting, the optimization problem corresponds now to finding the flattest function in the feature space, not in the input space. A simple illustration of the mapping of the input parameters to a feature space of higher dimension, the so-called kernel trick is shown in Figure 2.4 in a classification context.

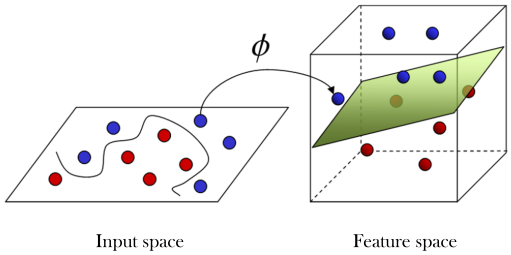


Figure 2.4: Illustration of the kernel trick (Zararsiz et al., 2012)

Among the existed kernel functions, Table 2.1 shows some commonly used functions.

Table 2.1: Usual kernel functions

Function names	Expressions
Polynomial function	$k(\boldsymbol{x}, \boldsymbol{x'}) = (\langle \boldsymbol{x}, \boldsymbol{x'} \rangle + r)^d$ with $r \in \mathbb{R}$ and $d \in \mathbb{N}$
Sigmoid function	$k(\boldsymbol{x}, \boldsymbol{x'}) = \tanh(\kappa \langle \boldsymbol{x}, \boldsymbol{x'} \rangle + \gamma)$ with $\kappa, \gamma \in \mathbb{R}^+$
Radial basis function (gaussian)	$k(\boldsymbol{x}, \boldsymbol{x'}) = \exp(-\frac{\|\boldsymbol{x} - \boldsymbol{x'}\|^2}{2\sigma^2})$ with $\sigma \in \mathbb{R}^+$

2.3.3 Boosted regression trees or gradient boosting

The boosted regression trees (BRT) method is a ML method (Friedman et al., 2000; Schapire, 2003) that differs fundamentally from the traditional ones that produce a single "best" model. The BRT method enables to fit and combine a large number of relatively simple weak prediction models, typically decision trees (called here "weak learner"), generating then a single strong learner in a forward step-wise manner. It allows to improve the prediction performance (Elith et al., 2006; Leathwick et al., 2006) and optimize an arbitrary loss function. The BRT method usually outperforms the random forest method (Hastie et al., 2009; Madeh Piryonesi and El-Diraby, 2021). Basically, the BRT method uses two algorithms: regression trees and boosting algorithm.

The tree-based models enable to partition the input parameter space into rectangles, using a series of rules to identify regions having the most homogeneous responses to parameters. They then fit a constant to each region. This constant value represents the mean response for observations in that region, assuming normally distributed errors. More details about regression trees can be found in (Loh, 2011). Decision trees are popular because they are easy to visualize and analyze results thanks

to their architecture and their ability to deal with different data types (numeric, binary, categorical, etc.) and different scaled input variables (no need for standardizing data). The hierarchical structure of a tree enables to automatically model the interactions between input parameters. Despite these advantages, trees are not usually very accurate. They have difficulties for modeling smooth functions even ones as simple as a straight-line response. Moreover, the tree structure depends on the training data, and small changes on them can result in very different series of splits (Hastie et al., 2009). These difficulties of decision trees could introduce uncertainty into their interpretation and limit their predictive performance.

The boosting algorithm is introduced as an iterative functional gradient descent algorithm that proceeds iteratively, such that models (e.g. decision trees) are fitted to the training data, then, an appropriate approach is gradually used to improve the poorly predicted observations. Boosting algorithms vary in how they quantify the lack of accuracy and select settings for the next iteration. This is why the method is often known as gradient boosting (GB).

Consider a loss function that represents the loss in predictive performance. This function should be differentiable and can be arbitrary and takes different forms (squared error, absolute error, Huber, quantile, etc.). For ease of exposition, we consider the mean squared error loss function as follows:

$$\mathcal{F}(y, \tilde{y}) = \frac{1}{P} \sum_{i=1}^{P} (\tilde{y}(\boldsymbol{x}_i) - y(\boldsymbol{x}_i))^2 \tag{2.69}$$

Boosting is a numerical optimization technique that consists in adding a new tree, at each step, that minimizes the loss function. Indeed, if we consider a GB algorithm with M steps and a predictive model $\tilde{y}^m$ at each step m ($1 \leq m \leq M$), the first "weak learner" $\tilde{y}^1$ is very basic and is simply the average of the observations ($\tilde{y}^1 = \bar{y}$). It is therefore not very accurate, but it serves as a basis for the rest of the algorithm. Then, for $m \geq 2$ the algorithm add a new estimator $h^m(\boldsymbol{x}_i)$ to improve $\tilde{y}^m$, such that:

$$\tilde{y}^{m+1}(\boldsymbol{x}_i) = \tilde{y}^m(\boldsymbol{x}_i) + h^m(\boldsymbol{x}_i) = y_i \tag{2.70}$$

or equivalently,

$$h^m(\boldsymbol{x}_i) = y_i - \tilde{y}^m(\boldsymbol{x}_i) \tag{2.71}$$

The last term representing the difference between the prediction and the reality is called the residual such that each $\tilde{y}^{m+1}$ attempts to correct the errors of its predecessor $\tilde{y}^m$. The main particularity of the GB method is to fit a prediction model at each step m ($m \geq 2$) but for the residuals instead of the data itself.

2.3.4 Artificial neural network

The ANN is designed to emulate the mechanism of the human brain and is generally considered a black box model due to its complexity. It is defined as an oriented graph, composed of a set of artificial neurons (so-called nodes). Recently, ANN has gained popularity in the literature due to the massive use of Deep Learning. This method shows an impressive performance in many applications in machine learning and pattern recognition, compared to classical methods. However, Deep Learning generally requires a big volume of data, which is not always the case especially when a low-data limit case is considered.

Indeed, ANN is characterized by their architecture composed of three different kinds of layers: the input layer, the hidden layer(s), and the output layer, as shown in Figure 2.5 for a multi-layer perceptron (MLP) with only one hidden layer, an input vector of size s and a single output. The learning algorithm of an MLP involves a forward-propagation step followed by a backward-propagation step. The fundamental concepts of MLP algorithms are discussed extensively in the literature (Ruhmelhart et al., 1986; Lin and Lee, 1996).

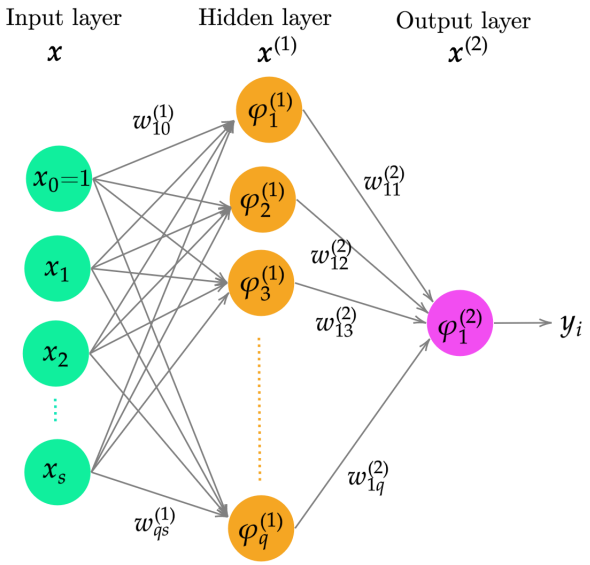


Figure 2.5: Illustration of a MLP architecture

The layers are connected in a sequential manner. The first layer consists of the input features distributed over the nodes. Each input node sends information to the next layer of nodes via the network's edges. Each edge has a numeric weight, denoted by $w_{kj}^{(l)}$, such that l, k, j indicate the ANN layer number, the neuron k of the next layer, and the neuron j of the previous layer, respectively. It is worth

noting that the weight $w_{k0}^{(1)}$, for $k = 1, \cdots, q$, can be seen as a bias and that weights are initialized as small random numbers and updated later. Then, the hidden layer extracts some of the most relevant patterns from the inputs by applying nonlinear activation functions on the linear combination of the inputs, and then sends them to the next layer for further analysis. Hence, the forward-propagation phase propagates information forward to the output layer through the hidden layer(s). In each successive layer, every neuron builds a linear combination of its inputs and then applies an activation function on these combinations in order to capture the presence of nonlinear relationships between the inputs and to compute its output. Finally, the final output is shown in the output layer.
The activation functions have different types and are often chosen from the expressions illustrated in Table 2.2:

Table 2.2: Usual activation functions

Function names	Expressions
Identity function	$\varphi(x) = x$
Sigmoid function	$\varphi(x) = \frac{1}{1+e^{-ax}}$, with $a \in \mathbb{R}^+$
Hyperbolic tangent	$\varphi(x) = \dfrac{1 - e^{-2x}}{1 + e^{-2x}}$
Threshold function	$\varphi(x) = \begin{cases} 1 & \text{if} x \geq a, \text{with } a \in \mathbb{R}^+ \\ 0 & \text{otherwise} \end{cases}$
Piecewise-linear function	$\varphi(x) = \begin{cases} 1 \text{ if } x \geq \frac{1}{2} \\ x \text{ if } -\frac{1}{2} < x < \frac{1}{2} \\ 0 \text{ otherwise} \end{cases}$

Sigmoid, hyperbolic tangent, and linear functions are particularly suitable for learning algorithms with backpropagation techniques (described later) since they are analytically differentiable. More recently, rectified linear units (ReLUs), leaky rectified linear units (leaky ReLUs), and exponential linear units (ELUs) become more common.

Finding the "optimum" values of weights w that minimize prediction error is critical for building a successful model. The backward-propagation algorithm does this by converting the ANN into a learning algorithm by learning from errors. It starts at the output layer and propagates recursively the error of the output layer through the hidden and input layers. To find the optimum value for w, small adjustments are achieved, and the impact on prediction errors is examined. Finally, we get the optimal values of w when further weight changes do not reduce errors. The optimization approach can be provided by different algorithms such as gradient descent, Broyden–Fletcher–Goldfarb–Shanno, or

Levenberg-Marquardt methods (Haykin, 1998). The gradient descent method is often employed to quantify prediction errors and solve a minimization problem:

$$\mathcal{F}(\boldsymbol{w}) = \frac{1}{2P}\sum_{i=1}^{P}\left(y(\boldsymbol{x}_i,\boldsymbol{w}) - \tilde{y}(\boldsymbol{x}_i,\boldsymbol{w})\right)^2 \tag{2.72}$$

During the process, the weights are updated to reduce this error, such that if the algorithm is assumed to be at the rth iteration, the new values of the weight w_{kj}^{l} are given by the gradient descent update at the r+1 iteration:

$$w_{kj}^{(l)}|_{r+1} = w_{kj}^{(l)}|_{r} - \alpha\frac{\partial\mathcal{F}}{\partial w_{kj}^{(l)}} \tag{2.73}$$

where $\alpha > 0$ is known as the learning rate and should be fixed. More details about the algorithm are provided in (Ruhmelhart et al., 1986; Wasserman, 1989; Haykin, 1998).

2.3.5 Radial basis functions interpolation

The RBF interpolation is a popular method for the approximation of scattered data for high-dimensional problems because of its good approximation properties and ease of implementation. It provides a general and flexible way of interpolation even for unstructured data where it is often impossible to apply polynomial or spline interpolation. RBF is used in many fields, ranging from statistics to the approximation of partial differential equations (Buhmann, 2000; Wendland, 2004; Jakobsson et al., 2009). Consider again the function y to be interpolated at the set of data points $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{P}$, all in $\mathbb{R}^s$, and consider φ, a positive real function. The interpolation space consists of all functions of the form

$$\tilde{y}(\boldsymbol{x}) = \sum_{i=1}^{P} w_i\varphi(\|\boldsymbol{x}-\boldsymbol{x}_i\|) + g(\boldsymbol{x}) \tag{2.74}$$

where w_i are real coefficients and g is a function in $\prod_m(\mathbb{R}^s)$ with $\prod_m(\mathbb{R}^s)$ the space of polynomials of degree at most m in $\mathbb{R}^s$. The function φ is called radial basis function and can take different forms, as shown in Table 2.3, and $\|.\|$ denotes the Euclidean norm. The polynomial space should be chosen so that is larger or equal to the order of the basis function (this guarantees the existence of a solution to the interpolation problem).

Table 2.3: Common radial basis functions

RBF	Expressions
Linear	$\varphi(r) = r$
Gaussian	$\varphi(r) = e^{-(\varepsilon r)^2}$
Multiquadric	$\varphi(r) = \sqrt{1+(\varepsilon r)^2}$
Inverse quadratic	$\varphi(r) = \frac{1}{1+(\varepsilon r)^2}$
Thin plate spline	$\varphi(r) = r^2 ln(r)$

Some basis functions introduce a parameter ε, the so-called shape parameter, to scale the basis function. The data points, $\boldsymbol{x}_i$, are sometimes called centers, and the basis functions are radial around these centers. The coefficients $\boldsymbol{w} = [w_1, \cdots, w_P]^T$ and the polynomial are then chosen so that $\tilde{y}$ interpolates y exactly at the data points such that:

$$\tilde{y}(\boldsymbol{x}_i) = y(\boldsymbol{x}_i) \text{ for } 1 \leq i \leq P \tag{2.75}$$

and $\boldsymbol{w}$ should satisfy

$$\sum_{i=1}^{P} w_i \tilde{g}(\boldsymbol{x}_i) = 0 \tag{2.76}$$

for all polynomials $\tilde{g} \in \prod_m(\mathbb{R}^s)$ with $D = \dim(\prod_m(\mathbb{R}^s))$, and consider $\{b_j\}_{j=1}^{D}$ the polynomial basis. Then, the polynomial g can be expanded on this basis by:

$$g(\boldsymbol{x}) = \sum_{j=1}^{D} \beta_j b_j(\boldsymbol{x}_i) \tag{2.77}$$

with $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_D]$ is the vector of the polynomial basis coefficients for the function g. This leads to a linear system of equations for the coefficients $\boldsymbol{w}$ and $\boldsymbol{\beta}$ which takes the form

$$\begin{pmatrix} \boldsymbol{M} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{pmatrix} \tag{2.78}$$

where $\boldsymbol{M}$ is the interpolation matrix befined by $M_{ik} = \varphi(\|\boldsymbol{x}_i - \boldsymbol{x}_k\|)$ for $1 \leq i, k \leq P$, $\boldsymbol{B}$ is the $P \times D$ matrix composed by the polynomial basis, and $\boldsymbol{y} = [y_1, \cdots, y_P]^T$ the vector of objective function values. More details about the problem resolution, the invertibility of the matrix $\boldsymbol{M}$, and the existence and uniqueness of the solution can be found in (Jakobsson et al., 2009) and references therein.

2.4 Dimensionality reduction for manifold learning

In many areas of science and engineering, datasets are defined in high-dimensional space with a large number of features or dimensions into which the computational complexity is very high. However, very often, data are found to lie on manifolds of reduced dimension d embedded within the high-dimensional space $\mathbb{R}^D$ in which the problems are defined. The extraction of these embedded manifolds allows to discard redundancy and simplify further operations and hence reducing significantly the computational complexity. This fact is at the heart of MOR techniques. For instance, the POD technique extracts first this manifold and then proceeds to solve problems by exploiting the low-dimensional manifold ($d \ll D$). On the contrary, PGD constructs the manifold and its approximation at the same time.

By following the same rationale in the case of a parametric model and ML problems, dimensionality reduction allows the extraction of the number of informative and uncorrelated parameters (that depend linearly or nonlinearly on the original model parameters) enabling to obtain more useful representations of the information in data for subsequent operations such as interpolation, classification, visualization, denoising, or outlier detection. This way of doing things becomes extremely useful when solving a parametric problem. This is because the lower the number of significant parameters is, the simpler its parametric metamodeling becomes: its offline construction and online manipulation.

Accordingly, we may distinguish two major classes of dimensionality reduction methods: linear and nonlinear. The former includes the classical PCA (Pearson, 1901; Jolliffe, 2005), multidimensional scaling (Cox and Cox, 2008), linear discriminant analysis (Ye et al., 2004), and independent component analysis (Stone, 2002). Linear techniques attempt to find new parameters that obey some optimization criteria and can be expressed as a linear combination of the original ones. Consequently, they often fail when the input data have curved or nonlinear structures. To overcome this issue, many nonlinear dimensionality reduction methods have been developed. These methods fall into global and local categories. The global nonlinear dimensionality reduction methods include the kernel PCA (Mika et al., 1998), kernel LDA (Lu et al., 2003), and kernel Fisher discriminant analysis (Mika et al., 1999), which attempt to preserve structure at all scales by mapping the original input data into a feature space by a "global" nonlinear mapping where inner products in the feature space can be computed by a kernel function in the input space without explicitly knowing the nonlinear mapping (Schölkopf et al., 1998; Baudat and Anouar, 2000; Park and Park, 2005). On the other hand, local methods try to preserve the local structure of the data by mapping close data points in the initial high-dimensional space into close points in the low-dimensional representation. Among these manifold learning techniques that belong to this category of nonlinear dimensionality reduction, we mention locally linear embedding (LLE) (Roweis and Saul, 2000), local tangent space alignment (Zhang and Zha, 2003), laplacian eigenmaps (Belkin and Niyogi, 2001), as well as the self-organizing manifold mapping (Kitani et al., 2011). The main key point behind manifold learning methods is the assumption that the input data lie on a low-dimensional manifold embedded in a high-dimensional space. Therefore, the need to learn the underlying intrinsic manifold geometry is required to address the problem of dimensionality reduction. Hence, instead of looking for an optimum linear subspace, the lately mentioned manifold learning methods try to discover an embedding function, locally defined, that describes the intrinsic similarities of the data.

In the framework of this work, manifold learning techniques were considered to address the nonlinear

geometry of the gear into which interpolation is complex to handle. Initially, a 3D geometry was defined in the cartesian system of coordinates into which an embedding manifold of lower dimension (< 3) with uncorrelated parameters was looked for to simplify the spatial interpolation procedure. The choice of the employed technique was made based on some criteria:

- the ability to recover the intrinsic geometric structure of nonlinear data manifolds;
- the computational efficiency;
- the number of free parameters to be considered;
- the ease of implementation.

By considering these criteria, the LLE method was chosen to be applied thanks to its efficiency and simplicity as well. The next section provides a brief overview of this method.

2.4.1 Locally linear embedding

The LLE is a nonlinear dimensionality reduction method proposed by Roweis and Saul (2000). 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 (Saul and Roweis, 2003; Yotov et al., 2005; Wu and Wu, 2018). Hence, nonlinear dimensionality reduction can be simplified into local linear dimensionality reduction where each data point 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.

Consider a set of P real-valued input points of D dimensional vectors, $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{P}$. 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 Figure 2.6:

- step 1: assign k nearest neighbors to each data point $\boldsymbol{x}_i$ (Figure 2.6a). The k nearest neighbors are identified using the 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 (Figure 2.6b). 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:

$$\epsilon(\boldsymbol{W}) = \sum_{i=1}^{P} \left|\left| \boldsymbol{x}_i - \sum_{j=1}^{P} w_{ij} \boldsymbol{x}_j \right|\right|^2 \tag{2.79}$$

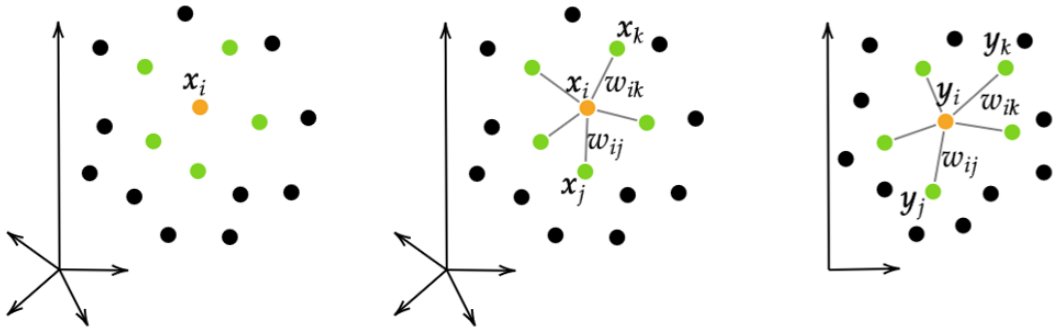
where $\boldsymbol{W} \in \mathbb{R}^{P \times P}$ is the weight matrix that includes the weights of linear reconstruction of all data points using their neighbors.

The minimization problem should satisfy two constraints: (1) the weights of linear reconstruction for each point satisfy $\sum_j w_{ij} = 1$ for $i = 1, ..., P$; (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.
Furthermore, it is worth noting that in the case where there are more neighbors than input dimensions ($k > D$), each data point can be perfectly reconstructed from its neighbors, however, the local reconstruction weights are no longer uniquely defined. In this case, some further regularization must be added to break the degeneracy, see (Saul and Roweis, 2003) for more details.

- step 3: compute the low-dimensional embedding vectors, $\boldsymbol{Y} = \{\boldsymbol{y}_i\}_{i=1}^{P}$, based on the idea that LLE preserves the local linearity from neighbors and the corresponding reconstruction weights (Figure 2.6c). 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:

$$\Phi(\boldsymbol{Y}) = \sum_{i=1}^{P} \left|\left| \boldsymbol{y}_i - \sum_{j=1}^{P} w_{ij} \boldsymbol{y}_j \right|\right|^2 \tag{2.80}$$

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



(a) (b) (c)

Figure 2.6: Steps of locally linear embedding: (a) assign k nearest neighbors, (b) reconstruct with linear weights, and (c) map to embedded coordinates

For more details about the resolution of minimization problems, the reader can refer to (Saul and Roweis, 2000, 2003). Compared with other nonlinear dimensionality reduction methods, LLE has only three parameters to be determined, which are the number of neighbors k, the dimension of the reduced order subspace, and the regularization parameter. Hence, the algorithm implementation is quite simple.

Recent works show new variants of LLE that can be more effective than the standard one. The interested reader can refer to (Ghojogh et al., 2020) for more details.

2.5 Summary and technical choices

2.5.1 Intrusive versus non-intrusive MOR

Projection-based reduced-order models and methods of residual minimization are called intrusive. This is because the knowledge of the governing equations (PDE and ODE expressions) representing the physical system is required to build the reduced-order model. In most cases, the modification of the FE source code must be carried out. However, this approach cannot be pursued for most commercial software for which the source code is not available and can not be modified by the user. These severe restrictions motivate the use of non-intrusive methods. The latter only require data computed by the full-order model.

Interpolated reduced-order model is a very popular non-intrusive approach, especially for parametric problems. Indeed, despite ignoring the physical model governed by mathematical equations, interpolated reduced-order models show good results by providing predictions of the QoI for new input

parameters not too far from the initial training points (Degroote et al., 2010). In this thesis, the FE source code of the used software was not available. For this reason, only non-intrusive approaches were considered. In particular, PODI models were investigated in order to build the parametric solutions to the multiphysics IH problem.

2.5.2 Metamodel or surrogate model

The parametric solutions that will be provided by the non-intrusive POD with interpolation or ML techniques for regression are called metamodels or surrogate models. They represent models, built from samples or training data, relating some input parameters $\boldsymbol{x}$ to some QoI y. A metamodel M aims at approximating the mapping between the input parameters and the QoI from the relation:

$$y \approx M(\boldsymbol{x}) \tag{2.81}$$

In this manuscript, the input parameters $\boldsymbol{x}$ are the IH process parameters, while the QoI y correspond to some IH outputs (temperature field and austenitic phase evolution for instance) measured at a set of sparse sensors inserted at different positions on the surface and within the considered workpiece.
The procedure of metamodeling requires four main steps at the offline stage: design of experiments, data generation, data pre-processing (dimension reduction, missing value imputation, etc.), and metamodel construction and validation. At the end of the process, we provide a metamodel into which fast and inexpensive predictions can be generated at the online stage for new input parameters, within the studied interval, that have not been used before during training. Figure 2.7 summarized the different steps involved in the process of metamodeling.

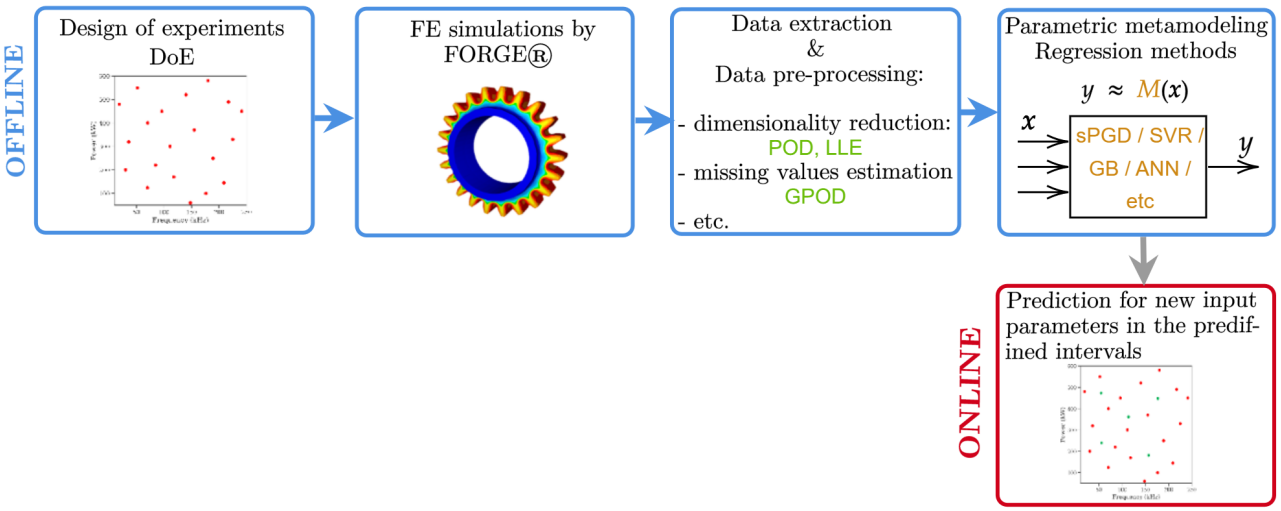


Figure 2.7: Steps of metamodeling approach

2.5.2.1 Design of experiments

For building metamodels, different kinds of DoE can be used. We can find model-free designs and adaptive designs; the former generate an a priori distribution of samples based on geometrical properties of the parameter space, such as the distance between samples, and the latter use information given by the evaluations of the QoI to propose relevant new samples improving the model in a precise objective. Since the nature of the model is unknown, a good DoE must fulfill two fundamental requirements (Santner et al., 2003):

- the samples should fulfill the parametric space in order to observe the response of the model at various conditions;
- each dimension must be explored as far as possible over the range of its variation. Therefore, two samples should not share a common value for a given dimension.

The next step is to appropriately choose the values of the input parameters and the number of experiments to build the DoE. The choice is not obvious because it depends on the chosen DoE and its specific characteristics. Indeed, it is worth pointing out that we are interested here in one-shot space-filling DoE with model-free designs, i. e. the parametric space is sampled with a given number of experiments defined a priori and without relying on information provided by the evaluations of the QoI. Among the most popular DoE, we can mention full factorial and fractional factorial designs

(Gunst and Mason, 2009; Zaman et al., 2020), maximin and minimax design (Johnson et al., 1990), central composite design (Djoudi et al., 2007), Taguchi design (Taguchi, 1987), Monte Carlo (Metropolis and Ulam, 1949), and LHS (Stein, 1987). The LHS technique is commonly used in data-driven non-intrusive reduced-order modeling because of its ability to fulfill the parameter space with a reduced number of samples. In the current work, the LHS DoE was considered due to the mentioned reasons. Basically, if the number of experiments is set to P, each dimension is split into P equiprobable subintervals where the samples are randomly distributed in each subinterval such that each sample is the only one in each axis-aligned hyperplane containing it. Fixing the sampling size P depends on the dimension of the input space, the regression methods to be applied, the complexity of the FE solver, and the cost of generating data. Hence, this task remains difficult as far as the complexity and properties of the FE solver are unknown. The generation of the DoE was achieved in Python version 3.8.

2.5.2.2 Data generation

All experiments of the DoE were executed on the FE FORGE® software. The simulation and model implementation are detailed in chapter 3.

2.5.2.3 Data pre-processing

All data extracted from FORGE® were treated with Python using Numpy and Pandas library to manipulate dataframes and matrices and to build numerical algorithms. The developed codes for dimensionality reduction (POD), manifold learning (LLE), and missing values estimation (GPOD) were also achieved with Python.

2.5.2.4 Metamodels construction and validation

To construct metamodels, different regression methods (sPGD, SVR, GB, ANN) were tested. To do that, Scikit-learn which is a free Python library dedicated to ML was employed (Pedregosa et al., 2011). All models, except the ANN and sPGD, were developed in Python using the Scikit-learn library. The sPGD code was provided in a separate sPGD library in Python and the ANN code was developed in Python using the Tensorflow and Keras libraries. For interpolation methods, the SciPy library in Python was employed. Once the metamodel is trained, its quality can be evaluated by using error metrics. Among others, we can mention the mean square error (MSE) or root MSE (RMSE), the mean absolute error (MAE), the normalized root mean square error (NRMSE), the root mean square

percentage error (RMSPE), and the mean absolute percentage error (MAPE), which are frequently used to measure the differences between the real and predicted values (Botchkarev, 2018). In the manuscript, different metrics were applied and introduced in separate sections. Some of those metrics are already implemented in the Scikit-learn library in Python and others were coded as functions. Finally, the Matplotlib library was used for the visualization of data and results.

2.6 Conclusion

This chapter presents a review of some MOR techniques. Those methods are different, and their application depends on the studied problem. Due to the restriction related to the intrusive MOR methods, non-intrusive methods were chosen to be applied. This approach is based on data that can be provided numerically or experimentally. In this manuscript, synthetic data provided by FE codes were provided by using FORGE® software. The physical modelisation and the development of the numerical model for the IH process will be detailed in the next chapter.

Chapter 3

Physical modeling and numerical simulation of induction heating process

Outline

3.1 Introduction 59
3.2 Mathematical formulation of physical phenomena 59
3.2.1 Modeling of electromagnetic phenomena 60
3.2.2 Modeling of thermal phenomena 64
3.2.3 Modeling of metallurgical phenomena 66
3.2.4 Modeling of mechanical phenomena 69
3.3 Coupling between physics 71
3.4 Numerical simulation of induction heating process 73
3.4.1 Principle of the IH simulation 73
3.4.2 Identification of material properties 76
3.4.3 Model development in FORGE® software 78
3.4.4 Numerical results 87
3.5 Conclusion 88

3.1 Introduction

The IH process is one of the most effective processes for a variety of applications thanks to its advantages. It is the object of many works to better understand its mechanism and the physics behind it. Indeed, physical models describing this process have been developed and integrated into several numerical simulation tools. However, regarding the several mutual interactions between the involved physics, one should be aware of the limitations of the considered physical models, assumptions, and possible errors. Therefore, an appropriate theoretical model that best describes the process should be chosen. Indeed, among the multiple FEM codes that propose valid formulations and coupling capabilities for the IH process, we are interested in FORGE® FEM code, which was used in this study to provide the data for developing the metamodels.
This chapter aims to show the physical modelisation and numerical simulation of the studied process. The first part of this chapter presents the basic physical concepts and assumptions related to the IH process. The second part summarizes the mutual coupling between physics, and the last part is devoted to the FE numerical simulation of the IH process on a spur gear. The definition of the used solvers and their coupling was first shown. Then, the identification of the material properties and the development of the model in FORGE® were achieved. Finally, some results were illustrated and interpreted.
Throughout the study, the objects are static and no relative movement between the gear and the inductor was considered.

3.2 Mathematical formulation of physical phenomena

As previously stated, the IH process is governed by several physical phenomena, namely electromagnetism, heat transfer, metallurgy, and mechanics. The modeling of the physical phenomena involved is a task of major importance. All the models historically developed are based on an understanding of the exact mechanisms that take place during heat treatment. In this section, we are interested in presenting the existing equations that describe the physical phenomena involved during IH.

Throughout the process, the models are based on several assumptions:

- the steel of the gear is considered homogeneous and isotropic;

- due to the phase transformation during the process, the material has a multiphase nature, and

homogenization must be applied. For this purpose, a linear law of mixtures is adopted. Therefore, each material property or parameter is the result of a linear mixture of the corresponding parameters at the present phases. These parameters are weighted by the volume fraction y_i of each phase i.

For notational convenience, boldface characters are used for vectors and non-bold characters for scalars and numbers.

3.2.1 Modeling of electromagnetic phenomena

3.2.1.1 Maxwell's equations

The electromagnetic phenomena that vary in time and space are governed by Maxwell's four local equations such that the behavior of magnetic fields ($\boldsymbol{B}, \boldsymbol{H}$), electric fields ($\boldsymbol{E}, \boldsymbol{D}$), charge density ($\rho$), and current density ($\boldsymbol{J}$) are described. Maxwell's equations have already been widely discussed in the literature (Stansel, 1949; Rodigin, 1950; Slukhotskii and Ryskin, 1974; Lehner, 2010), they are written in the following form:

- Maxwell-Gauss:

$$\nabla \cdot \boldsymbol{D} = \rho \tag{3.1}$$

- Maxwell-Faraday:

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{3.2}$$

- Maxwell-Ampere:

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t} \tag{3.3}$$

- Gauss's law for magnetic flux:

$$\nabla \cdot \boldsymbol{B} = 0 \tag{3.4}$$

where:

$\boldsymbol{B}$ is the imposed magnetic flux (Tesla);

$\boldsymbol{D}$ is the electric displacement field ($\mathrm{C.m^{-2}}$);

$\boldsymbol{E}$ is the electric field ($\mathrm{V.m^{-1}}$);

$\boldsymbol{H}$ is the magnetic field ($\mathrm{A.m^{-1}}$);

$\boldsymbol{J}$ is the electric current density ($\mathrm{A.m^{-2}}$);

ρ the volumetric density of electric charge ($\mathrm{C.m^{-3}}$).

In the context of induction heating, where frequencies are generally between 50 Hz and 1 MHz, the

quasi-steady-state approximation was considered. It consists in assuming that the propagation time of the electromagnetic waves is negligible compared to the signal period. Therefore, it enables us to neglect the displacement currents $\frac{\partial \boldsymbol{D}}{\partial t}$ in the Maxwell-Ampere equation (Cardinaux, 2008). By considering this hypothesis, the Maxwell-Gauss equation is decoupled from the others and can be written as follows:

$$\nabla \cdot \boldsymbol{D} = 0 \tag{3.5}$$

3.2.1.2 Constitutive relations

The system of Maxwell's equations is completed with the constitutive relations describing the macroscopic properties of the considered media to solve an electromagnetic problem. They relate $\boldsymbol{B}$ to $\boldsymbol{H}$, $\boldsymbol{D}$ to $\boldsymbol{E}$ and $\boldsymbol{J}$ to $\boldsymbol{E}$. Therefore, the evolution of the electromagnetic fields can be calculated in any media.

Magnetization law

For isotropic media, the magnetic excitation $\boldsymbol{H}$ is related to the magnetic induction $\boldsymbol{B}$ by the following constitutive relation:

$$\boldsymbol{B} = \mu \boldsymbol{H} \tag{3.6}$$

It is common to represent the magnetic permeability as:

$$\mu = \mu_r \mu_0 \tag{3.7}$$

where μ_0 is the vacuum permeability and takes a value of $4\pi \times 10^{-7}$ H.m^{-1}, μ_r is dimensionless and represents the relative magnetic permeability. For paramagnetic and diamagnetic media, μ_r is a constant and very close to 1. However, for ferromagnetic media, the relation between the fields $\boldsymbol{B}$ and $\boldsymbol{H}$ is no longer linear: the relative magnetic permeability μ_r is a function of the norm of $\boldsymbol{H}$ and the temperature T as follows:

$$\boldsymbol{B} = \mu_r(\|\boldsymbol{H}\|, T)\mu_0 \boldsymbol{H} \tag{3.8}$$

The temperature dependence of the magnetic permeability for a ferromagnetic steel is important, especially during the Curie transition: the material becomes non-magnetic with a constant relative permeability close to one. The electromagnetic profiles in the workpiece will be considerably modified.

Dielectric polarization law

The general form of the dielectric polarization law is :

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + P \tag{3.9}$$

where ε_0 is the dielectric permittivity of vacuum and P is the electric polarization. This relation is commonly used with a simplified writing

$$\boldsymbol{D} = \varepsilon(\|\boldsymbol{E}\|, T)\boldsymbol{E} \tag{3.10}$$

where ε and T are the dielectric permittivity and the temperature of the media, respectively.

Ohm's law

Ohm's law is written as follows:

$$\boldsymbol{J} = \sigma \boldsymbol{E} \tag{3.11}$$

where σ is the electrical conductivity.

The electromagnetic parameters of materials (magnetic permeability and electrical conductivity) are temperature-dependent. Therefore, it can be complex to obtain precise laws of evolution of these parameters, especially the magnetic permeability. Indeed, the evolution of this latter also depends on the magnetic field as shown in Eq.(3.8), and it is influenced by the nature of the material (ferromagnetic or non-magnetic) knowing that ferromagnetic materials become non-magnetic above the Curie temperature (Cardinaux, 2008).

Next, in order to get a well-posed problem for the bounded domain considered in the context of our study, boundary conditions are required. In the literature, two types of boundary conditions have been proposed (Biro and Preis, 1989): for the $\boldsymbol{H}$ field and the $\boldsymbol{B}$ field on the boundaries Γ_H and Γ_B, respectively. They are presented as follows:

$$\boldsymbol{B} \cdot \boldsymbol{n} = 0 \text{ on } \Gamma_B \tag{3.12}$$

$$\boldsymbol{H} \times \boldsymbol{n} = \boldsymbol{0} \text{ on } \Gamma_H \tag{3.13}$$

with $\boldsymbol{n}$ the outward unit normal vector. On Γ_H the tangential component of the magnetic field is zero, while on Γ_B the normal component of the magnetic flux is zero.

Indeed, for the problem resolution, certain assumptions have to be taken into consideration:

- the quasi-steady-state approximation is considered;
- the electromagnetic properties of the material are generally considered isotropic;
- we neglect the hysteresis cycle of ferromagnetic conductors and the energy dissipated by magnetization. This is because it is not the predominant heating mode, but the assumption may still be questionable in the case of certain materials;
- auto-induction phenomena are not considered.

Returning to Maxwell's equations, since the magnetic flux $\boldsymbol{B}$ satisfies Eq.(3.4), it can be expressed in terms of a magnetic potential vector $\boldsymbol{A}$ as follows:

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} \tag{3.14}$$

and then, from Eq.(3.14) and (3.2), we get:

$$\nabla \times (\boldsymbol{E} + \frac{\partial \boldsymbol{A}}{\partial t}) = 0 \tag{3.15}$$

By integrating Eq.(3.15), we obtain:

$$\boldsymbol{E} = -(\frac{\partial \boldsymbol{A}}{\partial t} + \nabla V) \tag{3.16}$$

where V is the electric scalar potential. Therefore, the combination of Maxwell's equations, the constitutive relations of materials, and the last defined relations allow expressing the electromagnetic problem in the quasi-steady formulation in the following form :

$$\begin{cases} \nabla \times (\frac{1}{\mu} \nabla \times \boldsymbol{A}) + \sigma \frac{\partial \boldsymbol{A}}{\partial t} = -\sigma \nabla V \\ -\nabla \cdot (\sigma \nabla V) = 0 \quad \text{on the inductor} \end{cases} \tag{3.17}$$

It is worth noting that the quantity $-\sigma \nabla V$, also denoted by $\boldsymbol{J}_s$, is the source current density in the inductor, i.e. the inductor is the only region with a prescribed source current or source voltage. Therefore, the electric scalar potential vanishes everywhere except for the inductor.

We have now a system of two equations separating the two unknowns $\boldsymbol{A}$ and V. The uniqueness of the solution can be ensured by specifying a gauge condition (Biro and Preis, 1990) (Eq.(3.18)) and boundary conditions on the domain boundary Γ. Only one type of boundary condition is chosen to be applied to the entire domain boundary and which is particularly adapted to our case (Eq.(3.19)):

$$\nabla \cdot \boldsymbol{A} = 0 \tag{3.18}$$

$$\boldsymbol{A} \times \boldsymbol{n} = \boldsymbol{0} \quad \text{on } \Gamma \tag{3.19}$$

This last condition, corresponding to a contact condition with a perfect conductor, allows considering two types of boundaries on our domain (Hömberg, 2004):

- the gear/inductor system is embedded in the air and the air boundary is considered far enough from the electromagnetic source to avoid cutting the field lines;
- the geometric and electromagnetic systems that have a symmetry plane, the magnetic field lines are then tangent in the boundary plane.

Therefore, the final system of equations is composed of the equations (3.17), (3.18), and (3.19). The common approach for solving the problem is by using harmonic approximation. This proceeds to simplify the mathematical model by assuming that the currents have a steady-state quality and thus the electromagnetic field quantities are harmonically oscillating functions with a single frequency. More details can be found in (Bristiel, 2001; Cardinaux, 2008). Finally, from the magnetic potential field $\boldsymbol{A}$ obtained by solving the electromagnetic problem, the heat source term due to the Joule losses is calculated as illustrated in Eq.(3.20). This quantity will be added as a volume source term to the second member of the heat transfer equation. Therefore, in the FE framework, it needs to be evaluated at every integration point of each mesh element of the workpiece and the inductor as well. This value is constant at each volume mesh element. Since the electromagnetic characteristic time step is often smaller than the thermal one, an averaged heat source over one or several periods of the electromagnetic field is calculated on the workpiece as follows:

$$\dot{Q}_{EM} = \frac{\boldsymbol{J}^2}{\sigma} \quad \Rightarrow \quad \dot{Q}_{EM} = \frac{1}{\tau}\int \sigma |\frac{\partial \boldsymbol{A}}{\partial t}|^2 dt \tag{3.20}$$

where τ is the period of the electromagnetic source term.

3.2.2 Modeling of thermal phenomena

The thermal problem is a well-known branch of science, governed by the first and second principles of thermodynamics which can be described in the form of the heat transfer equation:

$$\rho C_p \frac{dT}{dt} = -\nabla \cdot \boldsymbol{q} + \dot{Q} \tag{3.21}$$

where T is the temperature, ρ is the volume density of the material, C_p its specific heat, q the heat flux, and $\dot{Q}$ the heat source associated with the different couplings. In general, for heat treatments, the material properties are strongly affected by the temperature and the metallurgical phases. Therefore, the thermal material properties can be expressed as a linear mixture of the properties at each phase:

$$\rho = \sum_{i=1}^{nb\ phases} \rho_i y_i,\ Cp = \sum_{i=1}^{nb\ phases} Cp_i y_i,\ \sum_{i=1}^{nb\ phases} y_i = 1$$

where y_i is the rate of the phase presented in the material.

The heat source $\dot{Q}$ can have different origins associated with the different couplings. It can be expressed as a sum of 3 terms as follows:

$$\dot{Q} = \dot{Q}_{EM} + \dot{Q}_{meca} + \dot{Q}_{metal} \tag{3.22}$$

- The first term represents the heat generation by the Joule effect during induction, as illustrated in Eq.(3.20).

- The second one represents the heat generation by the mechanical dissipation

$$\dot{Q}_{meca} = \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} \tag{3.23}$$

- The last one models the latent heat absorbed or released during metallurgical transformations. It is usually assumed as the change in the transformation enthalpy

$$\dot{Q}_{metal} = \sum_{i=1}^{nb\ phases} \Delta H_i(T)\dot{y}_i \tag{3.24}$$

where $\boldsymbol{\sigma}$, $\dot{\boldsymbol{\varepsilon}}$, $\Delta H_i(T)$ are the stress tensor, the strain rate tensor, and the enthalpy of transformation of the phase i. During heating, the latent heat of transformation of the phase i is negative $-\Delta H_i$ (endothermic transformation). ΔH_i is approximated by a polynomial of T as expressed in Eq.(3.25).

$$\Delta H(T) = \int_{T_0}^{T} \rho C_p dT \tag{3.25}$$

The heat losses by the Joule effect are considered the principal and the most impactful heat source compared to the other heat sources for the IH process. Therefore, the last two terms could be neglected during heating.

The heat transfer can be done by 3 modes: conduction, convection, and radiation.

3.2.2.1 Heat transfer by conduction

This is a spontaneous exchange of energy by contact from a region of higher temperature to a region of lower temperature. This mode of transfer obeys Fourier's law which links the heat flux to the thermal gradient in the material, via the material property: thermal conductivity λ.

$$\boldsymbol{q} = -\lambda \nabla T \tag{3.26}$$

As with the other material properties, the thermal conductivity is expressed as a linear mixture of its value at each phase.

3.2.2.2 Heat transfer by convection

This mode of transfer occurs in the presence of a flowing fluid, which may be natural or forced. The outgoing heat flow follows the law :

$$\phi_c = -\boldsymbol{q}.\boldsymbol{n} = h(T_s - T_{ext}) \tag{3.27}$$

where ϕ_c, h, T_s, T_{ext} are the convective heat flux, the heat exchange coefficient, the surface temperature, and the external temperature, respectively.
The heat exchange coefficient h depends on the material, the fluid, and the relative speed of the fluid at their interface. This is why we distinguish natural convection from forced convection:
- Natural convection is a type of flow in which the fluid motion is not generated by any external source but by some parts of the fluid being heavier than others. It appears spontaneously when the buoyancy exceeds a limit depending on the material (Rayleigh number).
- Forced convection is due to a movement imposed by an external mechanical system. The convection heat transfer coefficient is then much higher.

3.2.2.3 Heat transfer by radiation

Heat transfer by radiation is a non-contact energy transfer. Indeed, it also exists in the vacuum such as solar radiation, for instance. The transfer of energy is done in an electromagnetic way at very high frequencies. It follows the law of Stefan-Boltzmann:

$$\phi_r = -\boldsymbol{q}.\boldsymbol{n} = \varepsilon\sigma_{SB}(T_s^4 - T_{ext}^4) \tag{3.28}$$

with σ_{SB} is the Stefan-Boltzmann constant (5.6703×10^{-8} W.m^{-2}.K^{-4}), ε the emissivity of the radiating material, T_s is the surface temperature, and T_{ext} is the temperature of the emitting body. For steels, they are generally considered as black bodies (all radiation frequencies contribute to the thermal equilibrium) with an emissivity equal to 0.88.

3.2.3 Modeling of metallurgical phenomena

3.2.3.1 Introduction to the phase transformation diagram

The purpose of heat treatments is to provide desired changes in the microstructure and therefore in the properties of steel parts. In order to understand the principles of heat treatment of steels, it is necessary to look at the microstructure of iron. At usual temperatures, iron can present different crystalline

structures (ferrite, cementite, and austenite). These solid structures can be identified according to the increasing temperature using an equilibrium diagram as shown in Figure 3.1.

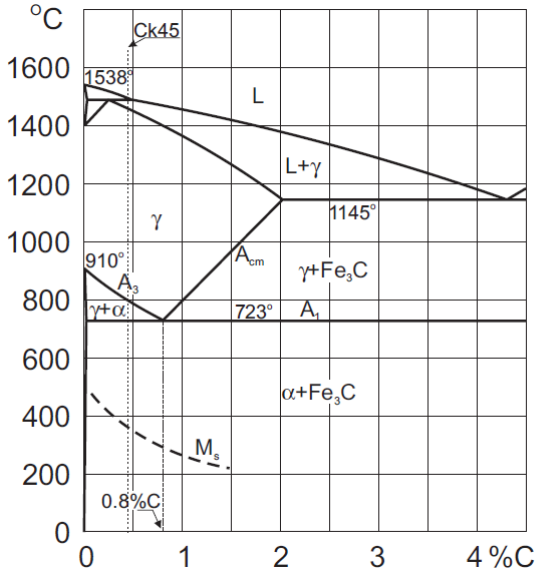


Figure 3.1: Iron-Carbon equilibrium diagram (Geijselaers, 2003)

However, the formed constituents are out of the thermodynamic equilibrium conditions during the IH process. This is because the transformations take place in non-equilibrium conditions in a wide range of temperatures which depend on the heating rates. Therefore, we can no longer rely on equilibrium diagrams but on transformation kinetics that describes, out of equilibrium, the formation rates and the respective proportions of the new phases. Among the models that describe the microstructural evolution of steel as a function of time during heating, we can distinguish the isothermal method based on the TTA (Temperature-Time-Austenitization) diagrams which present the diagrams established under isothermal conditions (the steel is heated rapidly to the temperature in question and held there for a certain predetermined time) and CHT (Continuous Heating Transformation) diagrams established with given heating rates (the steel is heated continuously at different heating rates).

3.2.3.2 Transformation kinetics of austenetic phase during heating

As previously mentioned, the austenitization consists of transforming the different metallurgical phases present at low temperature into austenite by the addition of thermal energy. The austenite transformation depends on the initial steel microstructure. In our case, we consider a ferrite-pearlite initial microstructure.

The transformations generally start at temperature $Ac1$ and complete austenitization is reached at temperature $Ac3$ ($Ac1$ and $Ac3$ are thermodynamic values depending on the material). Then, the homogenization of the austenite by carbon diffusion is obtained above $Ac3$, and the increase of the grain size above a temperature of $Ac3 + \Delta T$ is obtained as well. The last two points are complex to handle and are not considered in the numerical model.

Nevertheless, for rapid heating, phase transformations may occur at higher temperatures. It is therefore necessary to be able to propose models for the phase transformation kinetics that meet these requirements. The IH process is characterized by internal power densities which can generate a high heating rate (more than 1200°C/s). The classical approach, based on TTA diagrams, can be used which is already implemented in FORGE® software. This approach is based on the Johnson-Mehl-Avrami formalism which demonstrates that the austenite formation is decomposed into a germination phase followed by a growth phase before homogenization, and it allows to represent both slow and fast transformations (William and Mehl, 1939). For this model, the growth of the austenitic phase is represented by the Avrami law:

$$y_\gamma = y_\gamma^{max}[1 - exp(-b\ t^n)] \tag{3.29}$$

where y_γ^{max} is the maximum austenite rate that can be reached at a given temperature, t is the time elapsed since the beginning of the growth, and b and n are temperature-dependent material parameters of Avrami's law determined from the TTA diagram. An example of an isothermal TTA diagram for 35CrMo4 steel is illustrated in Figure 3.2. This diagram gives isovalues of austenitic transformation rates (for example the beginning and the end of transformation) on a time-temperature plane.

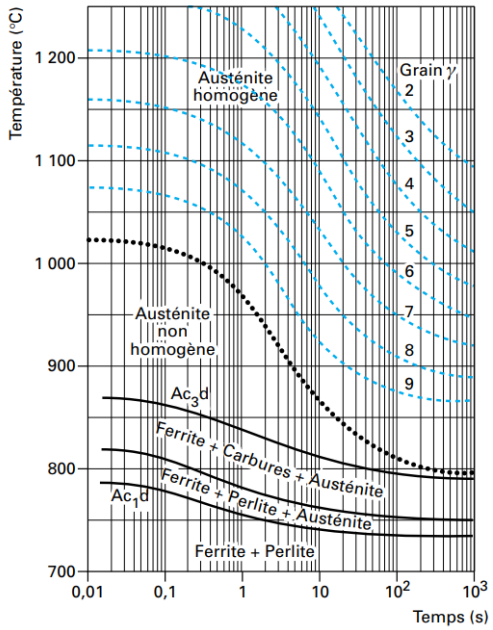


Figure 3.2: Isothermal TTA diagram of a low-alloy steel of the type 35CrMo4 (0.34 %C - 1.07 %Cr - 0.17 %Mo) (Orlich et al., 1973)

Otherwise, the continuous heating diagrams constitute an efficient model to define the transformation properties at different heating rates. Starting from CHT diagrams, the total amount of transformed phase can be expressed as:

$$y_\gamma = \begin{cases} 0 \quad \text{if} \quad T < Ac1(\dot{T}) \\ \dfrac{T - Ac1(\dot{T})}{Ac3(\dot{T}) - Ac1(\dot{T})} \quad \text{if} \quad Ac1(\dot{T}) \leq T \leq Ac3(\dot{T}) \\ 1 \quad \text{if} \quad T > Ac3(\dot{T}) \end{cases} \tag{3.30}$$

where y_γ, T, $Ac1(\dot{T})$ and $Ac3(\dot{T})$ are the volume fraction of austenite, the temperature, and the transformation temperatures, function of the heating rate, respectively.

Whatever the used austenitization model, the software will calculate the formation of austenite at the expense of the phases furthest from the equilibrium state. That is, if a material consisting of martensite, bainite, pearlite, and ferrite is heated to austenitization, the first phase to disappear will be martensite, then bainite, followed by pearlite, and finally, ferrite will turn into austenite.

Note: Due to the complexity of the treatment, the effect of the austenite grain size is not considered.

3.2.4 Modeling of mechanical phenomena

In the case of heat treatments, the total deformation is small and often in the elastic deformation range. Yet, paradoxically, the induced irreversible deformations, even small, are the main cause of residual

stresses and distortions. Therefore, the numerical simulation of the process requires an accurate description of the constitutive law in order to predict the material deformation. This description should be representative of the rheology of the material over a wide range of temperatures to consider the different involved physical phenomena.

Indeed, most metals exhibit a significant viscous behavior once the temperature exceeds 4/10 of the melting temperature. This temperature is often reached and exceeded for most heat treatments. Therefore, a thermo-elasto-plastic law cannot take into account time-dependent deformation (strain rate). The model implemented in FORGE® for heating admits that the total strain rate is the sum of thermal, mechanical, and metallurgical strain rates as follows:

$$\dot{\boldsymbol{\varepsilon}}^{tot} = \dot{\boldsymbol{\varepsilon}}^{el} + \dot{\boldsymbol{\varepsilon}}^{in} + \dot{\boldsymbol{\varepsilon}}^{th} + \dot{\boldsymbol{\varepsilon}}^{tr} \tag{3.31}$$

3.2.4.1 Elastic strain rate $\dot{\varepsilon}^{el}$

The elastic deformation $\boldsymbol{\varepsilon}^{el}$ is related to the stress tensor by Hooke's law via Young's modulus E and Poisson's ratio ν

$$\boldsymbol{\varepsilon}^{el} = \mathbb{C}^{el}(T) : \boldsymbol{\sigma} \tag{3.32}$$

with $\mathbb{C}^{el}$ is the fourth-order tensor that represents the inverse of the elasticity tensor. The elastic strain rate tensor can be then written as

$$\dot{\boldsymbol{\varepsilon}}^{el} = \mathbb{C}^{el} : \dot{\boldsymbol{\sigma}} + \dot{T}\frac{\partial \mathbb{C}^{el}}{\partial T} : \boldsymbol{\sigma} \tag{3.33}$$

The properties E and ν show a little difference when passing from one metallurgical phase to another. Therefore, their temperature-dependent evolutions can be the same for all phases.

3.2.4.2 Inelastic strain rate $\dot{\varepsilon}^{in}$

This deformation is decomposed into a plastic part $\boldsymbol{\varepsilon}^{p}$ and a viscous one $\boldsymbol{\varepsilon}^{vp}$. The plastic deformation is calculated from the classical theory of plasticity (yield surface f, von Mises plasticity criterion as well as the considered hardening laws). The viscoplastic deformation presents the time-dependent deformation (governed by elasto-plastic and elasto-viscoplastic flow laws).

To characterize the inelastic behavior of the material according to the evolution of transformations, FORGE® proposes to characterize each metallurgical phase as a function of temperature. Indeed, an additive law is used to describe the behavior of each phase, as follows:

$$\sigma_{eq} = \sigma_0 + H\varepsilon^n + K\dot{\varepsilon}^m \tag{3.34}$$

such that σ_0 is the yield strength, K (consistency) and m (velocity sensitivity coefficient) are the viscosity-related parameters, H is the strain hardening modulus and n is the hardening coefficient. All these rheological parameters σ_0, H, n, K and m are phase and temperature dependent. ε is the average equivalent strain of the considered phase and $\dot{\varepsilon}$ is the strain rate.

3.2.4.3 Thermal strain rate $\dot{\boldsymbol{\varepsilon}}^{th}$

This deformation describes the expansion or contraction of the material. It is linearly related to temperature by the coefficient of thermal expansion α.

$$\dot{\boldsymbol{\varepsilon}}^{th} = \alpha \dot{T} I \tag{3.35}$$

where I is the identity tensor. The coefficient α is also temperature-dependent and defined for each phase.

3.2.4.4 Phase change strain rate $\dot{\boldsymbol{\varepsilon}}^{tr}$

The phase transformation is accompanied by a change in volume due to the change in crystal structure from one phase to another. This phase change is modeled by an expansion parameter applied to the instantaneously transformed phase ratio. The volume change that occurs during the metallurgical transformation induces a spherical deformation written as follows (Denis et al., 1987):

$$\dot{\boldsymbol{\varepsilon}}^{tr} = \dot{\varepsilon}_0^{tr} I \tag{3.36}$$

where $\dot{\varepsilon}_0^{tr} = \sum_{i}^{nb\ phases} \dot{\varepsilon}_{0,i}^{tr}\ y_i$ such that $\dot{\varepsilon}_{0,i}^{tr}$ should be identified for each phase i.

3.3 Coupling between physics

According to what was explained so far, there are mutual interactions between all the physical fields related to the IH: electromagnetism, heat transfer, metallurgy, and mechanics. These interactions need to be well established to provide numerical models describing very well the physical reality.
Indeed, the influence of some couplings is more important than others. Hence, it makes sense to neglect the influence of some couplings and to consider the most significant ones for the numerical simulation. The set of couplings that will be taken into account in our simulation is shown in Figure 3.3.

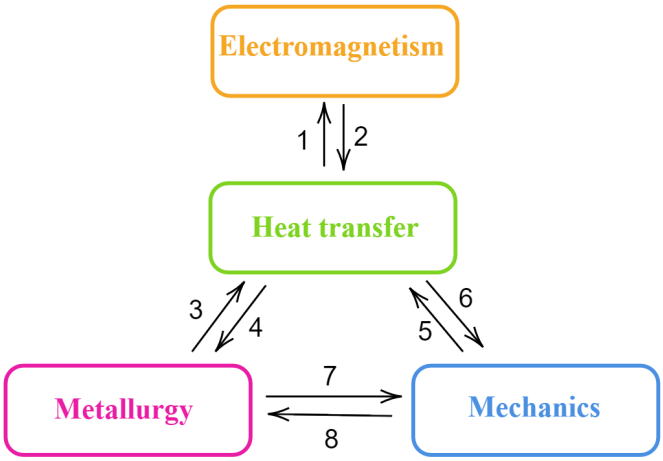


Figure 3.3: Multiphysics-based coupling

The different couplings are summarized as follows:

1. The temperature dependence of the electromagnetic parameters. The level of dependency is managed by a coupling parameter.
2. Power dissipated by the Joule effect.
3. The latent heat of transformation due to the metallurgical mechanisms.
4. The thermal agitation which is, in our case, the source of energy generating metallurgical transformations. This coupling is taken into account by the temperature-dependent Avrami parameters deduced from the isothermal TTA diagram and the transformation kinetics from CHT diagrams determined at a certain heating rate.
5. The mechanical power dissipated as heat characterized by the state of stress and strain rates.
6. Temperature dependence of mechanical properties. It is taken into account through both Hooke's law and the rheological law as well as the parameters related to the couplings.
7. Volume variations due to phase transformations.
8. The effect of mechanical stress on metallurgical transformations.

3.4 Numerical simulation of induction heating process

Numerical simulation represents a powerful tool capable of taking into account some or all of the involved physical phenomena. However, it requires a precise description of the latter, starting from the calculation of the power densities induced in the workpiece by solving the electromagnetic problem, passing to the description of heat transfer, metallurgical transformations kinetics, and mechanical deformations that occur, typically, in a very short time. FORGE® FEM code, is developed by Cemef and commercialized by Transvalor and offers the possibility to solve a complete multiphysics simulation. Therefore, it provides a better understanding of the overall behavior of the process and allows quantifying the effect of the included parameters.

3.4.1 Principle of the IH simulation

Even though a strong coupling between the solvers describing the involved physics, in the IH process, represents an interesting approach from the theoretical point of view, it becomes uninteresting when different ranges of characteristic times are considered for the different solvers, generating a high computational time. This is because a common time step for all the coupled phenomena should be defined, this value must be adapted to the smaller time constant, which leads to prohibitive computation times. Indeed, for the IH process, the range of frequencies is comprised between 50 Hz and 1 MHz. Consequently, the period of the electromagnetic phenomenon has always values lower than 0.02 sec. However, the time process defined for the transient thermal phenomenon is typically never lower than 0.1 sec. Therefore, to ensure a good compromise between the computation time and result accuracy, it is preferable to decouple the solvers. In FORGE®, the IH process is simulated by a weak coupling between the electromagnetic solver and the thermo-metallo-mechanical one. Each solver has to be run separately and the calculations will take place cyclically by ensuring the communication between them as shown in Figure 3.4.

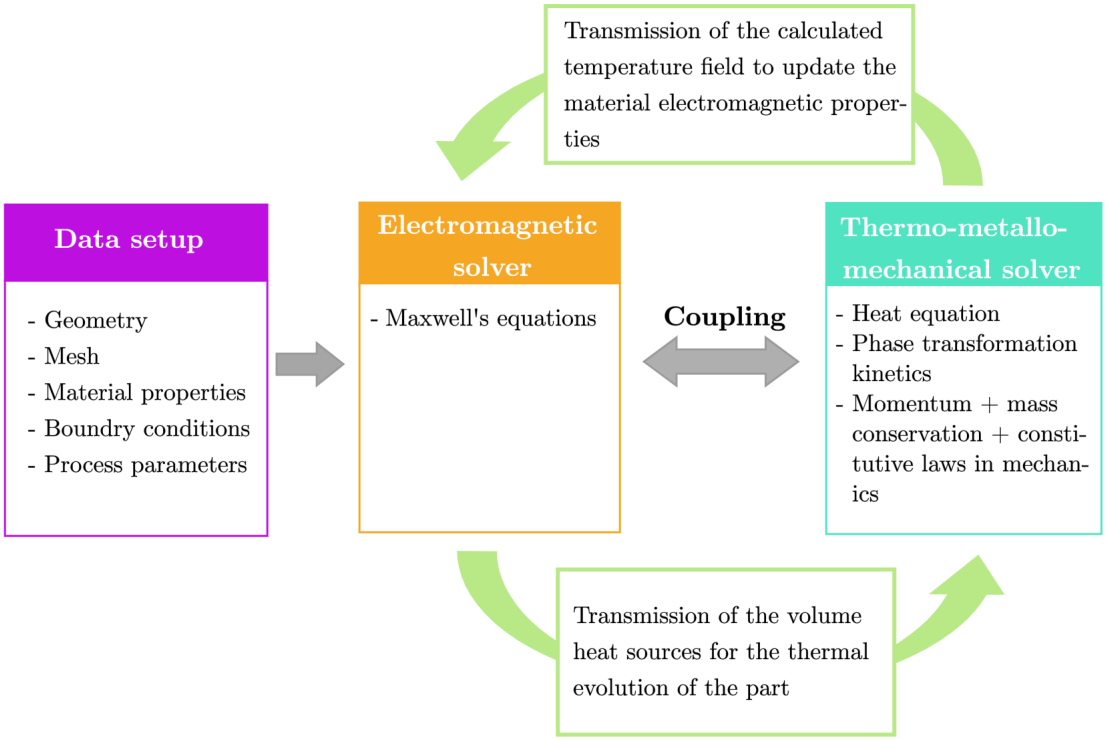


Figure 3.4: Global principle of induction heating process

Firstly, the data setup (geometries, mesh, boundary conditions, material properties, and process parameters) should be prepared for each solver, then the simulation chain is launched. The electromagnetic solver is used to obtain the amount of heat resulting from the electromagnetic field generated within the workpiece. The solver solves iteratively Maxwell's equations at each electromagnetic time step by splitting the electromagnetic period (the period of the alternating current supplied to the inductor) into a series of time steps (N increments) as shown in Figure 3.5.

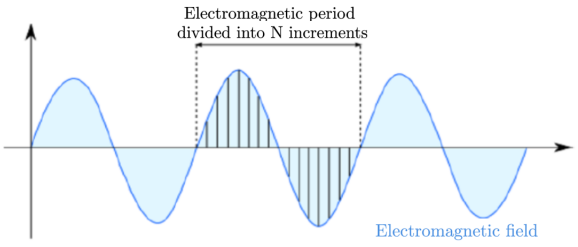


Figure 3.5: Discretization of an electromagnetic period (TRANSVALOR, 2020)

The accuracy of the results as well as the computation time depend on the number of increments of

the electromagnetic period, which can be easily managed in FORGE®.
At the convergence of the electromagnetic calculation, the amount of heat is provided to the thermo-metallo-mechanical solver as an internal heat source to calculate the temperature field by solving the heat transfer equation. Indeed, a coupling is provided by solving the mechanical system of equations as well as the transformation kinetics of austenite by using a CHT diagram, such that new extra terms of internal heat source are added to Joule losses as shown in Eq.(3.22).

During IH simulation, the results of the two solvers are exchanged according to some coupling predefined criteria such that the temperature field is provided again to the electromagnetic solver to update the electromagnetic properties whose evolutions modify the calculated magnetic field and the heat sources. The transition from the thermo-metallo-mechanical solver to the electromagnetic one is managed in FORGE® by different possible coupling criteria. To ensure coupling, we can impose a maximum temperature variation, a number of time increments, or a maximum time variation.
Here, the chosen criterion is a temperature $T_{coupling}$ such that when a temperature evolution ΔT exceeds this value, the circular chain mechanism restarts the electromagnetic calculation. The algorithm of the coupled computation is presented in Figure 3.6.

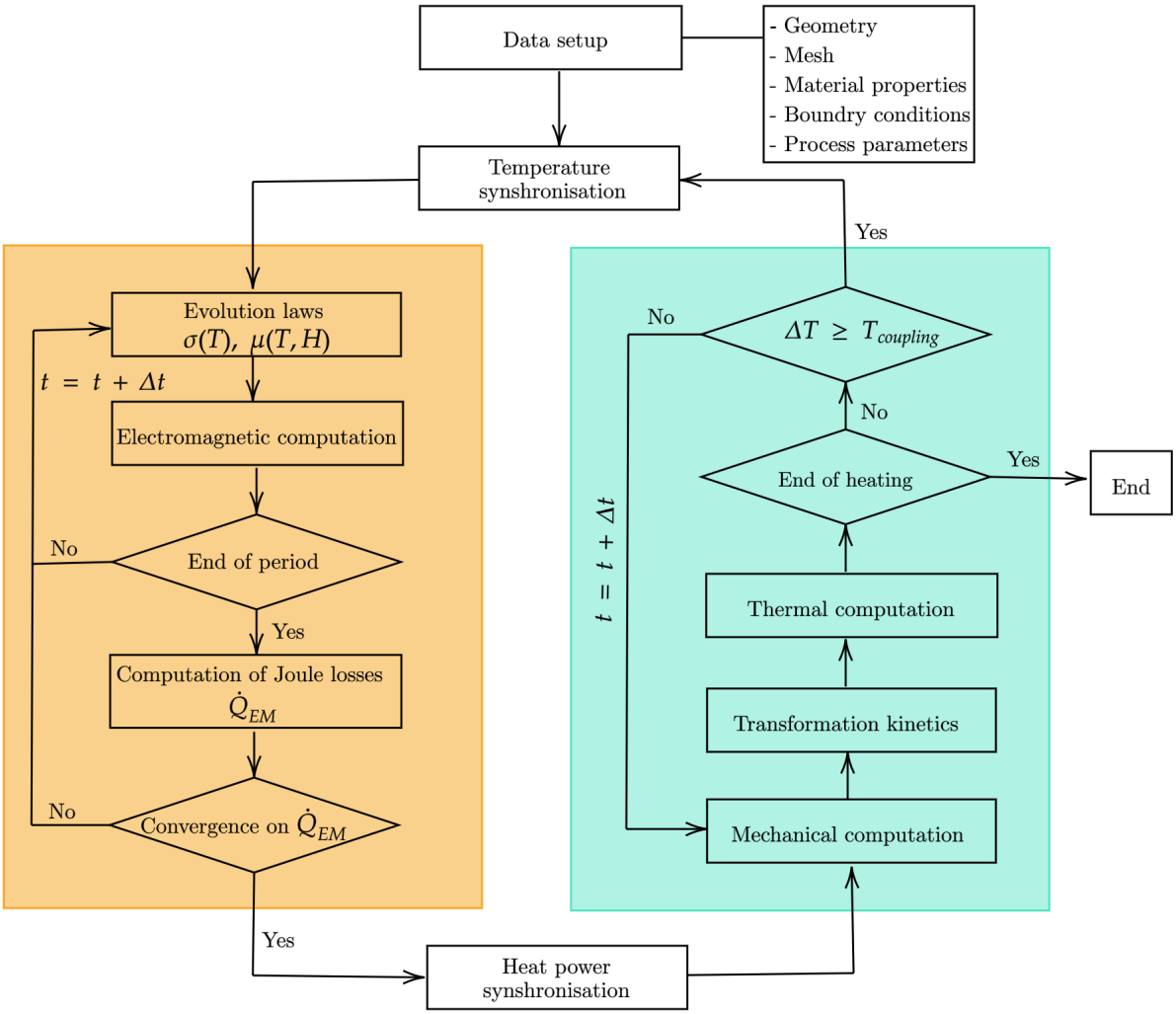


Figure 3.6: Chaining algorithm between the two solvers

3.4.2 Identification of material properties

Each of the involved physical fields is governed by a set of material properties. Those properties are generally dependent on the temperature, the metallurgical phases, and the magnetic field H for permeability. Even if the physical formulations used in the model are appropriate to obtain a reasonable result, the material properties need to be properly characterized. In our model, one can either use tabulated data or analytical expressions approaching physical phenomena.

3.4.2.1 Chemical Composition

For surface induction heat treatment applications, the used steel must have good hardenability (the ability of an alloy to benefit from the effects of hardening). Low-alloyed steels with a medium carbon

content of 0.4-0.55% are commonly used and particularly suitable for such processes on gears requiring good hardenability and mechanical properties. In this study, C45 steel was considered. Its chemical composition is presented in Table 3.1.

Table 3.1: Chemical composition of the employed C45 steel (FPDBase, 2019)

Element	C	Mn	Si	Ni	Cr	Mo	P	S	Fe
Weight %	0.46	0.65	0.2	0.2	0.2	0.05	0.023	0.023	bal.

3.4.2.2 Electromagnetic properties

The electromagnetic properties are little affected by phase transformation, unlike thermal and mechanical properties. Therefore, no need to identify those properties for each phase. The identification of the electric resistivity and the relative magnetic permeability is clearly explained in Appendix A.

3.4.2.3 Thermal properties

Thermal properties are essential for describing the temperature distribution in the workpiece during the induction heat treatments. Since the material can be seen as a mixture of different phases, each thermal property is temperature and phase-dependent. As mentioned earlier, the macroscopic behavior of the material is described by a linear rule of mixing.

The three thermal properties required for the simulation are the density, the thermal conductivity, and the specific heat, whose evolution with temperature is taken from the literature (Iron et al., 1953; Geijselaers, 2003; Cstroeski, 2003; Magnabosco et al., 2006; Lee et al., 2010b; Ding and Shin, 2012; Barba et al., 2018; Gorni, 2019) and from databases such as thermo-Calc (Lee et al., 2010a) and JMatPro (JmatPro, 1999). In this work, the material properties of ferrite and perlite phases were considered equal, while a big difference can be noticed for austenite. The identification of the density, the thermal conductivity, and the specific heat is clearly explained in Appendix B.

3.4.2.4 Metallurgical properties

It is worth reminding that the heating is located on the surface, and consequently the heating rate is different between the surface and the bulk material beneath. As previously stated, FORGE® can rely on TTA diagrams to describe the austenitic transformation kinetics. In this study, the CHT diagram, provided in the literature by Orlich (Orlich et al., 1973), was considered. The latter is well adapted to our case of study because of its dependency on the heating rates and is introduced in a tabular form. Figure 3.7 shows the CHT diagram for the continuous heating of C45.

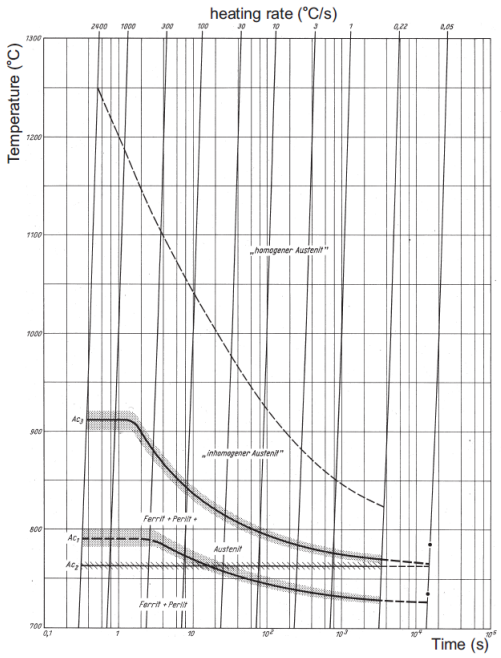


Figure 3.7: CHT diagram of C45 (Orlich et al., 1973)

3.4.2.5 Mechanical properties

As for thermal properties, the mechanical properties are temperature and phase-dependent. From a numerical point of view, the mechanical properties to be known are the Young's modulus and the Poisson ratio for the elastic part, the thermal expansion coefficients for the thermal part, the volume changes during phase transformations for the metallurgical part, and finally the properties related to the viscoplastic constitutive law. All the mechanical properties are illustrated in Appendix C.

3.4.3 Model development in FORGE® software

In this section, we describe the model implemented in both solvers. All the developments and simulations were carried out with the version FORGE® Nxt3.0. The reference workpiece is a 22-teeth spur-gear of C45 steel and the employed inductor is a copper single-turn rectangular-section coil. Its main dimensions are illustrated in Table 3.2.

Table 3.2: Gear and inductor dimensions

Gear	number of teeth: 22 module: 2.5 mm tip diameter: 60 mm reference diameter: 55 mm root diameter: 48.75 mm width: 10 mm pressure angle: 20°
Inductor	rectangular section 8 × 5 mm internal diameter: 65 mm

3.4.3.1 Electromagnetic solver

Geometry and mesh

The electromagnetic model is composed of the gear, the single-turn inductor, and the surrounding air. For computational time savings, a quart of the gear tooth was considered and two symmetry planes were imposed for the electromagnetic model as illustrated in Figure 3.8. The CAD of the different components was generated on Freecad software and imported into FORGE® in the form of .stl extension files.

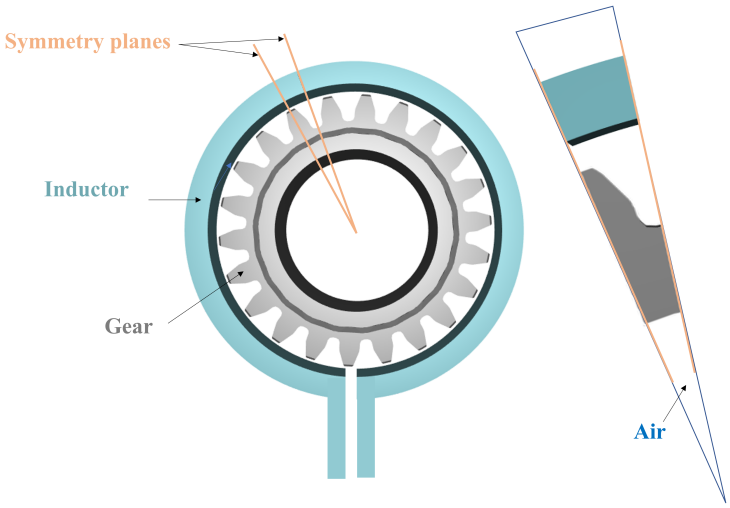


Figure 3.8: Geometry of the electromagnetic model

Then, attention should be paid to the quality of the mesh such that the choice of the average mesh size for the electromagnetic calculation must be motivated by the user's priorities. Indeed, we have to ensure a compromise between the results quality and the computation time. It is essential to:

- have a sufficiently fine mesh in the area of interest; the common interfaces between the gear, the inductor, and the air;

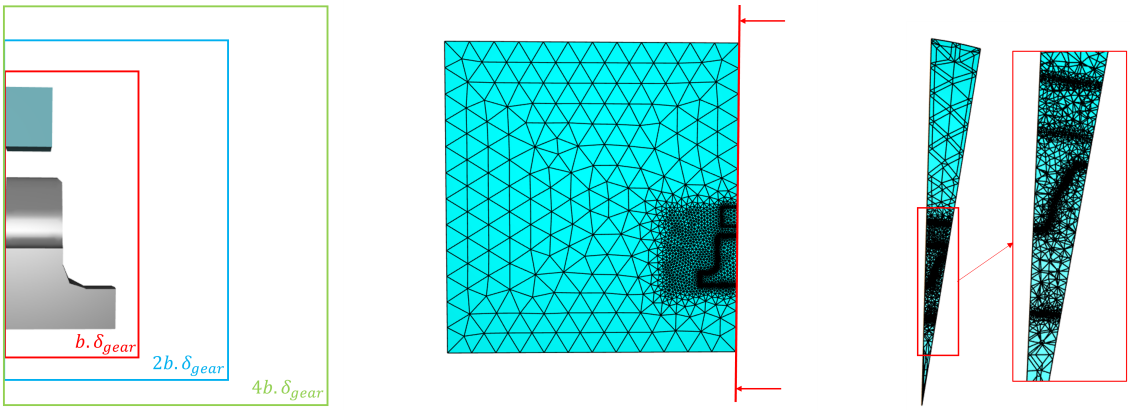
- increase the mesh size in the out-of-focus areas.

For that purpose, FORGE® proposes a methodology capable of defining the surface mesh size of the workpiece and the inductor as a function of the skin depth given in Eq.(1) as follows:

$$\delta_{gear} = \frac{a}{\sqrt{f} \times f_{gear}} \tag{3.37}$$

$$\delta_{inductor} = \delta_{gear} \times f_{inductor} \tag{3.38}$$

where f is the current frequency in Hz, δ_{gear} and $\delta_{inductor}$ are the surface mesh size of the gear and the inductor, respectively. f_{gear} and $f_{inductor}$ are user-defined factors and a is a skin depth factor. This latter is an average value defined for each material to represent the dependency of the skin depth on material properties. For the C45 steel, a takes a constant value of 80. Then, the mesh size increases progressively through the volume using a box factor as shown in Figure 3.9a. The global mesh consisting of the gear, inductor, and ambient air is shown in Figure 3.9b and a cross-sectional view is illustrated in Figure 3.9c.



(a) (b) (c)

Figure 3.9: Electromagnetic solver mesh: (a) box method, (b) global mesh, and (c) cross-sectional mesh view

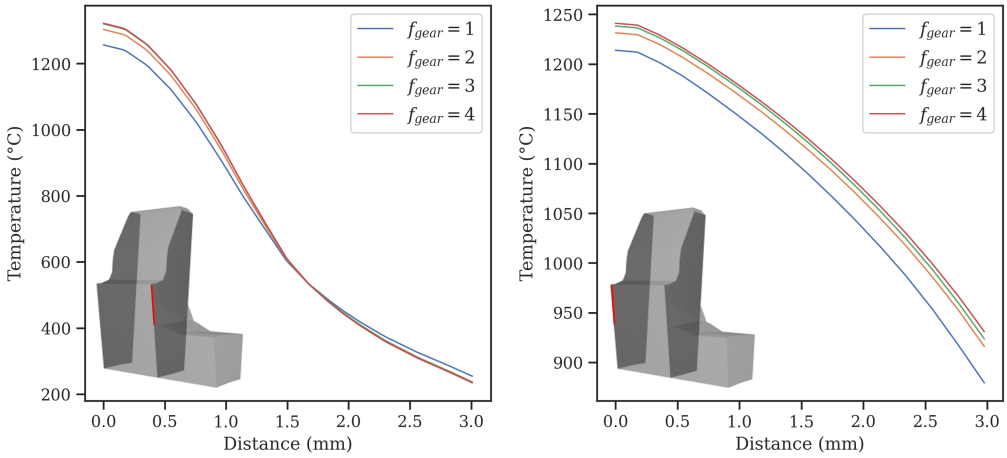
The mesh box technique proposed by FORGE® consists in defining geometric zones (parallelepipedic in particular) into which the mesh size is defined. A box factor $b = 5$ was defined. For our three-dimensional geometries, only tetrahedral mesh elements can be generated by FORGE®. To optimize the problem, four different electromagnetic meshes were tested while keeping the same mesh for the

thermo-metallo-mechanical solver. Simulations with a frequency value of 170 kHz and a power of 100 kW were executed.

The evolution of the in-depth temperature field at different positions in the gear tooth was plotted. The curves are illustrated in Figure 3.10. Besides, the different configurations and their computational results are presented in Table 3.3. It can be noticed that increasing the f_{gear} by a unity, the number of mesh elements increases significantly such that almost 18-times more elements are observed by going from f_{gear} equal 1 to 4. Consequently, the computation time increases when decreasing the mesh size such that a computation time of 8 hours is obtained for $f_{gear} = 4$. However, for the temperature evolution, no significant variation is observed for the three factors 2, 3, and 4 (see Figure 3.10) and the curves for the four positions are close. However, a significant gap is observed between $f_{gear} = 1$ and the others. Therefore, increasing the mesh size within a certain level does not lead to high variations in temperature.

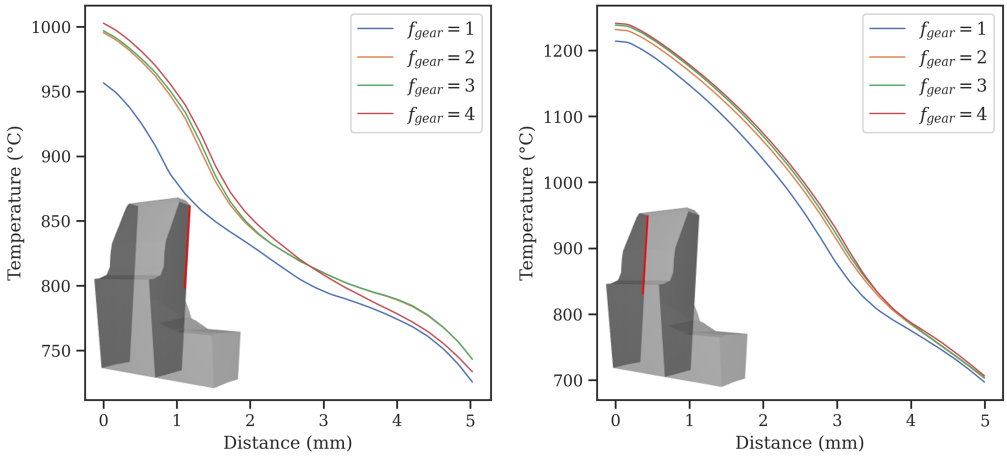
Table 3.3: Effect of the electromagnetic mesh size on computation time

f_{gear}	1	2	3	4
$f_{inductor}$	1			
Number of elements	90167	349525	864347	1609850
Number of cores	12			
Computation time	1h 13min	1h 36min	3h 21min	8h



(a) Edge of the gear tooth root

(b) Mid-width of the gear tooth root



(c) Edge of the gear tooth tip

(d) Mid-width of the gear tooth tip

Figure 3.10: Evolution of the in-depth temperature - Mesh convergence in the electromagnetic solver

Based on the obtained results, the third mesh with f_{gear} set to 2 was considered to ensure good accuracy with a reasonable computation time. However, it is worth pointing out that the obtained results in terms of computation time and number of elements can be highly modified by setting f_{gear} to 2 while considering a higher frequency. This is due to their high correlation.
Finally, a remeshing option was checked in order to adapt the mesh with the volume change of the workpiece obtained at the thermo-metallo-mechanical solver.

Material properties

For the electromagnetic solver, the material properties to be implemented are the relative magnetic permeability μ_r and the electric resistivity ρ. Those properties were identified for the C45 gear as

shown in section 3.4.2.2. The defined properties for the inductor and air are illustrated in Table 3.4.

Table 3.4: Electromagnetic properties of the inductor and surrounded air

Components	Material	μ_r	ρ (Ω.m)
Inductor	copper	1 (non-magnetic)	2×10^{-8}
Air	standard air	1 (non-magnetic)	insulator

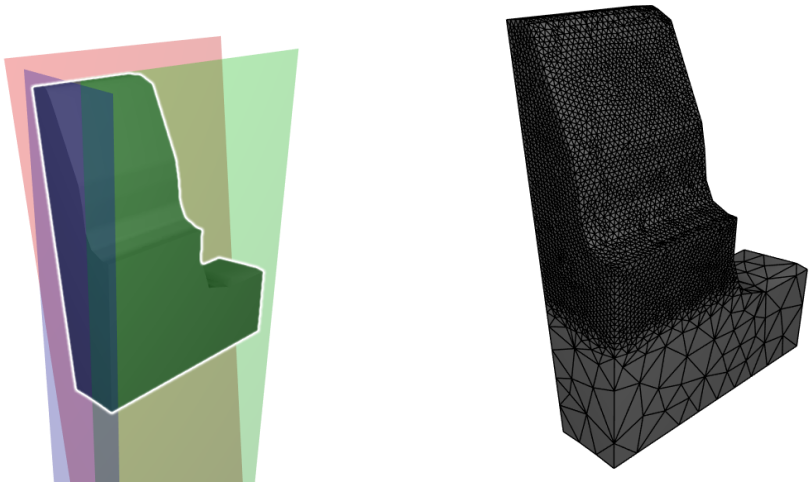
Process parameters

The process parameters that must be known to simulate the IH process are the frequency and the power delivered to the inductor. These two parameters were mainly considered in this work. Other numerical parameters should be introduced, such as the time discretization of the electromagnetic period. FORGE® proposes three possibilities for the choice of the number of increments per period; fast (40 increments), standard (60), and precise (80 increments). Among the three, the standard one was chosen.

3.4.3.2 Thermo-metallo-mechanical solver

Geometry and mesh

The thermo-metallo-mechanical model is only composed of the gear, where the temperature field has to be calculated. While the electromagnetic model has two symmetry planes, the model in the second solver has three symmetry planes, as shown in Figure 3.11a. For this model, a fine mesh is required in the HAZ for a better interpolation of the power density imported from the electromagnetic solver and for a good prediction of the phase distribution during a metallurgical transformation. Uniform mesh size was adopted in the HAZ while a coarser mesh of 1 mm was considered for the rest of the workpiece as shown in Figure 3.11b.



(a) (b)

Figure 3.11: Thermo-metallo-mechanical model: (a) geometry with symmetry planes (b) gear mesh

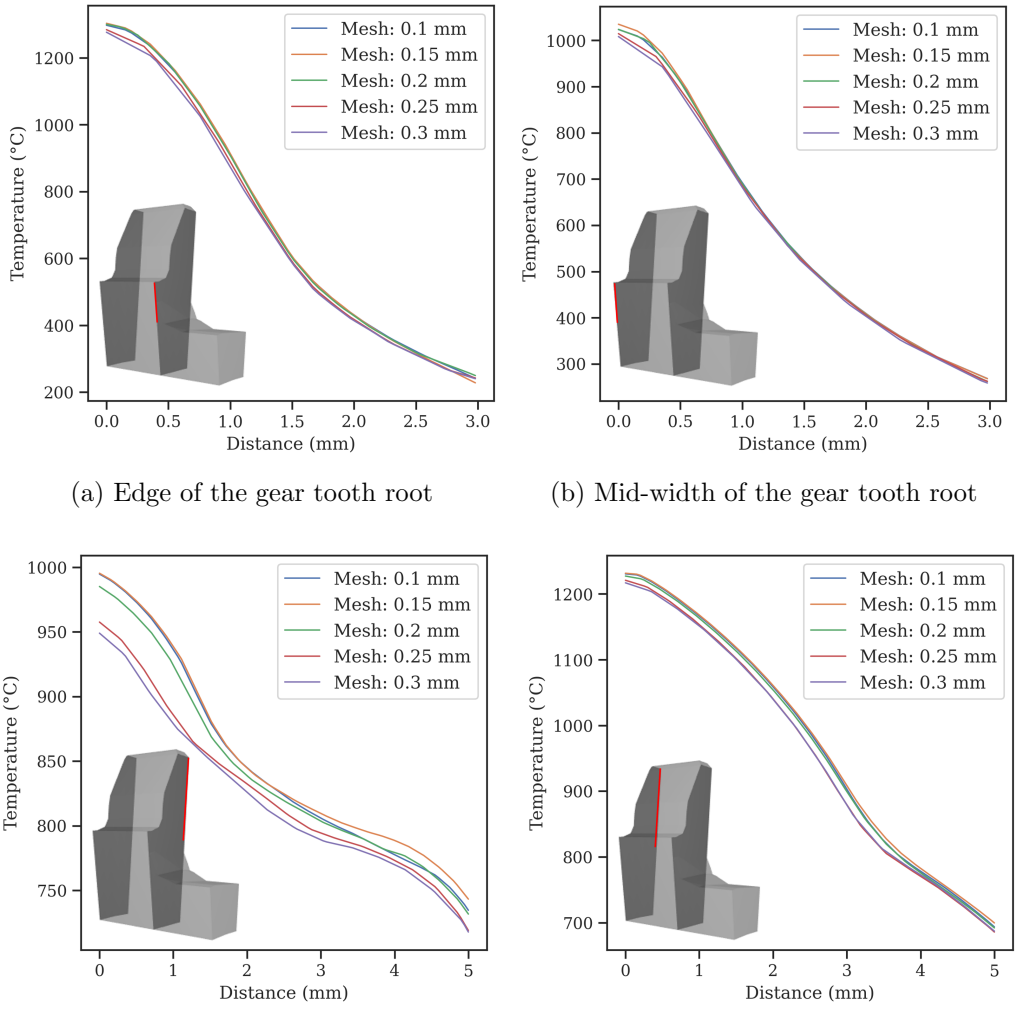
A convergence study was carried out for this solver as well to optimize the computational time by adopting a good mesh. It is worth mentioning that a volume mesh with tetrahedral elements was created for this solver. For the electromagnetic solver, the mesh parameters were kept the same ($f_{gear} = 2$), and the values of frequency and power were the same as those used in section 3.4.3.1. Here, five different mesh sizes were tested on the gear tooth.

The different configurations and their computational results are shown in Table 3.5. It can be noticed that increasing the mesh size, from 0.1 to 0.3 mm by 0.05 mm, leads to a considerable decrease in the number of mesh elements and the computation time as well. However, this decrease in computation time is not as important as the decrease in the number of mesh elements.

Figure 3.12 shows the evolution of the in-depth temperature field at different positions in the gear tooth. We observe that the curves with the different meshes are close for Figues 3.12a, 3.12b and 3.12d with a small difference which is less than 30°C at the surface. However, a considerable gap is observed in Figure 3.12c at the edge of the gear tooth tip, especially between the first three mesh sizes and the last two ones.

Table 3.5: Effect of the mesh size on computation time

mesh size (mm)	0.1	0.15	0.2	0.25	0.3
Number of elements	603435	179120	76226	40440	24390
Number of cores	12				
Computation time	2h 18min	1h 36min	1h 26min	1h	55min



(a) Edge of the gear tooth root

(b) Mid-width of the gear tooth root

(c) Edge of the gear tooth tip

(d) Mid-width of the gear tooth tip

Figure 3.12: Evolution of the in-depth temperature - Mesh convergence in the thermo-metallo-mechanical solver

Based on the obtained results, the mesh size is set to 0.15 mm to ensure good accuracy with a reasonable computation time.

Material properties

For the thermo-metallo-mechanical solver, all the material properties defined in sections 3.4.2.3, 3.4.2.4, and 3.4.2.5 were implemented in a material file of extension .tmf and imported in FORGE®.

Process parameters and boundary conditions

During induction heating, two modes of heat transfer are involved, namely convection and radiation. The amount of heat lost by the latter depends on the convection exchange coefficient h ($W.m^{-2}.°C^{-1}$)

and the emissivity of the material ε. In this work, a natural convection coefficient h of value 10 $W.m^{-2}.°C^{-1}$ was considered to describe the heat exchange between the gear and the air during heating and an emissivity of value 0.8. Then, the time process should be identified, knowing that the time step is automatically managed by FORGE® for this solver.

For the initial condition, the initial temperature was set to 20°C. For boundary conditions, the movement of the workpiece was fixed in some directions. For that purpose, a new object called a manipulator was added. This latter is a plane attached to the gear from one side as shown in Figure 3.13 such that the velocity along the axis Z was set to zero while keeping the workpiece free to expand along the X and Y axis. In fact, it is worth pointing out that only velocities or loads could be imposed for the manipulator.

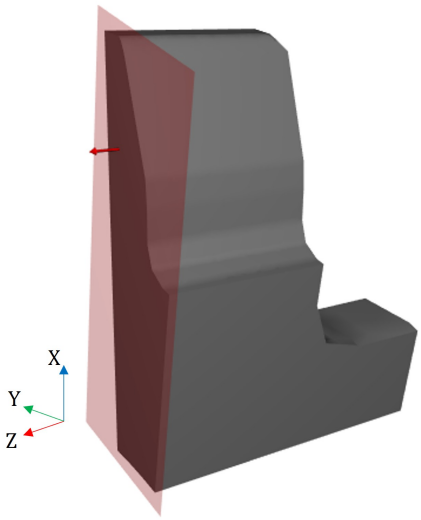


Figure 3.13: Manipulator to define the boundary conditions

Finally, among the very important parameters for the simulation, we have the coupling parameter between the two solvers: once this value is reached, the calculated temperature field will be delivered to the electromagnetic solver to update the electromagnetic material properties and so on. As mentioned before, a temperature value was chosen for coupling. On one side, a low value of coupling temperature (100°C or less) provides better accuracy of the predicted results but with higher computational time. In fact, the lower value of coupling temperature is, the more data transfer is, leading to a higher number of electromagnetic computations. On the other side, a very high coupling temperature enables a very rapid computation, assuming intrinsically the non-dependency of the electromagnetic material properties with temperature. For our simulations, a value of 150°C was chosen to ensure a reasonable

calculation time with good accuracy of the predicted results while taking into account the thermo-dependence of the electromagnetic properties.

3.4.4 Numerical results

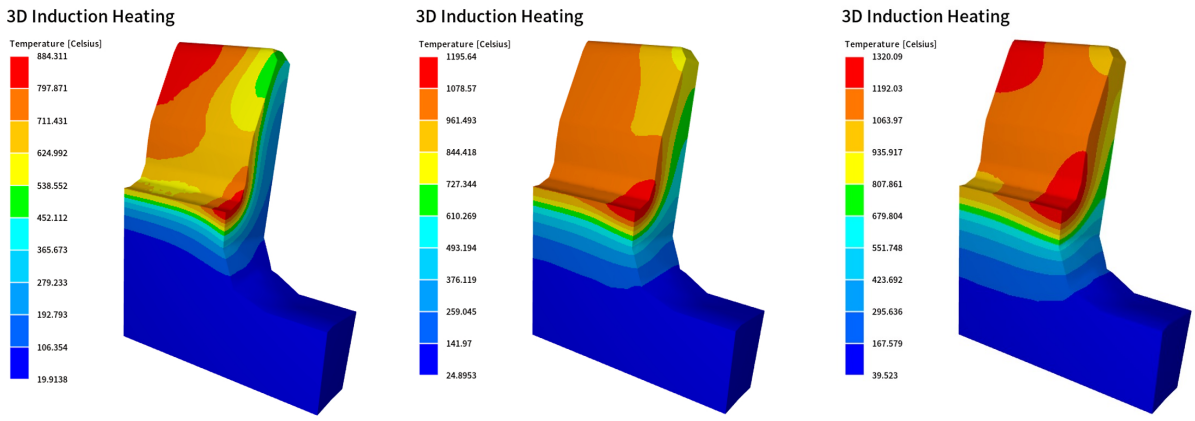
The numerical results that will be shown in this section were obtained with the process parameters illustrated in Table 3.6.

Table 3.6: Process parameters

Parameters	Values
Frequency (kHz)	170
Power (kW)	100

3.4.4.1 Temperature distribution

Figure 3.14 shows the distribution of the temperature at three time steps. It can be seen that the higher heat is concentrated in the superficial layer of the workpiece and that the maximum temperatures are recorded at the surface, which verifies the skin effect. Moreover, it was noticed that the temperature distribution is not homogeneous between the tip and root of the tooth.
In addition, it was observed that in the tooth root region, the edges are more heated than the mid-width. An opposite trend, however, was obtained in the tooth tip region. Such trend may be different when other values of process parameters (frequency and power) are considered.

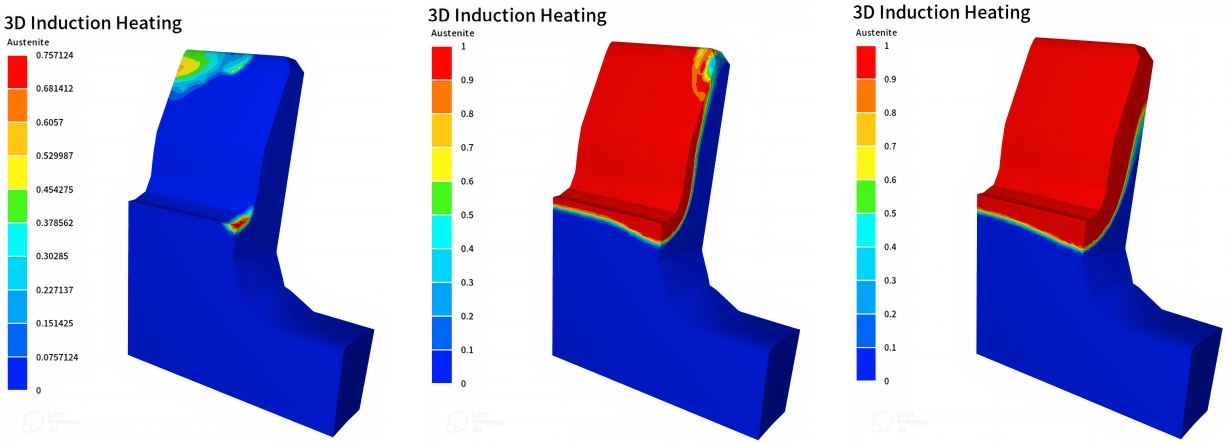


(a) (b) (c)

Figure 3.14: Isovalues of temperature distribution at times 0.1, 0.3, and 0.5 sec

3.4.4.2 Austenite phase distribution

Figure 3.15 shows the distribution of the austenite phase at three time steps. It was noticed that the austenite starts to form at the edges of the tooth root and mid-width of the tooth tip then spreads gradually. Any increase in temperature that reaches and exceeds $Ac1$ is accompanied by an immediate appearance of the austenite phase until a contour profile is obtained that covers the entire surface of the tooth at 0.5 sec.



(a) (b) (c)

Figure 3.15: Isovalues of austenite phase distribution at times 0.1, 0.3, and 0.5 sec

3.5 Conclusion

The numerical simulation of the multiphysics IH problem is complex and requires a good understanding of the involved physical phenomena, their mathematical formulation, mutual interactions, and the included parameters. All those points were evoked in this chapter. Indeed, precise control of the QoI is required to prevent process-induced defects such as overheating or underheating. The numerical FE simulation using FORGE® code is a valuable tool to tackle the complex IH process. The model development was illustrated step by step; creating geometries, meshing strategy with a convergence study to choose the best mesh, defining the material properties, and finally defining the boundary conditions and the process parameters. The obtained numerical results for temperature and austenite phase distribution demonstrate the skin effect, as well as the difference between the edge and the mid-width results on the gear geometry. Since the strategy for generating data is now clearly explained, creating metamodels using synthetic FE solutions provided by FORGE® code makes the object of the next chapter.

Chapter 4

Local parametric metamodels for induction heating process

Outline

4.1 Introduction 90
4.2 Numerical design of experiments - DoE 90
4.3 Data generation and sensors positions 91
4.4 Modeling the temperature field evolution 93
4.4.1 First proposed approach 94
4.4.2 Second proposed approach 101
4.5 Modeling the austenite-time evolution 108
4.5.1 Methodology 108
4.5.2 Results and discussion 110
4.6 Conclusion 114

4.1 Introduction

The procedure of metamodeling requires four basic steps at the offline phase as shown in Figure 2.7: design of experiments, data generation, data pre-processing (dimension reduction, missing value imputation, etc.), metamodel construction and validation. Each step deserves particular attention in order to provide an accurate metamodel at the end of the process, into which fast and inexpensive predictions can be generated at the online stage for new input parameters, within the considered parametric space, that have not been used during the training step.
This chapter aims to create metamodels for two QoI related to the IH process. Those metamodels were provided separately at some sparse sensors in the gear tooth. We start with introducing the adopted DoE and data generation methodology. Then, two separate sections are provided for each QoI, where the remaining metamodeling steps, specific to each one, are detailed. Finally, the obtained results are shown.

4.2 Numerical design of experiments - DoE

The efficiency of IH in the workpiece depends on several parameters, including heating system parameters (power supply capacity, frequency, heating time, etc.), gear material and geometry, inductor design, etc. Each parameter has its degree of influence on the final results. In this work, two process parameters were considered as listed in Table 4.1.

Table 4.1: Input parameters and their lower and upper limits

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

According to the literature, an adequate distribution of heat according to the contour of the teeth can be performed by a preheating at a medium frequency (MF) and a final heating at a high frequency (HF) in the sequential heating approach. However, in this study, the single frequency approach was considered to study the performance of the MOR techniques, while covering both MF and HF cases regarding the interval of the input parameters illustrated in Table 4.1. Moreover, the limits of the parametric input spaces are based on industry expertise. However, changing these input parameters could highly impact the process outputs. Hence, a data-driven non-intrusive MOR approach was adopted and parametric metamodels for the time evolution of temperature and austenite phase were

constructed to evaluate this effect.

However, data-driven engineering often requires an important amount of data. This constitutes one of its main drawbacks and, at the same time, one of its newest and powerful characteristics. For many engineering applications, such as the IH process, an important amount of data is not easy to obtain and it is very expensive. One may ask which DoE is appropriate for studying the IH process and which combination of input parameters should be adopted. Many kinds of DoE could be adopted, however, the choice was made to provide a good space-filling of the input space parameters in order to extract meaningful information about the input-output relationship, even in the low-data limit case. As previously stated, the LHS DoE was chosen because of its capacity to provide a reasonable representation of the whole multidimensional space. Therefore, a reduced and sparse sampling was adopted according to the LHS design of experiments, and a total of 20 full-order solutions was generated by solving the full-order FE models using the commercial FE software FORGE®, as shown in Figure 4.1. Indeed, for two input parameters, we exceed the minimum required for a linear regression and we can evaluate nonlinear regression methods with only 20 data.

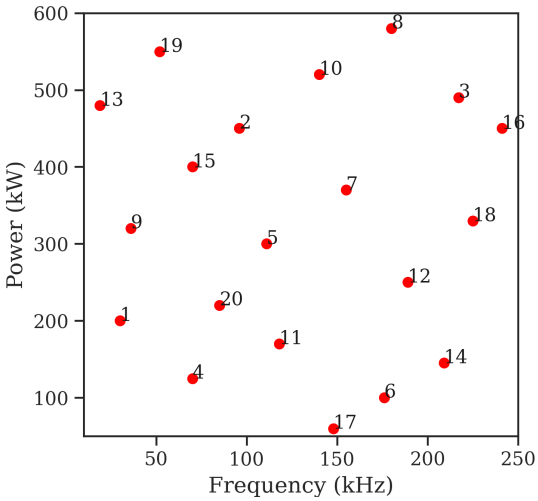


Figure 4.1: LHS design of experiments

4.3 Data generation and sensors positions

The computationally expensive FE simulations were computed for the IH process at the DoE samples. The fully-detailed procedure for numerical simulation is shown in chapter 3. All the computations were performed on a modern 64-bit workstation (40-core, 2 GHz, 256 GB RAM).

As mentioned before, heating simulation was carried out using two coupled solvers available in FORGE®. The first one solves Maxwell's equations based on the electromagnetic period, which is divided into a

series of 60time increments, to give rise to a heating power. The second one computes the temperature field by solving the heat equation using the calculated heating power and then the metallurgical phases by using transformation kinetics models. 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.

The temperature and the austenite phase evolution were extracted at 15 specific spatial points, obtained by implementing sensors on the studied geometry, as shown in Figure 4.2. These sensor positions were chosen in accordance with the position of pyrometers and thermocouples generally used in experiments to measure the temperature evolution at the surface and in-depth, respectively.

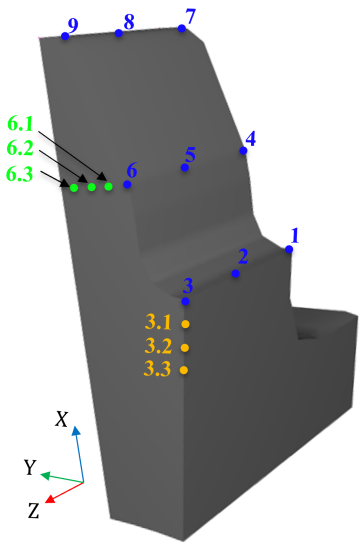


Figure 4.2: Sensors positions. Blue points represent the surface sensors while the green and orange ones represent those located at a certain depth

All simulations were supposed to run with the same time process, initially chosen to be equal to 0.5 sec, which is sufficient for such a rapid heating process. However, some simulations cannot reach this time, generating hence incomplete datasets which can highly affect the data analysis. The main reasons behind this issue are:

- The process algorithm shown in Figure 3.6 is repeated until the end of the heating cycle, however, it can also be stopped if the temperature or the number of the thermal time steps reaches a high predefined value. As previously stated, a dense time stepping was found for simulations having higher values of frequency and/or power. Indeed, the time step decreases even more when

running the simulation with a high variation of temperature. Figure 4.3 shows a visualization of the time evolution of the number of time steps and temperature for simulation 1 and simulation 18 from the DoE (Figure 4.1), representing very different input parameters. It can be noticed that the number of time steps is proportional to the temperature evolution during the same time interval. The number of time steps for simulation 18 is almost 4 times higher than the one for simulation 1. Therefore, the time step is very different between the two cases. When the number of time steps is accompanied by a decrease in the time step, this may generate a huge increase in the computation time and even a computation stop.

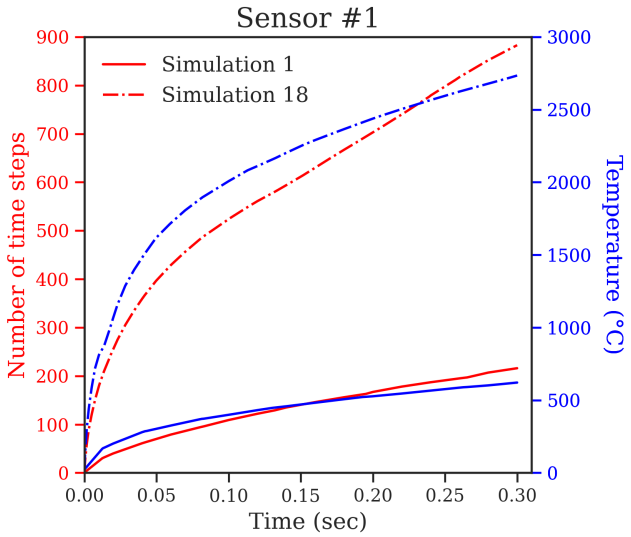


Figure 4.3: Illustration of the time evolution of the number of time steps and temperature for simulation 1 and simulation 18 at sensor #1

- all material properties were defined until 1400°C. Consequently, extrapolated values for higher temperatures could affect the results.

4.4 Modeling the temperature field evolution

According to the DoE, the temperature evolution changes considerably with the input parameters as illustrated in Figure 4.4 for sensor #5 and 3 different simulations from the DoE. It can be noticed that the temperature curves are monotonic but they show different stats. The maximum temperature at 0.5 sec is about 800°C for the left curve, while it takes about 1300°C for the other two curves. Although these latter curves have the same final temperature for different input parameters, they experience different trends. The large transition slope, observed for simulation 19 (between 0.05 and

0.2 sec) is related to the material properties change with respect to the evolution of the metallurgical phases during heating.

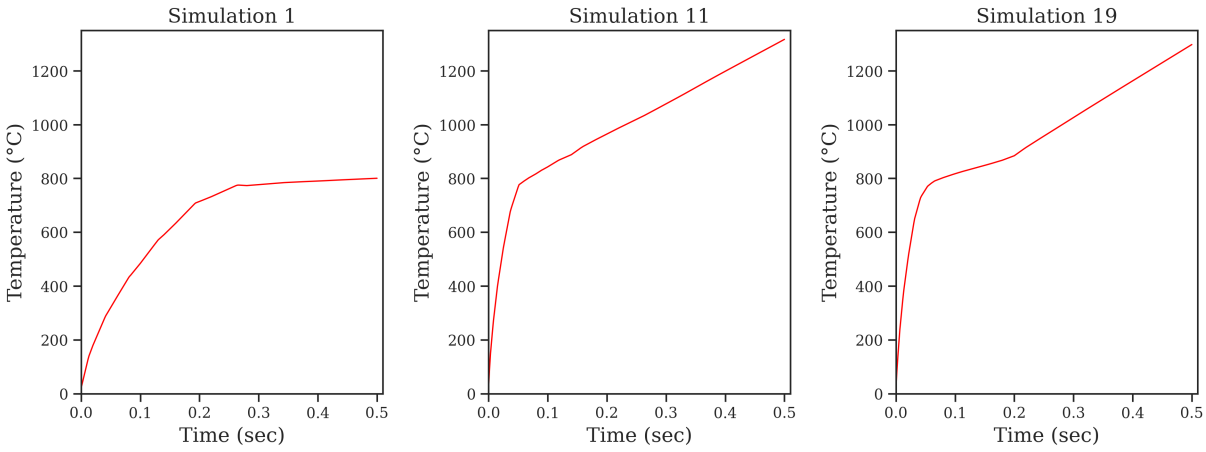


Figure 4.4: Illustration of the temperature field evolution

Therefore, the developed metamodels would eventually help to rapidly choose the input parameters that provide the desired results for a specific sensor position and consequently identify the parameters that could generate an underheating or overheating for the considered gear geometry. It is worth pointing out that the obtained FE results could exceed the melting temperature for some choices of frequency and power, however, the same values can provide the desired result for a different geometrical configuration with a larger gear or higher air gap between the gear and the inductor. This point constitutes an interesting research focus.

In this section, we focused on developing metamodels for the temperature-time evolution by applying a dimensionality reduction on a set of computed snapshots, collected at some sparse sensors in the space domain, by using POD. Then, regression methods were assumed to be applied on the POD modal coefficients. However, due to the obtained incomplete datasets with different data-lengths, POD could not be applied directly on data. Hence, two different strategies were adopted. These latter are explained in the next two sections.

4.4.1 First proposed approach

4.4.1.1 Methodology

The very first proposed idea consisted in considering different final time processes instead of a constant value and considering this latter as a third input parameter to model the temperature. However,

different time data-lengths were obtained by using different process times. First, data normalization and alignment with respect to the final time process for each snapshot were applied in order to transfer all quantities (time, temperature, austenite, etc.) into the same discretization; however, this method could introduce pollution in the regression. Indeed, a simple illustration of the proposed idea is illustrated in Figure 4.5.

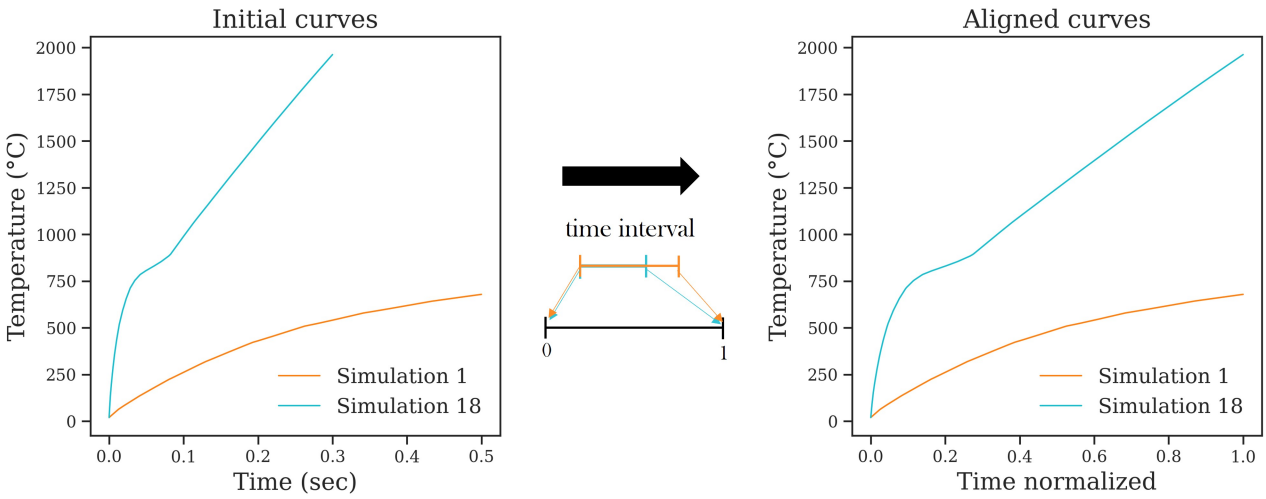


Figure 4.5: Data alignment

Representing curves that show potentially different physical events in the same normalized time interval is not a conceptually well-posed problem.

To overcome this issue, the proposed solution consists in truncating all temperature curves to a constant maximal temperature value and noting that the temperature evolution is monotonic, the role of time and temperature was inversed. Therefore, the time was modeled as a function of the temperature, where the maximum value of the latter was kept the same for all the experiments. A constant final temperature was chosen exclusively for each sensor depending on its position and according to the temperature distribution on the workpiece. It is worth mentioning that for some simulations, the chosen final temperature wasn't reached with the final time process equal to 0.5 sec. Therefore, those FE simulations were launched to continue the computation for a higher time process in order to reach the required temperature value. However, despite having a constant final temperature, each dataset does not have the same length; therefore, an alignment was applied. Once the new datasets for time as a function of the temperature were established, the POD-based reduced order model was applied to get the low-dimensional representation of the initial solutions, followed by nonlinear regression techniques.

This approach is illustrated in Figure 4.6

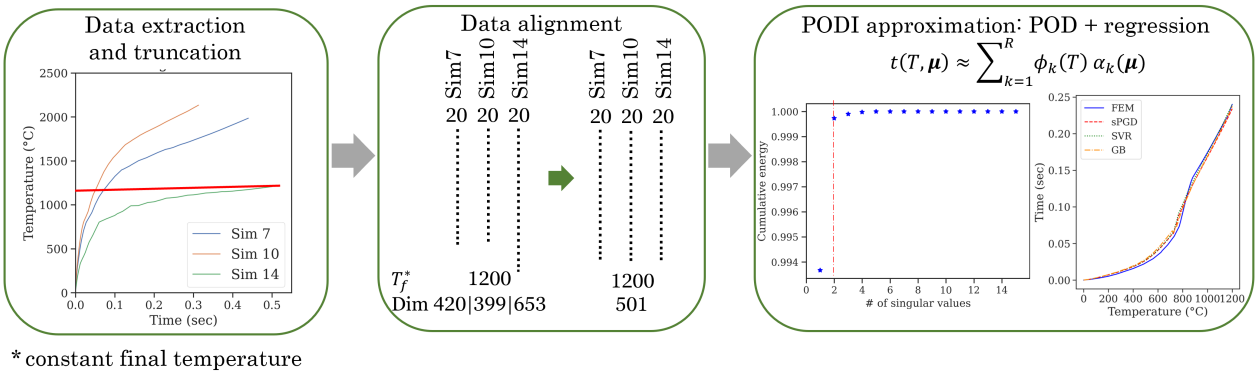


Figure 4.6: Illustration of the first proposed approach

4.4.1.2 Results and discussion

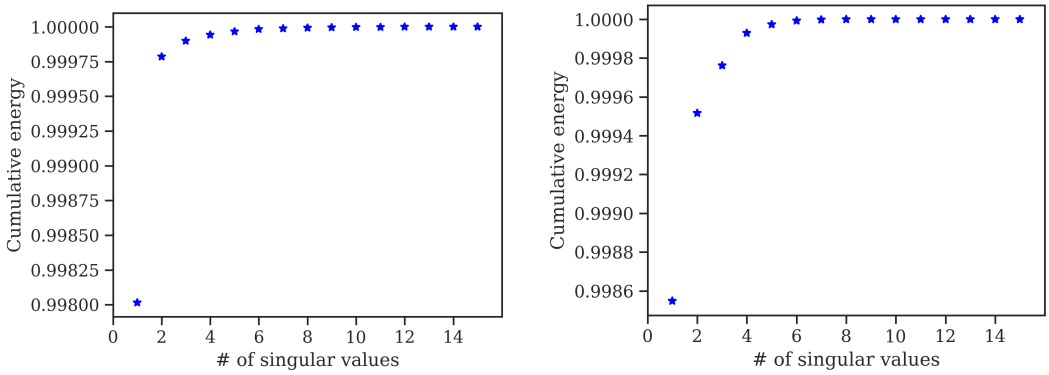
Once the data truncation and alignment were established for each sensor, simulations were split into two subsets. Simulations 2, 9, 11, 12, and 16 inside the DoE were used for testing the models while the other 15 simulations were taken into account to construct the parametric solution. In other words, 75% of data were used to build the models and 25% of them to evaluate their prediction accuracy.

The PODI was then performed on the 15 time-temperature curves $\boldsymbol{u}_i$ (snapshots) forming the snapshot matrices $\boldsymbol{M}_k$ for each sensor k. The average of time 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 the others. The Eq.(2.2) becomes:

$$\boldsymbol{u}_i \approx \bar{\boldsymbol{u}} + \sum_{k=1}^{R} \alpha_{ki}\boldsymbol{\phi}_k \quad \text{for} \quad i = 1, \ldots, P \tag{4.1}$$

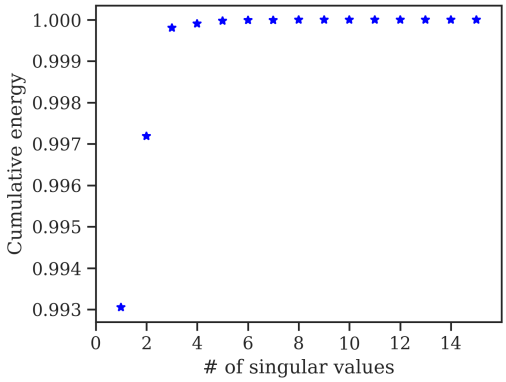
By applying the SVD, the left singular vectors of the snapshot matrices were truncated to the two or three first singular vectors (POD modes), depending on the sensor snapshots. This choice was made such that more than 99.9% of the cumulative energy was recovered, as shown in Figure 4.7 at 4 different sensors.

It is worth pointing out that only results for sensors (#1, #5, #9, and #3.2) in reference to Figure 4.2 are shown for the sake of clarity.

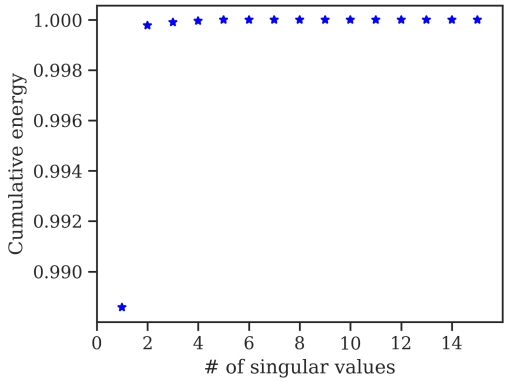


(a) Sensor #1

(b) Sensor #5



(c) Sensor #9



(d) Sensor #3.2

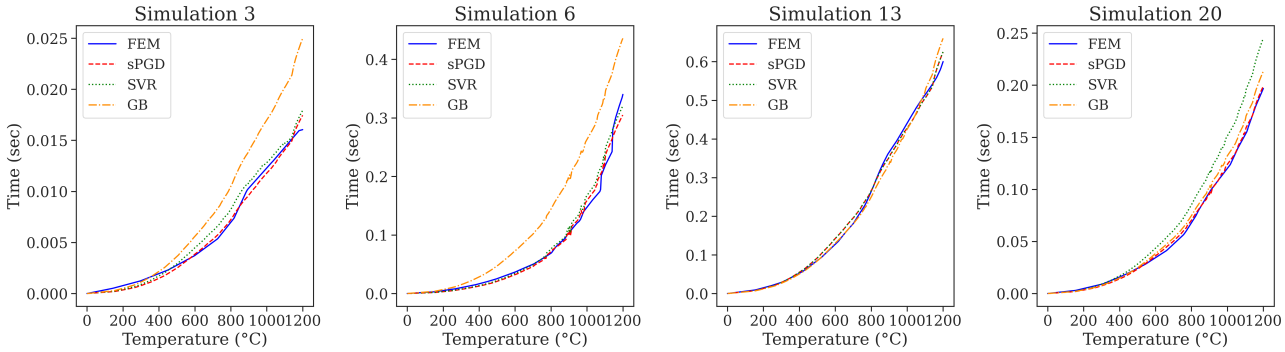
Figure 4.7: Cumulative energy computed from 15 snapshots of the time at 4 different sensors

Then, by representing the time function u only with its projection onto the POD basis vectors, assumed to be invariant with respect to the input parameters, the approximated function $\tilde{u}$ is fully defined by approximating the reduced state vectors of the snapshots data, the so-called POD modal coefficients. These latter were calculated using Eq.(2.13) and three different nonlinear regression methods were used to develop a model for the POD modal coefficients in order to enable predictions for untried parameters. The employed nonlinear regression methods are sPGD, SVR, and GB, where a comparison of the obtained results was achieved.

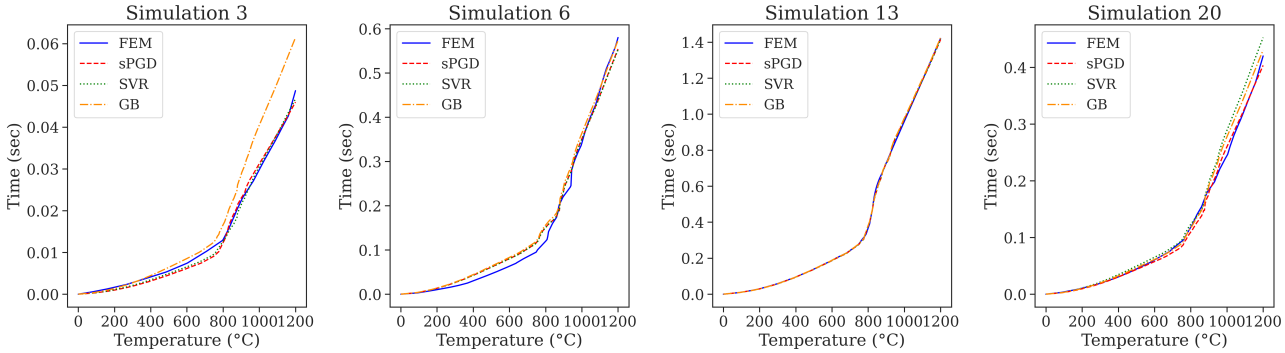
In fact, $15 \times R$ regression models were built, one for each POD modal coefficient and each sensor, such that R (2 or 3) is the number of the retained POD modes. It is commonly known that time is inversely proportional to frequency and electric power. As a result, the inverse of the input parameters was considered when constructing the regression models. Besides, the standardization of the input parameters was applied to avoid problems related to units and different scaled parameters. This last step is not necessary for methods based on decision trees, like the GB method.

As mentioned in chapter 2 section 2.3, all regression methods are controlled by hyperparameters that need to be well-chosen to get a better prediction and avoid overfitting. For regression methods included in the Scikit-learn library in Python, such as SVR and GB, a tool based on cross-validation (Refaeilzadeh et al., 2009), called GridSearchCV, was applied to choose the best combination of hyperparameters among a list of tested parameters. For sPGD, such tool is not available, hence multiple configurations were tested to find the best prediction.

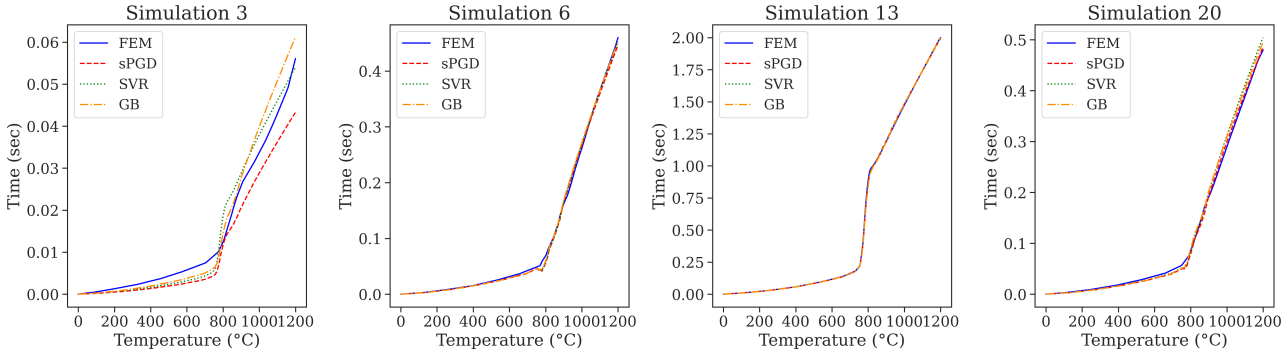
Once the data-fit models were created for each reduced coordinate, the final approximation of time was reconstructed using Eq.(4.1). Figure 4.8 and Figure 4.9 show the time-temperature evolution obtained by the full-order FEM and the regression methods for 4 sensors and 4 training and testing simulations, respectively. It can be noticed that for most of the training cases, the FE curves and the predicted ones by the 3 regression methods are almost overlapped even for very different time ranges. However, the GB method provided a less accurate approximation for some training snapshots. For the testing data, the overfitting generated by the GB method is observed since good prediction was provided for training data and not for the new untried ones. On the other hand, FE curves are in good agreement with approximated ones with sPGD and SVR for all sensors and the 4 testing simulations, where no significant difference was observed.



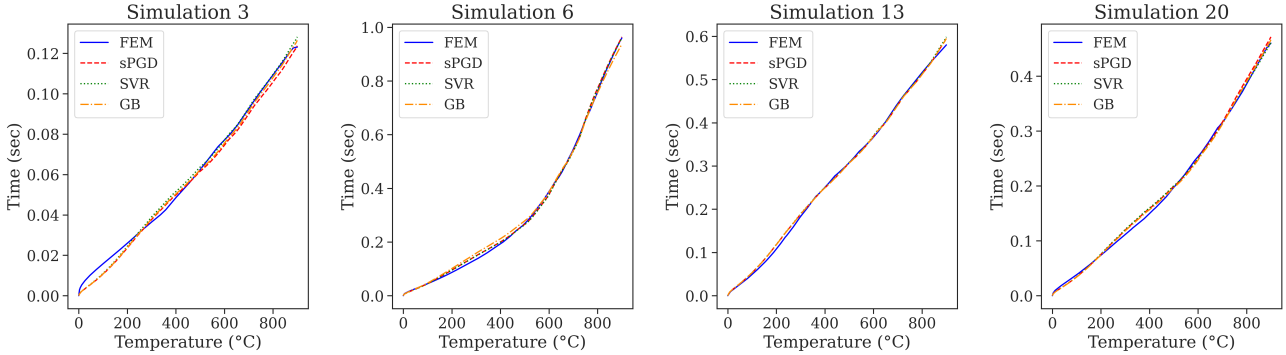
(a) Sensor #1



(b) Sensor #5

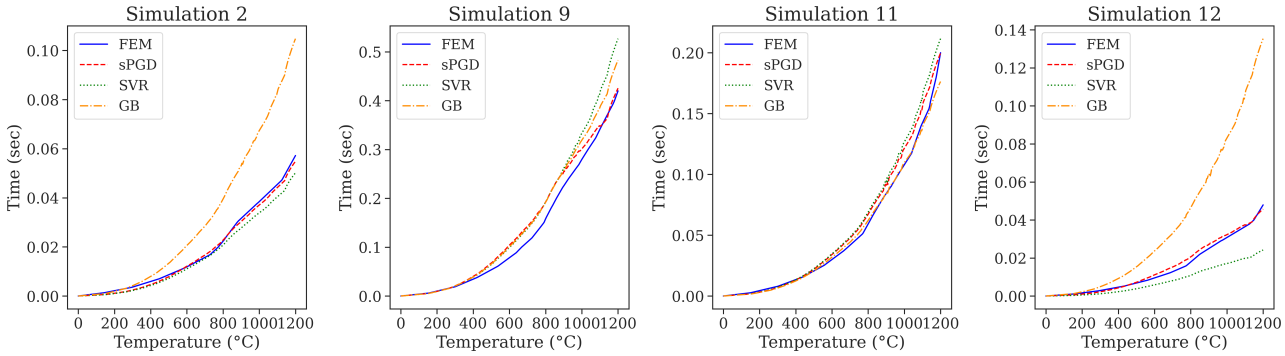


(c) Sensor #9

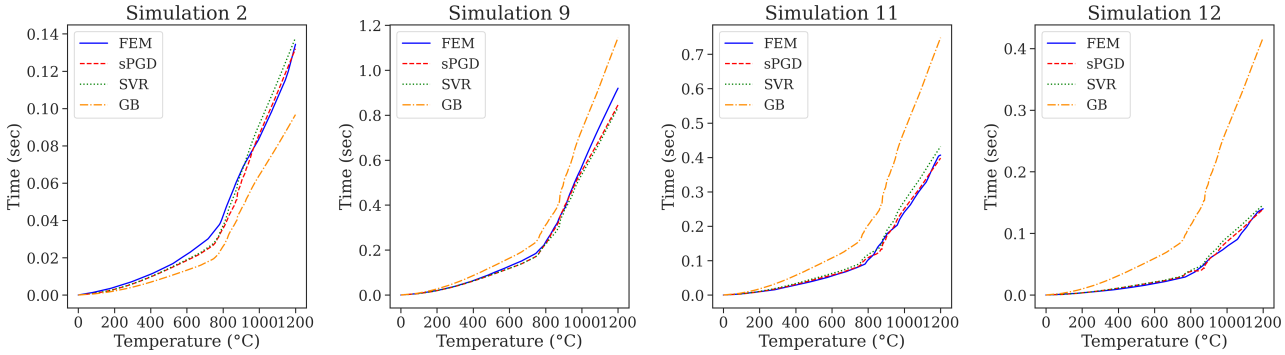


(d) Sensor #3.2

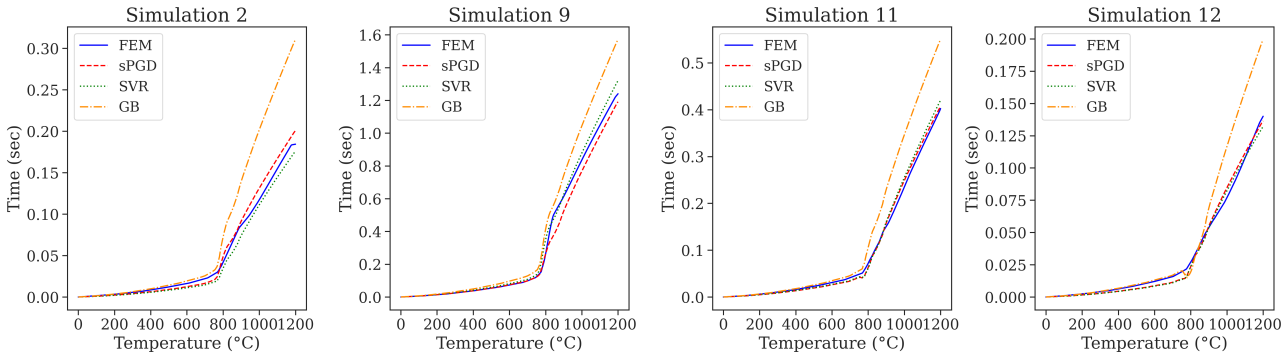
Figure 4.8: Comparison between the reference FE and the regression models of the time-temperature evolution at 4 sensors and for 4 training simulations



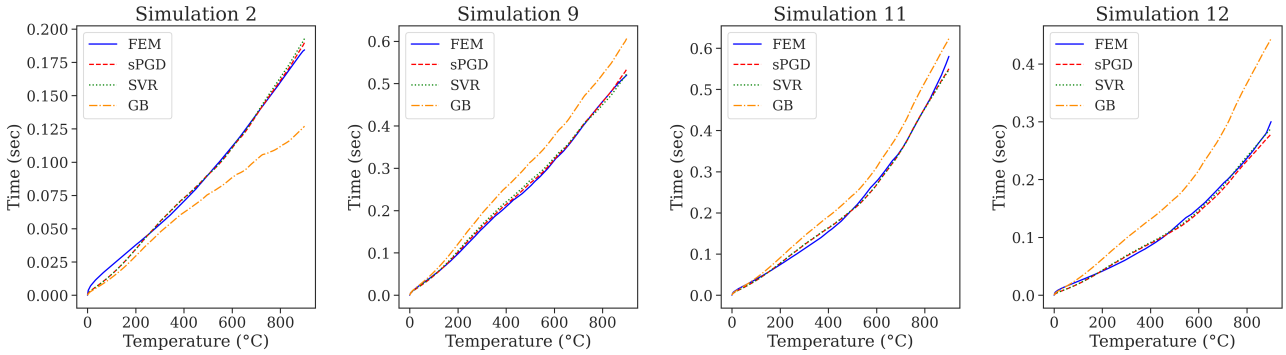
(a) Sensor #1



(b) Sensor #5



(c) Sensor #9



(d) Sensor #3.2

Figure 4.9: Comparison between the reference FE and the regression models of the time-temperature evolution at 4 sensors and for 4 testing simulations

Additionally, the relative error to measure the prediction accuracy for the training and the testing datasets is presented in Table 4.2, where the computed error is defined as:

$$Error(\%) = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{\sqrt{\int(\boldsymbol{t}_i^{pred}-\boldsymbol{t}_i^{FEM})^2 dT}}{\sqrt{\int(\boldsymbol{t}_i^{FEM})^2 dT}}\right)\times 100 \tag{4.2}$$

where n, $\boldsymbol{t}_i^{FEM}$, and $\boldsymbol{t}_i^{pred}$ are the number of data points (simulations), the vectors of the FE solutions, and the predicted solutions, respectively. The obtained results are in good agreement with the graphical results such that good prediction was obtained with sPGD and SVR with an error less than 10% for most of them. Globally, the best approximations were provided by the sPGD, followed by the SVR and then the GB method. The GB method shows overfitting and very high errors for the testing dataset.

Table 4.2: Relative errors of time evolution using 3 regression methods

	sPGD		SVR		GB	
Error (%)	train	test	train	test	train	test
Sensor #1	4.8	8.1	15.6	20.2	32.4	65.0
Sensor #5	5.1	6.1	8.0	8.7	15.7	89.3
Sensor #9	6.0	10.3	5.5	8.9	6.7	44.5
Sensor #3.2	2.7	3.3	2.5	3.0	3.8	22.6

4.4.2 Second proposed approach

4.4.2.1 Methodology

The second proposed approach consists in completing the missing values when considering the same time process for all snapshots. Indeed, 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, were different from one simulation to another. Among the total number of simulations, 70% of them had completely known data vectors because they reached the end of computation and the other ones had a gappiness percentage varying between 15% and 65%.

Then, the modeling of the temperature as a function of time was achieved. This approach was proposed because it seems to be more suitable and more general for spatial interpolation, elaborated in the next chapter. The applied methodology is illustrated in Figure 4.10.

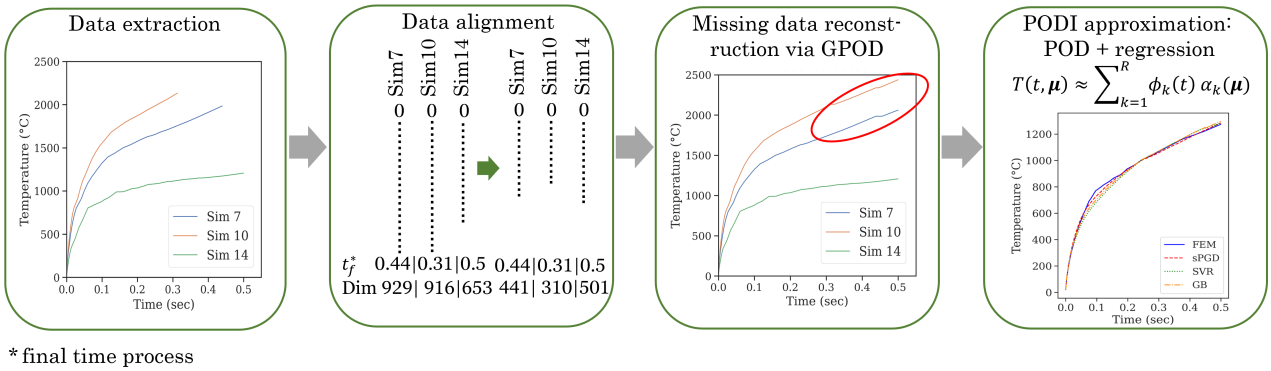
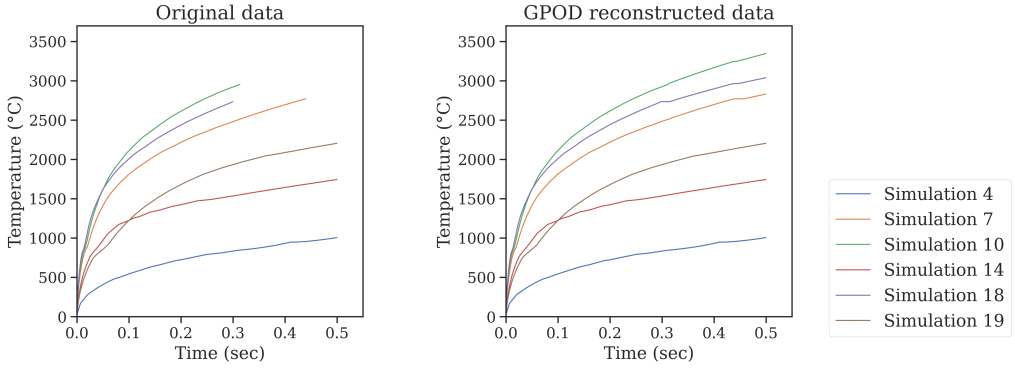


Figure 4.10: Illustration of the second proposed approach

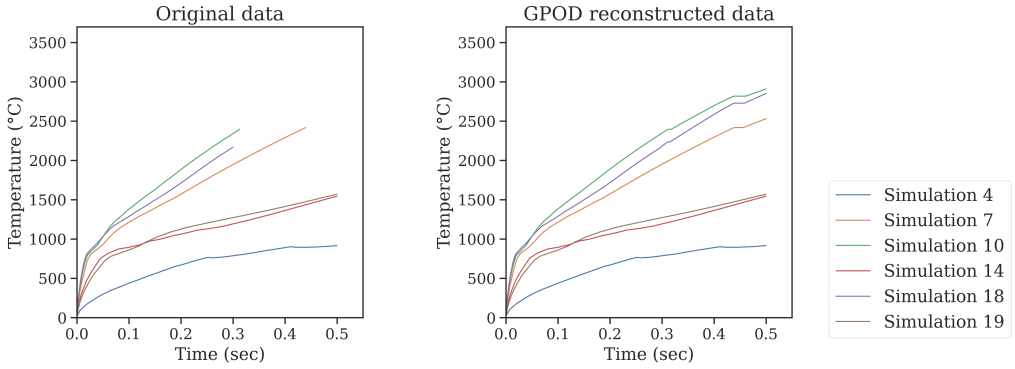
4.4.2.2 Results and discussion

In order to reconstruct an approximate solution for the incomplete data, the extended Everson-Sirovich procedure described in section 2.2.1.3 was applied. The number of modes was initialized to $R = 1$ and increased iteratively, enabling an improved full-field reconstruction accuracy, such that 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 Figure 4.11 for 6 different simulations.

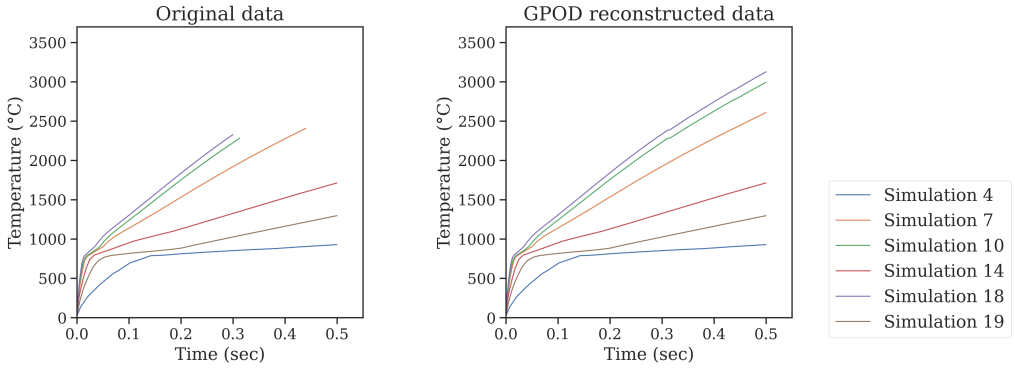
Since the gappy fields are completely unknown, comparing the approximated gappy values with the "true" ones can not be achieved. It is worth mentioning that the reconstructed missing data of temperature are unreachable in reality since they exceed the melting temperature of C45 steel. However, having 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 of 501-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.



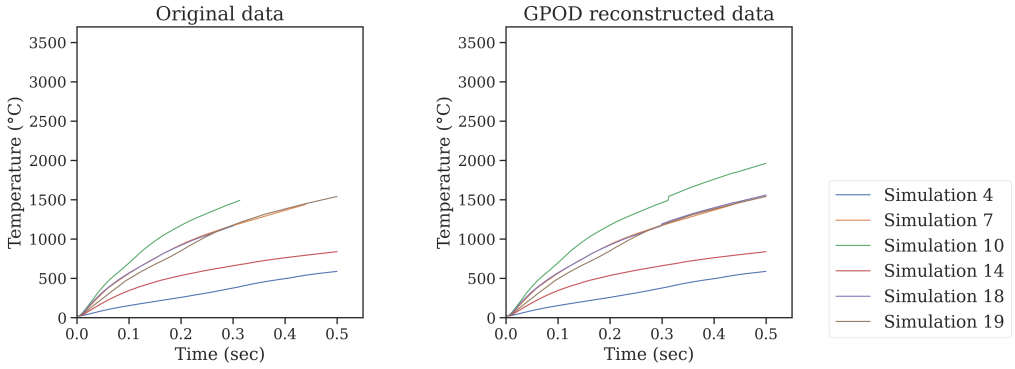
(a) Sensor #1



(b) Sensor #5



(c) Sensor #9

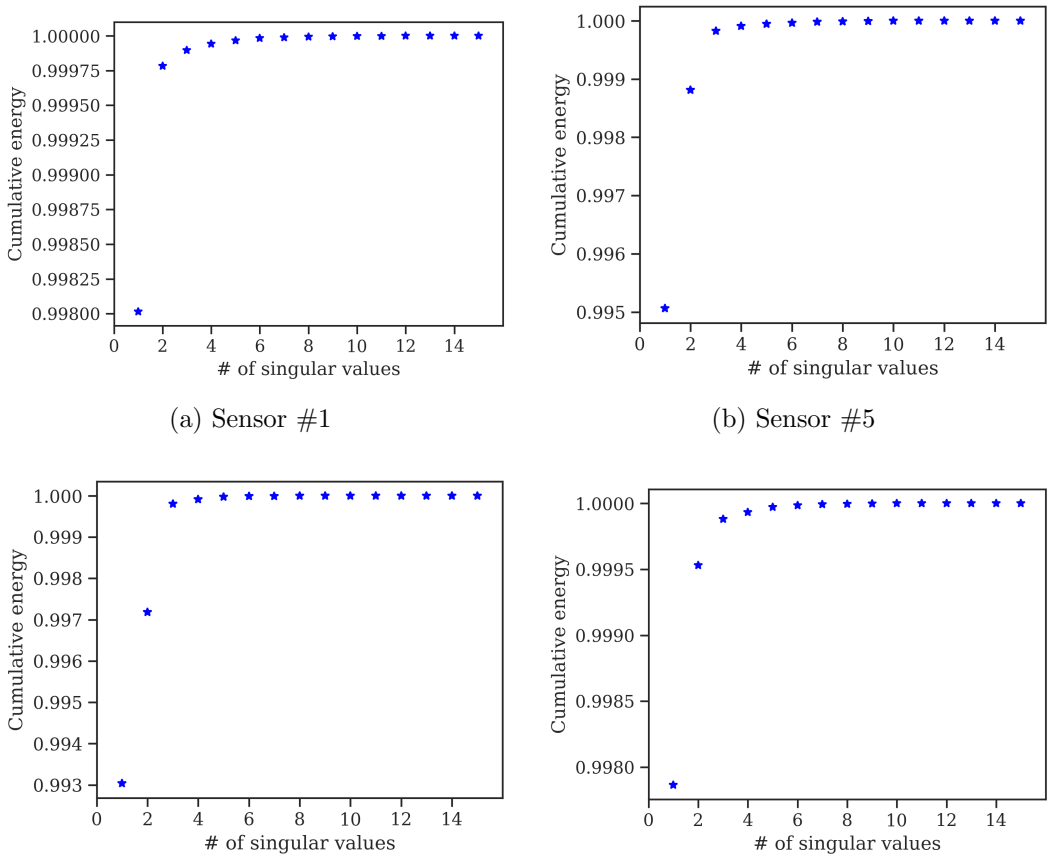


(d) Sensor #3.2

Figure 4.11: Original and reconstructed data for surface and internal sensors

Once the gappy data reconstruction was established for each sensor, simulations were split into two subsets. Similarly to section 4.4.1, simulations 2, 9, 11, 12, and 16 inside the DoE were used for testing while the other 15 simulations were used for training the models.

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. 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 snapshot data. This choice was made such that more than 99.9% of the cumulative energy was recovered, as shown in Figure 4.12 at 4 different sensors.



(a) Sensor #1

(b) Sensor #5

(c) Sensor #9

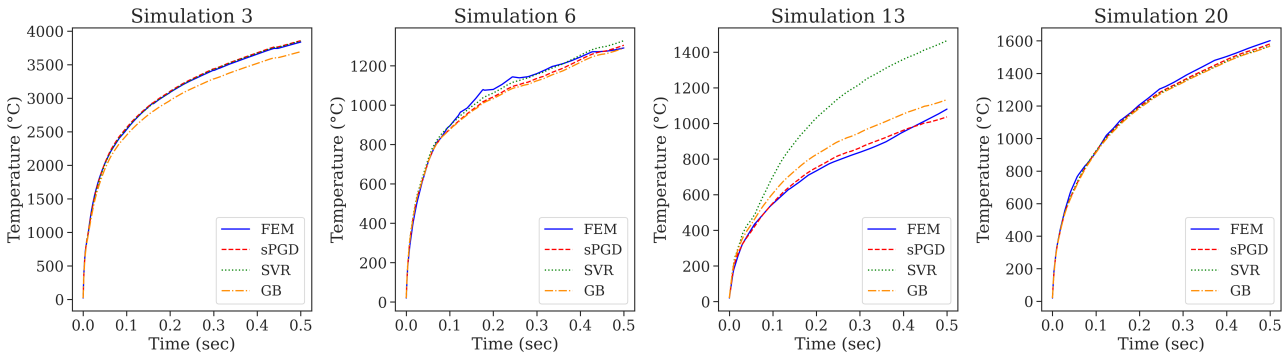
(d) Sensor #3.2

Figure 4.12: Cumulative energy computed from 15 snapshots of the thermal field at 4 different sensors

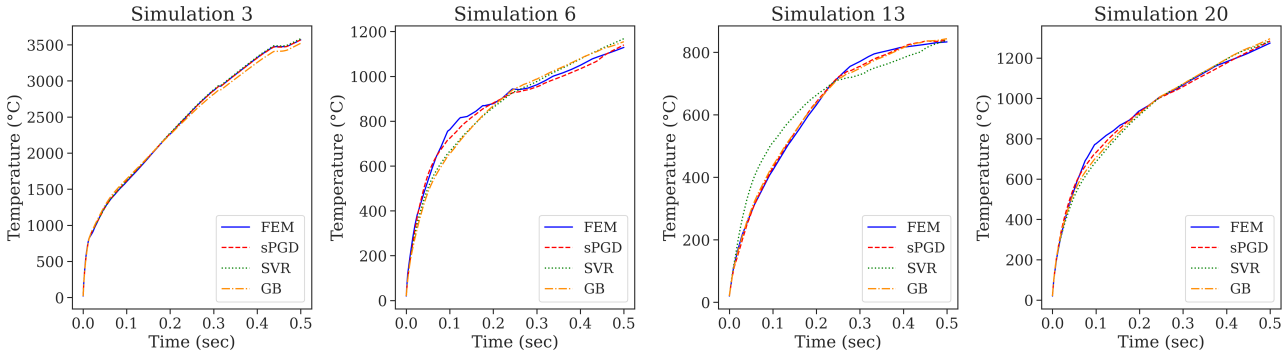
Models for the POD modal coefficients and for each sensor were then constructed by using the three previously employed regression methods, and by considering the frequency and power as input param-

eters for the models. The final approximation of temperature was reconstructed by using Eq.(4.1) with the predicted values of POD modal coefficients. Figure 4.13 and Figure 4.14 show the temperature-time evolution obtained by the full-order FEM and the regression methods for the 4 sensors and the 4 training and testing simulations, respectively. For the training simulations, it was noticed that the sPGD succeeded to well approximate the solution for the presented cases, showing a maximal value of temperature that varies from 700°C to 4000°C, in comparison with the SVR and GB methods. The approximation obtained with these two latter methods shows a similar trend as the FE results and they are in good agreement for most training simulations. However, they show difficulties in approximating the temperature curves that have, comparatively, a low final temperature such as the simulation 13. These results can be justified by the choice of hyperparameters that was achieved by cross-validated grid-search over the specified parameter values in a way to find the best set of parameters that provides the minimum error. However, these hyperparameters can perform very well with some data and not with others.

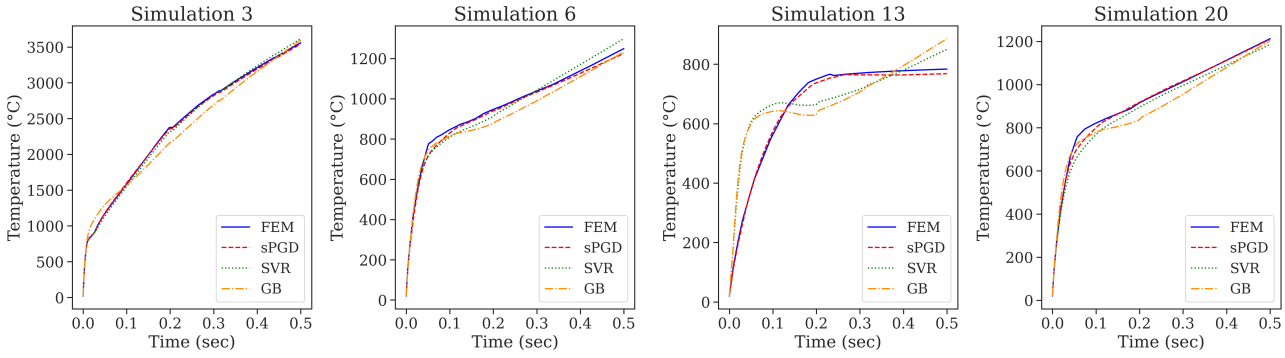
For the testing simulations shown in Figure 4.14, the FE curves are in good agreement with the approximated ones with the sPGD and the SVR methods for all sensors but for only 3 testing simulations. However, simulation 9 shows a very less accurate prediction which can be explained by the obtained results for the training simulations that have a similar trend to simulation 9. Moreover, the GB method comparatively shows the least accurate predictions.



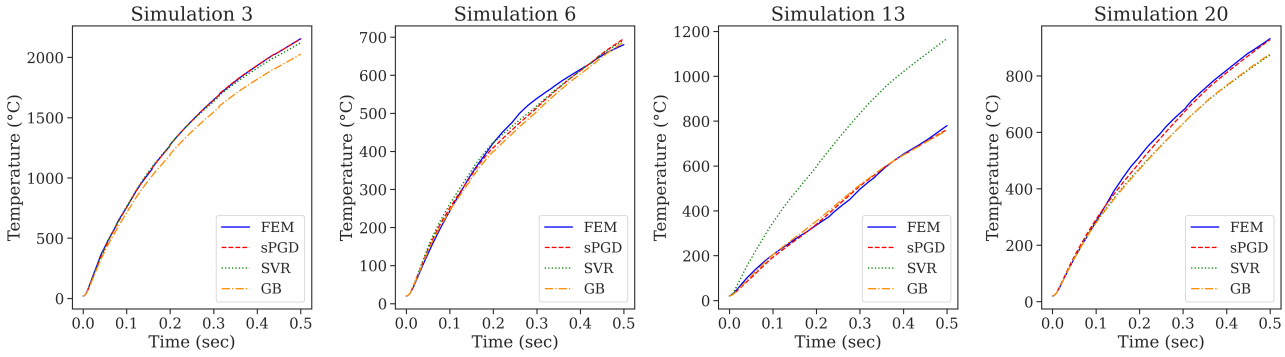
(a) Sensor #1



(b) Sensor #5

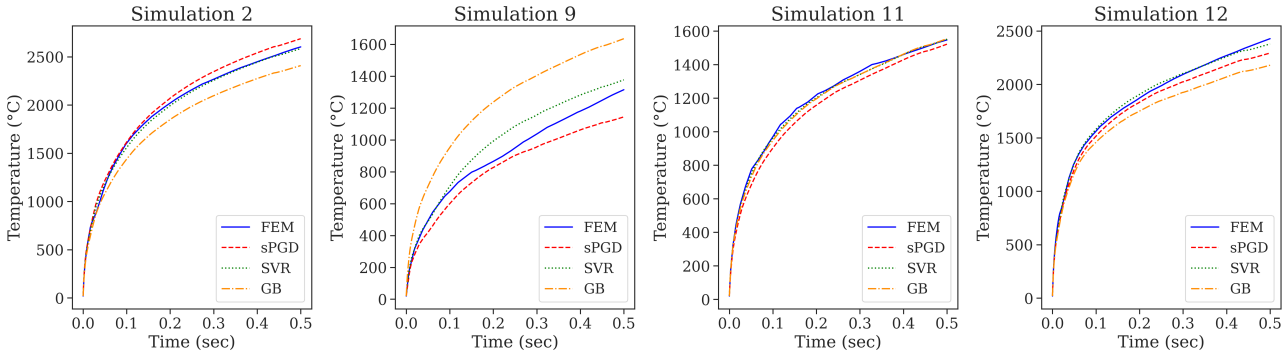


(c) Sensor #9

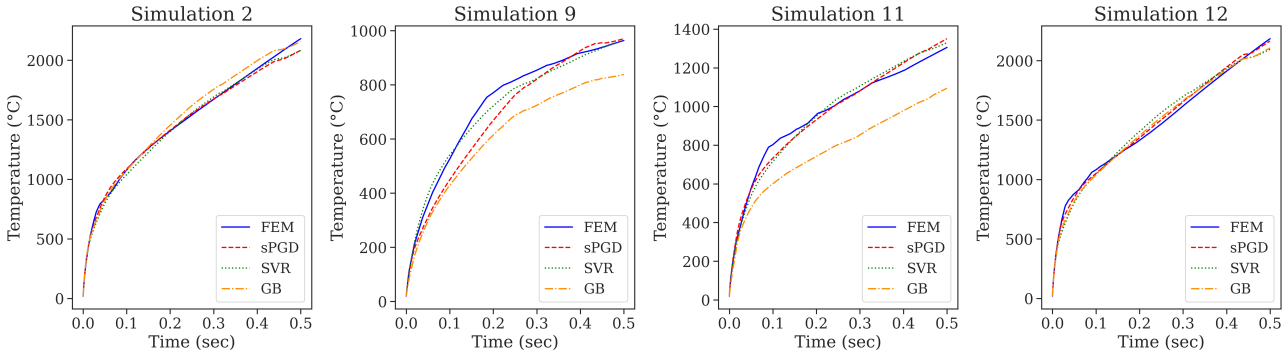


(d) Sensor #3.2

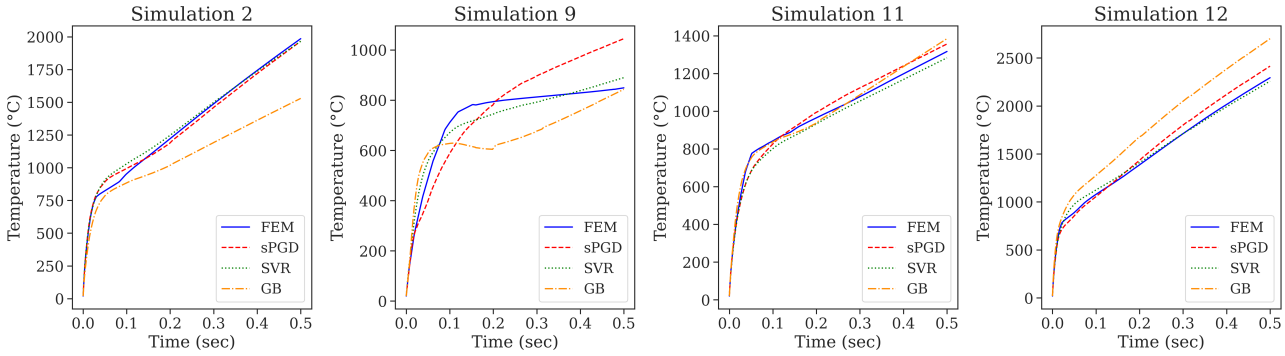
Figure 4.13: Comparison between the reference FE and the regression models of the temperature-time evolution at 4 sensors and for 4 training simulations



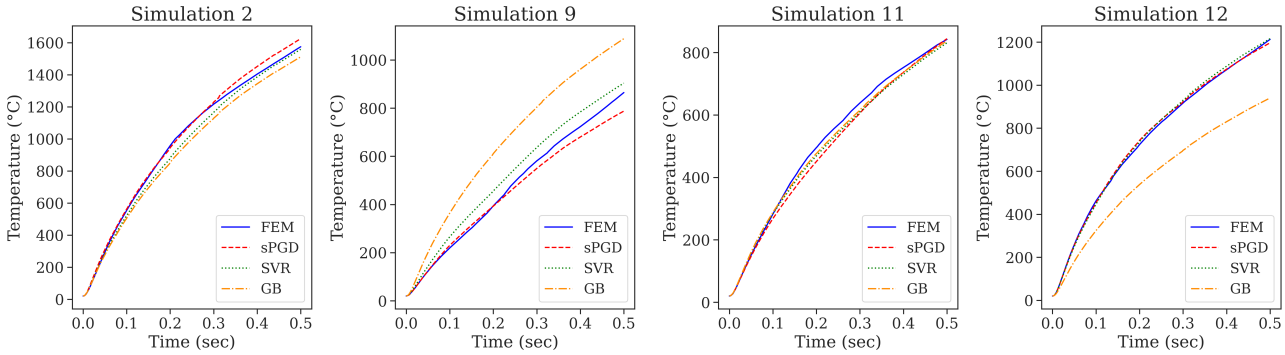
(a) Sensor #1



(b) Sensor #5



(c) Sensor #9



(d) Sensor #3.2

Figure 4.14: Comparison between the reference FE and the regression models of the temperature-time evolution at 4 sensors and for 4 testing simulations

The relative error to measure the prediction accuracy for the training and the testing datasets is presented in Table 4.3, where the computed error is defined by:

$$Error(\%) = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{\sqrt{\int(\boldsymbol{T}_i^{pred}-\boldsymbol{T}_i^{FEM})^2dt}}{\sqrt{\int(\boldsymbol{T}_i^{FEM})^2dt}}\right)\times 100 \tag{4.3}$$

where n, $\boldsymbol{T}_i^{FEM}$, and $\boldsymbol{T}_i^{pred}$ are the number of data points (simulations), the vectors of the FE solutions, and the predicted solutions, respectively.

As shown in Table 4.3, the lowest errors were obtained with sPGD and do not exceed 2% for the training data and 6% for the testing data at the 4 considered sensors. Moreover, the SVR method performs well with the testing data with errors less than 5.5% for the 4 sensors. The testing errors with SVR are very close to those obtained with sPGD despite having higher errors for training than sPGD. Besides, the GB method gives higher errors that exceed 10% for the testing data while having lower errors for the training data, resulting therefore an overfitting.

These results indicate that the sPGD and the SVR methods can provide relatively accurate models to predict the temperature-time evolution under the imposed values of input parameters even in the sparse low-data limit case, unlike the GB method which doesn't perform well with a reduced amount of data.

Table 4.3: Relative errors of temperature-time evolution using 3 regression methods

	sPGD		SVR		GB	
Error (%)	train	test	train	test	train	test
Sensor #1	1.3	5.6	4.8	2.1	5.2	14.3
Sensor #5	1.1	3.4	4.6	3.3	3.7	10.1
Sensor #9	1.3	5.3	4.0	3.1	5.3	13.5
Sensor #3.2	2.0	3.1	9.2	4.1	3.6	18.6

4.5 Modeling the austenite-time evolution

4.5.1 Methodology

During heating, austenitization consists in transforming the different metallurgical phases present at low temperatures into austenite. Several models for the characterization of phase transformation kinetics can be found in the literature as shown in chapter 3. In this work, a continuous TTA diagram or CHT was employed. However, it is worth pointing out that only data generated by FORGE® are needed in this work and the knowledge of the models is not even required.

Besides, depending on the input parameter changes, austenite phase evolution can be observed in three different stats as shown in Figure 4.15.

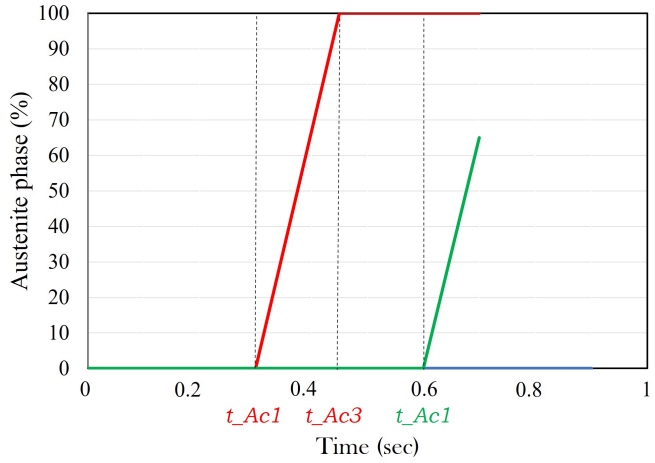


Figure 4.15: Three possible states of austenite phase evolution

As can be seen, the phase transformation starts at time t_Ac1, and the austenitization is completed at time t_Ac3. These instants correspond to the beginning and end transformation temperatures, respectively. In some cases, due to a short processing time or low values of process parameters, austenite phase evolution does not reach a complete austenitization (green curve) or there is no austenite phase (blue curve). Therefore, a classification method can be applied on snapshots in order to avoid dealing with different trend curves such that snapshots for each sensor can be classified into two classes:

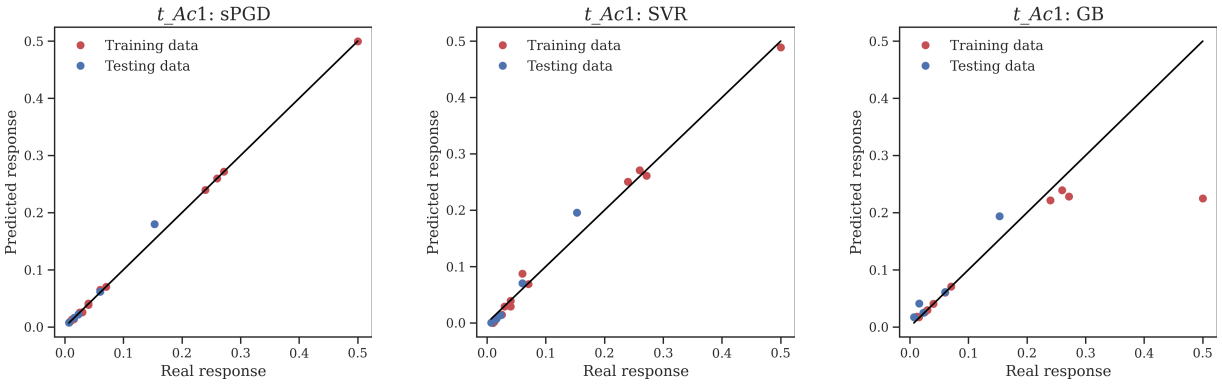
- class 1: no austenite phase transformation was produced, the austenite phase remains at zero until the end of the process (blue curve in Figure 4.15);
- class 2: there is an austenite phase transformation; two cases are envisaged:
  - 100% of austenite is reached at the final time step (red curve in Figure 4.15);
  - less than 100% of austenite was obtained after heating (green curve in Figure 4.15).

However, due to the importance of the metallurgical transformation in addition to the reduced quantity of data (20 snapshots), snapshots provided in section 4.4.1, with time processes higher than 0.5 sec for some simulations, were considered to avoid dealing with data where no austenitic transformation was produced. In order to model the austenite phase evolution, the proposed idea consists first in collecting the characteristic time instants t_Ac1 and t_Ac3 and next applying nonlinear regression methods to build a model of these two quantities. The same regression methods; sPGD, SVR, and

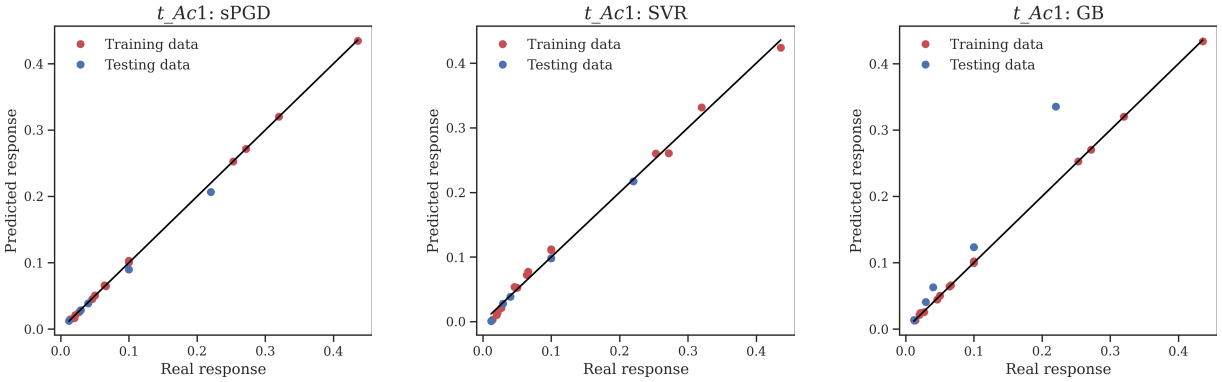
GB were employed and a comparison of the obtained results was illustrated in the next section.
It is worth pointing out that the inverse of frequency and power were considered as input parameters, given the inverse proportionality with respect to time.

4.5.2 Results and discussion

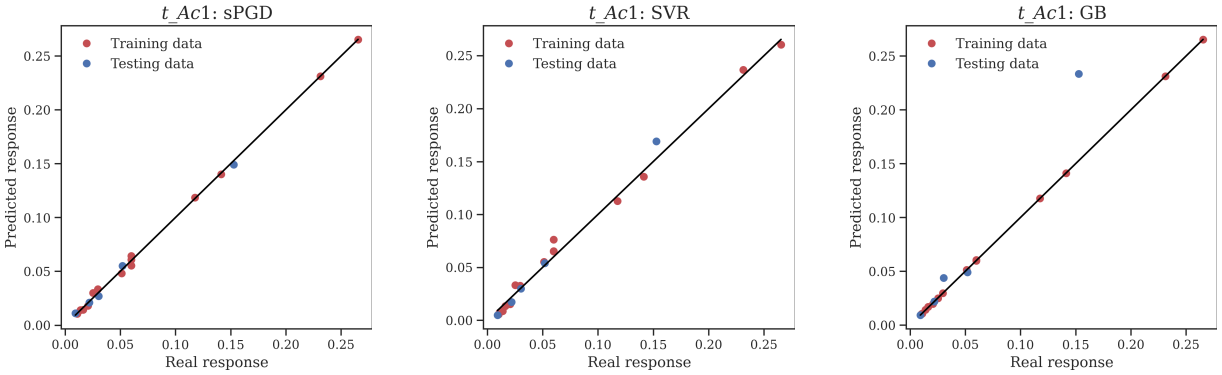
A standardization of the input parameters was first applied and then the datasets were divided into training and testing subsets (75% of data were used to build the models and 25% to evaluate their prediction accuracy). Figure 4.16 shows the FE versus the predicted values of t_Ac1 for sensors (#1, #5, #9, #3.2), using sPGD, SVR, and GB regression methods and for both training and testing datasets. The red points correspond to the 15 simulation data used for training the models while the blue ones correspond to the 5 data used for testing. The black line represents the perfect prediction such that when points are close to this latter, the model provides a good fit to the data. Indeed, the dispersion of these points with respect to the black line gives a visual indicator of error. The sPGD model seems to be more accurate than SVR and GB methods for predicting t_Ac1. The SVR method shows a good performance as well. However, the GB method is less accurate for predicting new data that are not used for training. This interpretation is observed for almost all sensors.



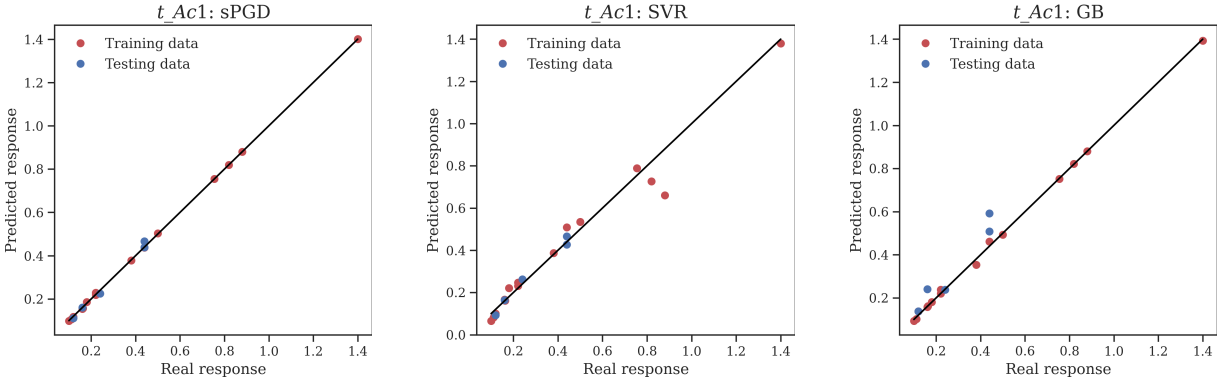
(a) Sensor #1



(b) Sensor #5



(c) Sensor #9



(d) Sensor #3.2

Figure 4.16: Reference FE versus predicted values of t_Ac1 for 4 sensors

In order to measure the prediction accuracy, the relative error for predicting a quantity y was applied such that:

$$Error(\%) = \left(\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{pred} - y_i^{FEM})^2}}{max(\boldsymbol{y}^{FEM}) - min(\boldsymbol{y}^{FEM})} \right) \times 100 \tag{4.4}$$

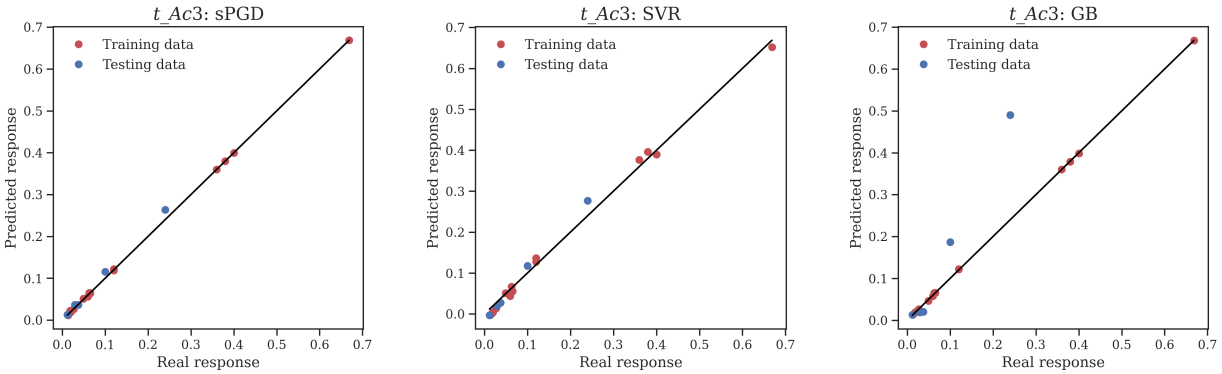
where $\boldsymbol{y}^{FEM}$ is the vector of the FE solutions, $\boldsymbol{y}^{pred}$ is the one predicted by the model, and n is the total number of data points (i.e. the number of elements in the vectors $\boldsymbol{y}^{FEM}$ and $\boldsymbol{y}^{pred}$).

The errors for the training and testing datasets are shown in Table 4.4. The obtained errors confirm the graphical interpretation. It can be noticed that the sPGD method outperforms the SVR and GB methods with testing errors that do not exceed 10% for all sensors. The errors for the testing data with the SVR method are acceptable as well, however, the errors obtained with the GB method show an overfitting despite the optimization of the hyperparameters.

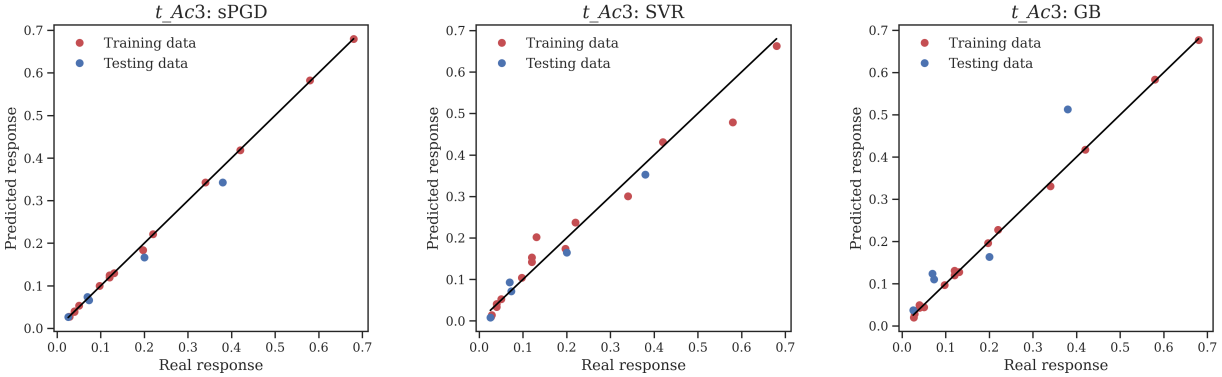
Table 4.4: Error for t_Ac1 using the three regression methods

	sPGD		SVR		GB	
Error (%)	train	test	train	test	train	test
Sensor #1	0.7	9.5	1.8	13.9	14.6	14.9
Sensor #5	0.1	4.1	2.3	2.7	8.1	31.3
Sensor #9	0.4	1.1	1.5	2.8	0.1	25.5
Sensor #3.2	0.2	3.0	6.2	8.1	8.4	17.9

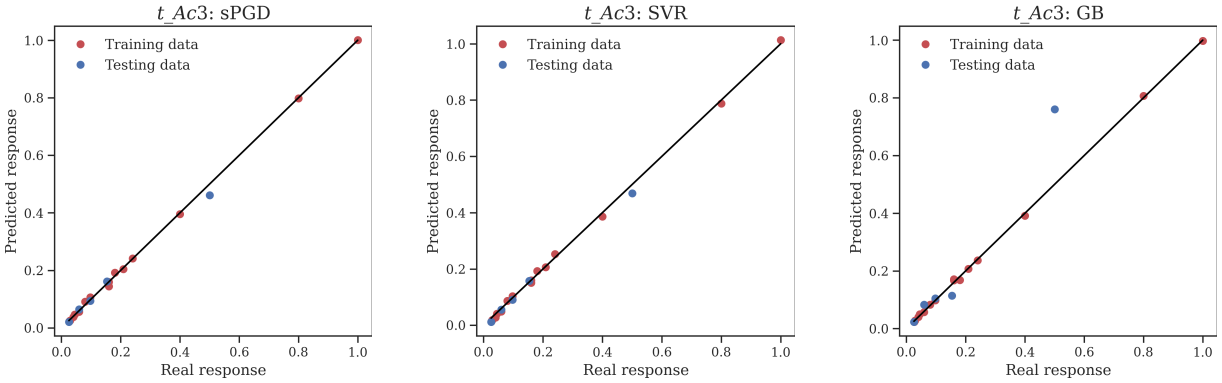
By using the same rationale, Figure 4.17 shows the FE versus the predicted values of t_Ac3 for the same sensors. A good approximation for t_Ac3 was also observed for all sensors with sPGD and SVR methods and for both training and testing points, however, less accurate prediction for testing points was observed with the GB regression method.



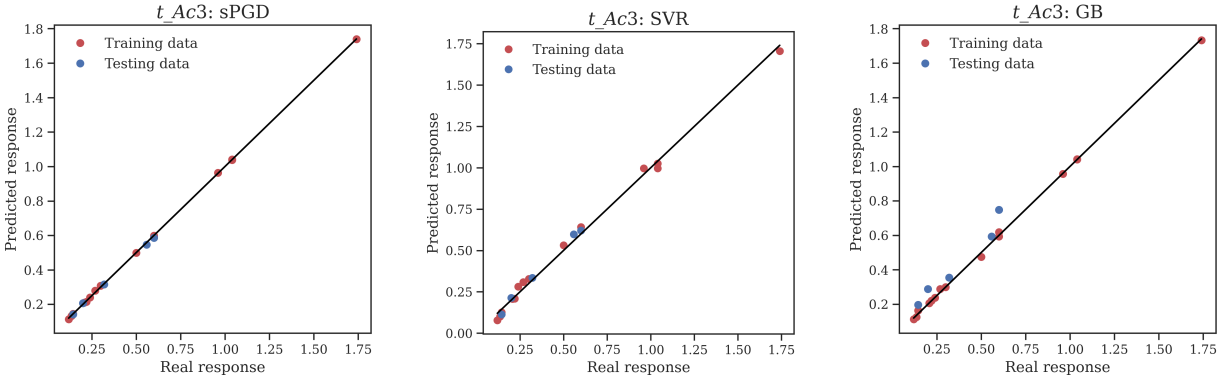
(a) Sensor #1



(b) Sensor #5



(c) Sensor #9



(d) Sensor #3.2

Figure 4.17: Reference FE versus predicted values of t_Ac3 for 4 sensors

The percentage of error for the training and the testing datasets is shown in Table 4.5. The lowest errors were observed with the sPGD method for both training and testing data and for all sensors. The obtained errors with the SVR method are less than 10% for the testing data and for the 4 sensors. For the GB method, an overfitting was obtained with very low training errors compared with the testing ones.

Table 4.5: Error for t_Ac3 using the three regression methods

	sPGD		SVR		GB	
Error (%)	train	test	train	test	train	test
Sensor #1	0.2	5.1	2.0	8.9	0	27.4
Sensor #5	0.5	6.1	5.6	6.6	0.8	19.3
Sensor #9	0.2	4.2	1.0	3.3	0.6	24.8
Sensor #3.2	0.3	1.7	2.0	5.7	26.4	18.4

In summary, models for t_Ac1 and t_Ac3 were created. They provide a complete approximation of the austenite phase transformation history as follows:

- 0% of the austenite phase is produced for all $t < t_Ac1$, which means that no phase transformation is initiated yet;
- 100% of the austenite phase is produced for all $t > t_Ac3$, which means that a complete austenitization is reached;
- the austenite phase evolves linearly between 0% and 100% for all $t_Ac1 < t < t_Ac3$.

4.6 Conclusion

Parametric metamodels for the multiphysics IH process were developed, for 15 sparse sensors located at different positions in the gear tooth, to predict the evolution of temperature and austenite phase under different values of process parameters chosen from the predefined intervals. To achieve this goal, a set of synthetic data for different values of frequency and power was first generated according to the LHS design of experiments.

For modeling the temperature field, two methodologies were proposed. The first one consists in truncating all snapshots such that only data below a constant final temperature value were considered, and the time as a function of temperature was modeled. Three regression methods (sPGD, SVR, and

GB) were then performed to create metamodels for the low-dimensional representation of the initial snapshot data provided by POD. A good prediction accuracy was obtained with the sPGD and SVR methods. However, the GB method showed overfitting and very high errors for the testing data.

The second proposed methodology consists in completing the missing values to reach a constant final time process for certain snapshots by using the GPOD method. Next, dimensionality reduction by POD coupled with the three regression methods was applied to develop models to the POD modal coefficients related to the temperature-time evolution. The proposed approach was successfully applied, and good approximations were provided for each sensor using the sPGD and the SVR methods for most of the simulations, except the simulations showing a lower level of temperature. However, the GB method showed a bad prediction accuracy.

For modeling the austenite phase, the beginning and complete austenitization time instants (t_Ac1 and t_Ac3) were extracted and modeled using the three regression methods. The best approximation was provided by sPGD, followed by SVR with a little less accurate prediction, and the worst prediction was provided by the GB method.

The obtained results indicate that the sPGD and the SVR methods can provide relatively accurate models to predict the temperature and the austenite phase evolution under the imposed values of input parameters even in the sparse low-data limit case, unlike the GB method that doesn't perform well with a reduced amount of data.

The developed approaches have the advantage of a real-time prediction that is highly suitable for active control of the process parameters. However, models for some sparse points in the space domain were built so far. Extending the solutions to address all the HAZ represents the subject of the next chapter.

Chapter 5

Full-field approximation in the space domain

Outline

5.1 Introduction 117

5.2 Spatial approximation for the temperature field 117

5.2.1 First approach: manifold learning and PODI for full-field spatial reconstruction 117

5.2.2 Second approach: GPOD for full-field spatial reconstruction 130

5.3 Spatial approximation for the austenite phase 136

5.4 Conclusion 141

5.1 Introduction

It is always interesting, from the industrial point of view, to know how the QoI evolves in the HAZ just by knowing its values at some sparse spatial sensors. Given that models for the temperature and austenite phase evolution were constructed for some sensors (illustrated in Figure 4.2), extending the approximate solutions for more spatial points in the HAZ of the gear tooth was required. This chapter proposes two different methodologies to approximate a thermal full-field from sparse data available in the space domain. Besides, a general model to predict the austenite phase in the space domain was developed as well. Results and discussions were also provided.

Note: in this chapter, the evaluation and validation of the proposed methodologies were achieved for two simulations from the existing DoE, illustrated in Figure 4.1 of chapter 4. All the results were shown for simulations 1 and 14 such that medium and high frequency values were considered.

5.2 Spatial approximation for the temperature field

This section is devoted to the interpolation of the temperature field, where two approaches were suggested and detailed in the next two sections.

5.2.1 First approach: manifold learning and PODI for full-field spatial reconstruction

This first approach enables to deal with the complex geometry of the gear by mapping the 3D cartesian coordinates of the geometry to a subspace of lower dimension where a better exploitation of data could be insured. The main advantage of the proposed methodology is its adaptability for other problems with complex geometries satisfying certain conditions. In this section, interpolation at the extreme surface of the workpiece was first studied. Then, volume interpolation was carried out at a certain depth. The proposed approach is applicable for a given configuration, i.e. given values of current frequency and power.

5.2.1.1 Surface interpolation

As shown in Figure 4.2, 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 in addition to the nonlinear geometry of the studied half-gear tooth make the interpolation difficult to achieve. To overcome these issues, a methodology based on the manifold learning by LLE of the surface mesh coordinates, coupled with

PODI was proposed to provide an interpolation much easier to handle and to allow fast prediction for new spatial coordinates. The methodology is illustrated in Figure 5.1.

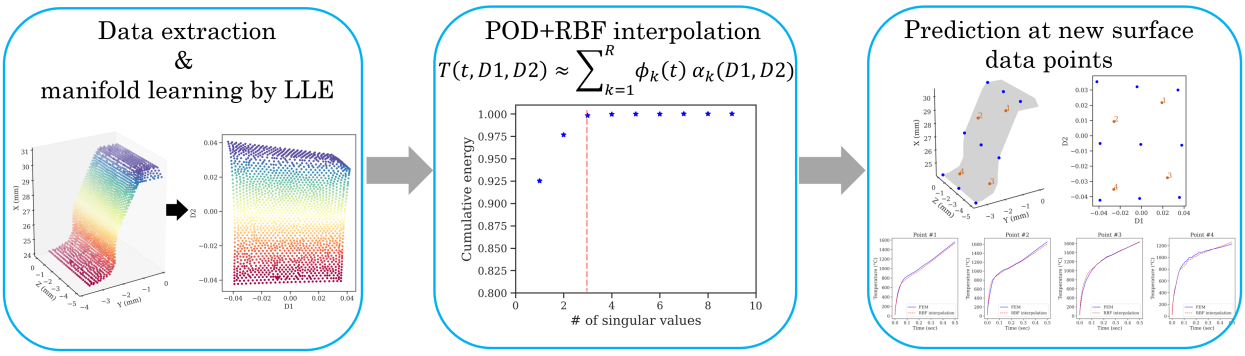
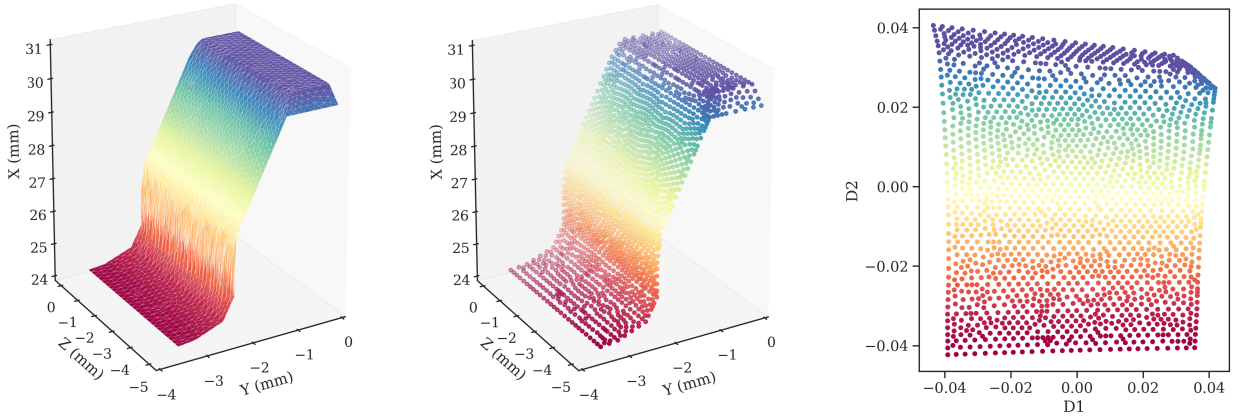


Figure 5.1: Methodology for surface interpolation

Data extraction and manifold learning by LLE

Firstly, the half-gear tooth surface was isolated from the rest of the workpiece in the FE model as shown in Figure 5.2a. The surface mesh coordinates were then extracted as shown in Figure 5.2b such that the scattered points represent the mesh nodes of the half-tooth surface (1802 points). Then, nonlinear dimensionality reduction by LLE was applied to the scattered data, as described in section 2.4.1, to map the 3D spatial coordinates into a 2D embedding manifold while preserving the neighborhood. The k nearest neighbors, set to 35, were determined by euclidean distance. This choice of k was made in accordance with the density of data points, and it was assumed to be enough to approximate each data point. Recent developments for optimally choosing k can be found in (Ghojogh et al., 2020). The result of LLE is illustrated in Figure 5.2c. It can be seen that the three-dimensional coordinates were mapped into a single global coordinate system of two dimensions, where the color coding illustrates the neighborhood-preserving mapping.

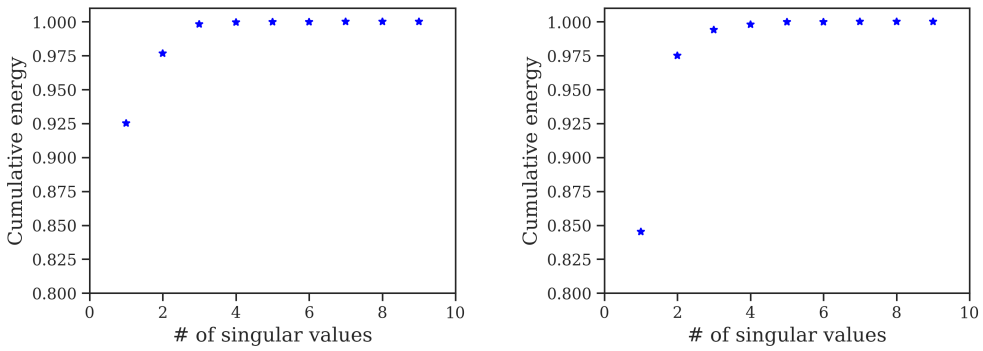


(a) (b) (c)

Figure 5.2: LLE approach: (a) isolated surface geometry (b) mesh nodes in 3D coordinates (c) embedded coordinates in the 2D manifold

PODI application

In this step, the PODI was applied on the snapshot matrix composed of the temperature-time evolution per sensor (the 9 available sensors) as columns. In practice, those snapshots should be obtained at the online stage for new selected values of input parameters (frequency and power) by using metamodels developed in chapter 4, section 4.4.2. However, simulations 1 and 14 from the initial DoE were considered here. By applying the POD on the snapshot matrix for each simulation in turn, reduced bases composed by three POD modes, for both simulations 1 and 14, were retained to recover 99% of cumulative energy and to produce accurate reconstructions, as shown in Figure 5.3.

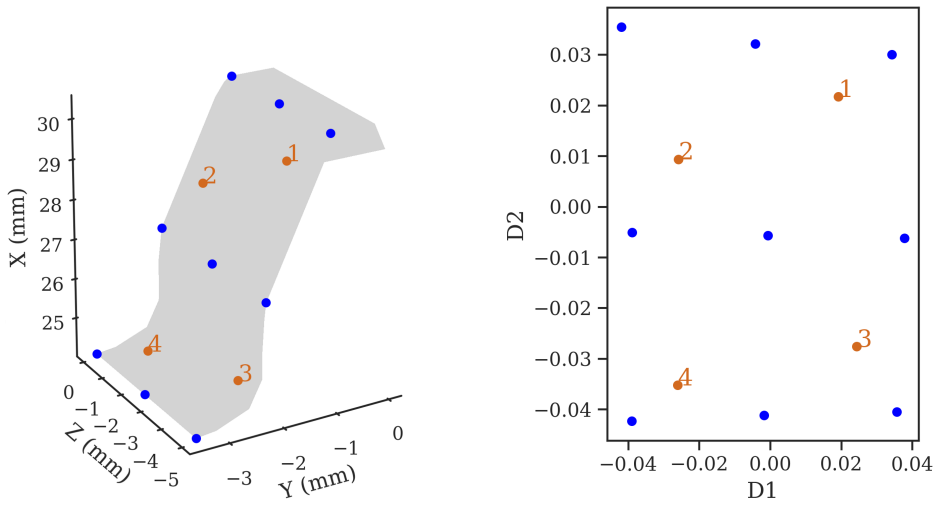


(a) Simulation 1 (b) Simulation 14

Figure 5.3: Cumulative energy computed from 9 snapshots of the thermal field for 2 different simulations

Then, the POD modal coefficients were interpolated by using RBF interpolation which provides a general and flexible way of interpolation in multidimensional spaces, even for unstructured data (Buhmann, 2000; Wendland, 2004; Jakobsson et al., 2009). 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 Figure 5.4a and 5.4b, respectively. The validation of the applied methodology was provided at 4 new positions (orange points), shown in Figure 5.4.



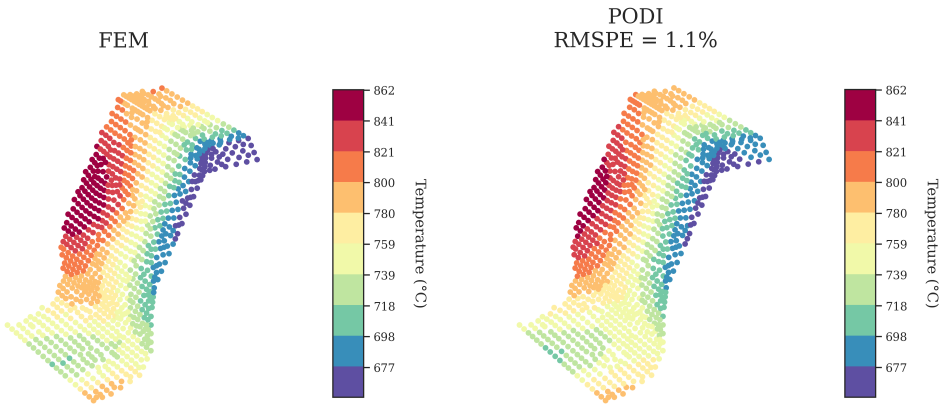
(a) (b)

Figure 5.4: 3D surface sensors (a) and their corresponding coordinates in the embedded manifold (b)

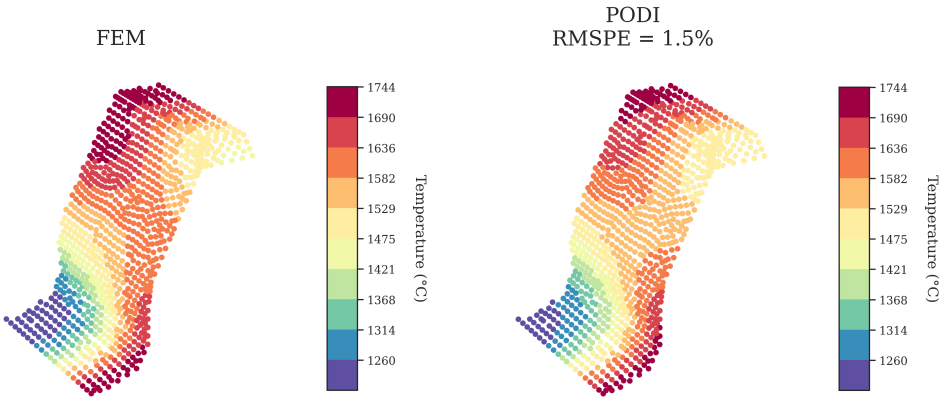
Results for surface interpolation

A comparison between the FEM and the PODI approximation of the temperature field at the last time step (t = 0.5 sec) is shown in Figure 5.5, for simulation 1 (Figure 5.5a) and simulation 14 (Figure 5.5b), and for the full gear surface. It was noticed that the results obtained with PODI show a similar trend with the FE results for both simulations. The calculated RMSPE that do not exceed 1.5% agree with this interpretation. However, this metric is an average over the whole surface results, and the approximation accuracy could vary from one zone to another. An error gap of about 20°C was observed for simulation 1 at mid-width and its surroundings. For simulation 14, the approximated results are underestimated in some zones, especially at the flank of the gear tooth. This is due to

the lack of information in the areas where no sensors are available. Therefore, predictions could be enhanced by adding more sensors.



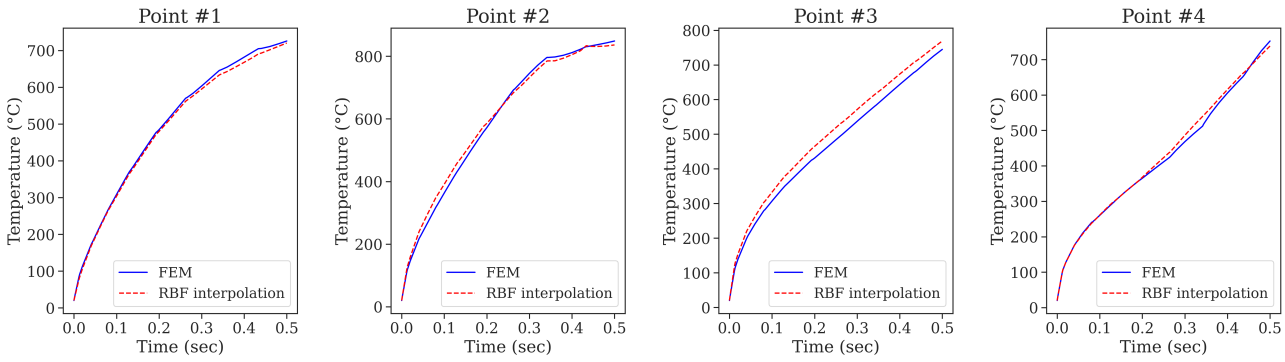
(a) Simulation 1



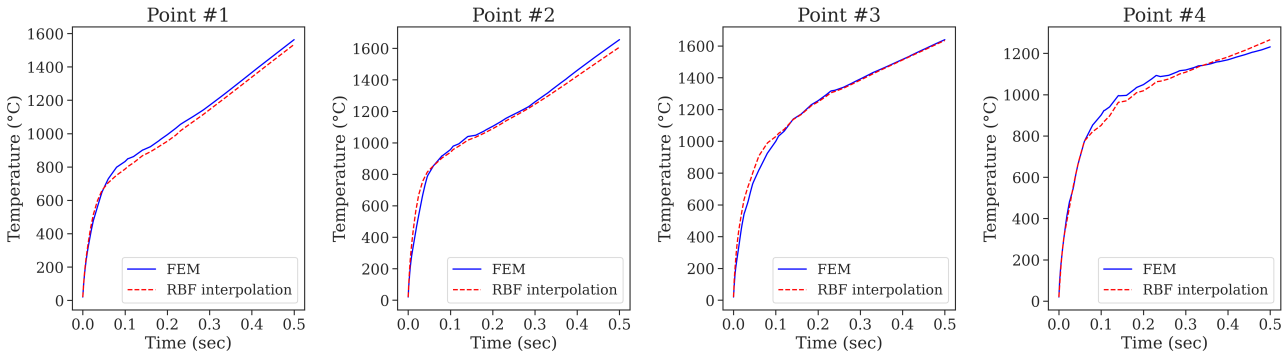
(b) Simulation 14

Figure 5.5: Comparison between the FEM and the PODI approximation of the surface temperature distribution at t = 0.5 sec

In addition, a comparison between the FEM and the PODI approximation of the temperature-time evolution is shown in Figure 5.6, for simulation 1 (Figure 5.6a) and simulation 14 (Figure 5.6b), and for the test points illustrated in Figure 5.4. It can be seen that the two curves are very close for all test points and almost overlapped for many of them.



(a) Simulation 1



(b) Simulation 14

Figure 5.6: Comparison between the FEM and the PODI approximation of the temperature-time evolution at 4 test positions and for simulations 1 and 14

In addition, the relative error defined by Eq.(5.1) was used to determine the estimation accuracy for the test points and for the two simulations.

$$Error(\%) = \frac{\sqrt{\int (\boldsymbol{T}^{FEM} - \boldsymbol{T}^{pred})^2 dt}}{\sqrt{\int (\boldsymbol{T}^{FEM})^2 dt}} \times 100 \tag{5.1}$$

As shown in Table 5.1, the obtained relative errors were less than 6% at the 4 test points. These results indicate that the applied methodology for surface interpolation was robust and provided accurate results.

Table 5.1: 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	2.5	5.8	2.3
Simulation 14	2.8	3.0	2.6	2.3

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 affects the accuracy of the interpolation and a good choice of the position of sensors is highly recommended.

5.2.1.2 Volume interpolation

According to Figure 4.2, 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 study a more general case, where interpolation in internal parts and for different zones of the gear tooth can be carried out. The proposed methodology is also applicable to the already existing sensors, however, interpolation is restricted to the spatial region limited by the available sensors.

The proposed methodology is composed of several steps as described in Figure 5.7

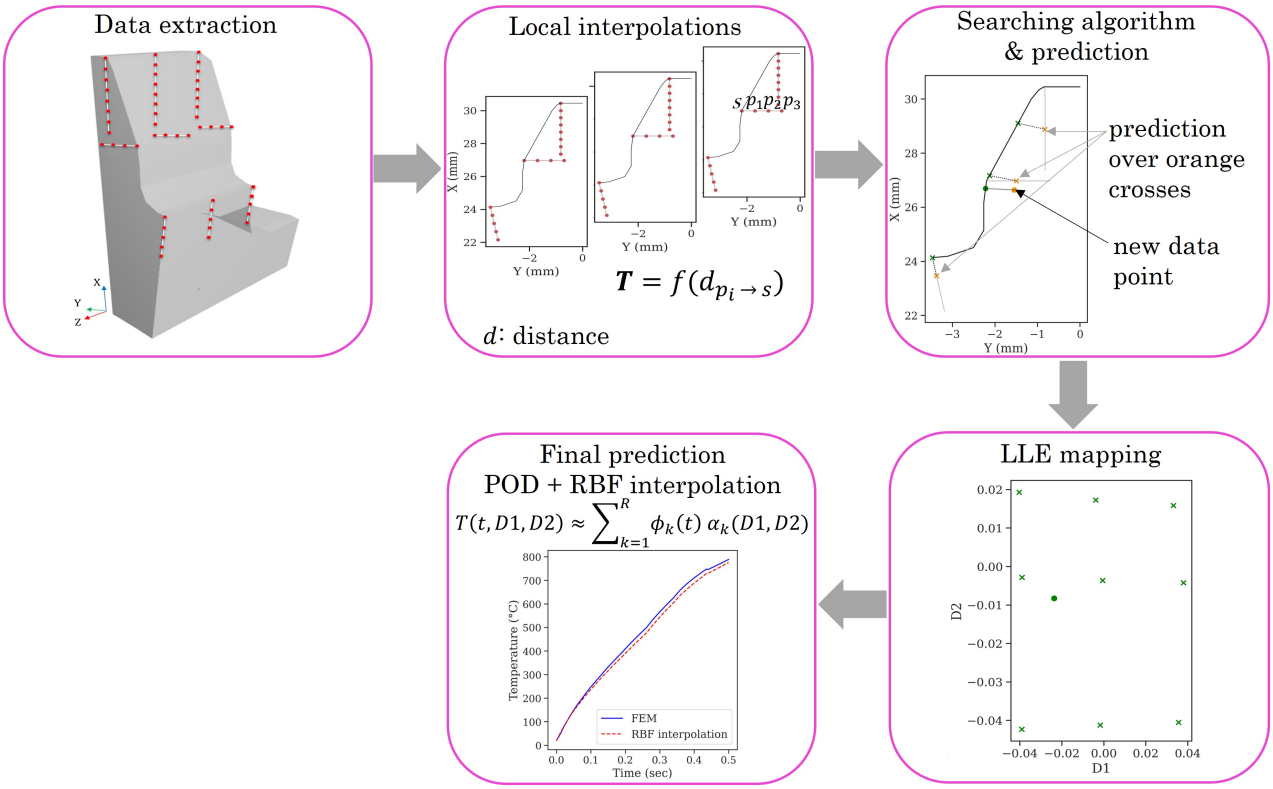


Figure 5.7: Methodology for volume interpolation

Data generation

We start by collecting data in sensors by a post-processing on the simulation result files such that no new complete computations are required. The considered sensors are illustrated in Figure 5.8 with red points such that runs of few minutes were achieved to extract the results of temperature for the set of sensors (passing through the same grey line) at a time.

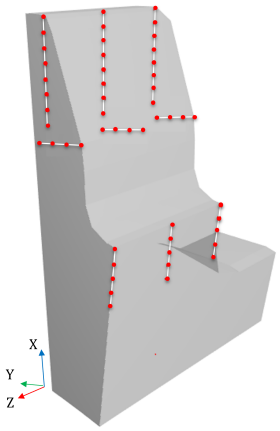


Figure 5.8: Position of sensors for volume interpolation

Local interpolations

Since the solution is currently known for all sensors in Figure 5.8, an interpolation function for the temperature 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 RBF 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 Figure 5.8). This step was tested for 2 new points illustrated in Figure 5.9.

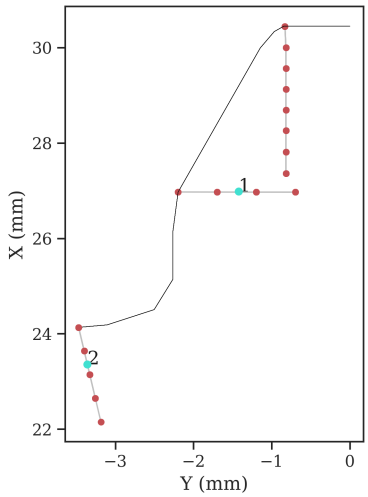
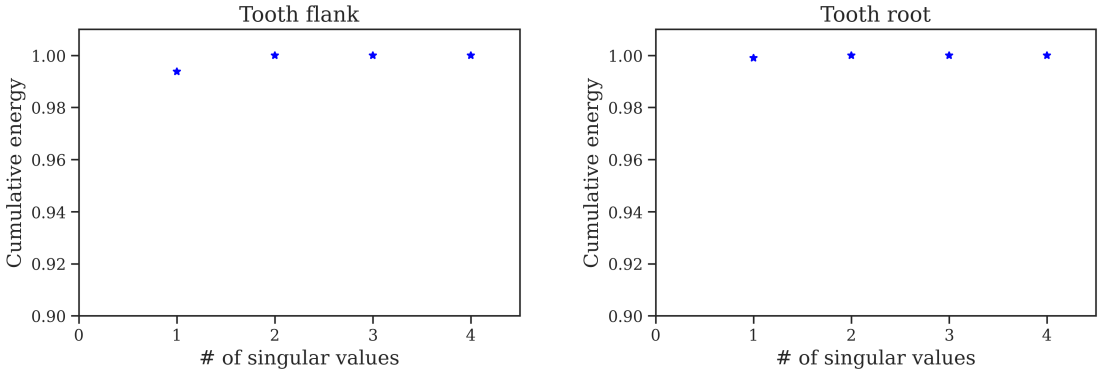


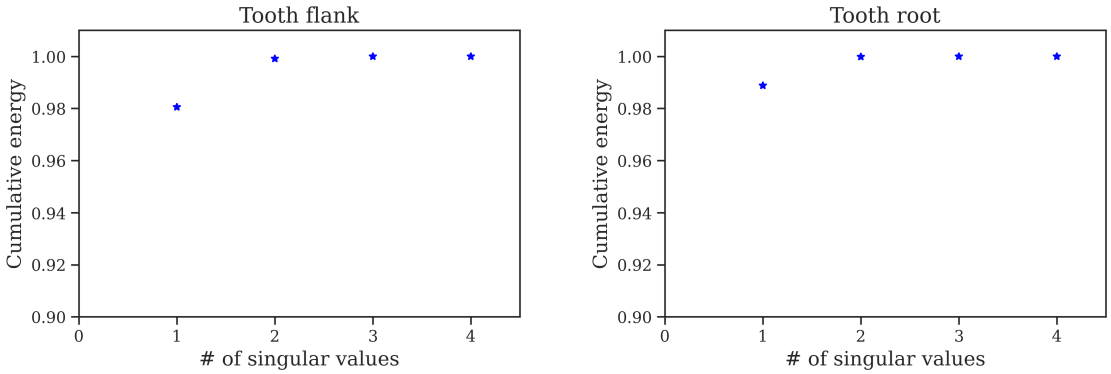
Figure 5.9: Test points for local interpolation at mid-width (Z = 0)

By applying the POD on the snapshot matrices, the first 2 singular vectors were retained to recover

more than 99.9% of the cumulative energy, as shown in Figure 5.10 for the 4 snapshots in the tooth root and flank and for two simulations 1 and 14.



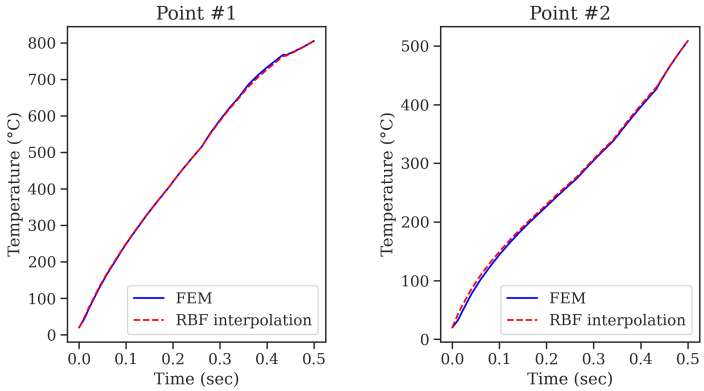
(a) Simulation 1



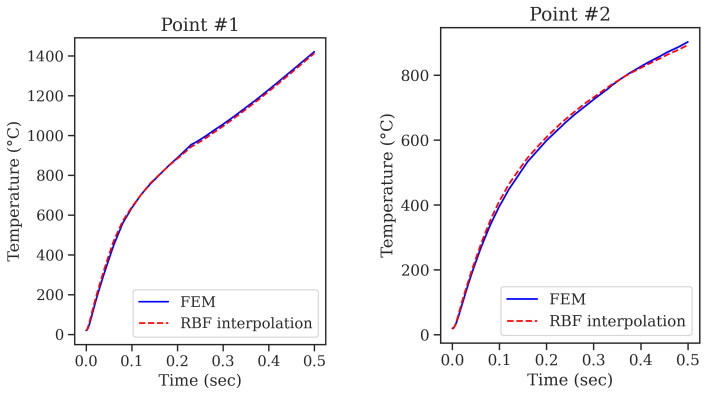
(b) Simulation 14

Figure 5.10: Cumulative energy computed from 4 snapshots of the thermal field for 2 different simulations

A comparison between the FEM and the PODI approximation of the temperature-time evolution is shown in Figure 5.11, for simulation 1 and simulation 14, and for the test points illustrated in Figure 5.9. It can be seen that the curves are perfectly overlapped for the 2 test points and both simulations.



(a) Simulation 1



(b) Simulation 14

Figure 5.11: Comparison between the FEM and the linear RBF interpolation of the temperature-time evolution at 2 test positions and for simulations 1 and 14

The relative error was calculated and listed in Table 5.2. The obtained errors for both test points do not exceed 1.5% for both simulations 1 and 14.

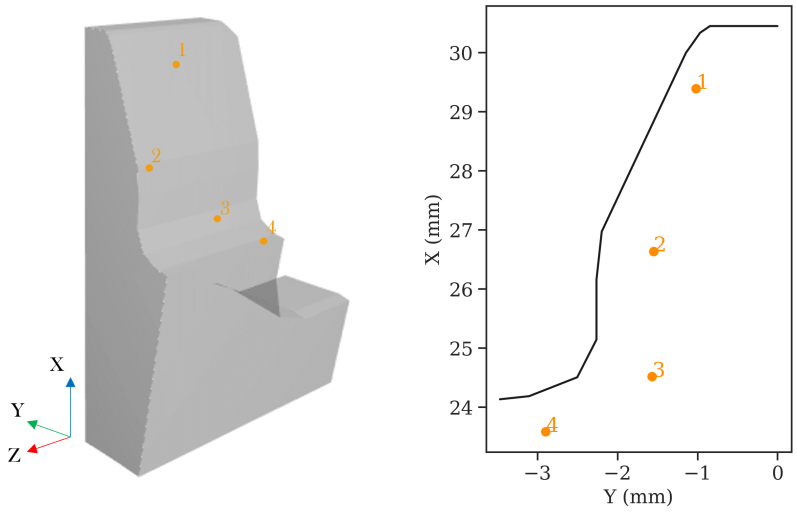
Table 5.2: Relative errors of the temperature-time evolution using POD with linear RBF interpolation

Error (%)	point #1	point #2
Simulation 1	0.5	1.4
Simulation 14	1.0	1.4

Equivalent points detection by searching algorithm

Now, 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, in order to evaluate the performance of the methodology, 4 test points located at certain depths and at different Z coordinates were

considered, as shown in Figure 5.12a, where Figure 5.12b shows the projection of the test points onto the plane Z = 0 for better visualization.



(a) (b)

Figure 5.12: Test points (a) in the 3D coordinate system (b) projected onto the plane Z = 0

For each test point, its associated points in the 9 gray lines in Figure 5.8 satisfying the equality of the minimum distance from the surface were determined. The research of these points is based on an algorithm inspired by the Dichotomic search as illustrated in Figure 5.13. For the sake of clarity, the illustration is made for one segment, but the same steps were applied to the other ones.

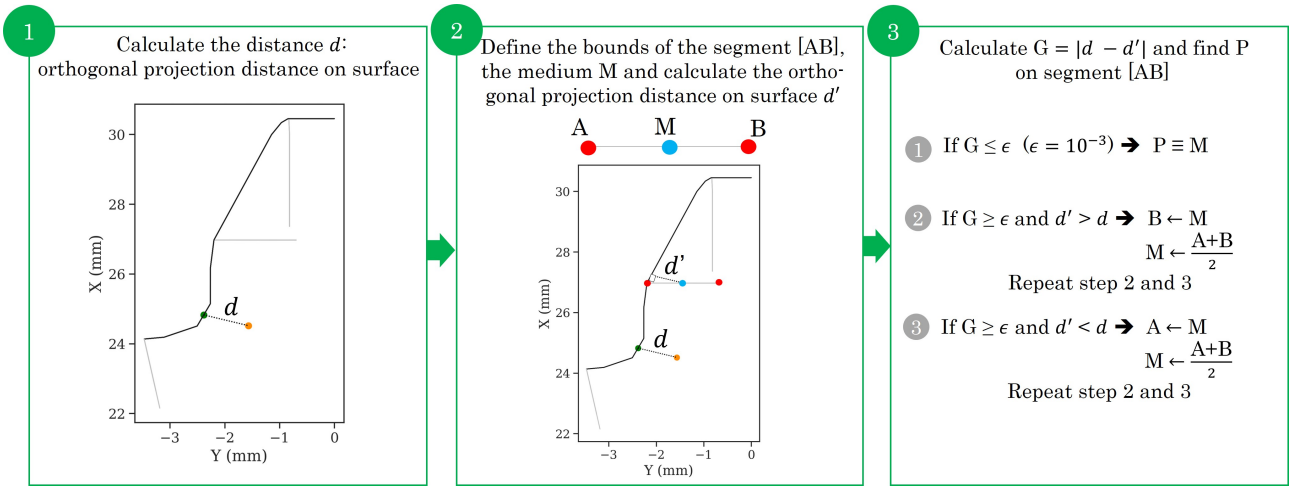


Figure 5.13: Illustration of the searching method

Figure 5.14 shows the results of the applied search algorithm for the 4 test points and at the mid-width of the tooth. This is 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.

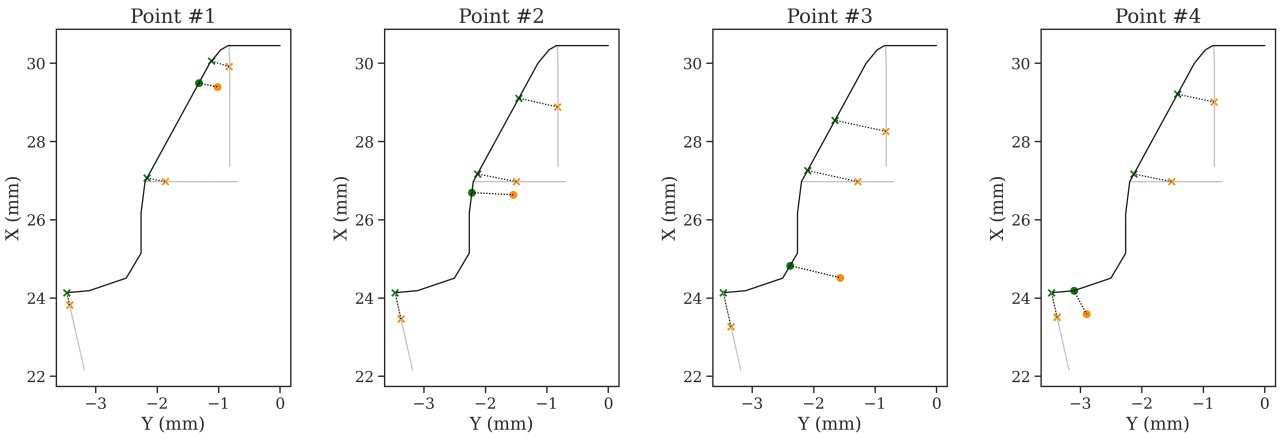


Figure 5.14: Test points and their associated ones on the lines containing the available sensors

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 crosses on the gray lines (Figure 5.14) 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 evolution on the orange crosses was computed by using the previously created interpolants.

Mapping by LLE

The coordinates of the orthogonal projection of all orange crosses in Figure 5.14 onto the surface, represented by the green crosses, were extracted. A total of 9 points coordinates were provided for each test point. Those 3D surface coordinates were then mapped into the 2D embedding manifold previously generated by LLE for surface interpolation, as shown in section 5.2.1.1. The representation of the mapped crosses for the 4 test points is illustrated in Figure 5.15, where the projection of the test points on the surface was also mapped, and represented in the 2D space by green points.

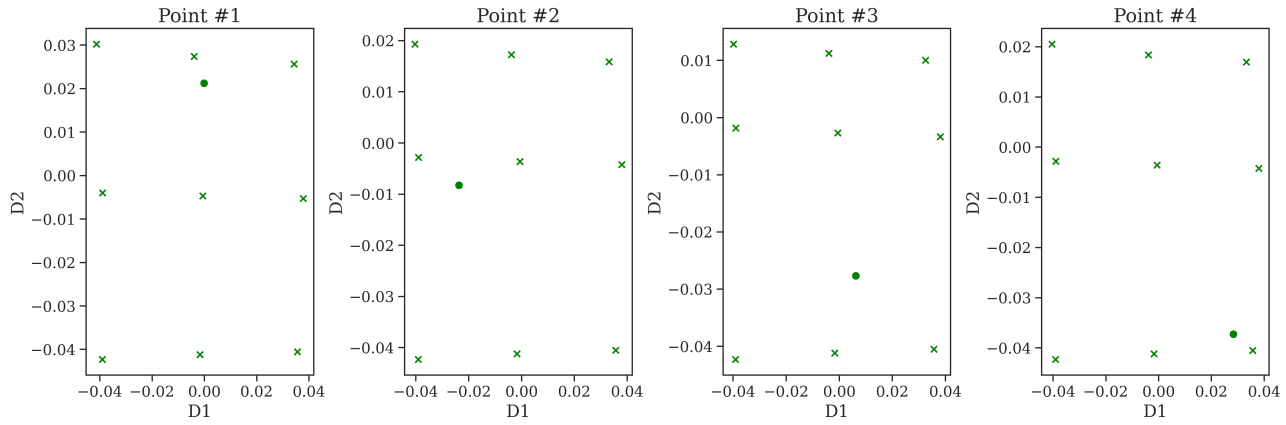


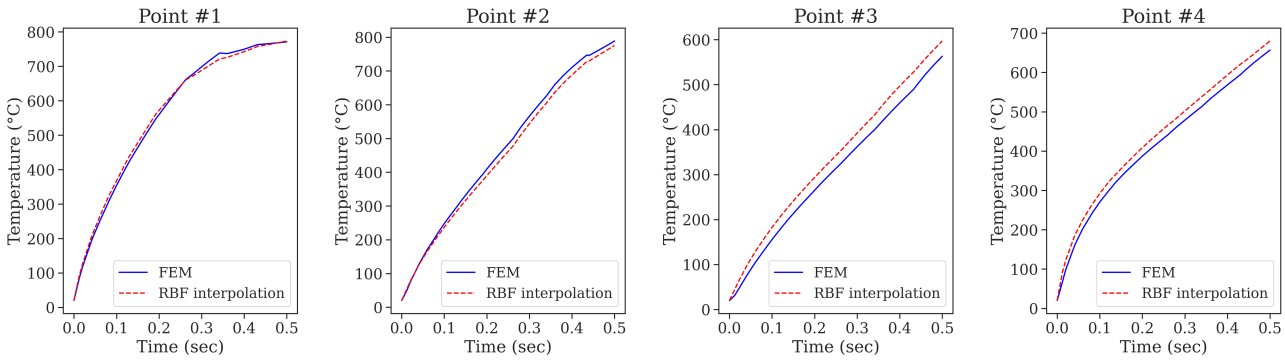
Figure 5.15: Embedded coordinates in the 2D manifold

PODI application

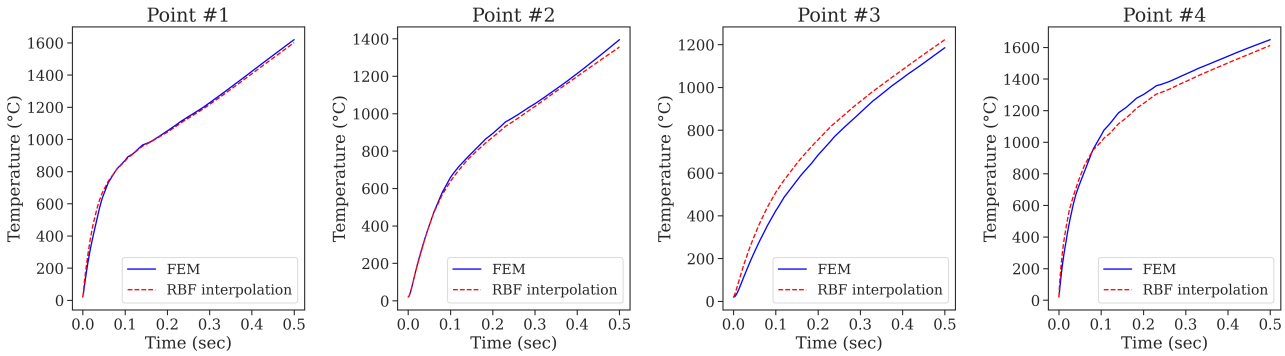
Following the same rationale as the surface interpolation, the PODI was applied on the snapshot matrices containing, as columns, the just predicted temperature-time evolution at the orange cross points in Figure 5.14 for each test point, where the applied strategy was again tested for simulation 1 and 14 from the DoE. Similarly to the surface interpolation, the applied POD on the snapshot matrices yielded a reduced basis having three POD modes, for both simulations and for each test point, to recover more than 99% of cumulative energy. Then, the POD modal coefficients were interpolated by using RBF interpolation with a multiquadric basis function, as a function of the coordinates in the 2D space obtained by LLE.

Results for volume interpolation

After reconstruction, a comparison between the FEM and the PODI approximation of the temperature-time evolution is shown in Figure 5.16, for simulation 1 (Figure 5.16a) and simulation 14 (Figure 5.16b) and for the 4 test points. It can be seen that all the curves show the same trend and a good correlation was obtained. However, a gap between curves was observed for points #3 and #4. This could be explained by the high thermal gradients in the curvature zone of the gear tooth and the position of the green points with respect to the green crosses.



(a) Simulation 1



(b) Simulation 14

Figure 5.16: Comparison between the FEM and the PODI approximation of the temperature-time evolution at 4 test points and for simulations 1 and 14

The relative error defined by Eq.(5.1) was also calculated to determine the estimation accuracy. The obtained errors, listed in Table 5.3, do not exceed 10% for both simulations. These results prove the efficiency of the applied strategy for volume interpolation.

Table 5.3: 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	1.7	3.3	8.8	5.1
Simulation 14	1.7	1.8	7.6	3.8

5.2.2 Second approach: GPOD for full-field spatial reconstruction

5.2.2.1 Data extraction

There are mainly two ways to extract data from the simulation: the first one, used so far, consists in defining some sensor positions on the workpiece by pre-processing during data setup before the simulation execution or by post-processing on the simulation results. Therefore, file results could be

generated containing the history of all possible output quantities (temperature, displacement, metallurgical phase rates, heating power, etc.) for each imposed sensor. The main advantage consists in providing a file per sensor, showing the time evolution of all required QoI simultaneously. However, this strategy becomes ineffective when many sensors (≥ 30) are addressed. This is because the significant increase in execution and storage costs.

The second way enables the extraction of data at all mesh nodes and for all physical quantities related to the IH process but at a single time instant. In other words, if a simulation is run at 100-time steps, data can be exported manually step by step in time by the user, which is very time consuming. Therefore, depending on the case of study, the user can choose the first or the second one.

In this section, we used the second method since a full-field approximation using only some known data at some sparse sensors. To achieve this goal, the temperature field was extracted at all spatial points for most simulations from the DoE, but for only 5 particular time instants; from 0 to 0.5 sec with a time step of 0.1 sec. For simulations that do not reach the final time process, data were extracted until the closest value to 0.5 sec.

5.2.2.2 GPOD application

The GPOD with a completely known snapshot set was applied to reconstruct missing values for new simulation only by knowing the solution at some sensors. Firstly, the extracted data were treated to construct a snapshot matrix composed of temperature values. The POD was then performed on the snapshot matrix $\boldsymbol{M} = [\boldsymbol{T}_1\boldsymbol{T}_2\cdots\boldsymbol{T}_N]$ with a total of 105 snapshots such that $\boldsymbol{T}_i = \boldsymbol{T}(F_i, P_i, t_i)$ for $i = 1, \cdots, N$, and with $\boldsymbol{T}, F_i, P_i, t_i$ are the vector of temperature values at every spatial node, the frequency, the power, and the time for the snapshot i, respectively. As previously stated, the average over the snapshot set was first computed and the average 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. By applying the POD, the left singular vectors of the snapshot matrix were truncated to the 6 first singular vectors such that more than 99.9% of the cumulative energy was recovered, as shown in Figure 5.17.

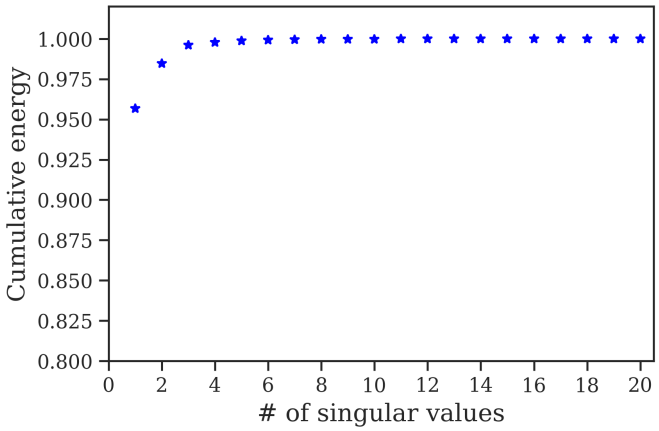
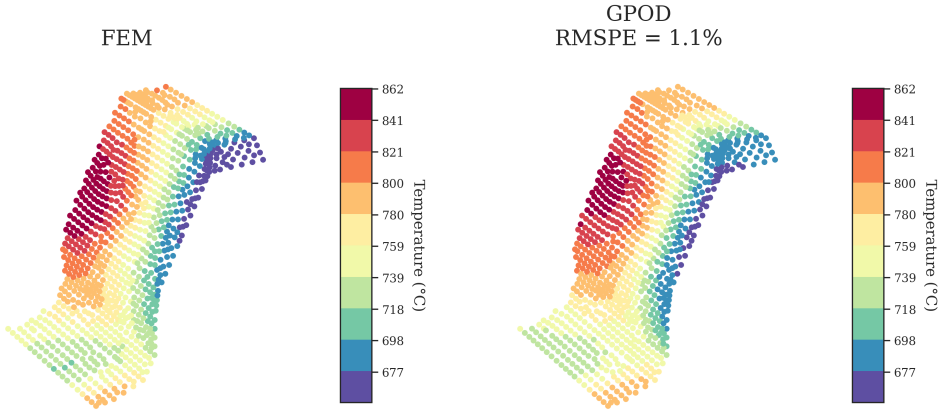


Figure 5.17: Cumulative energy of the thermal field computed from snapshots at different input parameters and time steps

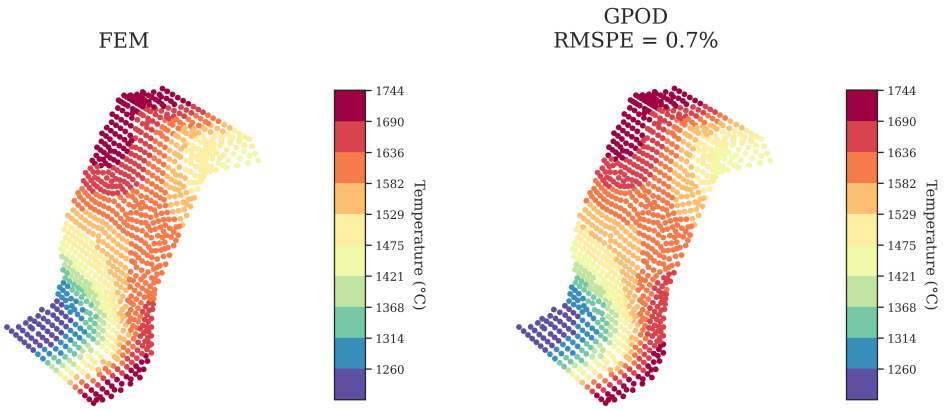
The constructed POD basis can be used to reconstruct a new vector with missing elements as shown in section 2.2.1.3 of chapter 2. The main idea of this approach is to be able to approximate the temperature at any point for a given value of frequency, power, and time instant only by knowing the temperature at sensors illustrated in Figure 4.2, where metamodels had already been developed. Therefore, for a new gappy vector, we defined its corresponding mask vector that gives 1 to vector elements corresponding to the available sensor positions and 0 elsewhere. Then, by using the 6 POD basis vectors, we computed the POD coefficients by minimizing the error between the original and repaired vectors, computed over the existing data elements.

5.2.2.3 Results and discussion

As for section 5.2.1, the approach was tested on simulation 1 and simulation 14, and the evaluation of the method for some particular surface and volume points was achieved as well. A comparison between the FEM and the GPOD approximation of the temperature field at the last time step (t = 0.5 sec) and for the full gear surface is first shown in Figure 5.18, for simulation 1 (Figure 5.18a) and simulation 14 (Figure 5.18b). It was noticed that the approximated results obtained with GPOD are in good agreement with the FE results for both simulations, especially for simulation 14 where no significant difference was observed and an RMSPE of 0.7% was obtained. For simulation 1, an RMSPE of 1.1% was obtained, however, an error gap of at most 20°C was observed at the flank of the gear tooth and the tip of the gear edge.



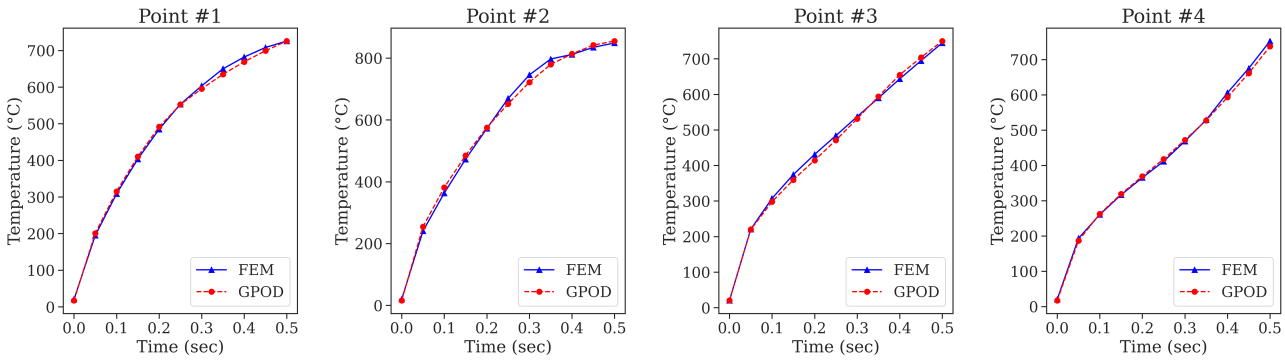
(a) Simulation 1



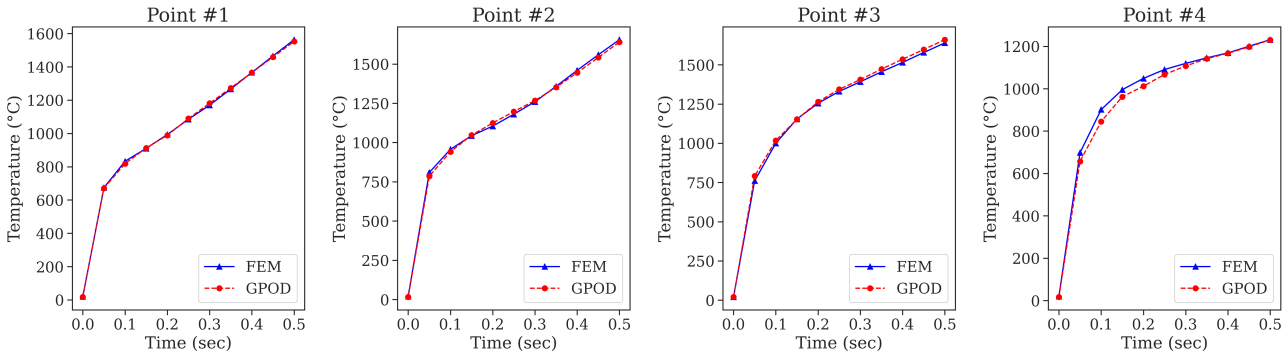
(b) Simulation 14

Figure 5.18: Comparison between the FEM and the GPOD approximation of the surface temperature distribution at t = 0.5 sec

Then, the approach was also evaluated for the surface points, illustrated in Figure 5.4a. A comparison between the FEM and the GPOD approximation of the temperature-time evolution is shown in Figure 5.19a for simulation 1 and Figure 5.19b for simulation 14. The temperature was evaluated at 11-time instants shown by markers. It can be seen that the two curves obtained by the FEM and the GPOD method are very close, even overlapped for some test points.



(a) Simulation 1



(b) Simulation 14

Figure 5.19: Comparison between the FEM and the GPOD approximation of the temperature-time evolution at 4 test positions and for simulations 1 and 14 - surface points

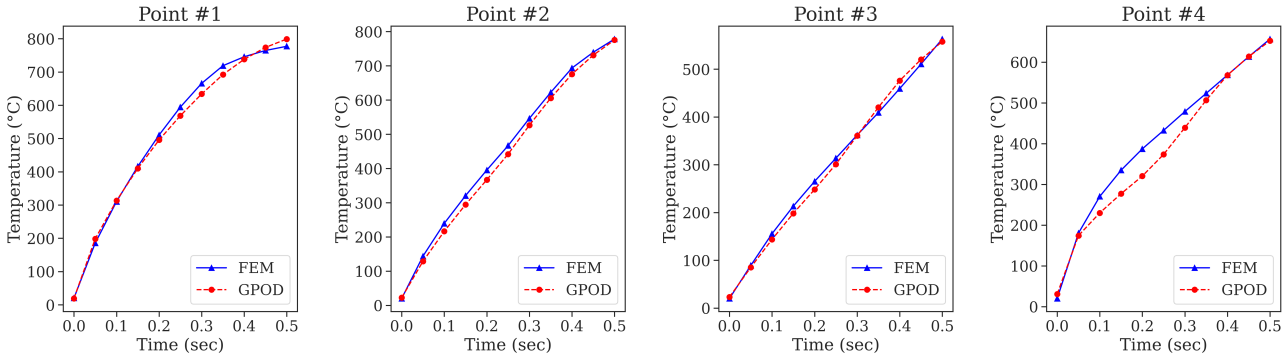
The relative error defined by Eq.(5.1) was calculated to determine the estimation accuracy. The obtained errors, listed in Table 5.4, do not exceed 3% for both simulations, which are very satisfactory results.

Table 5.4: Relative errors of the temperature-time evolution using GPOD approximation - surface points

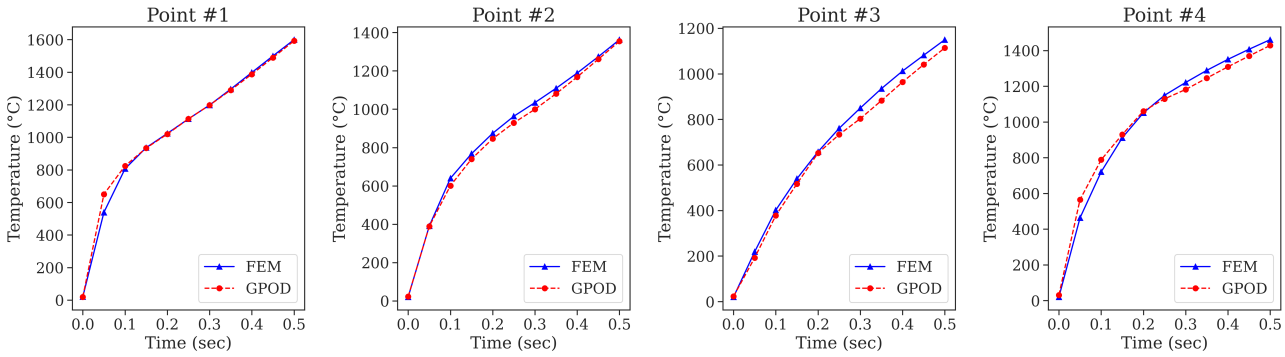
Error (%)	point #1	point #2	point #3	point #4
Simulation 1	1.5	2.1	2.0	1.6
Simulation 14	0.6	1.2	1.3	2.7

The volume points for validation are illustrated in Figure 5.12. A comparison between the FEM and the GPOD approximation of the temperature-time evolution is shown in Figure 5.20 for simulation 1 (Figure 5.20a) and simulation 14 (Figure 5.20b). It can be noticed that the trends are observed at the 11-time instants and the approximated curves are in good agreement with the FE ones for all test points, except point #4 of simulation 1. These results prove the efficiency of the reduced basis to

describe the behavior of the full system as well as to show that having results at few sensors (about 0.1% of the total number of mesh nodes) enables the reconstruction of the full field with sufficient accuracy. Therefore, having data for more sensors could increase significantly the approximation accuracy.



(a) Simulation 1



(b) Simulation 14

Figure 5.20: Comparison between the FEM and the GPOD approximation of the temperature-time evolution at 4 test positions and for simulations 1 and 14 - volume points

The relative error was also calculated and the obtained errors, listed in Table 5.5, show the effectiveness of the GPOD method to estimate the full field. The relative errors do not exceed 9% for simulation 1 and 5% for simulation 14.

Table 5.5: Relative errors of the temperature-time evolution using GPOD approximation - volume points

Error (%)	point #1	point #2	point #3	point #4
Simulation 1	3.1	3.9	3.3	8.6
Simulation 14	3.1	2.7	4.5	4.3

The GPOD, used for approximating the temperature evolution at any spatial point and for some time

instants, provides good results for surface and volume points as well.

5.3 Spatial approximation for the austenite phase

In this section, a generalization of the developed parametric metamodels of the austenite phase was proposed. In section 4.5 of chapter 4, models for the characteristic time instants t_Ac1 and t_Ac3 were proposed locally at some sparse sensors as a function of the inverse of frequency and power. Here, the space domain was addressed in addition to process parameters, and a more general approximation was provided. Regression models for the characteristic time instants t_Ac1 and t_Ac3 were developed such that the vector of input parameters is defined as follows:

$$\boldsymbol{\mu} = \{\frac{1}{F}, \frac{1}{P}, D1, D2, \delta\} \tag{5.2}$$

such that $F, P, D1, D2$, and δ are the frequency, the power, the two coordinates in LLE subspace, and the depth of the considered spatial points in the sense of its orthogonal projection to the surface, respectively. By applying the three regression methods; sPGD, SVR, and GB on the training set (225 training points randomly chosen from a total of 300 (i.e. 75% of data for training models)) and by optimizing hyperparameters by using the GridSearchCV tool in the Scikit-learn library, we got the results illustrated in Figure 5.21 and Figure 5.22 for t_Ac1 and t_Ac3, respectively.

The two Figures show the real versus the predicted values of t_Ac1 and t_Ac3 using sPGD, SVR, and GB regressions. The red points correspond to the data used to build the model while the blue ones correspond to the data used to evaluate its accuracy. We remind that when points are close to the black line, the model provides a good fit to data. Indeed, regarding the dispersion of data with respect to the black line, both t_Ac1 and t_Ac3 models are accurate. For t_Ac1, no significant difference is observed between the three methods. For t_Ac3, the sPGD model seems to be the most accurate, comparatively.

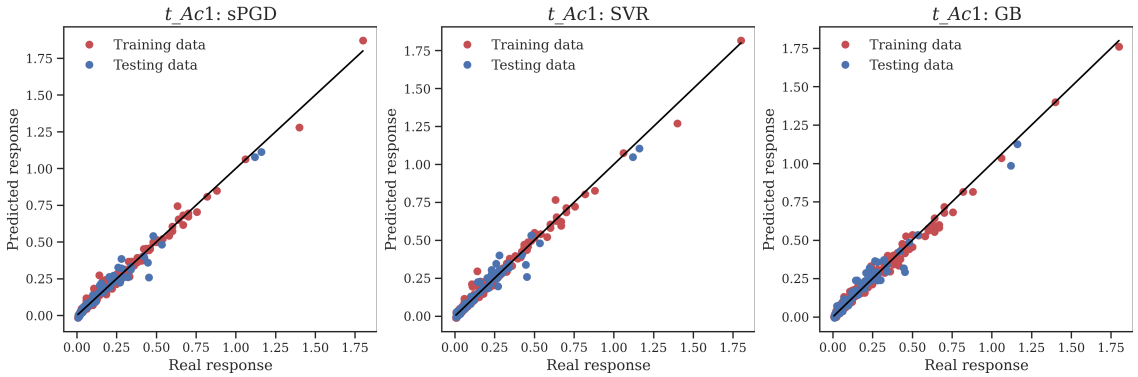


Figure 5.21: Reference FE versus predicted values of t_Ac1

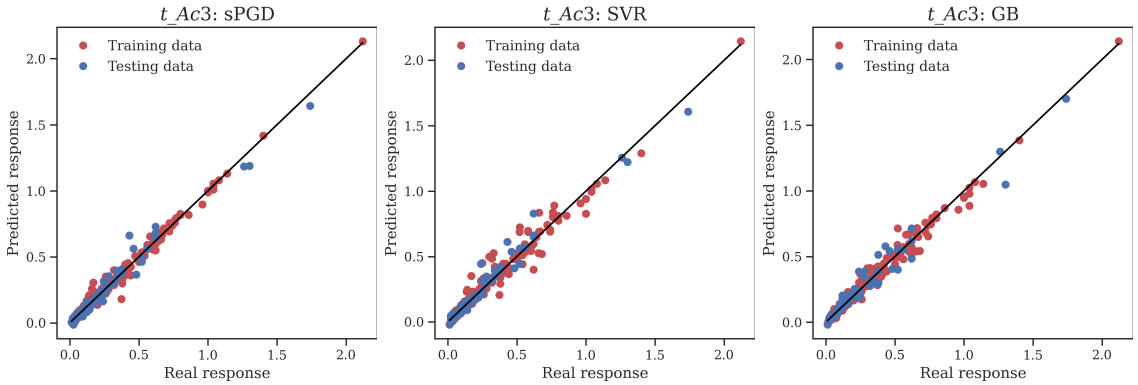


Figure 5.22: Reference FE versus predicted values of t_Ac3

The relative error to measure the prediction accuracy for the training and the testing datasets is presented in Table 5.6, where the computed error is defined by:

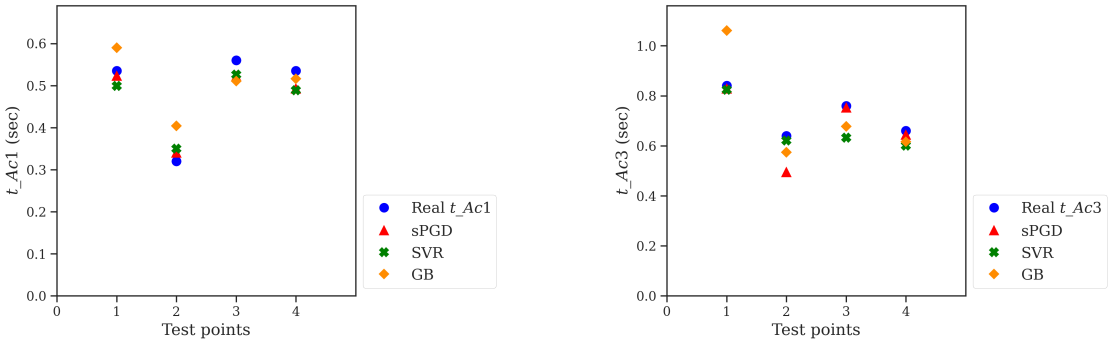
$$Error(\%) = \left(\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{pred} - y_i^{FEM})^2}}{max(\boldsymbol{y}^{FEM}) - min(\boldsymbol{y}^{FEM})} \right) \times 100 \tag{5.3}$$

where n, $\boldsymbol{y}^{FEM}$, $\boldsymbol{y}^{pred}$, y_i^{FEM}, and y_i^{pred} are the number of data points, the vector of the FE solutions, the vector of the predicted solutions, the ith element of vector $\boldsymbol{y}^{FEM}$ and the ith element of vector $\boldsymbol{y}^{pred}$, respectively. As shown in Table 5.6, low errors that do not exceed 5% were obtained for both t_Ac1 and t_Ac3 and with the three regression methods. Globally, no significant difference was observed between the three methods. But still to check their performances on new particular test points.

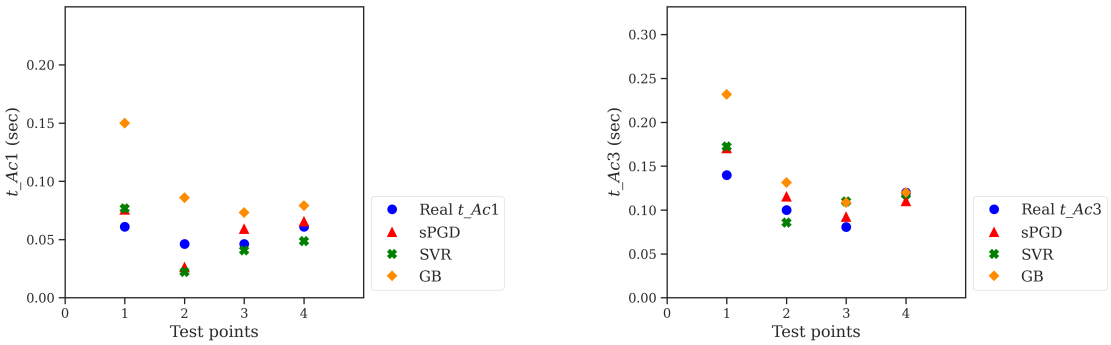
Table 5.6: Error for t_Ac1 and t_Ac3 using three regression methods

	sPGD		SVR		GB	
Error (%)	train	test	train	test	train	test
t_Ac1	1.3	3.3	1.4	3.6	1.4	3.9
t_Ac3	1.4	2.8	2.5	3.0	1.8	3.4

Therefore, to validate the predictive models for new positions, t_Ac1 and t_Ac3 were predicted for surface and volume points. The surface ones are illustrated in Figure 5.4a and a comparison between the FE and the predicted values is shown in Figure 5.23 for simulation 1 and simulation 14 by using the three regression methods. The observed results are comparable, however, a better prediction was obtained by sPGD followed by SVR and GB, respectively, for t_Ac1 and t_Ac3 and for both simulations.



(a) Simulation 1



(b) Simulation 14

Figure 5.23: Comparison between the FE and the predicted values of t_Ac1 and t_Ac3 at 4 surface test points and for simulations 1 and 14

The relative error was also calculated for each test point. However, the most popular measures of error such as the MAPE and the RMSPE are not particularly suitable for zero or close-to-zero true values, which is the case for some values of t_Ac1 and t_Ac3. This is because of the division by the true values

in MAPE and RMSPE expressions. To overcome such issue, several measures of prediction accuracy were developed such as symmetric mean absolute percentage error (SMAPE) (Makridakis, 1993), mean absolute scaled error (MASE) (Hyndman and Koehler, 2006), and mean arctangent absolute percentage error (MAAPE) (Kim and Kim, 2016). In this work, MAAPE was used and it is defined as:

$$MAAPE = \frac{1}{n}\sum_{i=1}^{n} arctan\left(\left|\frac{y_i^{pred} - y_i^{FEM}}{y_i^{FEM}}\right|\right) \times 100 \tag{5.4}$$

The obtained errors, listed in Table 5.7, are different from one method to another. Although predictions seem to be quite good in Figure 5.23, the calculated errors are not sufficiently good and largely exceed 10% for many cases. It is worth pointing out that a prediction is considered good enough when the errors do not exceed 10%. Such results can have different origins, but the most probable one is related to the choice of the metric. It can also be noticed that the errors for simulation 14 are much higher than the errors for simulation 1. This can be due to the very low surface values of t_Ac1 and t_Ac3 for simulation 14.

Table 5.7: Error for t_Ac1 and t_Ac3 using three regression methods - surface test points

Error (%)		point #1		point #2		point #3		point #4	
		t_Ac1	t_Ac3	t_Ac1	t_Ac3	t_Ac1	t_Ac3	t_Ac1	t_Ac3
Simulation 1	sPGD	2.3	1.4	6.3	22.4	6.8	1.0	7.9	2.4
	SVR	6.6	1.8	9.3	2.9	6.0	16.6	8.5	8.9
	GB	10.2	25.6	25.6	10.2	8.7	10.7	3.5	6.6
Simulation 14	sPGD	23.3	21.4	40.9	15.1	26.7	14.1	7.1	8.2
	SVR	25.0	22.7	47.9	14.2	11.8	34.7	20.1	1.8
	GB	96.9	58.0	70.8	30.3	52.6	33.3	28.7	0.03

The volume points used for validation are not the same as used previously. This is because only sensors in Figure 4.2 were considered in this study to construct the model. Hence, a total of 6 volume sensors, in the tooth root and flank, were considered and prediction for new points in the zone limited by these sensors was carried out. The volume test sensors are illustrated in Figure 5.24.

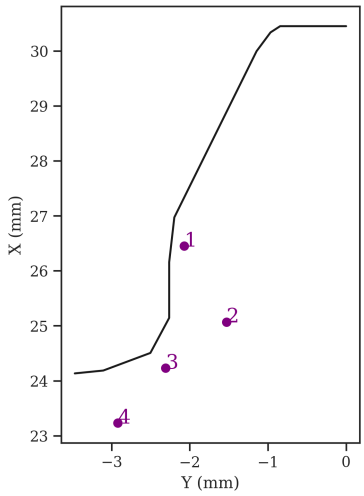
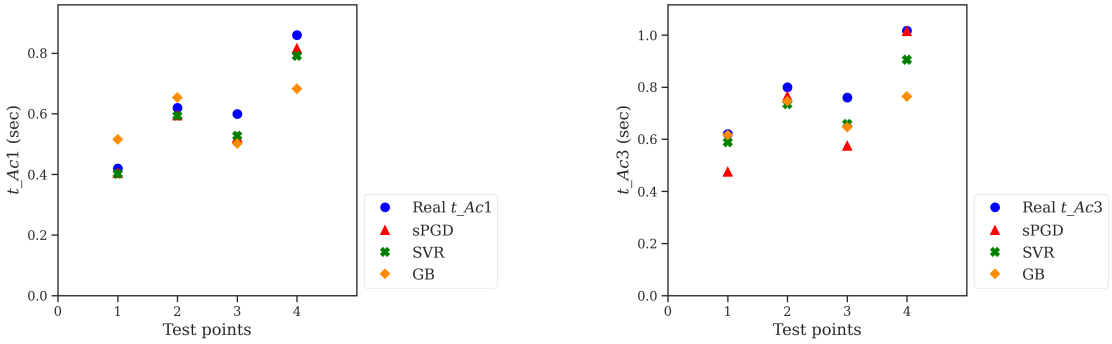
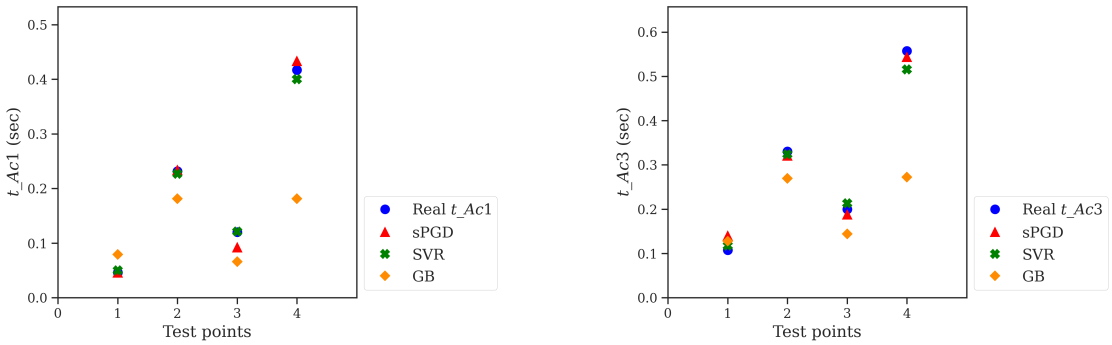


Figure 5.24: Volume test points

A comparison between the FE and the predicted values is shown in Figure 5.25 for simulation 1 and simulation 14 by using the three regression methods. Again, the observed results are comparable, however, the worst prediction is observed with the GB method for t_Ac1 and t_Ac3. The relative error was also calculated using the MAAPE metric. The obtained errors, listed in Table 5.8, are better than those obtained for the surface points prediction. The obtained errors are in good agreement with the visual interpretation from Figure 5.25. Besides, it can be noticed that among the 3 regression methods, the GB is less accurate than the other two methods.



(a) Simulation 1



(b) Simulation 14

Figure 5.25: Comparison between the FE and the predicted values of t_Ac1 and t_Ac3 at 4 volume test points and for simulations 1 and 14

Table 5.8: Error for t_Ac1 and t_Ac3 using three regression methods - volume test points

Error (%)		point #1		point #2		point #3		point #4	
		t_Ac1	t_Ac3	t_Ac1	t_Ac3	t_Ac1	t_Ac3	t_Ac1	t_Ac3
Simulation 1	sPGD	3.5	22.9	4.0	4.4	12.9	23.8	5.1	0.1
	SVR	4.1	4.9	4.2	8.1	12.0	13.2	7.9	10.3
	GB	22.4	0.6	5.3	6.7	16.1	14.8	20.3	24.2
Simulation 14	sPGD	1.0	28.7	0.6	3.0	23.1	6.1	3.7	2.4
	SVR	8.4	9.0	2.1	1.9	1.0	6.6	4.1	7.5
	GB	61.6	18.3	21.4	18.2	42.4	27.3	51.4	47.2

5.4 Conclusion

The extension of the approximate solutions, obtained at 15 sparse sensors located at different positions in the gear tooth, was conducted to address the HAZ. To achieve this goal, different approaches were proposed for approximating the temperature and the austenite phase. For the temperature field, two approaches were adopted:

- the first one is based on nonlinear dimensionality reduction coupled with POD with RBF interpolation. Indeed, an interpolation over the gear tooth surface was first addressed by applying the LLE method 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. Second, 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 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. The volume interpolation resulted in good approximations as well.

- the second one is based on the application of GPOD using a completely known set of snapshots. The main idea of this approach was to be able to approximate the temperature at any spatial point for a given value of frequency, power, and time instant by using the solution at few positions. To do that, the POD was first performed on the snapshot matrix formed by vectors of temperature at each mesh node as columns. The reduced basis was used to determine the modal coefficients associated with the new gappy vectors with only 15 known elements. The final full-field reconstruction was achieved and the approach was evaluated at some surface and volume sensors. Good approximations were provided for the test points.

For the austenite phase evolution, models for t_Ac1 and t_Ac3 were constructed while considering the process parameters (frequency and power) and geometrical parameters (LLE parameters and depths) as input parameters. The results of sPGD, SVR, and GB regression methods were compared. No significant difference was observed for modeling the two characteristic time instants and the three methods provide models with sufficient accuracy, less than 10%. However, when they were tested on new data points, sPGD and SVR outperform the GB method that provides high errors ($> 10\%$) for some test point positions and for both simulations, comparatively.

So far, parametric solutions were proposed to predict the history of temperature and austenite phase for any value of frequency and power in predefined intervals and for new spatial positions in the HAZ. All those models are allocated to a single geometry of gear. However, we were interested in studying the geometrical parameters and the effect of new gear dimensions. This makes the subject of the next chapter of the manuscript.

Chapter 6

Geometrical analysis

Outline

6.1 Introduction . . . 145
6.2 Motivation . . . 145
6.3 ANN-based approach . . . 145
6.3.1 Methodology . . . 145
6.3.2 Data generation and ANN modeling . . . 146
6.3.3 Results and discussion . . . 147
6.4 Intermediate quantity-based approach . . . 155
6.4.1 Methodology . . . 155
6.4.2 Results and discussion . . . 157
6.5 Conclusion . . . 165

6.1 Introduction

So far, a single reference geometry was considered to study the effect of process parameters at some sparse sensors and then at the HAZ. Given that changing geometrical parameters could has un impact on the thermal results, this chapter aims at proposing a methodology to deal with different geometrical parameters and creating a parametric solution for the temperature field for new gear dimensions. In this chapter, we are only focusing on the temperature field for studying the effect of geometrical parameters.

6.2 Motivation

Studying the effect of geometrical parameters could be achieved by applying the methodology adopted in chapter 4 by considering a DoE with different possible combinations of geometrical parameters such as gear module, number of teeth, air gap, etc. This methodology is often employed for optimizing mechanical component designs. In this chapter, we tried to know how the temperature field evolves for new geometries without using a new DoE. Such objective is not easy to achieve regarding the complexity and the nonlinearity of the IH process. Two approaches were proposed to predict at real time the temperature-time evolution for new gear dimensions by using data extracted from the reference gear geometry (used so far in the previous two chapters) and some data known for the new geometries.

6.3 ANN-based approach

6.3.1 Methodology

The proposed approach is motivated by the great performance of ANN and deep learning in the last decade when dealing with nonlinear problems with multi-dimensional inputs and outputs. Moreover, ANNs show the capacity to predict any continuous function with a multi-layer neural network ensured by the universal approximation theorem (Hornik, 1991), and to solve various problems: regression, classification, and dynamical processes with a great variety of architectures. Therefore, the proposed approach consists in providing many temperature-time evolution curves from the reference geometry, used so far, under the existing simulations with different values of frequency and power. Next, the ANN model was trained with those temperature curves truncated after few first time steps as inputs and the entire curves as outputs of the model as illustrated in Figure 6.1. Then, for a new geometry,

the FE simulation should be run to get the temperature at the same initial time steps and the rest of the curve could be predicted by the developed ANN model. Therefore, an interesting time saving could be provided.

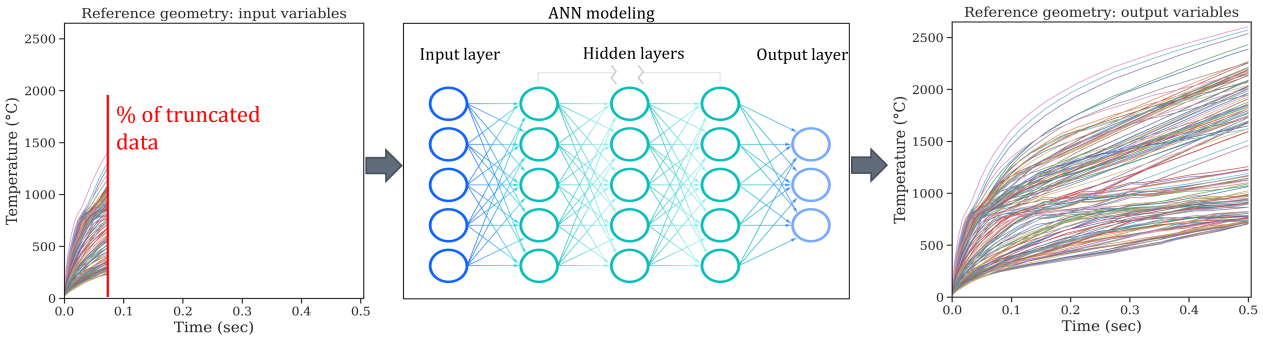


Figure 6.1: Methodology for full-field approximation using ANN

6.3.2 Data generation and ANN modeling

In the metamodeling context, the ANNs are usually recommended when the dataset is quite rich and the size of the training sampling is important (Simpson et al., 2001). For this reason, the temperature at all spatial data points was extracted as shown in section 5.2.2.1 of chapter 5 such that the temperature values were interpolated between the available values at each 0.1 sec by using a cubic interpolation. Therefore, curves were provided at all mesh nodes and for 501-time steps.

However, ANNs remain a complex method and a black box. Indeed, the architecture of the network must be chosen, such as the number of layers, the number of neurons per layer, the types of activation functions, etc. This in addition to the weights initialization, learning rate choice (α), the inputs and outputs standardization, etc. These choices determine the capacity of the neural network to predict accurately complex functions, however, no general rules exist. It is worth pointing out that 4 different cases were tested such that the percentage of truncated data was set to 5%, 10%, 15%, and 20%, respectively. In addition, only data at the gear surface were considered for modeling, however, the approach could be extended to consider the volume data as well.

Firstly, the inputs and outputs of the training samples were standardized to avoid different scaled data and improve the training stability of the model. This step is recommended with ANN modeling. Then, different sets of ANN hyperparameters were tested to choose a good one providing good accuracy. The ANN architecture and the employed hyperparameters were listed in Table 6.1.

Table 6.1: ANN architecture and hyperparameters

	Layers	# of nodes	Activation function	Initialization	α
5%	Input layer	25	-	-	-
	Hidden layers	50+250+250+50	ReLu	HeNormal	0.001
	Output layer	501	Linear	HeNormal	0.001
10%	Input layer	50	-	-	-
	Hidden layers	50+250+250+50	ReLu	HeNormal	0.001
	Output layer	501	Linear	HeNormal	0.001
15%	Input layer	75	-	-	-
	Hidden layers	50+150+150+50	ReLu	HeNormal	0.001
	Output layer	501	Linear	HeNormal	0.001
20%	Input layer	100	-	-	-
	Hidden layers	50+100+50	ReLu	HeNormal	0.001
	Output layer	501	Linear	HeNormal	0.001

6.3.3 Results and discussion

The developed models were tested for 4 new gear geometries listed in Table 6.2 such that only temperatures at few time steps were known. The choice of the geometrical parameters is not arbitrary and based on real experimental tests applied on the employed C45 steel.

Table 6.2: Process and geometrical parameters of the new geometries

Geometries	Module	# of teeth	Air gap	Frequency	Power
1	4	16	2.5	96	450
2	3	18	2.5	170	100
3	2	28	2.5	170	100
4	1.5	38	2.5	30	200

The relative errors for training and testing datasets are listed in Table 6.3 by using Eq.(4.3). The calculated relative errors represent the average overall surface points errors. The obtained errors show that an increase in the percentage of inputs results in a decrease in the relative error for the training and the testing data as well. Indeed, the error with 5% of input data is at least 2-times more than the error obtained with 20% of input data. In fact, the more input data are, the better prediction is.

Table 6.3: Relative errors (%) of the temperature-time evolution for the reference and the new geometries

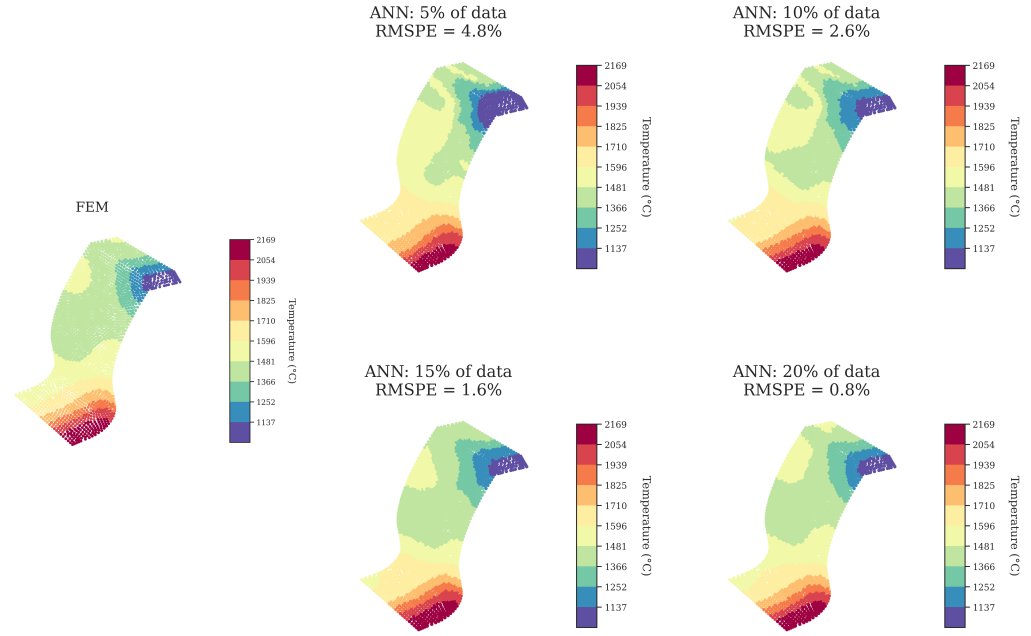
	Training data	Testing data			
	Reference geometry	Geometry #1	Geometry #2	Geometry #3	Geometry #4
5%	4.8	5.4	8.0	5.7	7.2
10%	3.2	3.6	3.8	3.2	7.1
15%	2.5	2.6	2.8	2.2	5.8
20%	1.5	1.7	2.1	2.2	3.7

Moreover, a comparison between the FEM and the ANN approximations of the temperature field at the middle and last time steps (t = 0.25 sec and 0.5 sec) is shown in Figures 6.2 and 6.3, respectively. This illustration shows the test cases in the full half-gear surface and for the 4 new geometries.

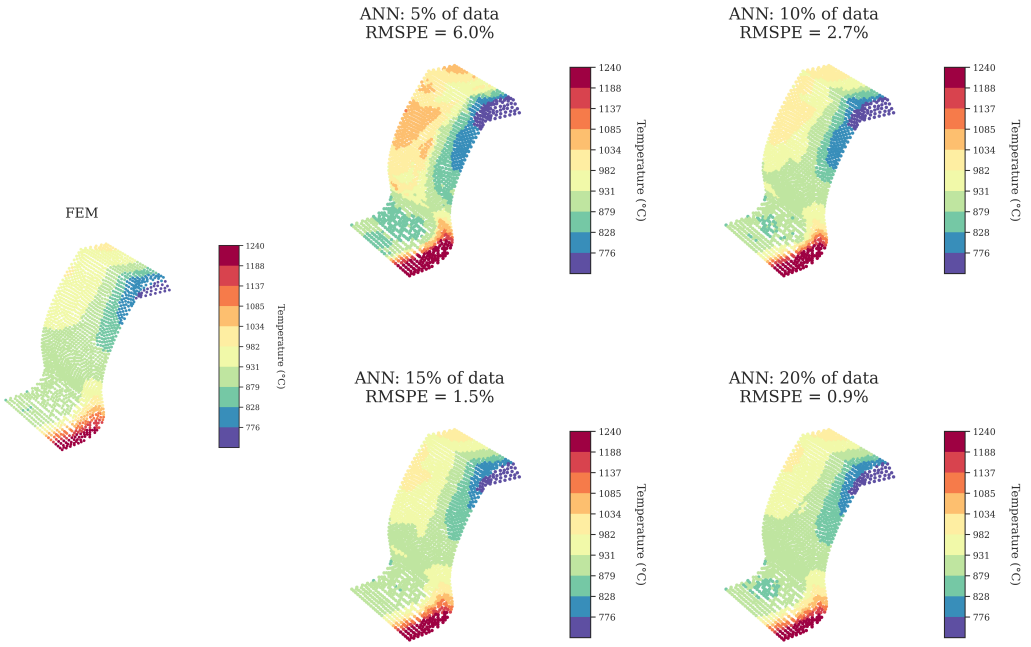
For the middle time step, the results obtained with ANN show a similar trend with the FE results for the 4 new geometries and for the 4 cases, with an RMSPE less than 6%. A much better prediction accuracy with RMSPE lower than 1.6% was observed with 15% and 20% of the input data compared to the prediction with 5% and 10%.

For the last time step, shown in Figure 6.3, the predictions by ANN are less accurate than those obtained at the middle time step. For 5% and 10% of data, the ANN model tends to overestimate the FE solution and provides different trends, especially for the model with 5% of the input data. The obtained RMSPE, close and higher than 10%, agrees with this interpretation. The increase of the percentage of input data decreases the RMSPE and consequently enhances the model accuracy such that the RMSPE is around 5% with 20% of input data curves.

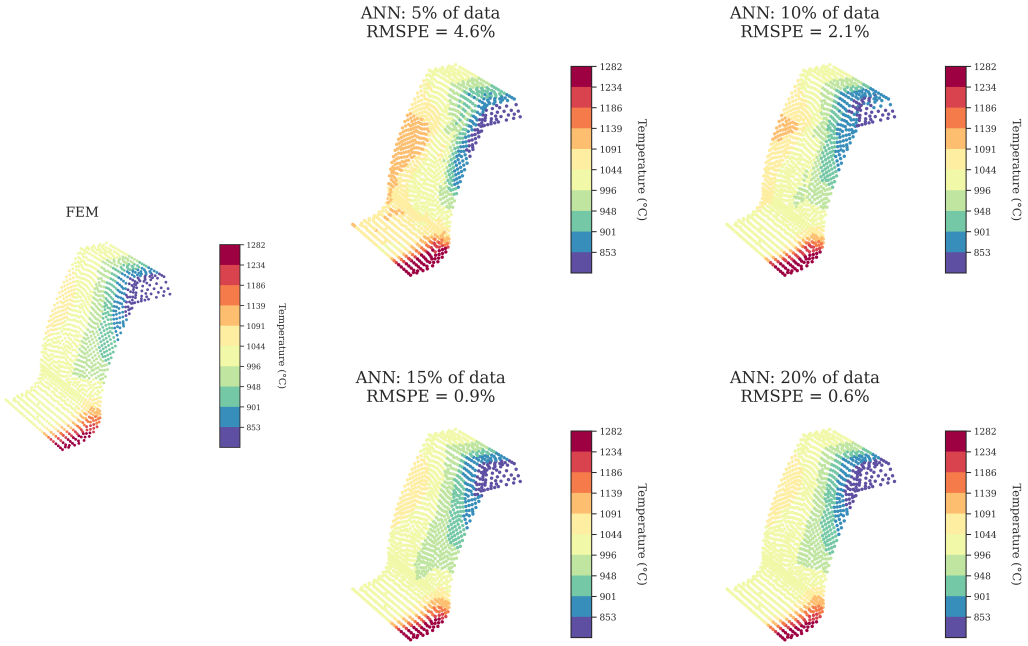
We can conclude that we can accurately approximate the temperature until the middle time step (0.25 sec) with ANN models of the 4 tested cases, however, the best prediction is certainly obtained by increasing the percentage of the truncated temperature curves. However, the prediction is less accurate at the last time step. This is probably due to the fact that some curves have similar or very close input data curves, but different ends for the training data which generates a lack of accuracy.



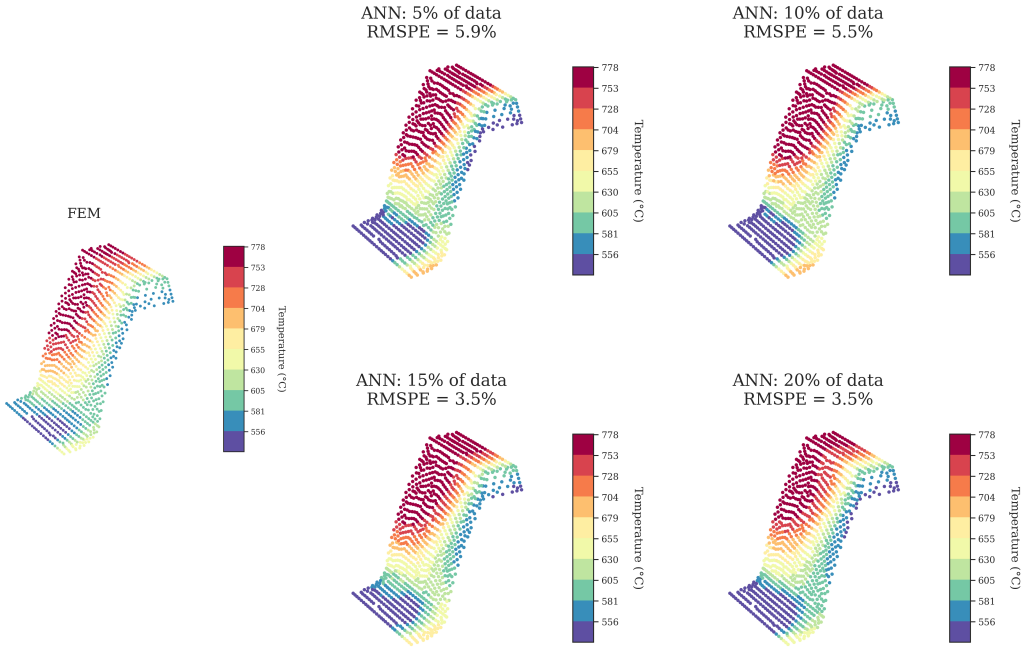
(a) Geometry #1



(b) Geometry #2

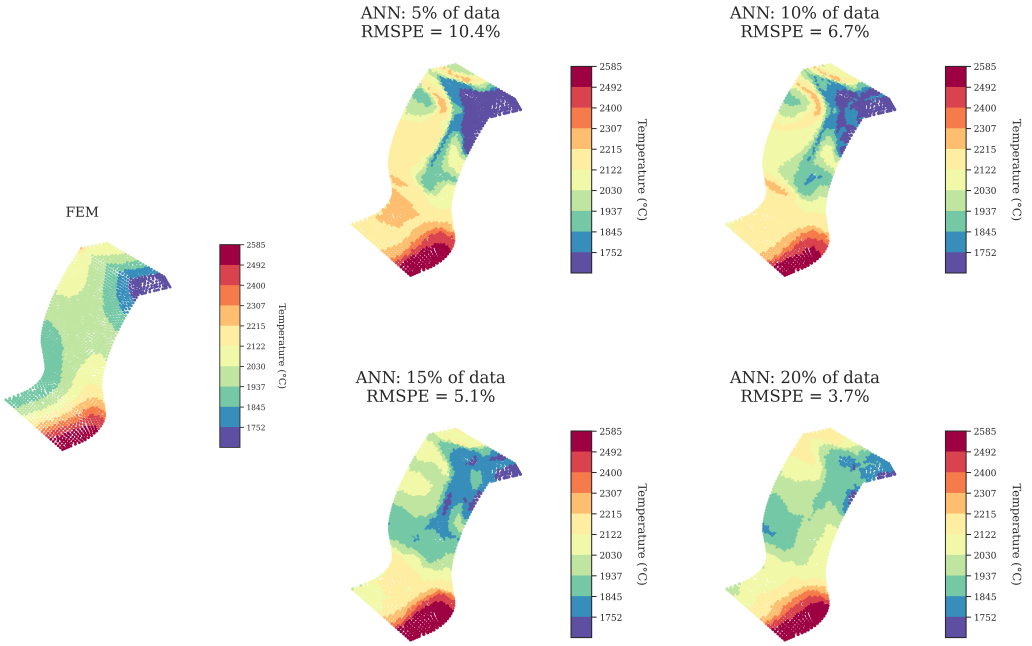


(c) Geometry #3

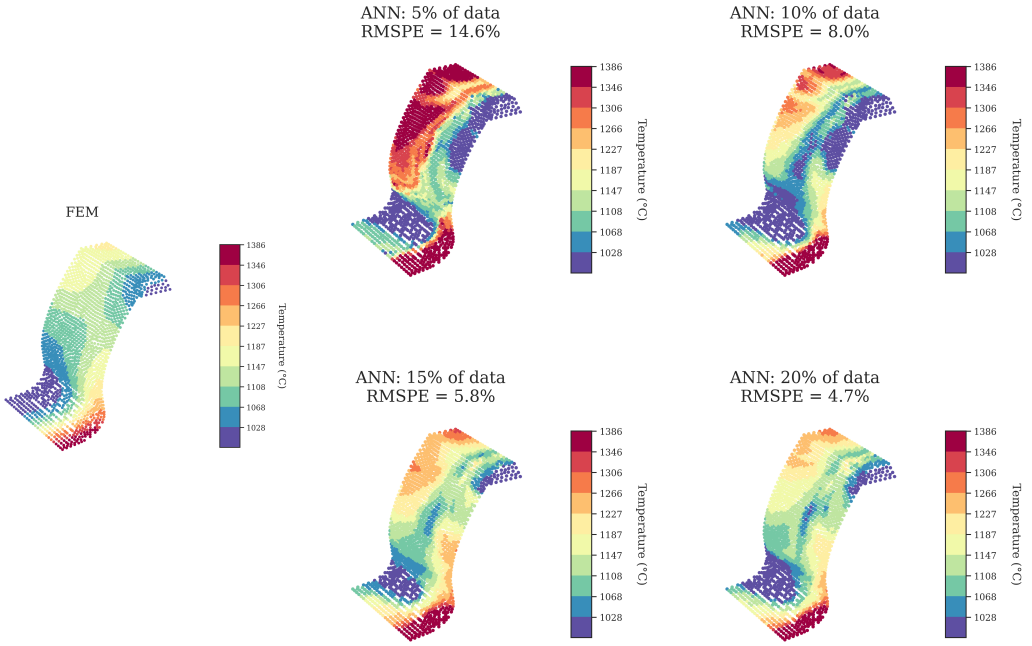


(d) Geometry #4

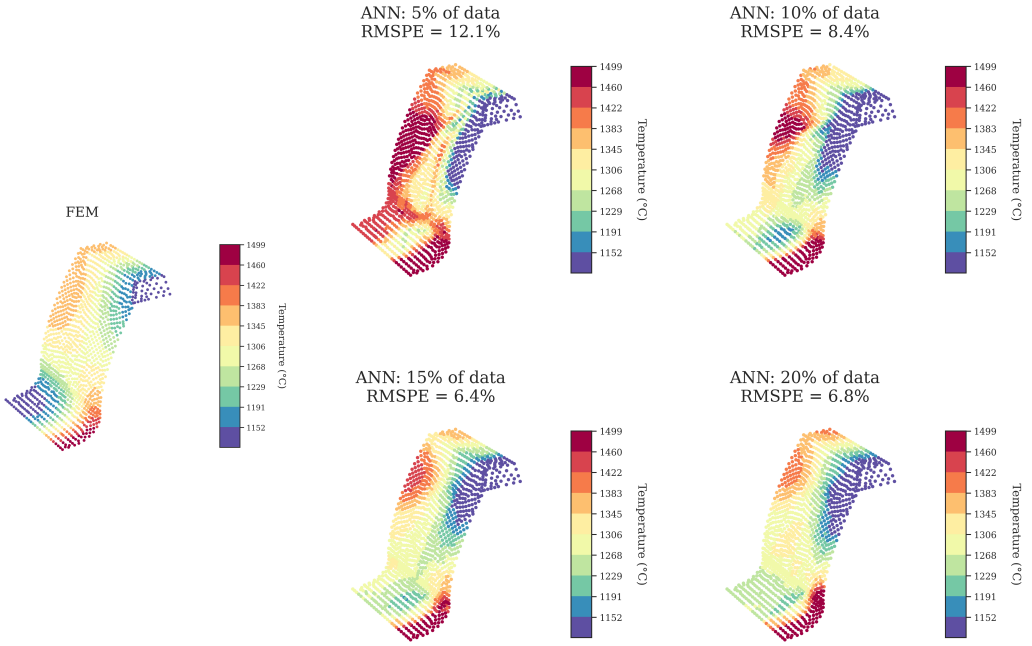
Figure 6.2: Comparison between the FEM and the ANN approximations of the surface temperature distribution for the new geometries at t = 0.25 sec



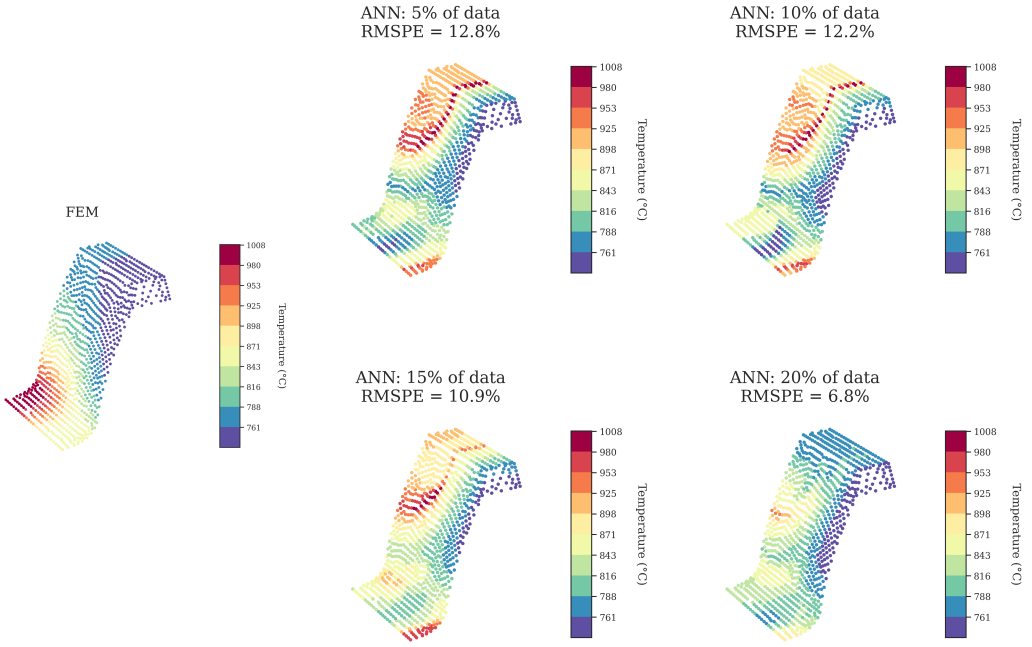
(a) Geometry #1



(b) Geometry #2



(c) Geometry #3

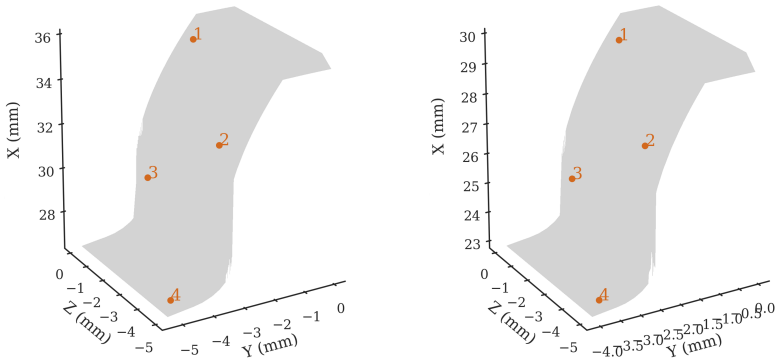


(d) Geometry #4

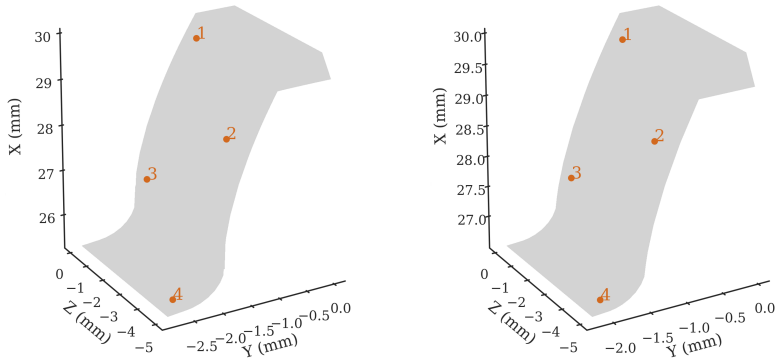
Figure 6.3: Comparison between the FEM and the ANN approximations of the surface temperature distribution for the new geometries at t = 0.5 sec

In addition, a comparison between the FEM and the ANN approximations of the temperature-time

evolution is shown in Figure 6.5, for the 4 new gear geometries and at the 4 test points illustrated in Figure 6.4. It can be noticed that the FE and the ANN curves show similar trends and the best predictions are observed with 15% and 20% of input data which is in accordance with the previous interpretations.

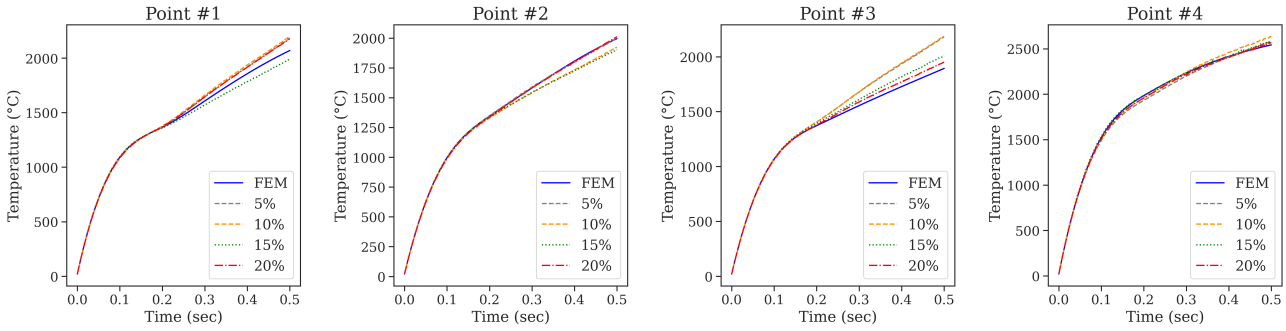


(a) Geometry #1 (b) Geometry #2

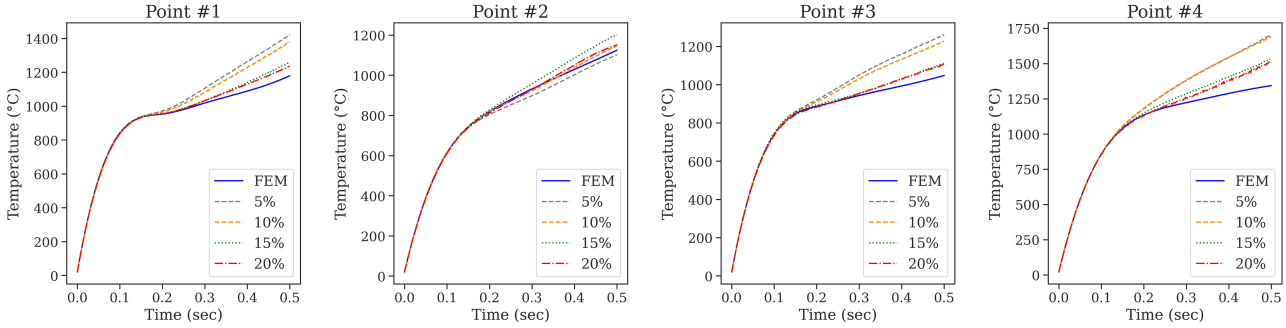


(c) Geometry #3 (d) Geometry #4

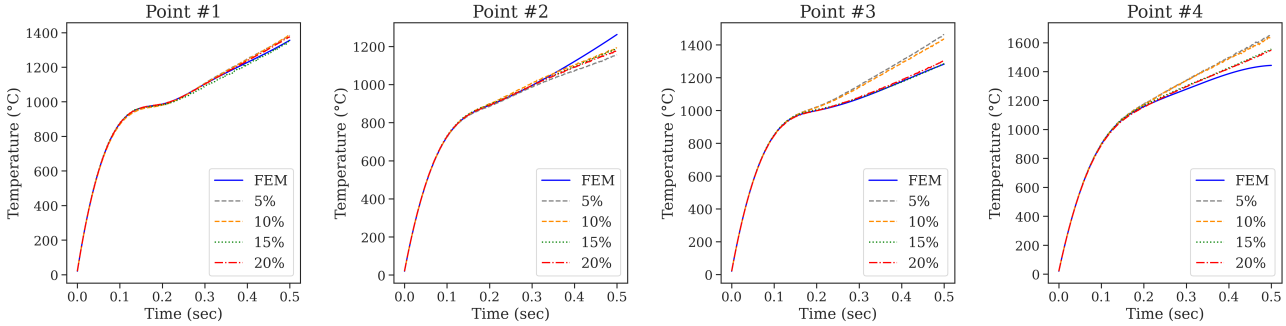
Figure 6.4: Test points at the new gear geometries



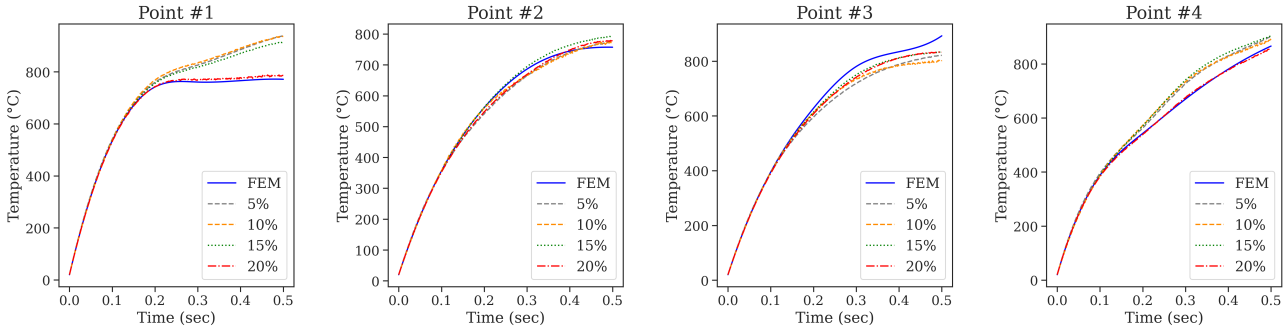
(a) Geometry #1



(b) Geometry #2



(c) Geometry #3



(d) Geometry #4

Figure 6.5: Comparison between the reference FEM and the ANN approximations of the temperature-time evolution for the 4 new gear geometries and at the 4 test points

The relative errors, defined by Eq.(5.1), were also calculated. The obtained errors, listed in Table 6.4, are in good agreement with the graphical results. The lowest errors are observed with 15% and 20% of input data and for all geometries.

Table 6.4: Relative errors (%) of the temperature-time evolution for the 4 new geometries and at the 4 test points

		Point #1	Point #2	Point #3	Point #4
Geometry #1	5%	4.4	0.8	11.1	1.4
	10%	4.6	0.2	9.6	1.5
	15%	0.3	7.0	2.6	0.9
	20%	2.9	0.4	1.4	0.7
Geometry #2	5%	11.6	1.8	14.8	15.0
	10%	7.4	0.6	1.0	11.3
	15%	3.4	2.3	3.0	5.9
	20%	2.7	1.7	2.8	5.8
Geometry #3	5%	2.6	1.9	8.8	6.4
	10%	1.6	4.7	3.5	4.0
	15%	0.4	2.6	0.6	2.8
	20%	0.7	2.9	0.7	2.7
Geometry #4	5%	10.5	2.0	6.9	6.7
	10%	9.4	2.7	4.1	8.0
	15%	7.8	2.3	3.2	5.8
	20%	1.1	2.0	3.9	0.7

6.4 Intermediate quantity-based approach

6.4.1 Methodology

According to the heat transfer equation, shown in Eq.(3.21), the temperature-time evolution depends on the heat source $\dot{Q}$ in the workpiece. This fact motivated us to assume that if this quantity is known at a given position of the gear, the temperature can be predicted at that point. However, the problem here is that $\dot{Q}$ is not a constant but it evolves with time. Therefore, the approximation of temperature should be achieved with full information about $\dot{Q}$. Indeed, instead of considering the vectors of $\dot{Q}$ to create a metamodel for the temperature by using the data of the reference geometry ($\dot{Q}$ of the initial DoE at all surface gear geometry) for training models, another representative quantity was selected. This quantity is the area of the $\dot{Q}$ curve, calculated by the trapezoidal rule. Indeed, we assumed that the area and the initial value of $\dot{Q}$ together with the geometrical coordinates enable to approximate the low-dimensional representation of the initial temperature-time evolutions, obtained by POD. Therefore, predicting the temperature evolution for new geometries by considering these

quantities could be achieved. However, a model for the area of $\dot{Q}$ was required. To do this, a step-by-step approach was proposed. This approach was adopted because a low-accuracy model was obtained when the area over the full-time interval was directly approximated by using as inputs the area over 5% of the time interval, the initial value of $\dot{Q}$ ($\dot{Q}_0$), and the geometrical coordinates. Therefore, the applied approach enables to learn incrementally from the previous output which also represents the area of the $\dot{Q}$ evolution curve but over a smaller time interval compared to the actual area to be modeled. However, it is worth pointing out that this approach has an inconvenient which is the cumulation of the error while going from one output to another. The proposed methodology is shown in Figure 6.6.

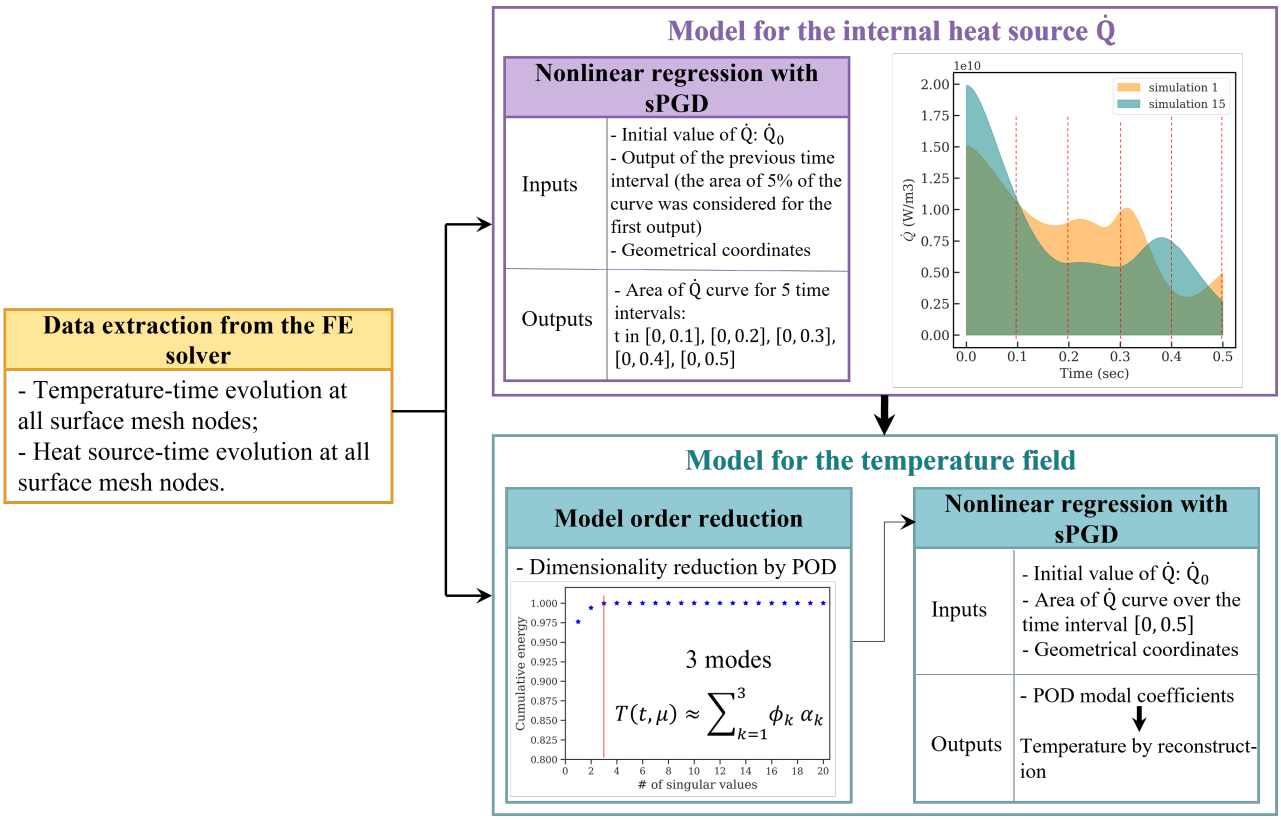


Figure 6.6: Methodology of the intermediate quantity-based approach

In this section, the geometrical coordinates were considered in a normalized form such that similar positions in different gear geometries have the same coordinates as shown in Figure 6.7. To do that, a normalization over the X, Y, and Z coordinates was applied according to the following expressions:

$$\boldsymbol{X_{norm}} = \frac{\boldsymbol{X} - \boldsymbol{X_{min}}}{\boldsymbol{X_{max}} - \boldsymbol{X_{min}}} \tag{6.1}$$

$$\boldsymbol{Y_{norm}} = \frac{\boldsymbol{Y} - \boldsymbol{Y_{min}}}{\boldsymbol{Y_{max}} - \boldsymbol{Y_{min}}} \tag{6.2}$$

$$\boldsymbol{Z_{norm}} = \frac{\boldsymbol{Z} - \boldsymbol{Z_{min}}}{\boldsymbol{Z_{max}} - \boldsymbol{Z_{min}}} \tag{6.3}$$

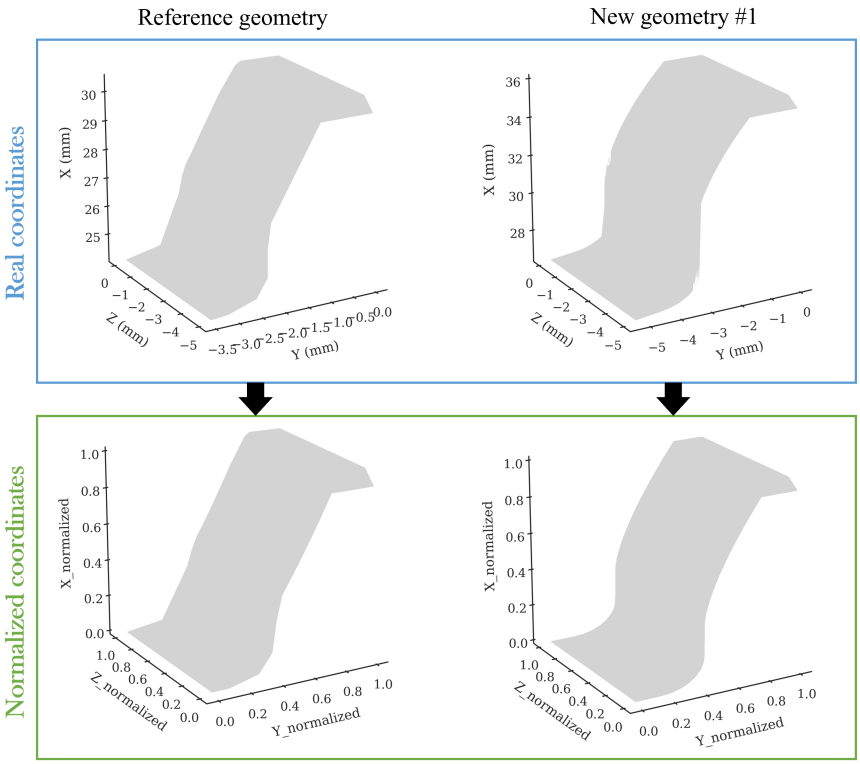


Figure 6.7: Real and normalized Cartesian coordinates

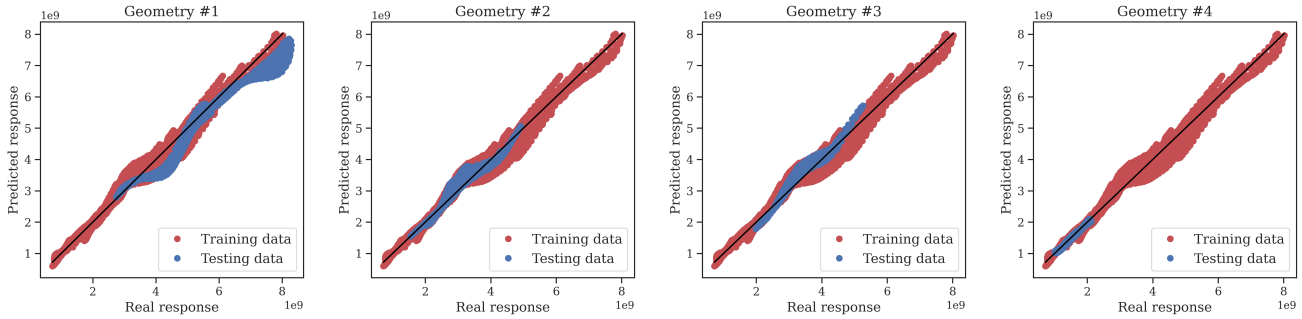
It is worth pointing out that similarly to the first approach, the simulation for a new gear geometry should be run for few time steps and the tested geometries are those listed in Table 6.2.

6.4.2 Results and discussion

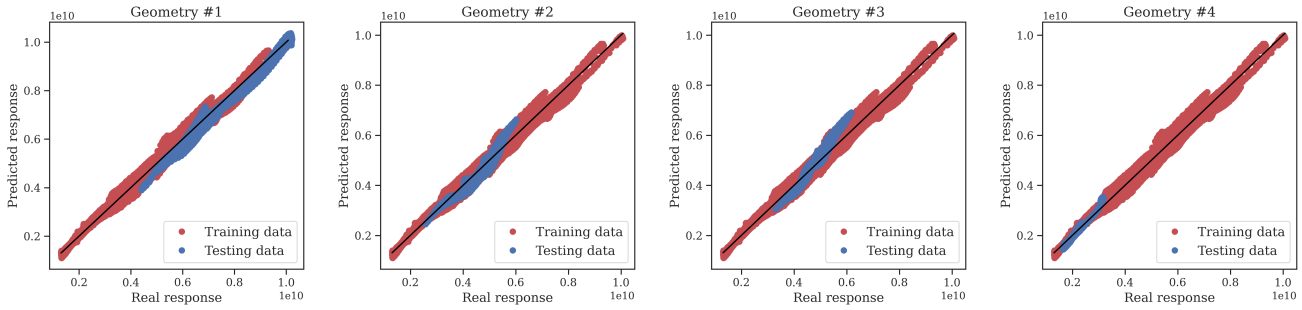
6.4.2.1 Modeling the area of the internal heat source evolution

In this section, models for 5 quantities representing the area of the $\dot{Q}$ evolution curve over different time intervals were developed. Figure 6.8 shows the FE versus the predicted values with sPGD of the 5 outputs and for the 4 new geometries. It is worth reminding that the data of the reference geometry

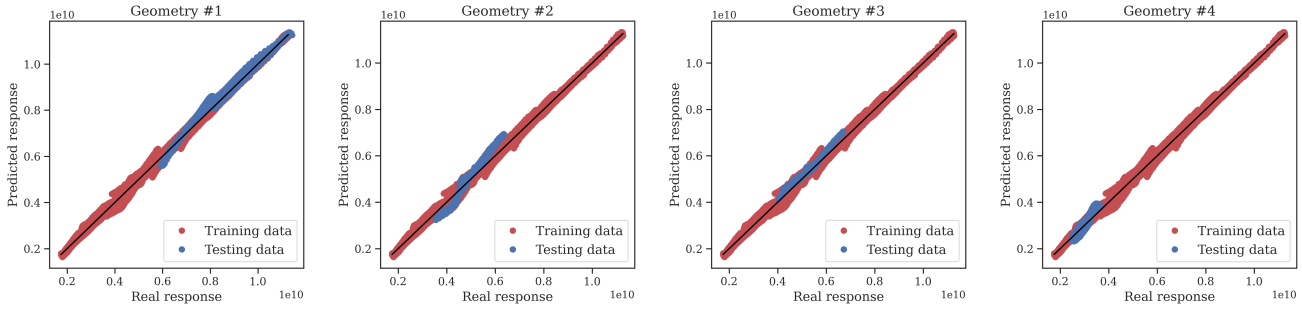
are called the training data and the ones of the new geometries are called testing data. It can be noticed that the models are more accurate for the last 3 outputs than the first two ones. This is because the maximum variation in $\dot{Q}$ is seen at the beginning of the curves. Globally, the distribution of the red and blue points is close to the black line corresponding to the perfect prediction for all outputs and all geometries.



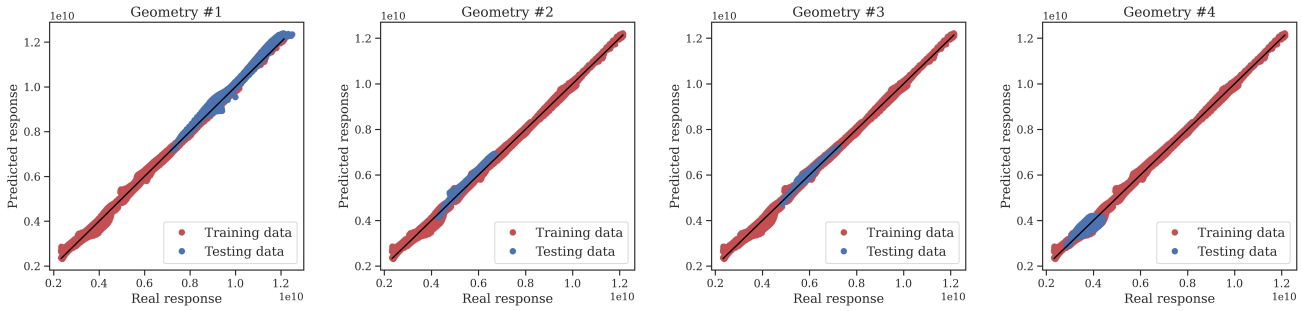
(a) Area of $\dot{Q}$ over the time interval $[0, 0.1]$



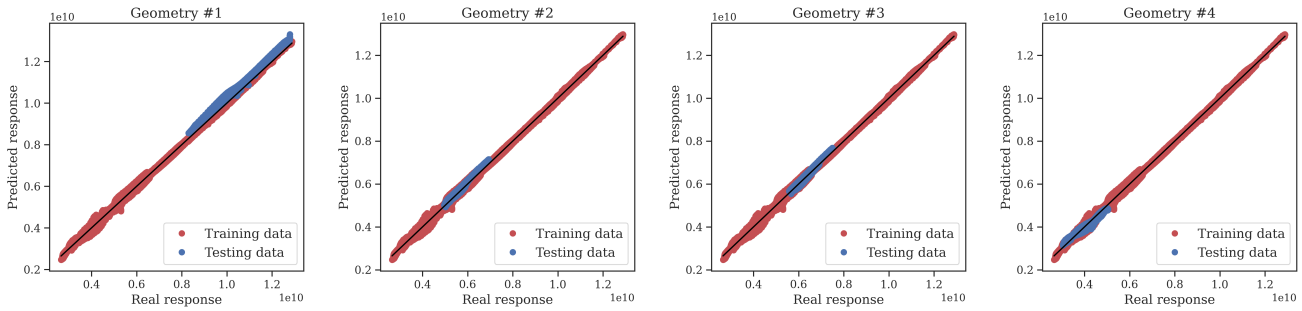
(b) Area of $\dot{Q}$ over the time interval $[0, 0.2]$



(c) Area of $\dot{Q}$ over the time interval $[0, 0.3]$



(d) Area of $\dot{Q}$ over the time interval $[0, 0.4]$



(e) Area of $\dot{Q}$ over the time interval $[0, 0.5]$

Figure 6.8: Reference FE versus predicted values of the area of $\dot{Q}$ over 5 different time intervals for the reference and the new geometries

In addition, the RMSPE for the training and the testing datasets are listed in Table 6.5. The obtained errors are in good agreement with the graphical results. The highest errors (between 4% and 6%) are observed with the two first outputs. Then, they decrease for the last 3 outputs such that the errors are less than 4% for the testing data, except for geometry #4 where the error is the highest, comparatively.

Table 6.5: RMSPE (%) of the area of $\dot{Q}$ for the reference and the new geometries

	Training data	Testing data			
	Reference geometry	Geometry #1	Geometry #2	Geometry #3	Geometry #4
Area of $\dot{Q}$ in $[0, 0.1]$	5.6	5.4	6.5	6.3	2.9
Area of $\dot{Q}$ in $[0, 0.2]$	4.3	4.5	4.5	6.6	5.3
Area of $\dot{Q}$ in $[0, 0.3]$	3.1	3.3	3.9	2.5	6.1
Area of $\dot{Q}$ in $[0, 0.4]$	2.3	2.3	3.2	1.3	5.7
Area of $\dot{Q}$ in $[0, 0.5]$	1.9	2.7	1.6	1.3	4.1

6.4.2.2 Modeling the temperature field evolution

Firstly, a snapshot matrix $\boldsymbol{M} = [\boldsymbol{T}_1 \boldsymbol{T}_2 \cdots \boldsymbol{T}_N]$ was constructed such that $\boldsymbol{T}_i = \boldsymbol{T}(F_i, P_i, \boldsymbol{x}_i)$, for $i = 1, \cdots, N$, is the vector of temperature values at every time step for the snapshot i, and the snapshots (columns of M) are defined with different values of frequency, power, and mesh node positions on the gear. The POD was then performed on the snapshot matrix $\boldsymbol{M}$ composed of more than 10000 snapshots such that the left singular vectors of $\boldsymbol{M}$ were truncated to the 3 first vectors to recover more than 99.9% of the cumulative energy, as shown in Figure 6.9.

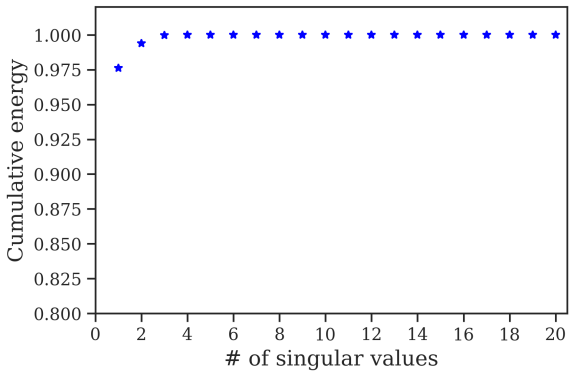
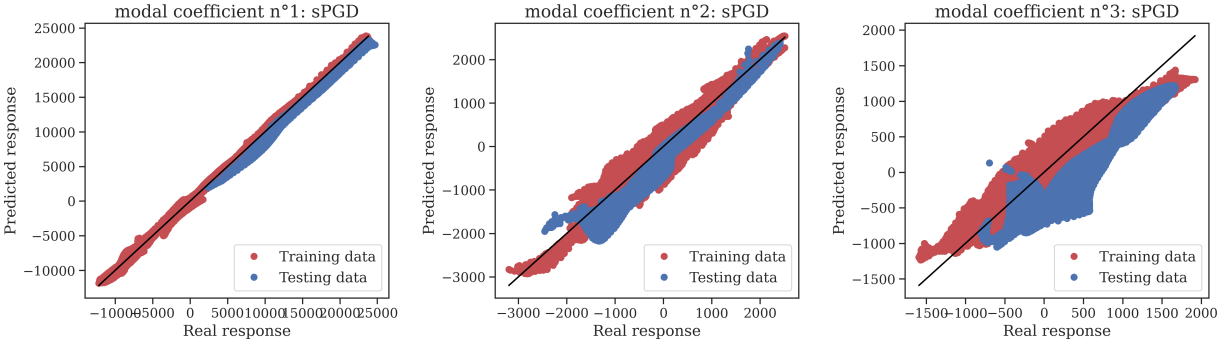
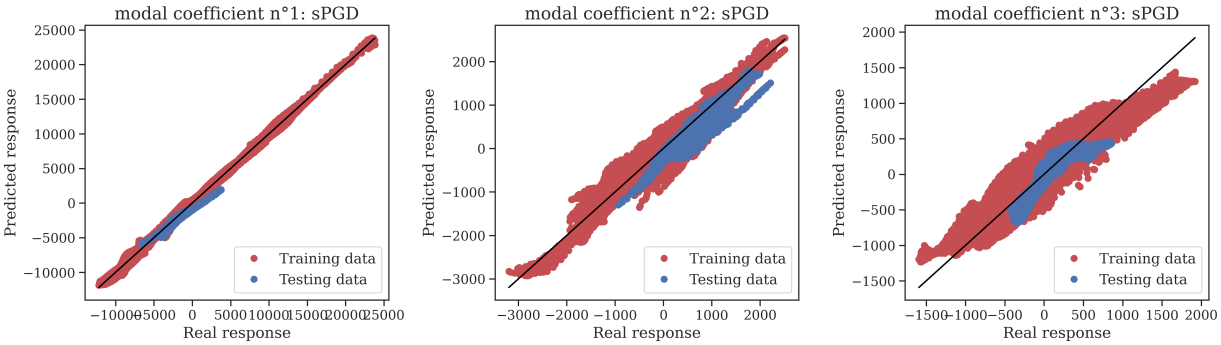


Figure 6.9: Cumulative energy of the thermal field computed from snapshots at different input parameters and spatial nodes

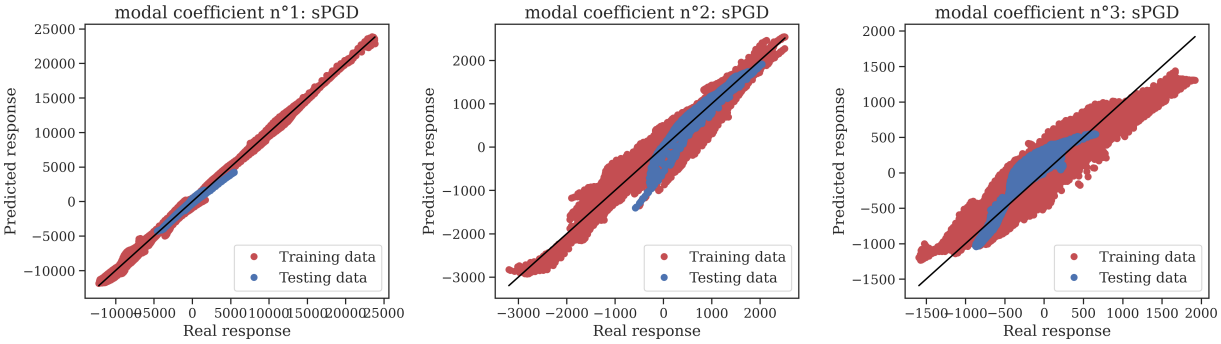
Models for the three POD modal coefficients were constructed by using the sPGD method regarding its good performance for approximating the temperature. As shown in Figure 6.6, the area of $\dot{Q}$ curve over the full-time interval $[0, 0.5]$, $\dot{Q}_0$, and the geometrical coordinates ($\boldsymbol{X_{norm}}$, $\boldsymbol{Y_{norm}}$, and $\boldsymbol{Z_{norm}}$) were considered as input parameters for the models. Data from the reference geometry were considered as training data to construct models and data with new geometries were considered to test them. However, two points have to be checked first: the capacity of the just mentioned input parameters to approximate the temperature field and the ability of the developed models to provide good predictions for new gear geometries. To verify these points, the FE versus the predicted values by sPGD of the POD modal coefficients are illustrated in Figure 6.10 for the training and testing datasets. It can be noticed that an accurate model was observed for the first POD modal coefficients corresponding to the first mode that retains more than 97% of the variance. Moreover, the training (red points) and testing (blue points) data are close to the black line for the 4 testing geometries. However, the models for the two other modal coefficients are less accurate than the first one but they still acceptable.



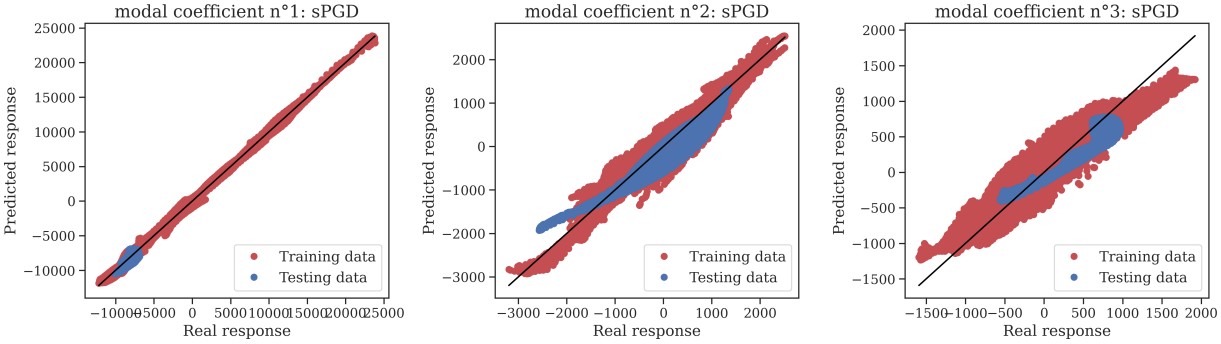
(a) Geometry #1



(b) Geometry #2



(c) Geometry #3



(d) Geometry #4

Figure 6.10: Reference FE versus predicted values of the low-dimensional representation of temperature for the reference and the new geometries

The final approximation of temperature was reconstructed by using Eq.(4.1) with the predicted values of POD modal coefficients. Table 6.6 shows the relative errors for the training and the testing datasets over all surface points by using Eq.(4.3). The obtained errors for the reference and the new geometries are within the same order of magnitude such that errors less than 6% were obtained.

Table 6.6: Relative errors (%) of the temperature-time evolution for the reference and the new geometries

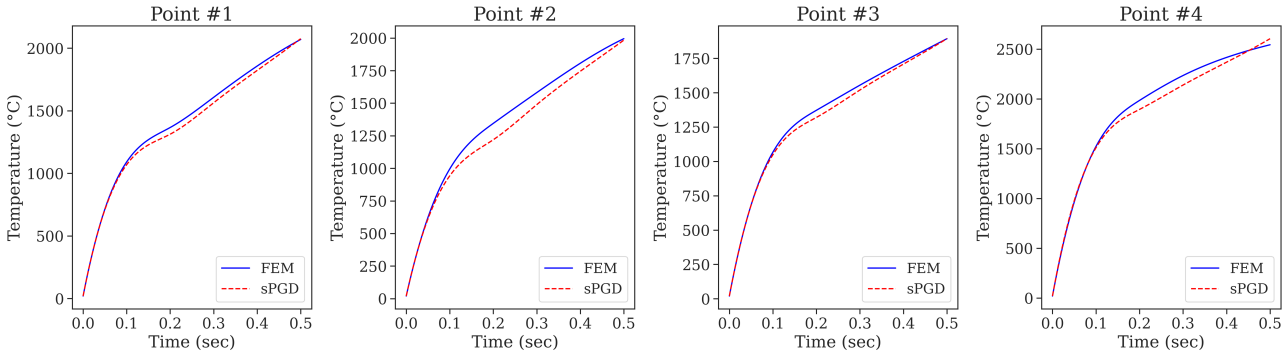
Training data	Testing data			
Reference geometry	Geometry #1	Geometry #2	Geometry #3	Geometry #4
2.3	3.6	2.5	1.5	5.5

In addition, Figure 6.11 shows the temperature-time evolution obtained by the FEM and the sPGD method at the 4 points illustrated in Figure 6.4 and for the 4 new geometries. It is worth mentioning that the prediction here is made with the FE values of $\dot{Q}$ to calculate the area and not the prediction from section 6.4.2.1. It can be noticed that good predictions were observed for most test points.

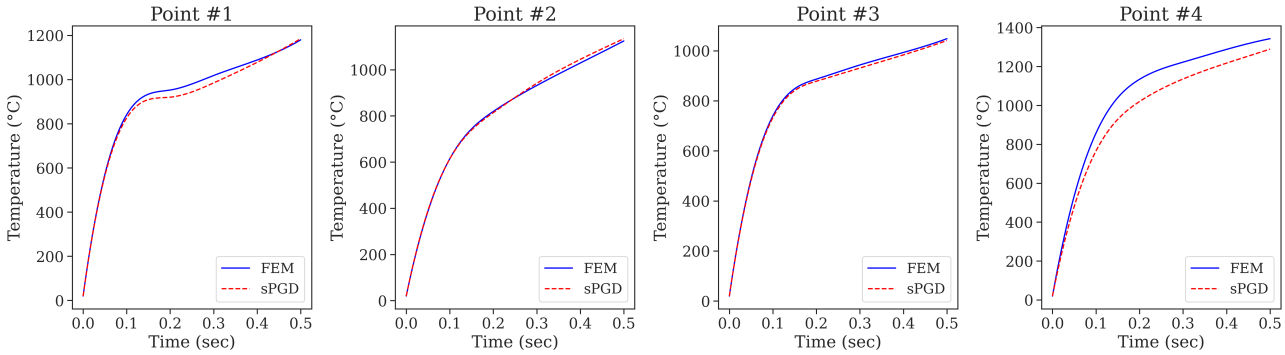
The relative errors associated with curves in Figure 6.11 are listed in Table 6.7. Globally, low errors, less than 5%, were obtained for most of them.

Table 6.7: Relative errors (%) of the temperature-time evolution for the 4 new geometries and at the 4 test points using the FE values of $\dot{Q}$ area

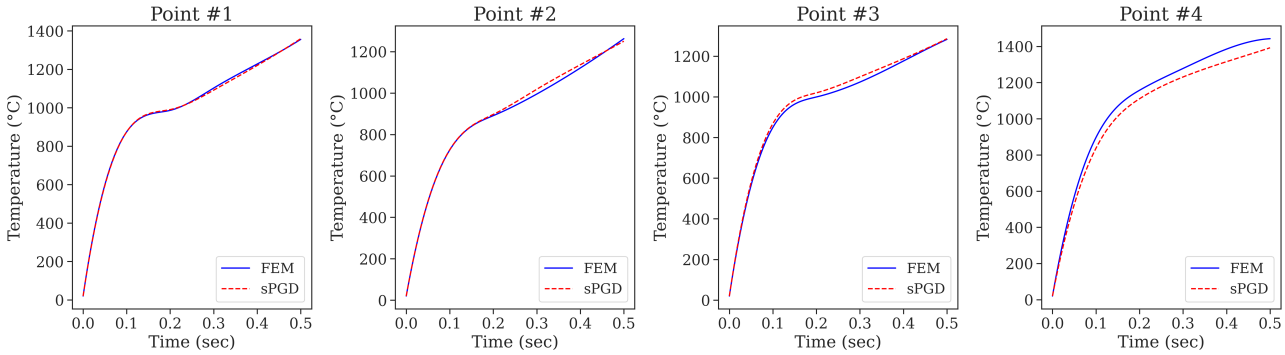
	Point #1	Point #2	Point #3	Point #4
Geometry #1	2.3	5.4	2.1	3.1
Geometry #2	2.3	1.0	1.0	7.6
Geometry #3	0.5	1.3	1.7	4.6
Geometry #4	3.9	9.1	2.9	3.9



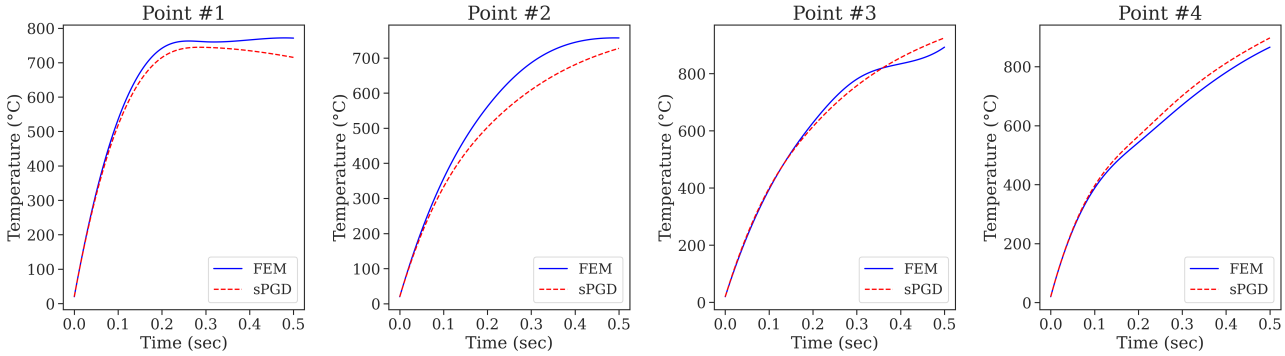
(a) Geometry #1



(b) Geometry #2



(c) Geometry #3

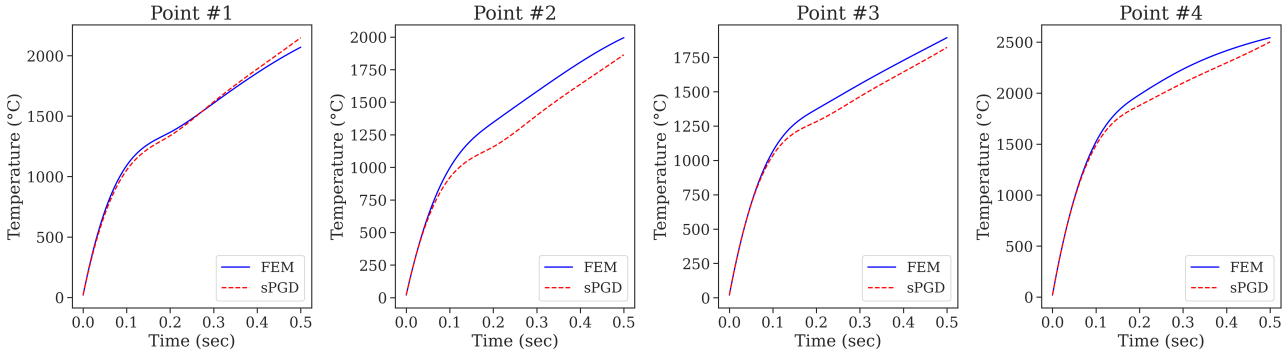


(d) Geometry #4

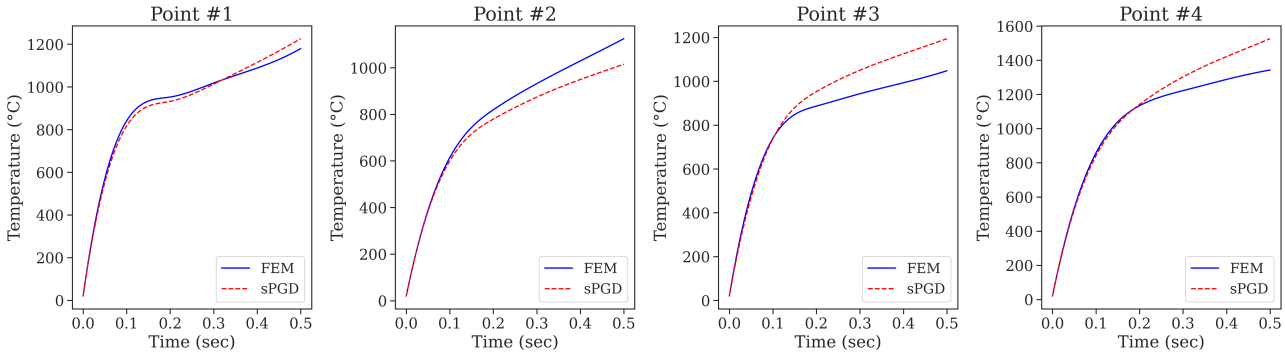
Figure 6.11: Comparison between the FEM and the sPGD approximations of the temperature-time evolution for the 4 new geometries and at the 4 test points using the FE values of $\dot{Q}$ area

By considering those results, the two points mentioned above are well verified. Now, the last step is to test the proposed methodology by using the prediction of the area of $\dot{Q}$ from the models developed in section 6.4.2.1 to approximate the temperature-time evolution.

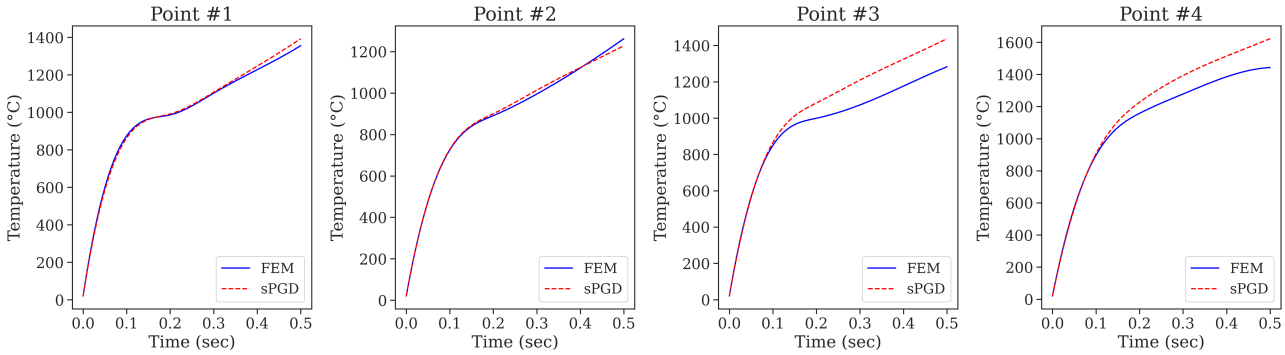
Figure 6.12 shows the temperature-time evolution obtained by the FEM and the sPGD method at the same test points and for the 4 geometries. The curves show similar trends, however, there are some differences between the FE solutions and sPGD approximations for some points. This discrepancy is due to the predicted input parameter and the actual temperature model accuracy as well.



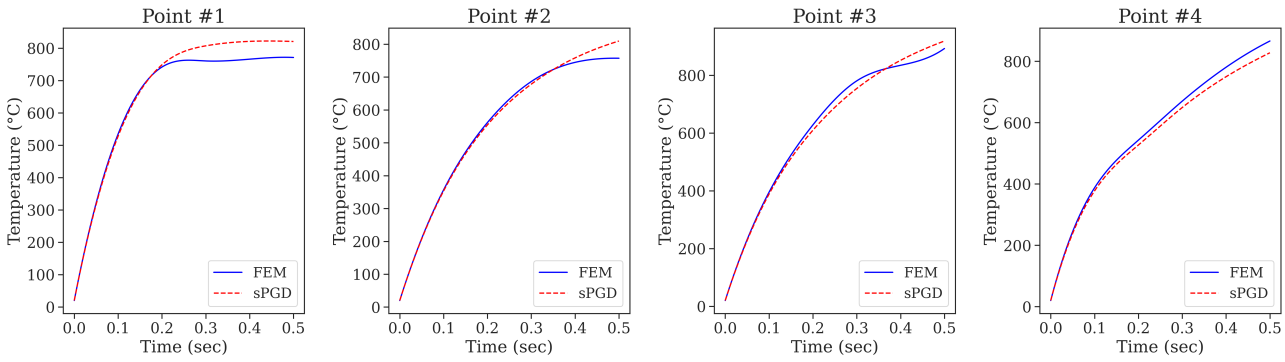
(a) Geometry #1



(b) Geometry #2



(c) Geometry #3



(d) Geometry #4

Figure 6.12: Comparison between the FEM and the sPGD approximations of the temperature-time evolution for the 4 new geometries and at the 4 test points using the predicted values of $\dot{Q}$ area

The relative errors associated with the curves shown in Figure 6.12 are listed in Table 6.8. The obtained errors are different from one geometry to another but they do not exceed 11%, and a very good prediction with an error less than 3% was obtained with several test points. Despite having the highest average error for the area of $\dot{Q}$ with geometry #4, the obtained errors for temperature are the lowest for these test points.

Table 6.8: Relative errors (%) of the temperature-time evolution for the 4 new geometries and at the 4 test points using the predicted values of $\dot{Q}$ area

	Point #1	Point #2	Point #3	Point #4
Geometry #1	2.2	10.3	5.1	4.7
Geometry #2	2.4	6.8	10.7	8.0
Geometry #3	1.4	1.3	10.8	8.4
Geometry #4	5.3	2.7	2.9	3.6

In summary, the temperature-time evolution can be accurately approximated by using the heat source quantity, however, if this latter has a nonlinear evolution in time, its use becomes complex.

6.5 Conclusion

Investigating the effect of geometrical gear parameters and developing metamodels for temperature-time evolution was conducted in this chapter. To achieve this goal, two approaches were proposed. the first approach consists in approximating the full temperature-time evolution by using the beginning of the curves. Indeed, the nonlinear ANN regression method with MLP was adopted and 4 models with different input percentages (5%, 10%, 15%, 20%) were developed. Those models were trained

by the available data from the reference geometry and then tested on 4 new gear geometries. This approach allows to accurately approximate the temperature until the middle time step (0.25 sec) for the 4 tested cases, however, the more input data we have, the better prediction we get. On the other hand, the predictions at the last time step are less accurate than those obtained at the middle time step. This is because some curves have similar or very close input curves, but different ends for the training data which generates a lack of accuracy.

The second approach relies on the use of another intermediate physical quantity to model the temperature. According to the heat equation, this quantity is the heat source which is the principal cause of heating. Indeed, models for the area of the heat source evolution curves were first developed and then used for modeling the temperature field by using the sPGD method. It was demonstrated that we can accurately predict the temperature everywhere in the gear surface and for any new geometry, only by considering the data of the reference geometry. These data were used for training the models. However, modeling the area of the heat source curve step-by-step by using only few initial values of this latter, obtained by running the simulation for a new geometry with 5% of the original time process, remains a challenge and gives rise to errors that certainly affect the approximation of the temperature.

Chapter 7

Conclusions and perspectives

MOR and parametric metamodeling represent valuable approaches to optimize and control heat treatment processes. Indeed, the FE numerical simulations are usually computationally expensive in terms of time and storage, however, metamodels reduce significantly the computational cost by providing very fast predictions with sufficient results accuracy. This thesis focuses on developing a methodology based on MOR techniques to generate metamodels for the temperature and austenite rate evolution during the IH process carried out on a spur-gear of C45 steel to allow real-time prediction over the parameter space for an optimization purpose. The complex gear geometry and the multiphysics-based coupling encountered during the process give rise to several challenges for metamodeling, as they can generate nonlinearity and heterogeneity of thermal distribution over the workpiece. Therefore, this work gives an insight into the possibility of using data-driven MOR approaches in multiphysics application.

7.1 Outcomes

A literature review was first provided to give an overview of the dimensionality reduction methods. Intrusive (Projection-based POD, PGD) and non-intusive (PODI, sPGD, SSL) MOR methods were shown. However, regarding the complexity of the IH process and the non accessibility to the FE source code describing the process in the commercial software, the non-intrusive approach was chosen to be applied. In this thesis, synthetic data provided by FE codes were generated by using FORGE® software and the PODI approach was employed in order to build the parametric-based solutions of the IH problem. Since the PODI is coupled with interpolation and regression methods, an overview of several ML and interpolation methods was also carried out.

In addition, the involved physical phenomena, their mathematical formulations, mutual interactions, and the included parameters were evoked in this work for a better understanding of the IH process. Then, the numerical FE simulation using FORGE® code was carried out such that the model development was illustrated step by step; creating geometries, meshing strategy with a convergence study, defining the material properties, defining the boundary conditions and the process parameters, and finally the coupling parameter. The obtained numerical results for temperature and austenite phase distribution demonstrate the skin effect as well as the difference between the edge and the mid-width results in the case of the gear geometry. Based on the FE model, a set of 20 synthetic data for different values of frequency and power, chosen from the predefined intervals, were first generated according to the LHS design of experiments. Then, parametric metamodels for the IH process were developed, for 15 sparse sensors located at different positions in the gear tooth, to predict the evolution of temperature and austenite phase.

For modeling the temperature field, two methodologies were proposed. The first one consists in truncating all snapshots such that only data below a constant final temperature value were considered, and the time as a function of temperature was then modeled. The second proposed methodology consists in completing the missing values to reach a constant final time process for certain snapshots by using the GPOD method. For the two methodologies, three regression methods (sPGD, SVE, and GB) were performed to create metamodels for the low-dimensional representation of the initial snapshot data provided by POD. The obtained results indicate that the sPGD and the SVR methods can provide relatively accurate models to predict the temperature and time evolutions under the imposed values of input parameters even in the sparse low data-limit case. However, the GB method doesn't perform well with a reduced amount of data.

For modeling the austenite phase, the beginning and the complete austenitization time instants (t_Ac1 and t_Ac3) were extracted and modeled using the three regression methods. The best approximation was provided by sPGD, followed by SVR with a little less accurate prediction, and the worst prediction was provided by the GB method.

The developed approaches have the advantage of a real-time prediction that is highly suitable for active control of the process parameters. To go further, the extension of the approximate solutions, obtained at 15 sparse sensors, was conducted to address the HAZ. To achieve this goal, different

approaches were proposed for approximating the temperature and the austenite phase. For the temperature field, two approaches were adopted: the first one can be applied on a single simulation (given values of frequency and power) and is based on nonlinear dimensionality reduction coupled with POD with RBF interpolation. Indeed, an interpolation over the gear tooth surface was first addressed by applying the LLE method 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. Second, 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 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, an orthogonal projection onto the surface, a mapping by LLE, and a POD with RBF interpolation. The volume interpolation resulted in good approximations as well.

The second approach is based on the application of GPOD with a completely known set of snapshots. The objective is to be able to approximate the temperature at any spatial point for a given value of frequency, power, and time instant by using the known solution at few positions. To do that, the POD was first performed on the snapshot matrix composed of vectors of temperature at each mesh node as columns. The reduced basis was then used to determine the modal coefficients associated with the new gappy vectors with only 15 known elements (sensors positions). The final full-field reconstruction was achieved and evaluated at some surface and volume sensors where good approximations were provided for the test points. This second approach is more general and can be used to find the solution even for new values of input parameters.

For the austenite phase evolution, models for t_Ac1 and t_Ac3 were constructed while considering the process parameters (frequency and power) and geometrical parameters (LLE parameters and depths) as input parameters. A comparison between the results of sPGD, SVR, and GB regression methods was achieved. No significant difference was observed for modeling the two characteristic time instants and the three methods provide models with sufficient accuracy, less than 10%. However, when they

were tested on new data points, sPGD and SVR outperform the GB method that provides high errors ($> 10\%$).

Regarding that all developed models are allocated to a single geometry of gear. We focused later on studying the effect of new gear dimensions by developing metamodels for temperature-time evolution by using data from the reference geometry and the same DoE. The first approach consists in approximating the full temperature-time evolution by using the beginning of the curves. Indeed, the nonlinear ANN regression method was adopted and 4 models with different input percentages (5%, 10%, 15%, 20%) were developed. Those models were trained by the available data from the reference geometry and then tested on 4 new gear geometries. This approach allows to accurately approximate the temperature until the middle time step for the 4 tested cases. However, it is worth pointing out that the more input data we have, the better approximation we get. On the other hand, the approximations at the last time step are less accurate than those obtained at the middle time step. This is because some curves have similar or very close input data curves, but different ends for the training data which generate a lack of accuracy.

The second approach for geometrical analysis relies on the use of another intermediate physical quantity to model the temperature. This quantity is the heat source which is the principal cause of heating according to the heat equation. Indeed, models for the area of the heat source evolution curves were first developed and then used for modeling the temperature field by using the sPGD method. It was demonstrated that we can accurately predict the temperature everywhere in the gear surface and for any new geometry, only by considering the data of the reference geometry for training the models. However, modeling the area of the heat source curve step-by-step by using only few initial values of this latter, obtained by running the simulation for a new geometry with 5% of the original time process, remains a challenge and gives rise to errors that certainly affect the approximation of the temperature.

Finally, we would like to mention that programming several scripts also represents an important part of the thesis outcomes. Those scripts for data-preprocessing, dimensionality reduction, metamodeling, and validation by using many methods and tools such as GPOD, LLE, regression methods (SVR, sPGD, ANN, etc.), and metrics represent a valuable tool for future use in industry.

7.2 Perspectives

In the continuity of this work, several suggestions could be put forward:

- Our study focuses on the single frequency approach, however, extending the applied methodology to address the dual-frequency approach which is more suitable for the contour hardening of gears could be interesting. Given that all the current work research is based on numerical data provided by FE simulation, the development of a FE model is required and the experimental validation of the latter is highly recommended;

- To ensure a better heat distribution over the gear during the heating, a rotating workpiece around an axis could be adopted. In that case, we pass from a static case to a dynamic one which is harder to manipulate experimentally and numerically. Experimentally, the use of thermocouples to measure the temperature evolution becomes extremely difficult. Numerically, extra parameters and operators should be added to the model to take into account the dynamic case and a suitable mesh should be adopted as well. From a MOR point of view, this represents an interesting challenge.

- Hybrid twin paradigm: in this thesis, metamodels were developed based on synthetic data from numerical simulation. Those data were generated by solving PDEs describing the physics. It is commonly known that most physical models rely on hypotheses and simplifications which can generate a gap between the numerical and real solutions. In this context, the hybrid twin paradigm allows to enrich the metamodels with data from experiments. Such approach can significantly increase the approximation accuracy. Therefore, this could help to re-adjust the metamodels and generate confidence intervals.

- Even though a one-shot LHS DoE representing a model-free design was adopted in this thesis, an adaptive sampling represents an interesting option and can enhance the metamodels. As previously stated, with a model-free design, the parametric space is sampled with a given number of experiments a priori defined and without relying on information provided by the evaluation of the QoI. However, the adaptive design uses information given by the evaluations of the QoI to fulfill the initial design space by proposing relevant new samples to improve the model.

- In this work, the temperature and austenite phase evolution were considered as physical QoI to be approximated, however, the inverse analysis can be achieved as well. Indeed, with this latter

approach, we can identify the process or/and material parameters that provide the desired QoI. Several approaches have been developed in the published literature and can be extended to our multiphysics IH process.

- The analysis of the whole induction hardening process including the heating and quenching parts with MOR techniques is an interesting study focus. We could benefit from the results of the heating part to study the second part and provide metamodels for hardness and residual stress profiles where mechanical fields are more impactful.

- Develop an industrial tool that enables the use of the developed models and tools for real-time visualization and prediction of the temperature field evolution and the austenite phase, locally and globally over the gear.

Résumé étendu

Contexte industriel et scientifique

Les industries aéronautique et automobile souhaiteraient à terme optimiser le processus de fabrication des structures mécaniques afin de répondre aux exigences industrielles en termes de coûts et de préoccupations environnementales, tout en conservant des propriétés et des performances mécaniques optimales. La production de composants industriels optimaux avec des propriétés mécaniques intéressantes nécessite un choix judicieux du matériau, cependant, le matériau le plus adapté est souvent coûteux. Par conséquent, afin de garantir un bon compromis entre la qualité et le coût du matériau, les traitements thermiques représentent une excellente option pour conférer au matériau une microstructure et propriétés mécaniques optimales correspondant aux différents critères de performance prédéfinis (Biasutti et al., 2012; Rudnev et al., 2017). Cependant, pour de nombreuses applications, seules les propriétés superficielles du matériau jouent un rôle important. Dans ce contexte, les traitements thermiques de surface par voie mécaniques, thermiques ou thermochimiques sont particulièrement adaptés pour améliorer la résistance à la fatigue et la résistance aux charges mécaniques externes imposées en modifiant les propriétés des zones critiques (Rudnev et al., 2017).

Sur un marché de plus en plus concurrentiel, les fournisseurs industriels français souhaitent remplacer les traitements thermochimiques conventionnels tels que la cémentation et la nitruration (Czerwinski, 2012) par des solutions plus économiques et plus respectueuses de l'environnement, avec une variabilité moindre. Une alternative intéressante aux traitements thermochimiques est le traitement thermique par induction. Ce dernier est généralement considéré comme la dernière étape de la chaîne de fabrication et est l'un des traitements thermiques les plus attrayants, largement utilisés dans les industries automobile et aérospatiale, notamment pour le procédé de durcissement superficiel (Barrere, 1992; Pascal, 2003; Candeo et al., 2011; Rokicki, 2017).

Durcissement superficiel par induction

Le durcissement superficiel par induction consiste à chauffer rapidement une pièce par induction électromagnétique jusqu'à une température au-delà de la plage de transformation de phase (austénitisation) et puis à la refroidir rapidement (trempe). L'austénitisation permet à la microstructure initiale de l'acier de se transformer progressivement en une solution solide de carbone appelée fer gamma (γ-Fe) ou austénite. Lors de la trempe, le refroidissement rapide de la pièce, portée à haute température, par douche ou immersion dans un fluide agité a pour effet de piéger le carbone dans le réseau cristallin du fer, formant ainsi la martensite. Par conséquent, le principal effet bénéfique de la trempe par induction est d'introduire une phase de martensite à grain fin ainsi qu'un champ de contraintes résiduelles compressives, induites sous l'action de la déformation plastique non uniforme induite par le gradient thermique et la variation de volume due à la transformation de phase (Durban, 1997; Denis, 1997; Grum, 2007; Rudnev et al., 2017), dans la couche superficielle sans affecter la métallurgie du matériau à coeur. En effet, la présence de la première améliore la dureté, la résistance à l'usure et la résistance à la fatigue de contact, tandis que la présence de la seconde agit pour inhiber l'initiation des fissures. En effet, le procédé de trempe par induction présente l'avantage d'offrir un temps de chauffe superficiel très court, un contrôle précis de la zone traitée, une bonne aptitude à l'intégration dans la chaîne de production, des possibilités étendues quant aux dimensions des pièces à traiter, de bonnes performances en fatigue, une bonne reproductibilité, et un mode opératoire compatible avec des exigences environnementales sévères, par rapport aux traitements thermochimiques (Chatterjee, 2003).

Bien que les étapes de chauffage et de refroidissement soient importantes et nécessitent une bonne maitrise pour avoir des performances mécaniques meilleures, l'étape de chauffe est d'un intérêt majeur et représente la partie la plus critique. En effet, chauffer une pièce sur une profondeur limitée et obtenir le champ de température désiré est une tâche délicate en raison des phénomènes de conduction thermique qui pourraient modifier les propriétés métallurgiques du matériau sous la couche durcie.

Dans ce manuscrit, l'étape de chauffage par induction a été considérée afin de l'optimiser et d'apporter une meilleure compréhension de cette dernière. En effet, la technologie d'induction suscite un intérêt croissant dans les applications industrielles (Lozinskii, 1969), médicales (Stauffer et al., 1984) et domestiques (Moreland, 1973) grâce à ses bonnes performances et à ses avantages par rapport aux autres techniques de chauffage classiques. De plus amples détails sur ses applications, ses développements et

ses défis sont fournis dans (Lucía et al., 2013).

Chauffage par induction

Principe du chauffage par induction

Le chauffage par induction est une application directe de deux lois physiques fondamentales : La loi de Lenz et l'effet Joule. En effet, en alimentant un inducteur avec un courant électrique alternatif de fréquence donnée (souvent entre 50 Hz et quelques centaines de kHz), un champ magnétique est créé autour de l'inducteur. Lorsqu'un matériau électriquement conducteur est plongé dans ces champs, il est traversé par un flux magnétique dont les variations induisent, selon la loi de Lenz, une force électromotrice qui donne naissance à des courants de Foucault (Gié Hubert, 1982). Ces courants induits produisent de la chaleur par effet Joule et la pièce s'échauffe. Le principe de chauffage est présenté dans la figure 1. Ce mécanisme est la seule source de production de chaleur pour les matériaux non magnétiques ou paramagnétiques (l'aluminium, le cuivre, et les aciers dont la température est supérieure à leur température de Curie par exemple). Pour les matériaux ferromagnétiques, la chaleur est également produite par la présence d'un effet d'hystérésis dans la courbe d'aimantation (induction magnétique B en fonction du champ magnétique H). Ce second mécanisme est généralement d'un effet beaucoup plus faible que le premier, mais le rapport reste difficile à quantifier car il est très dépendant du matériau et des paramètres du procédé.

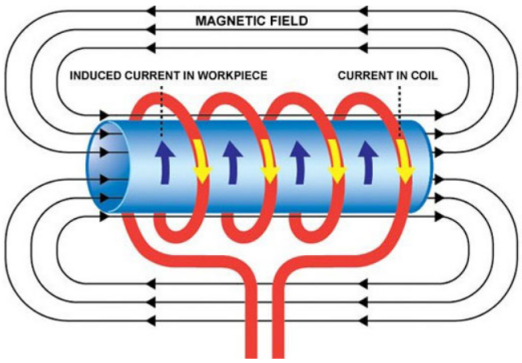


Figure 1: Principe du chauffage par induction (DUOLIN)

Effet de peau

Lorsqu'une pièce en matériau conducteur est traversée par un courant alternatif, la distribution du courant sur la section traversée n'est pas uniforme. La valeur maximale de la densité de courant est toujours située à la surface de la pièce et diminue à l'intérieur de celle-ci. Ce phénomène de non-

uniformité de la distribution du courant à travers une section du conducteur est appelé effet de peau. La zone de génération de chaleur est concentrée dans une fine couche sous la surface de la pièce. La profondeur de pénétration appelée δ est la quantité qui permet de quantifier l'importance de ce phénomène, elle est définie de manière habituelle comme la profondeur où la densité de courant est e-1 fois sa valeur à la surface, Figure 2. La formule théorique dans l'Eq.(1) présente l'ordre de grandeur de l'épaisseur de la peau (Lamb, 1883).

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}} \tag{1}$$

où f est la fréquence du courant, ρ est la résistivité électrique du matériau, et μ sa perméabilité magnétique. Il est facile d'interpréter que plus la fréquence est élevée, plus la profondeur de peau est faible. En effet, cette quantité est d'une importance primordiale pour le chauffage par induction, en particulier lorsqu'il est appliqué à des matériaux ferromagnétiques. Dans ce cas, la profondeur de pénétration n'est pas constante pendant le chauffage mais tend à varier en raison du champ magnétique et des propriétés électromagnétiques du matériau qui dépendent de la température.

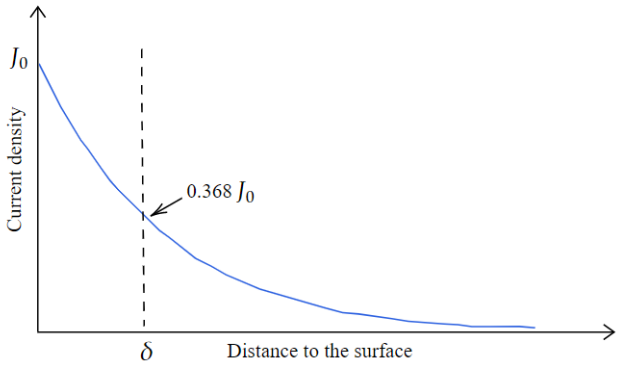


Figure 2: Représentation de la profondeur de peau

Effets de bord et d'extrémité

L'hétérogénéité de la température dans une pièce chauffée par induction est due à l'effet de peau pour les gradients entre la surface et le cœur, et aux effets de bord et de pointe pour les gradients de surface. Si on considère une pièce cylindrique immergée dans un champ homogène (comme à l'intérieur d'un long solénoïde inducteur), la présence de la pièce conductrice induit une distorsion du champ électromagnétique (Figure 3). Ces distorsions du champ, qui entraînent des variations de la densité de courant induit et des distributions de puissance, sont appelées "effets de bord". Dans les cas industriels réels, les inducteurs solénoïdes peuvent rarement être considérés comme infiniment

longs. Dans ces conditions, il existe un espace entre les lignes de champ à chaque extrémité : c'est l'effet d'extrémité. Les effets d'extrémité et de bord sont très similaires et s'influencent souvent l'un l'autre. Ils sont essentiellement régis par quatre variables : l'effet de peau, la profondeur de la pièce traversée par l'inducteur (si elle ne traverse pas complètement l'inducteur), le rapport entre le diamètre intérieur de l'inducteur et le diamètre de la pièce, et le facteur d'espacement des spires. Ce dernier facteur représente le rapport entre l'espace entre deux spires et la largeur d'une spire dans l'inducteur. Dans le cas d'un inducteur à une seule spire, le facteur est égal à 1. Une mauvaise combinaison de ces quatre facteurs peut entraîner une surchauffe ou une souschauffe de l'extrémité de la pièce. Il est donc nécessaire de les contrôler afin d'obtenir la distribution du champ de température souhaitée.

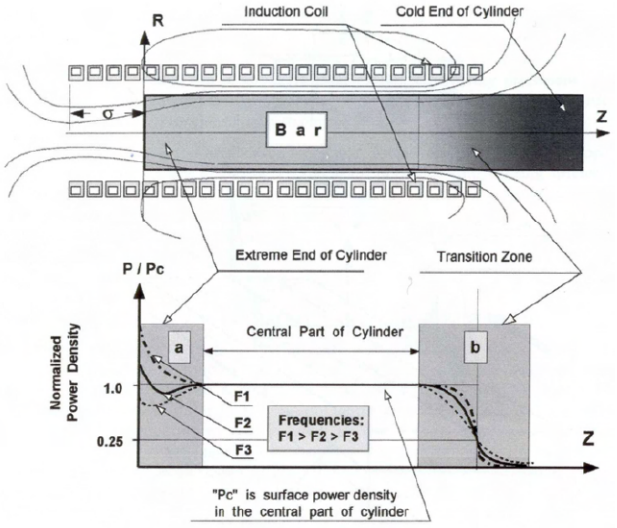


Figure 3: Lignes de champ et distribution de puissance le long d'un cylindre immergé dans un champ magnétique créé par un inducteur solénoïde (Rudnev et al., 2017).

Problématiques et défis

L'optimisation du procédé de chauffage par induction permet d'obtenir de meilleures caractéristiques de chauffage et, par conséquent, de meilleures propriétés mécaniques après la trempe. Cependant, la principale difficulté derrière le contrôle et l'optimisation du procédé est liée à sa multidisciplinarité. En effet, il fait intervenir plusieurs domaines de la physique (électromagnétisme, transfert de chaleur, métallurgie et mécanique). Cette nature multiphysique combinée à des vitesses de chauffage très élevées (supérieur à 2000°C/s pour cette étude) montre une grande complexité pour obtenir des composants avec les propriétés mécaniques souhaitées.

La littérature montre que des méthodes expérimentales, analytiques et numériques ont été développées

pour étudier le procédé de chauffage. Les objectifs et les défis qui motivent le développement de ces méthodes sont les suivants :

- réduire le temps de développement ;
- réduire les coûts en faisant bon produit dès la première fois, en réduisant les “coûts non récurrents” tels que les coûts des outils et des prototypes et les “coûts récurrents” tels que les coûts de production ;
- comprendre les mécanismes physiques impliqués afin de mieux contrôler les paramètres d’influence et de pouvoir ainsi mieux prédire les quantités physiques d’intérêt;
- anticiper les éventuels défauts et risques de non-qualité.

Il convient de souligner que nous nous intéressons, dans le cadre de notre projet, aux composants de transmission de puissance, pignons d’engrenage en particulier, qui sont considérés comme des pièces de géométrie complexe. Ils sont couramment utilisés depuis plus de trois mille ans dans presque tous les secteurs d’activités ; automobile, aéronautique, naval, horlogerie, et autres. Cependant, en fonction de leur domaine d’utilisation, le contrôle de la dissipation thermique à l’intérieur du pignon et le choix de la stratégie thermique la plus appropriée doivent être réalisés. Plusieurs paramètres doivent donc être pris en considération :

- la position relative de l’inducteur par rapport à la pièce (entrefer, longueurs respectives, etc.) ;
- le courant fourni à l’inducteur ;
- la fréquence d’alimentation et l’effet de peau qui caractérisent la distribution des courants induits dans la pièce : plus la fréquence augmente, plus les courants induits sont concentrés sur la surface. Typiquement, les inducteurs sont alimentés par des courants alternatifs dont la fréquence varie de quelques dizaines de Hertz à plusieurs centaines ou milliers de Hertz. En effet, des générateurs à double fréquence (Esteve et al., 2006; Zgraja, 2019) ont également été développés pour alimenter l’inducteur, séquentiellement ou simultanément, avec deux fréquences différentes afin d’obtenir des profondeurs de pénétration différentes et de permettre le chauffage du contour. Cette approche est utilisée pour la trempe de pièces à géométrie de surface irrégulière comme les pignons (Rivat, 2002) de sorte que les moyennes fréquences (3 à 10 kHz) favorisent le chauffage du pied de dent, tandis que les hautes fréquences ($<$ 400 kHz) favorisent le chauffage du sommet et du flanc de dent (voir Figure 4) ;

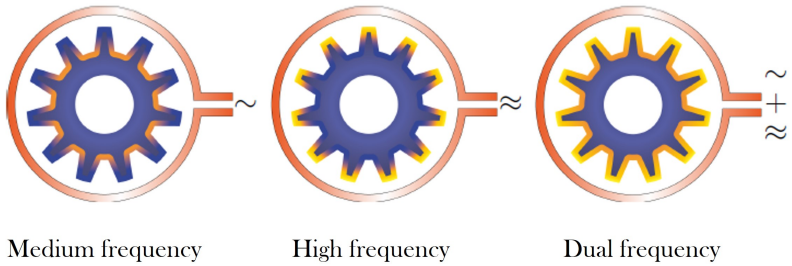


Figure 4: Effet de la fréquence sur le procédé de chauffage par induction (adapté de (Wolfgang and Hans-Joachim, 2002))

- le temps de procédé ;
- le type d'inducteur (géométrie, taille, technologie). La géométrie des inducteurs peut être très variée, allant d'un monospire à des inducteurs complexes multispires ;
- la géométrie du pignon. Les roues dentées peuvent avoir un diamètre allant de quelques dizaines de millimètres dans l'automobile à quelques mètres dans le naval par exemple, et un module allant de 1.5 à 12 mm ;
- le matériau et la microstructure initiale du pignon à chauffer.

Le développement du procédé repose traditionnellement sur l'expertise industrielle basée sur une approche d'essais et d'erreurs. En effet, de nombreuses études portant sur l'analyse de la sensibilité, la sélection des paramètres optimaux du procédé, et l'optimisation pour des objectifs bien définis (maximisation, minimisation ou prévision d'une quantité physique d'intérêt) ont été réalisées pour la trempe par induction (chauffage suivi d'une trempe). Dans Misra et al. (2014), la profondeur effective (la distance vers l'intérieur de la surface de la pièce jusqu'à une dureté spécifique) et les valeurs de dureté ont été analysées pour différents paramètres dans des plages fixes de moyenne fréquence, de vitesse d'avance, de pression de trempe, et de température. Un modèle de régression a été développé pour prédire ces quantités physiques en se basant sur la méthode des surfaces de réponse (MSR) couplée à un plan d'expérience de type composite centré. Dans Onan et al. (2015), des études d'optimisation ont été menées pour évaluer l'effet de l'alimentation électrique, de la vitesse de balayage, et de l'entrefer entre la pièce et l'inducteur sur le contour de durcissement induit. Le plan d'expérience orthogonal L27 de Taguchi a été adopté pour les essais expérimentaux. Les données obtenues ont été analysées à l'aide de l'analyse de la variance (ANOVA). D'autres études visant à optimiser le procédé en évaluant l'effet de des paramètres influents ont été développées, voir par exemple (Kohli and Singh, 2011).

Cependant, l'approche expérimentale est non seulement longue, mais elle nécessite un nombre important d'essais expérimentaux pour une plage de validation restreinte, compte tenu des diverses conditions et paramètres contrôlant le procédé. Même si les phénomènes physiques décrivant le procédé sont relativement bien connus, il est difficile de déterminer leurs effets sur des formes complexes de composants industriels ou d'inducteur. En effet, la distribution de la chaleur peut varier d'une zone à l'autre, ce qui entraîne des altérations métallurgiques et mécaniques qui compliquent l'optimisation du procédé.

Cela a ouvert la voie à des études analytiques du procédé. De nombreux modèles analytiques ont été développés au cours du dernier siècle. Les premiers modèles analytiques pour le calcul des champs électromagnétiques et thermiques unidimensionnels et même bidimensionnels dans le système de chauffe ont été développés par Vologdin (1939), suivi par Rodigin, Curtis, Stansel et d'autres (Curtis, 1944; Stansel, 1949; Rodigin, 1950; Brunst, 1957; Simpson, 1960; Slukhotskii and Ryskin, 1974). Les solutions analytiques sont basées sur de nombreux concepts tels que l'expansion de série, la transformation de Fourier, la séparation des variables, les fonctions de Bessel et d'autres fonctions spéciales. Les modèles analytiques reposent sur des analyses mathématiques complexes et moins coûteuses, mais basés sur des simplifications et des hypothèses qui affectent leurs précisions lorsqu'ils sont appliqués à des géométries complexes et avec des propriétés matériaux non linéaires. Dans ce cas, les techniques numériques constituent une approche prometteuse pour résoudre les modèles physiques linéaires et non linéaires liés aux couplage électromagnétiques, thermiques, et les autres physiques impliquées dans le procédé. Ces techniques peuvent être appliquées à n'importe quelle géométrie complexe avec des propriétés matériaux linéaires ou non linéaires. Parmi ces méthodes, nous citons la méthode des différences finies pour résoudre les problèmes de transfert de chaleur (Özişik et al., 2017), la méthode des éléments de frontière pour calculer les champs électromagnétiques tridimensionnel (3D) à haute fréquence pendant le chauffage par induction dans (Kagami and Fukai, 1984; Muhlbauer et al., 1993), la méthode des volumes finis pour résoudre les problèmes électromagnétiques dans (Beckstein et al., 2017), et la méthode des éléments finis (MEF) (Hutton, 2003) qui est bien connue et largement expérimentée pour résoudre tous les aspects physiques inclus dans le procédé de chauffage (Pascal, 2003; Cardinaux, 2008; Aliferov et al., 2010; Spezzapria et al., 2012; Jin, 2015).

Le développement de ces méthodes ainsi que l'amélioration des capacités informatiques ont conduit

au développement de nombreux outils de simulation numérique avancés. Ces outils sont capables de modéliser le procédé de chauffe induction, de résoudre les équations physiques derrière et de fournir un champ thermique précis, à la fois dans l'espace et dans le temps. Par conséquent, une meilleure compréhension des phénomènes physiques et de leurs interactions est assurée. Plusieurs travaux sur la simulation du procédé ont été réalisés dans la littérature. Des géométries de pièces simples avec une seule fréquence ont été d'abord étudiées, comme dans (Jacot et al., 1996) pour des pièces à symétrie axiale, en utilisant la MEF et la méthode des éléments de frontière pour résoudre le problème électromagnétique. Dans Bay et al. (2003), un modèle combinant les phénomènes électromagnétiques, thermiques, et mécaniques pour une pièce à symétrie axiale a été développé. Ensuite, grâce à l'approche bi-fréquence fournie à un inducteur commun, le traitement de contour sur des pièces de forme complexe telles que les pignons a été réalisé (Petzold, 2014). En outre, les logiciels FLUX 2D, FLUX 3D, et FLUX expert ont été utilisés dans (Delaleau et al., 2000) pour résoudre les équations électromagnétiques et thermiques d'un modèle de pignon 3D, mais sans tenir compte des transformations métallurgiques et des déformations mécaniques. Dans Mingardi (2013), un modèle 3D de pignon droit en acier (AISI 4340) a été développé pour simuler la chauffe à l'aide du logiciel FLUX 3D. L'effet des paramètres machine (fréquence moyenne pour le préchauffage, dimensions de l'inducteur, et présence d'un concentrateur de flux au-dessus et au-dessous du pignon) a été évalué sur la distribution finale de la température afin d'obtenir un contour de durcissement homogène et d'empêcher l'austénitisation de la dent entière. Il a été démontré que les concentrateurs de flux ont un rôle important dans la réduction des effets de bord et que la fréquence moyenne dans la phase de préchauffage (1 à 10 kHz, pendant 1 à 5 secondes) conduit à un contour plus approprié à la fin de la phase de chauffage à haute fréquence (200 kHz, pendant 0.2 seconde). Dans Hömberg et al. (2016), la simulation du durcissement par induction en bi-fréquence, y compris la transformation de phase et les effets mécaniques, a été élaborée pour un modèle de pignon 3D.

Cependant, malgré l'amélioration des capacités informatiques, la modélisation précise et réaliste en 3D reste un véritable challenge, vu le couplage mutuel des phénomènes physiques évoqués précedemment sur des formes géométriques complexes. Bien que les différents quantités physiques liés au procédé puissent être prédits, la modélisation numérique 3D souffre encore de certains inconvénients. Le coût de calcul élevé, la prise en charge de l'approche bi-fréquence, et la communication de données entre les solveurs, pour n'en citer que quelques-uns (Li et al., 2013; Sumithra and Thiripurasundari, 2017).

En outre, des analyses de sensibilité aux paramètres procédé, des prédictions des quantités physiques d'intérêt, et d'identification des paramètres optimaux du procédé ont été réalisées dans (Barglik et al., 2014; Senhaji, 2017) à l'aide de la simulation numérique. Les plans d'expériences utilisés, tels que le plan de Taguchi ou le plan factoriel, sont généralement construits en définissant dans un premier temps les paramètres d'entrée et leurs niveaux (2 ou 3 niveaux en général) et en construisant ensuite ce que l'on appelle la matrice d'expériences en envisageant différentes combinaisons possibles des niveaux imposés des paramètres. Ces essais sont ensuite simulés numériquement via des logiciels de calcul tels que FORGE®, ANSYS, COMSOL multiphysics et FLUX. Ensuite, l'ANOVA et la MSR sont effectuées sur les résultats obtenus à des fins d'optimisation. Ces techniques permettent de réduire le nombre d'expériences nécessaires à l'optimisation et de créer une surface de réponse en utilisant des polynômes d'ordre inférieur. Cependant, en fonction du nombre de paramètres, de leurs niveaux, et des interactions possibles entre eux, le nombre de simulations à effectuer peut être très élevé. Par conséquent, cette approche devient inefficace en ce qui concerne la procédure d'optimisation.

Une solution prometteuse, investiguée dans cette thèse de doctorat, consiste à utiliser les techniques de réduction de l'ordre des modèles (ROM), qui peuvent exprimer la solution d'un problème donné (exprimé par une équation différentielle partielle (EDP), par exemple) dans une base réduite qui assure une description satisfaisante du contenu physique ou mathématique du système. Cette base est souvent extraite de solutions expérimentales ou numériques du problème en question. Cette ROM peut être réalisée par la méthode de "proper orthogonal decomposition" (POD) ou par la méthode des bases réduites (Chinesta et al., 2017), par exemple. Ces techniques permettent de réduire considérablement la complexité de la résolution, de sorte que la complexité de la solution dépend de la taille de la base réduite plutôt que de la taille du modèle initial, résolu par la MEF dans la plupart des cas.

Même si l'utilisation d'une base réduite implique une certaine perte d'informations, elle permet des gains de temps de calcul impressionnants en fournissant des prédictions rapides. En outre, la solution prédite pour un problème donné est suffisamment précise tant que le cas considéré reste dans l'espace engendré par la base réduite. Cependant, une faible précision est attendue lorsque la solution requise ne peut pas être approximée de manière précise dans l'espace engendré par la base réduite. Pour améliorer la généralité tout en garantissant la précision, une technique intéressante, appelée "proper generalized decomposition" (PGD), permet de résoudre des problèmes complexes en construisant une représentation réduite du système étudié, tout en maintenant un niveau élevé de précision. La PGD re-

pose sur l'idée de décomposer la solution recherchée en une série de produits de fonctions élémentaires. Chaque fonction élémentaire est ensuite approximée indépendamment, ce qui permet une réduction significative de la dimension du problème (Chinesta et al., 2013). Cependant, malgré son efficacité, la mise en oeuvre de cette méthode peut être très intrusive, nécessitant la modification du code source du modèle physique étudié. Cela peut limiter son utilisation dans certains cas, notamment lorsque le code source du modèle est propriétaire ou indisponible.

Pour éviter ce caractère intrusif, des techniques non intrusives ont été proposées. Elles consistent à construire la solution du problème considéré (un problème paramétrique en général) en utilisant plusieurs solutions de haute fidélité générées pour différentes valeurs des paramètres du modèle selon un plan d'expériences. Parmi ces techniques, on peut mentionner la méthode POD avec interpolation (PODI) (Bui-Thanh et al., 2003a), où les méthodes de régression ou d'interpolation habituelles sont utilisées pour exprimer la dépendance des coefficients modaux de la POD par rapport aux paramètres d'entrée considérés. En effet, les régressions, qui font partie de "machine learning" (ML) supervisé, sont largement utilisées en ingénierie pour développer des modèles (Brunton and Kutz, 2019; Chinesta et al., 2020; Sancarlos et al., 2021; Sun et al., 2022). Par ailleurs, la méthode "sparse subspace learning " (SSL) (Borzacchiello et al., 2017) fait interpoler les solutions précalculées en considérant une base d'approximation hiérarchique, selon un échantillonnage hiérarchiquement structuré sur l'ensemble de l'espace paramétrique. La SSL peut fournir une bonne solution paramétrique, cependant, le volume de données, qui augmente de manière exponentielle avec le nombre de paramètres d'entrée et le niveau d'approximation hiérarchique, reste une vraie problématique. D'autre part, une technique de ROM parcimonieusement échantillonnée, appelée "sparse PGD" (sPGD), a été proposée dans (Ibañez et al., 2018).

Les modèles, développés à partir de ces techniques non intrusives basées sur les données et sans connaissance préalable des physiques, sont souvent appelées modèles de substitution ou métamodèles. Ces derniers permettent la simulation en temps réel, l'optimisation, l'analyse inverse, et le contrôle du procédé, au stade en ligne. Ainsi, du point de vue de l'ingénierie et de l'intelligence artificielle, les données peuvent apporter des connaissances de différentes manières :

- en produisant des modèles basés sur les données, reliant les sorties d'intérêt à des entrées non corrélées ;

- en enrichissant les modèles existants basés sur des lois physiques bien établies ;
- en extrayant des motifs présentant un contenu informationnel élevé ;
- en classifiant les tendances et les variables ;
- en réduisant la dimensionnalité, notamment lorsqu'un problème de grande dimension est traité ;
- en extrayant les corrélations existantes et en ne conservant que les informations utiles, suffisantes et explicatives.

Toutefois, plusieurs questions et défis découlent de ces développements :

- le choix de la stratégie d'échantillonnage pour obtenir les modèles les plus informatifs ;
- la variabilité des données ;
- le choix des techniques d'interpolation ou de régression pour une meilleure estimation de la réponse pour de nouvelles valeurs de paramètres (non utilisées lors de l'apprentissage). Il existe de nombreuses possibilités (approximations polynomiales, krigeage, méthodes de ML, etc.), cependant, il n'y a aucune garantie que la méthode choisie fournie de bons résultats sur les données étudiées ;
- la quantité de données à utiliser, étant donné que la collecte et le traitement des données sont souvent très coûteux. Par conséquent, le paradigme de "smart data" devrait remplacer les procédures basées sur les "big data" pour de nombreux cas d'application. Cependant, l'application de la régression sur peu de données montre une difficulté majeure quand il s'agit de problèmes multi-paramétriques avec des comportements non linéaires. Celui-ci conduit à ce que l'on appelle "curse of dimensionality", c'est-à-dire la croissance exponentielle du nombre de degrés de liberté et, de manière équivalente, du nombre de points d'échantillonnage nécessaires (ou points d'entraînement dans la terminologie ML) dans les espaces paramétriques ;
- l'imputation des valeurs manquantes. Pour certaines procédures de modélisation, les données manquantes doivent être complétées pour fournir une cartographie globale ou pour déduire des mesures dans des régions où les mesures ne peuvent pas être directement effectuées. De nombreux chercheurs travaillent sur les problèmes de données manquantes depuis des décennies, et de nombreuses approches statistiques (Yates, 1933; Little and Rubin, 2019) et non statistiques

basées sur la décomposition modale (Everson and Sirovich, 1995) ont été utilisées pour fournir une reconstruction suffisamment précise des données manquantes.

- etc.

Les listes ci-dessus révèlent que l'utilisation des données, aujourd'hui et à l'avenir, diffère radicalement de son utilisation passée. Les nouveaux développements exigent des compétences en matière de collecte, d'exploration et d'analyse de données. C'est pourquoi la plupart des défis mentionnés sont en développement continu.

Objectifs

Cette thèse s'inscrit dans le cadre d'un projet de recherche lancé en 2019 par l'Institut de Recherche Technologique (IRT) Matériaux Métallurgie Procédés (M2P) qui s'intitule : TRANSmission FUture GÉnération (TRANSFUGE). L'objectif global de ce travail de recherche, qui rassemble des partenaires industriels[1] et s'appuie sur l'expertise du laboratoire PIMM des Arts et Métiers ParisTech, est d'optimiser le procédé de chauffage par induction en développant une méthodologie basée sur les techniques de réduction de modèles et de métamodélisation paramètrique pour générer des solutions paramétriques pour certaines quantités physiques. Ses solutions paramétriques (ou métamodèles) doivent fournir des prédictions très rapides en temps réel des résultats avec un coût plus faible et une précision suffisante.
En particulier, les objectifs spécifiques de ce travail de recherche sont les suivants :

- développer un modèle numérique pour la simulation de chauffage par induction en utilisant le code de calcul EF FORGE®.

- développer des métamodèles locaux pour l'évolution de la température et de la phase austénitique, au niveau de quelques capteurs répartis sur la denture du pignon, à partir de données synthétiques collectées via des simulations numériques pour différentes valeurs de paramètres du procédé.

- proposer une méthodologie pour étendre la solution paramétrique connue à certaines positions de capteurs afin de traiter l'ensemble de la zone affectée thermiquement (ZAT).

[1] Safran, Stellantis, Renault, Naval Group, NTN-SNR, Valeo, Ascometal

- étudier l'effet des variations géométriques du pignon sur l'évolution de la température pendant le procédé.

Bilan des travaux

Une revue de la littérature a d'abord été réalisée pour donner un aperçu des méthodes de réduction de la dimensionnalité. Des méthodes intrusives (POD basée sur la projection, PGD) et non intrusives (PODI, sPGD, SSL) de ROM ont été présentées. Toutefois, compte tenu de la complexité du procédé et de l'inaccessibilité au code source EF décrivant le procédé dans les logiciels commerciaux, l'approche non intrusive a été choisie pour être appliquée. Dans cette thèse, des données synthétiques fournies par des codes EF ont été générées via le logiciel FORGE® et l'approche PODI a été employée afin de construire les solutions paramétriques pour les quantités physiques d'intérêt. Étant donné que l'approche PODI est couplée à des méthodes d'interpolation et de régression, un aperçu de plusieurs méthodes de ML et d'interpolation a également été réalisé.

En outre, les phénomènes physiques mis en jeu, leurs formulations mathématiques, leurs interactions mutuelles et les paramètres inclus ont été évoqués dans ce travail pour une meilleure compréhension du procédé. Ensuite, la simulation numérique par éléments finis (EF) en utilisant le code FORGE® a été réalisée de telle sorte que le développement du modèle a été illustré étape par étape : création des géométries, stratégie de maillage avec une étude de convergence, définition des propriétés des matériaux, définition des conditions aux limites et des paramètres du procédé, et enfin le paramètre de couplage. Les résultats numériques obtenus pour la distribution de la température et de la phase austénitique démontrent l'effet de peau ainsi que la différence entre les résultats aux extrémités et à mi-largeur du pignon. A partir du modèle EF, un ensemble de 20 solutions synthétiques (snapshots) pour différentes valeurs de fréquence et de puissance, choisies dans des intervalles prédéfinis, a d'abord été généré selon un plan d'expériences de type hypercule latin. Ensuite, des métamodèles paramétriques pour le procédé ont été développés, pour 15 capteurs épars situés à différentes positions sur la denture du pignon, afin de prédire l'évolution de la température et de la phase austénitique.

Pour modéliser le champ de température, deux méthodologies ont été proposées. La première consiste à tronquer tous les snapshots de sorte que seules les données inférieures à une valeur de température finale constante soient prises en compte, et le temps en fonction de la température a ensuite été mod-

élisé. La seconde méthodologie proposée consiste à compléter les valeurs manquantes afin d'atteindre une instant de temps final constant pour certains snapshots en utilisant la méthode "Gappy POD" (GPOD). Pour les deux méthodologies, trois méthodes de régression (sPGD, SVR et GB) ont été appliquées afin de créer des métamodèles pour la représentation à faible dimension des snapshots initiales fournies par POD. Les résultats obtenus indiquent que les méthodes sPGD et SVR fournissent des modèles relativement précis pour prédire les évolutions de la température et du temps avec les valeurs imposées des paramètres d'entrée, même avec peu de données. Par contre, la méthode GB génère des résultats beaucoup moins précis avec une quantité réduite de données.

Pour modéliser la phase austénitique, les instants de début et de fin d'austénitisation (t_Ac1 et t_Ac3) ont été extraits et modélisés à l'aide des trois méthodes de régression. La meilleure approximation a été fournie par la sPGD, suivie par la SVR avec une prédiction légèrement moins précise, et la prédiction la moins précise a été fournie par la méthode GB.

Les approches développées présentent l'avantage d'une prédiction en temps réel qui convient parfaitement au contrôle actif des paramètres du procédé. Pour aller plus loin, l'extension des approximations, obtenues au niveau de 15 capteurs épars, a été réalisée pour traiter toute la ZAT. Pour atteindre cet objectif, différentes approches ont été proposées pour l'approximation de la température et de la phase austénitique. Pour le champ de température, deux approches ont été adoptées : la première peut être appliquée à une seule configuration (simulation avec valeurs données de fréquence et de puissance) et repose sur une réduction de dimensionnalité non linéaire couplée à la POD avec interpolation à fonctions de base radiale (FBR). En effet, une interpolation sur la surface de la denture du pignon a d'abord été abordée en appliquant la méthode "locally linear embedding" (LLE) pour représenter les coordonnées cartésiennes 3D dans un sous-espace 2D tout en préservant le voisinage. Celui-ci permet de récupérer la structure non linéaire globale de la denture du pignon à partir d'ajustements localement linéaires. La méthode LLE a été suivie par la méthode POD avec interpolation FBR sur la matrice des snapshots contenant l'évolution de la température en fonction du temps sur les capteurs de surface disponibles. L'interpolation a été réalisée par rapport aux coordonnées 2D obtenus par LLE. Cette approche a été testée et approuvée pour de nouveaux points dans l'espace.

Ensuite, des points en volume ont été pris en compte pour traiter l'interpolation en volume. Pour ce faire, des interpolations locales par rapport à la distance à la surface ont d'abord été appliquées

sur un ensemble de points à différentes zones de la pièce en utilisant la méthode POD avec interpolation linéaire. Pour tout nouveau point situé à une certaine profondeur à laquelle la prédiction de la température est requise, plusieurs étapes ont été ensuite appliquées, notamment un algorithme de recherche intelligent, une projection orthogonale sur la surface, un mapping par la LLE, et une POD avec interpolation FBR. De bonnes approximations ont également été obtenues pour une interpolation en volume.

La deuxième approche est basée sur l'application de la GPOD sur un ensemble de snapshots complètement connus. L'objectif est de pouvoir prédire la température en tout point de l'espace pour une valeur donnée de fréquence, de puissance, et d'instant de temps en utilisant la solution connue en quelques positions. Pour ce faire, la POD a d'abord été réalisé sur la matrice des snapshots composée de vecteurs de température à chaque noeud de maillage comme colonnes. La base réduite a ensuite été utilisée pour déterminer les coefficients modaux associés aux nouveaux vecteurs dont seulement 15 éléments sont connus (positions des capteurs). La reconstruction finale du champ complet a été réalisée et évaluée pour certains capteurs en surface et en volume, et de bonnes approximations ont été fournies pour les points de test. Cette deuxième approche est plus générale et peut être appliquée pour prédire la solution même pour de nouvelles valeurs des paramètres d'entrée (fréquence et puissance).

Pour l'évolution de la phase austénitique, des modèles pour t_Ac1 et t_Ac3 ont été construits en considérant les paramètres du procédé (fréquence et puissance) et les paramètres géométriques (coordonnées 2D du LLE et profondeurs des points) comme paramètres d'entrée. Une comparaison entre les résultats des trois méthodes de régression a été menée. Aucune différence significative n'a été observée pour la modélisation des deux instants caractéristiques et les trois méthodes fournissent des modèles d'une précision suffisante, inférieure à 10 %. Cependant, les prédictions effectuées sur des nouveaux points de tests montrent que la sPGD et la SVR sont plus performantes que la méthode GB qui fournit des erreurs élevées (> 10 %).

Étant donné que tous les modèles développés sont attribués à une seule géométrie du pignon, nous nous sommes ensuite intéressés à l'étude de l'effet d'un changement dimensionnel du pignon sur l'évolution de la température, ceci sans passer par un nouveau plan d'expérience, mais en s'appuyant sur les résultats de la géométrie de référence. Pour ce faire, deux approches ont été proposées pour prédire l'évolution de la température dans des nouvelles géométries. La première approche est basée sur le

réseau de neurone en utilisant comme paramètres d'entrée quelques incréments initiaux des courbes temporelles de la température. Quatre modèles avec différents pourcentages de courbes temporelles (5 %, 10 %, 15 %, 20 %) ont été développés. Ces modèles ont été entraînés avec les données disponibles de la géométrie de référence, puis testés sur 4 nouvelles géométries du pignon. Cette approche permet d'approximer avec précision la température jusqu'à la moitié de la durée totale du procédé pour les 4 cas testés. Toutefois, il convient de souligner que la qualité des prédictions dépend de la quantité de données d'entrée utilisée. D'autre part, les approximations au dernier pas de temps sont moins précises que celles obtenues vers mi-procédé. Ceci est dû au fait que certaines courbes ont des débuts similaires ou très proches, mais des fins différentes pour les données d'entraînement, ce qui limite la capacité des modèles à prédire avec précision les températures finales.

La deuxième approche pour l'analyse géométrique repose sur l'utilisation d'une autre quantité physique intermédiaire pour modéliser la température. Cette quantité est la puissance de chauffe qui est la principale cause du chauffage selon l'équation de la chaleur. En effet, des modèles pour l'aire sous les courbes d'évolution de la puissance de chauffe ont d'abord été développés, puis utilisés pour modéliser le champ de température en utilisant la méthode sPGD. Il a été démontré que nous pouvons prédire avec précision la température partout sur la surface du pignon et pour toute nouvelle géométrie, en considérant uniquement les données de la géométrie de référence pour l'entraînement des modèles. Cependant, la modélisation de l'aire de la courbe de la puissance de chauffe d'une manière incrémentale en utilisant seulement quelques valeurs initiales de cette dernière, obtenues en exécutant la simulation pour une nouvelle géométrie avec 5 % du temps de procédé original, reste un défi et donne lieu à des erreurs qui affectent certainement l'approximation de la température.

Enfin, il est à noter que la programmation de plusieurs scripts représente également une partie importante des résultats de la thèse. Ces scripts pour le traitement des données, la réduction de la dimensionnalité, la métamodélisation et la validation, en utilisant de nombreuses méthodes et outils tels que la POD, la GPOD, la LLE, les méthodes de régression (SVR, sPGD, ANN, etc.) et les métriques, représentent un outil précieux pour une utilisation future dans l'industrie.

Perspectives

Dans la continuité de ce travail, plusieurs suggestions peuvent être proposées :

- Notre étude se concentre sur l'approche mono-fréquence, cependant, il serait intéressant d'étendre la méthodologie appliquée pour considérer l'approche bi-fréquence, qui est plus adaptée au durcissement de contour des pignons. Étant donné que tous les travaux de recherche actuels sont basés sur des données numériques fournies par la simulation EF, le développement d'un nouveau modèle EF est nécessaire, et la validation expérimentale de ce dernier est fortement recommandée;

- Pour assurer une meilleure répartition de la chaleur sur le pignon pendant la chauffe, on peut adopter une pièce qui tourne autour d'un axe avec une vitesse donnée. Dans ce cas, on passe d'un cas statique à un cas dynamique qui est plus difficile à manipuler expérimentalement et numériquement. Expérimentalement, l'utilisation de thermocouples pour mesurer l'évolution de la température devient extrêmement difficile. Numériquement, des paramètres et des opérateurs supplémentaires doivent être ajoutés au modèle pour prendre en compte le cas dynamique, et un maillage approprié doit également être adopté. D'un point de vue ROM, cela représente un défi intéressant.

- Application des jumeaux hybrides : dans cette thèse, les métamodèles ont été développés à partir de données synthétiques issues de la simulation numérique. Ces données ont été générées par la résolution des EDPs décrivant les physiques. Il est communément admis que la plupart des modèles physiques reposent sur des hypothèses et des simplifications qui peuvent générer un écart entre les solutions numériques et réelles. Dans ce contexte, le paradigme hybride permet d'enrichir les métamodèles avec des données expérimentales. Une telle approche peut considérablement augmenter la précision de l'approximation. Elle pourrait ainsi réajuster les métamodèles et générer des intervalles de confiance.

- Bien que cette thèse adopte une stratégie d'échantillonnage par hypercube latin représentant une conception indépendante du modèle, un échantillonnage adaptatif représente une option intéressante et peut améliorer les métamodèles. En effet, le hypercube latin échantillonne l'espace paramétrique avec un nombre donné d'expériences définies a priori, sans se fier aux informations provenant de l'évaluation de la quantité physique d'intérêt. Cependant, un échantillonnage adaptatif permet d'ajuster dynamiquement les points d'échantillonnage en proposant de nouveaux échantillons pertinents en fonction des informations obtenues lors des évaluations précédentes de la quantité physique d'intérêt. Cette approche nous permet de nous concentrer sur les régions de l'espace des paramètres les plus susceptibles de fournir des informations importantes

ou de meilleures performances. Elle optimise efficacement l'allocation des ressources informatiques en donnant la priorité à l'échantillonnage dans les zones qui contribuent le plus à réduire l'incertitude ou à améliorer la précision du modèle.

- Dans ce travail, l'évolution de la température et de la phase austénitique ont été considérées comme quantités physiques à approximer, mais des analyses inverses peuvent également être réalisées. Celui-ci permet d'identifier les paramètres du procédé et/ou du matériau qui fournissent les valeurs souhaitées. Plusieurs approches ont été proposées dans la littérature et qui peuvent être étendues à notre procédé multiphysique de chauffage par induction.

- L'analyse de l'ensemble du procédé de trempe par induction (chauffage + trempe) à l'aide des techniques de ROM est un sujet d'étude intéressant. Nous pourrions bénéficier des résultats de la partie de chauffe pour étudier la seconde partie et fournir des métamodèles pour les profils de dureté et de contraintes résiduelles.

- Développer un outil industriel qui permet l'utilisation des métamodèles et des outils développés pour la visualisation et la prédiction en temps réel de l'évolution du champ de température et de la phase austénitique, localement et globalement sur un pignon.

Contributions scientifiques

Les travaux de thèse ont été publiés dans plusieurs revues scientifiques et présentés lors de conférences internationales :

Articles dans des revues internationales

- Derouiche, K., Daoud, M., Traidi, K., Chinesta, F. (2022). Real-time prediction by data-driven models applied to induction heating process. International Journal of Material Forming, 15(4), 1-17.

- 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(5), 738.

Article dans un magazine français

Un article a été publié dans le magazine “Traitements & Matériaux” consacré aux sujets liés à la “Transformation des matériaux métalliques”

- Derouiche, K., Garois, S., Daoud, M., Traidi, K., Chinesta, F. (2021). Approches basées sur la métamodélisation et l’intelligence artificielle pour le traitement thermique par induction. Traitements et matériaux. Dossier Matériaux, surfaces & revêtements, 473, 14 décembre 2021.

Article de Conférence

Un article publié pour la conférence ESAFORM qui a eu lieu en avril 2021. Derouiche, K., Daoud, M., Traidi, K., Chinesta, F. (2021). A non-intrusive model order reduction approach for multi-physics parametrized problems-Application to induction heating process.

Conférences internationales

Au cours de mon doctorat, j’ai participé à deux conférences internationales :

- ESAFORM en avril 2021, Liège, Belgique
- COMPLAS en septembre 2021, Barcelone, Espagne

Plénière globale du projet TRANSFUGE

Cinq présentations ont été réalisées lors des plénières globales de l’IRT-M2P devant les partenaires industriels du projet TRANSFUGE.

Présentation dans une journée thématique

Une journée thématique intitulée “Méthodes avancées en simulation numérique” qui s’est déroulée en avril 2022 au Laboratoire de Mécanique de Paris-Saclay (LMPS), issu de la fusion du LMT et du laboratoire MSSMAT. Cette journée s’inscrit dans le cadre de l’année de la mécanique.

Bibliography

Aizerman, M. A. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control*, 25:821–837.

Akram, F., Batti Fadhil, A., Adil, A., and Essam, A.-B. (2009). Modeling of induction heating process of a conducting ferromagnetic materials. *Journal of Engineering*, pages 1–17.

Aliferov, A., Dughiero, F., and Forzan, M. (2010). Coupled magneto-thermal fem model of direct heating of ferromagnetic bended tubes. *IEEE Transactions on Magnetics*, 46(8):3217–3220.

Ammar, A., Huerta, A., Chinesta, F., Cueto, E., and Leygue, A. (2014). Parametric solutions involving geometry: A step towards efficient shape optimization. *Computer Methods in Applied Mechanics and Engineering*, 268:178–193.

Ammar, A., Mokdad, B., Chinesta, F., and Keunings, R. (2006). A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. *Journal of Non-Newtonian Fluid Mechanics*, 139:153—176.

Ammar, A., Mokdad, B., Chinesta, F., and Keunings, R. (2007). A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. part ii: Transient simulation using space-time separated representation. *Journal of Non-Newtonian Fluid Mechanics*, 144:98–121.

Ammar, A., Normandin, M., and Chinesta, F. (2010). Solving parametric complex fluids models in rheometric flows. *Journal of Non-Newtonian Fluid Mechanics*, 165:1588–1601.

Andrews, C., Davies, J., and Schwarz, G. (1967). Adaptive data compression. *Proceedings of the IEEE*, 55(3):267–277.

Astrid, P. (2004). *Reduction of Process Simulation Models: a proper orthogonal decomposition approach*. PhD thesis, Eindhoven: Technische Universiteit Eindhoven.

Astrid, P., Weiland, S., Willcox, K., and Backx, T. (2008). Missing point estimation in models described by proper orthogonal decomposition. *IEEE Transactions on Automatic Control*, 53(10):2237–2251.

Audouze, C., De Vuyst, F., and Nair, P. (2009). Reduced-order modeling of parameterized pdes using time–space-parameter principal component analysis. *International journal for numerical methods in engineering*, 80(8):1025–1057.

Ballarin, F., Manzoni, A., Quarteroni, A., and Rozza, G. (2015). Supremizer stabilization of pod–galerkin approximation of parametrized steady incompressible navier–stokes equations. *International Journal for Numerical Methods in Engineering*, 102(5):1136–1161.

Barba, P. D., Mognaschi, M. E., Bullo, M., Dughiero, F., Forzan, M., Lupi, S., and Sieni, E. (2018). Field models of induction heating for industrial applications. *Przeglad Elektrotechniczny*, 94:1–5.

Barglik, J., Smalcerz, A., Przylucki, R., and Doležel, I. (2014). 3d modeling of induction hardening of gear wheels. *Journal of Computational and Applied Mathematics*, 270:231–240.

Barrere, E. (1992). *Contribution à la modélisation des contraintes résiduelles lors de la trempe superficielle au défilé après un chauffage par induction*. PhD thesis, École nationale supérieure d'arts et métiers, campus de Bordeaux.

Baudat, G. and Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. *Neural computation*, 12(10):2385–2404.

Bay, F., Labbé, V., Favennec, Y., and Chenot, J.-L. (2003). A numerical model for induction heating processes coupling electromagnetism and thermomechanics. *International journal for numerical methods in engineering*, 58(6):839–867.

Beckstein, P., Galindo, V., and Vukčević, V. (2017). Efficient solution of 3d electromagnetic eddy-current problems within the finite volume framework of openfoam. *Journal of Computational Physics*, 344:623–646.

Belkin, M. and Niyogi, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in neural information processing systems*, 14.

Biasutti, F., Krause, C., and Lupi, S. (2012). Induction hardening of complex geometry and geared parts. *Heat Processing*, 3:60–70.

Bird, R. B., Curtiss, C. F., Armstrong, R. C., and Hassager, O. (1987). *Dynamics of polymeric liquids, volume 2: Kinetic theory*. Wiley, New York.

Biro, O. and Preis, K. (1989). On the use of the magnetic vector potential in the finite-element analysis of three-dimensional eddy currents. *IEEE Transactions on Magnetics*, 25(4):3145–3159.

Biro, O. and Preis, K. (1990). Finite element analysis of 3–d eddy currents. *IEEE Transactions on Magnetics*, 26(2):418–423.

Bognet, B., Bordeu, F., Chinesta, F., Leygue, A., and Poitou, A. (2012). Advanced simulation of models defined in plate geometries: 3d solutions with 2d computational complexity. *Computer Methods in Applied Mechanics and Engineering*, 201:1–12.

Bognet, B., Leygue, A., and Chinesta, F. (2014). Separated representations of 3d elastic solutions in shell geometries. *Advanced Modeling and Simulation in Engineering Sciences*, 1(1):1–34.

Bordeu, F., Ghnatios, C., Boulze, D., Carles, B., Sireude, D., Leygue, A., and Chinesta, F. (2015). Parametric 3d elastic solutions of beams involved in frame structures. *Advances in aircraft and spacecraft science*, 2(3):233–248.

Borzacchiello, D., Aguado, J. V., and Chinesta, F. (2017). Non-intrusive sparse subspace learning for parametrized problems. *Archives of Computational Methods in Engineering: State-of-the-Art Reviews*, page 1–24.

Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint arXiv:1809.03006*.

Bristiel, P. (2001). *Modelisation magnetothermique, metallurgique et mecanique de la trempe superficielle apres chauffage par induction appliquee aux vilebrequins*. PhD thesis, Ecole Nationale Supérieure d'Arts et Métiers Centre de Bordeaux.

Brunst, W. (1957). Induktive wärmebehandlung, insbesondere oberflächenhärtung, der stähle. In *Die induktive Wärmebehandlung*, pages 197–232. Springer.

Brunton, S. L. and Kutz, J. N. (2019). *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press.

Buhmann, M. D. (2000). Radial basis functions. *Acta numerica*, 9:1–38.

Bui-Thanh, T., Damodaran, M., and Willcox, K. (2003a). Proper orthogonal decomposition extensions for parametric applications in compressible aerodynamics. In *21st AIAA applied aerodynamics conference*, page 4213.

Bui-Thanh, T., Damodaran, M., and Willcox, K. (2003b). Proper orthogonal decomposition extensions for parametric applications in transonic aerodynamics (aiaa paper 2003-4213). In *Proceedings of the 15th AIAA Computational Fluid Dynamics Conference*.

Bui-Thanh, T., Damodaran, M., and Willcox, K. (2004). Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. *AIAA journal*, 42(8):1505–1516.

Cancès, E., Defranceschi, M., Kutzelnig, W., Le Bris, C., and Maday, Y. (Amsterdam, 2003). Computational quantum chemistry: a primer. *in Handbook of Numerical Analysis, Elsevier*, 10:3—270.

Candeo, A., Ducassy, C., Bocher, P., and Dughiero, F. (2011). Multiphysics modeling of induction hardening of ring gears for the aerospace industry. *IEEE Transactions on Magnetics*, 47(5):918–921.

Cao, B.-T., Freitag, S., and Meschke, G. (2016). A hybrid rnn-gpod surrogate model for real-time settlement predictions in mechanised tunnelling. *Advanced Modeling and Simulation in Engineering Sciences*, 3(1):1–22.

Cardinaux, D. (2008). *Etude et modéllisation numérique 3D par éléments finis d'un procédé de traitement thermique de toiles embouties après chauffage par induction: Application à un renfort de pied central automobile*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris.

Chakir, R., Maday, Y., and Parnaudeau, P. (2019). A non-intrusive reduced basis approach for parametrized heat transfer problems. *Journal of Computational Physics*, 376:617–633.

Chatterjee, M. (2003). Comparative study of carburizing vs. induction hardening of gears. *Industrial Gears*.

Chinesta, F. and Cueto, E. (2014). *PGD-based modeling of materials, structures and processes*. Springer.

Chinesta, F., Cueto, E., Abisset-Chavanne, E., Duval, J. L., and Khaldi, F. E. (2020). Virtual, digital and hybrid twins: a new paradigm in data-based engineering and engineered data. *Archives of computational methods in engineering*, 27(1):105–134.

Chinesta, F., Huerta, A., Rozza, G., and Willcox, K. (2017). Model reduction methods. *Encyclopedia of Computational Mechanics Second Edition*, pages 1–36.

Chinesta, F., Keunings, R., and Leygue, A. (2013). *The proper generalized decomposition for advanced numerical simulations: a primer*. SpringerBriefs in Applied Sciences and Technology. Springer, Cham.

Chinesta, F., Leygue, A., Bognet, B., Ghnatios, C., Poulhaon, F., Bordeu, F., Barasinski, A., Poitou, A., Chatel, S., and Maison-Le-Poec, S. (2014). First steps towards an advanced simulation of composites manufacturing by automated tape placement. *International journal of material forming*, 7(1):81–92.

Cordier, L. (2008). Two typical applications of POD: coherent structures eduction and reduced order modelling. Lecture series 2008-01 on post-processing of experimental and numerical data, Von Karman Institute for Fluid Dynamics, février 2008.

Cox, M. A. and Cox, T. F. (2008). Multidimensional scaling. In *Handbook of data visualization*, pages 315–347. Springer.

Cstroeski, J. (2003). *Boundary Element Methods for Inductive Hardening*. PhD thesis, University of Topengen, Germany.

Cueto, E., González, D., and Alfaro, I. (2016). *Proper generalized decompositions: an introduction to computer implementation with Matlab*. SpringerBriefs in Applied Sciences and Technology. Springer, New York.

Curtis, F. W. (1944). High-frequency induction heating.

Czerwinski, F. (2012). Thermochemical treatment of metals. In Czerwinski, F., editor, *Heat Treatment*, chapter 5. IntechOpen, Rijeka.

Degroote, J., Vierendeels, J., and Willcox, K. (2010). Interpolation among reduced-order matrices to obtain parameterized models for design, optimization and probabilistic analysis. *International Journal for Numerical Methods in Fluids*, 63(2):207–230.

Delaleau, C., Bonzano, G., De Sousa Dias, M., and Criqui, B. (2000). Une avancée dans le domaine de la trempe de contour applliquée aux dentures droites et hélicoidales: simulation et expérimentation.

Demo, N., Tezzele, M., Mola, A., and 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*, pages 111–121.

Denis, S. (1997). Prévision des contraintes résiduelles induites par traitement thermique et thermochimique. *Revue de Métallurgie*, 94(2):157–176.

Denis, S., Sjöström, S., and Simon, A. (1987). Coupled temperature, stress, phase transformation calculation model. numerical illustration of the internal stresses evolution during cooling of a eutectoid carbon steel cylinder. *Metallurgical Transactions A*, 18(7):1203–1212.

Derouiche, K., Daoud, M., Traidi, K., and Chinesta, F. (2022). Real-time prediction by data-driven models applied to induction heating process. *International Journal of Material Forming*, 15(4):1–17.

Derouiche, K., Garois, S., Champaney, V., Daoud, M., Traidi, K., and Chinesta, F. (2021). Data-driven modeling for multiphysics parametrized problems-application to induction hardening process. *Metals*, 11(5).

Ding, H. and Shin, Y. C. (2012). A metallo-thermomechanically coupled analysis of orthogonal cutting of aisi 1045 steel. *Journal of manufacturing science and engineering*, 134(5):1–12.

Djoudi, W., Aissani-Benissad, F., and Bourouina-Bacha, S. (2007). Optimization of copper cementation process by iron using central composite design experiments. *Chemical Engineering Journal*, 133(1-3):1–6.

Dolci, V. and Arina, R. (2016). Proper orthogonal decomposition as surrogate model for aerodynamic optimization. *International Journal of Aerospace Engineering*, 2016:1–15.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.

DUOLIN. Induction heating. `https://www.duolin.com/induction_heating/`. [Online; accessed 2020-04-21].

Durban, C. (1997). *Analyse et optimisation des paramètres influants sur le calcul prévisionnel des contraintes résiduelles conséquentes au chauffage superficiel par induction suivi d'une trempe*. PhD thesis, Paris, ENSAM.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129–151.

Esteve, V., Jordan, J., Dede, E., Sanchis-Kilders, E., and Maset, E. (2006). Induction heating inverter with simultaneous dual-frequency output. In *Twenty-First Annual IEEE Applied Power Electronics Conference and Exposition, 2006. APEC'06.*, pages 1505—1509. IEEE.

Everson, R. and Sirovich, L. (1995). Karhunen–loeve procedure for gappy data. *Journal of the Optical Society of America*, 12(8):1657–1664.

Fang, K.-T., Li, R., and Sudjianto, A. (2005). *Design and modeling for computer experiments*. Chapman and Hall/CRC.

FPDBase (2019). Material database of forming properties. *Transvalor SA, Sophia Antipolis, France*.

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Special invited paper. additive logistic regression: A statistical view of boosting. *Annals of statistics*, pages 337–374.

Geijselaers, H. J. M. (2003). *Numerical Simulation of Stresses due to Solid State Transformations : The Simulation of Laser Hardening*. PhD thesis, University of Twente, Enschede.

Ghojogh, B., Ghodsi, A., Karray, F., and Crowley, M. (2020). Locally linear embedding and its variants: Tutorial and survey. *ArXiv*, abs/2011.10925.

Girfoglio, M., Quaini, A., and Rozza, G. (2021). A pod-galerkin reduced order model for a les filtering approach. *Journal of Computational Physics*, 436:110260.

Gié Hubert, S. J.-P. (1982). Electromagnétisme. *Lavoisier - Techniques et Documentation*, 1.

Gonzalez, D., Ammar, A., Chinesta, F., and Cueto, E. (2010). Recent advances in the use of separated representations. *International Journal for Numerical Methods in Engineering*, 81(5):637–659.

Gorni, A. A. (2019). *Steel forming and heat treating handbook*. São Vicente SP, Brazil.

Grum, J. (2007). Overview of residual stresses after induction surface hardening. *International Journal of Materials and Product Technology*, 29(1-4):9–42.

Guénot, M., Lepot, I., Sainvitu, C., Goblet, J., and Coelho, R. F. (2013). Adaptive sampling strategies for non-intrusive pod-based surrogates. *Engineering computations*, 30(4):521–547.

Gunes, H., Sirisup, S., and Karniadakis, G. E. (2006). Gappy data: To krig or not to krig? *Journal of Computational Physics*, 212(1):358–382.

Gunst, R. F. and Mason, R. L. (2009). Fractional factorial design. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(2):234–244.

Hamdaoui, M., Le Quilliec, G., Breitkopf, P., and Villon, P. (2013). Surrogate pod models for parametrized sheet metal forming applications. *Key Engineering Materials*, 554:919–927.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *Boosting and Additive Trees*, pages 337–387. Springer New York, New York, NY.

Haykin, S. (1998). Neural networks: A comprehensive foundation, prentice hall ptr. *Upper Saddle River, NJ, USA*.

Henneron, T., Pierquin, A., and Clenet, S. (2019). Surrogate model based on the pod combined with the rbf interpolation of nonlinear magnetostatic fe model. *IEEE Transactions on Magnetics*, 56(1):1–4.

Hesthaven, J. S., Rozza, G., and Stamm, B. (2016). *Certified reduced basis methods for parametrized partial differential equations*. First edition. Springer International Publishing, SpringerBriefs in Mathematics, 978-3-319-22469-5.

Holmes, P., Lumley, J. L., Berkooz, G., and Rowley, C. W. (2012). *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257.

Hutton, D. V. (2003). *Fundamentals of finite element analysis*. First Edition, McGraw-Hill Science Engineering, Boston.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.

Hömberg, D. (2004). A mathematical model for induction hardening including mechanical effects. *Nonlinear Analysis: Real World Applications*, 5(1):55–90.

Hömberg, D., Liu, Q., Montalvo-Urquizo, J., Nadolski, D., Petzold, T., Schmidt, A., and Schulz, A. (2016). Simulation of multi-frequency-induction-hardening including phase transitions and mechanical effects. *Finite Elements in Analysis and Design*, 121:86–100.

Ibañez, R., Abisset-Chavanne, E., Ammar, A., Gonzalez, D., Cueto, E., Huerta, A., Duval, J. L., and Chinesta, F. (2018). A multidimensional data-driven sparse identification technique: the sparse proper generalized decomposition. *Complexity*, 2018:1–11.

Iron, B., Association, S. R., et al. (1953). *Physical Constants of Some Commercial Steels at Elevated Temperatures:(based on Measurements Made at the National Physical Laboratory, Teddington)*. Butterworths Scienctific Publations.

Jacot, A., Swierkosz, M., Rappaz, J., Rappaz, M., and Mari, D. (1996). Modelling of electromagnetic heating, cooling and phase transformations during surface hardening of steels. *Le Journal de Physique IV*, 6:C1–203–C1–213.

Jakobsson, S., Andersson, B., and Edelvik, F. (2009). Rational radial basis function interpolation with applications to antenna design. *Journal of computational and applied mathematics*, 233(4):889–904.

Jin, J.-M. (2015). *The finite element method in electromagnetics*. John Wiley & Sons.

JmatPro (1999). Practical software for materials properties. `https://www.sentesoftware.co.uk/`.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148.

Jolliffe, I. (2005). Principal component analysis. *Encyclopedia of statistics in behavioral science*.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492.

Kagami, S. and Fukai, I. (1984). Application of boundary-element method to electromagnetic field problems (short paper). *IEEE Transactions on Microwave Theory and Techniques*, 32(4):455–461.

Karatzas, E. N., Stabile, G., Nouveau, L., Scovazzi, G., and 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. *Computer Methods in Applied Mechanics and Engineering*, 347:568–587.

Karhunen, K. (1946). Zur spektraltheorie stochastischer prozesse. *Annales Academiæ Scientiarum Fennicæ. Series A. I*, 34.

Kim, S. and Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679.

Kirby, M. and Sirovich, L. (1990). Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1):103–108.

Kitani, E. C., Hernandez, E. M., Giraldi, G. A., Thomaz, C. E., and Corcoran, P. (2011). Exploring and understanding the high dimensional and sparse image face space: a self-organized manifold mapping. In *New Approaches to Characterization and Recognition of Faces*, pages 225–238. InTech.

Kohli, A. and Singh, H. (2011). Optimization of processing parameters in induction hardening using response surface methodology. *Sadhana*, 36(2):141–152.

Kosambi, D. (1943). Statistics in function space. *Journal of Indian Mathematical Society*, 7:76–88.

Laakso, S. V. A. (1959). Heat matters when matter heats – the effect of temperature-dependent material properties on metal cutting simulations. *Acta Metallurgica*, 7:59–60.

Ladeveze, P. and Chamoin, L. (2011). On the verification of model reduction methods based on the proper generalized decomposition. *Computer Methods in Applied Mechanics and Engineering*, 200(23-24):2032–2047.

Ladevèze, P. (1999). *Nonlinear computational structural mechanics*. Springer New York, NY.

Lamb, H. (1883). On electrical motions in a spherical conductor. *Philosophical Transactions of the Royal Society*, 174:519–549.

Lawrence, S., Giles, C. L., and Tsoi, A. C. (1997). Lessons in neural network training: Overfitting may be harder than expected. In *AAAI/IAAI*, pages 540–545. Citeseer.

Leathwick, J., Elith, J., Francis, M., Hastie, T., and Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding new zealand: an analysis using boosted regression trees. *Marine ecology progress series*, 321:267–281.

Lee, S.-J., Erik, J., Pavlina, C., and Van Tyne, J. (2010a). Kinetics modeling of austenite decomposition for an end-quenched 1045 steel. *Materials Science and Engineering A*, 527:3186—3194.

Lee, S.-J., Pavlina, E. J., and Van Tyne, C. J. (2010b). Kinetics modeling of austenite decomposition for an end-quenched 1045 steel. *Materials Science and Engineering: A*, 527(13-14):3186–3194.

Lehner, G. (2010). *Elektromagnetische Feldtheorie*. Springer.

Li, K., Huang, T.-Z., Li, L., and Lanteri, S. (2021). Non-intrusive reduced-order modeling of parameterized electromagnetic scattering problems using cubic spline interpolation. *Journal of Scientific Computing*, 87.

Li, W., Yuan, Z., and Chen, Z. (2013). Adaptive mesh morphing method for numerical analysis of electromagneto-mechanical coupling using lagrangian approach. In *Proc. 19th Int. Conf. Comput. Electromagn. Fields COMPUMAG*, pages 1–2.

Liang, Y., Lee, H., Lim, S., Lin, W., Lee, K., and Wu, C. (2002). Proper orthogonal decomposition and its applications—part i: Theory. *Journal of Sound and vibration*, 252(3):527–544.

Liberge, E. (2008). *Réduction de modèles par POD Galerkin pour les problèmes d'intéraction fluide structure*. PhD thesis, Université de La Rochelle.

Limousin, V., Delgerie, X., Leroy, E., Ibañez, R., Argerich, C., Daim, F., Duval, J. L., and Chinesta, F. (2019). Advanced model order reduction and artificial intelligence techniques empowering advanced structural mechanics simulations: application to crash test analyses. *Mechanics & Industry*, 20(8):804.

Lin, C.-T. and Lee, C. G. (1996). *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*. Prentice-Hall, Inc.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. Third Edition. Hoboken, NJ, USA: Wiley.

Loeve, M. (1948). Functions aleatoires du second ordre. *Processus stochastique et mouvement Brownien*, pages 366–420.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23.

Lombardi, E., Bergmann, M., Camarri, S., and Iollo, A. (2009). *Low-order models: optimal sampling and linearized control strategies*. PhD thesis, INRIA.

Lozinskii, M. G. (1969). *Industrial Applications of Induction Heating*. 1st English ed., New York, NY, USA: Pergamon.

Lu, J., Plataniotis, K. N., and Venetsanopoulos, A. N. (2003). Face recognition using kernel direct discriminant analysis algorithms. *IEEE transactions on Neural Networks*, 14(1):117–126.

Lucía, O., Maussion, P., Dede, E. J., and Burdío, J. M. (2013). Induction heating technology and its applications: past developments, current technology, and future challenges. *IEEE Transactions on industrial electronics*, 61(5):2509–2520.

Lumley, J. (1967). The structure of inhomogeneous turbulence. *In atmospheric turbulence and wave propagation. Ed. AM Yaglom, VI Tatarski*, pages 166–178.

Ly, H. V. and Tran, H. T. (2001). Modeling and control of physical processes using proper orthogonal decomposition. *Journal of Mathematical and computer modelling*, 33(1-3):223–236.

Madeh Piryonesi, S. and El-Diraby, T. E. (2021). Using machine learning to examine impact of type of performance indicator on flexible pavement deterioration modeling. *Journal of Infrastructure Systems*, 27(2):04021005.

Magnabosco, I., Ferro, P., Tiziani, A., and Bonollo, F. (2006). Induction heat treatment of a iso c45 steel bar: Experimental and numerical analysis. *Computational materials science*, 35(2):98–106.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International journal of forecasting*, 9(4):527–529.

Mangasarian, O. L. (1994). *Nonlinear programming*. SIAM.

McCormick, G. P. (1983). *Nonlinear programming: Theory, algorithms, and applications*. John Wiley & Sons, Inc.

Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, 44(247):335–341.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K.-R. (1999). Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, pages 41–48. Ieee.

Mika, S., Schölkopf, B., Smola, A., Müller, K.-R., Scholz, M., and Rätsch, G. (1998). Kernel pca and de-noising in feature spaces. *Advances in neural information processing systems*, 11.

Mingardi, D. (2013). Numerical models for induction hardening of gears. Master's thesis, Master thesis, Università degli Studi di Padova, Italy.

Misra, M. K., Bhattacharya, B., Singh, O., and Chatterjee, A. (2014). Multi response optimization of induction hardening process-a new approach. *IFAC Proceedings Volumes*, 47(1):862–869.

Moreland, W. (1973). The induction range: Its performance and its development problems. *IEEE Transactions on Industry Applications*, 9, no. 1:81—-85.

Muhlbauer, A., Muiznieks, A., and Lessmann, H.-J. (1993). The calculation of 3d high-frequency electromagnetic fields during induction heating using the bem. *IEEE Transactions on Magnetics*, 29(2):1566–1569.

Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2016). *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons. New York, NY, USA, 2009. ISBN 0471581003.

Nadal, E., Leygue, A., Chinesta, F., Beringhier, M., Rodenas, J., and Fuenmayor, F. J. (2015). A separated representation of an error indicator for the mesh refinement process under the proper generalized decomposition framework. *Computational Mechanics*, 55(2):251–266.

Newman, A. J. (1999). *Modeling and reduction with applications to semiconductor processing*. PhD thesis, University of Maryland, College Park.

Nguyen, M.-N. and Kim, H.-G. (2022). An efficient podi method for real-time simulation of indenter contact problems using rbf interpolation and contact domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 388:114215.

Nils J., N. (1965). Learning machines: Foundations of trainable pattern-classifying systems.

Onan, M., Baynal, K., and Ünal, H. İ. (2015). Determining the influence of process parameters on the induction hardening of aisi 1040 steel by an experimental design method. *Indian Journal of Engineering & Materials Sciences*, 22:513–520.

Orlich, J., Rose, A., and Wiest, P. (1973). Iii zeit temperatur austenitisierung schaubilder. *Atlas zur Wärmebehandlung von Stähle*.

Özişik, M., Orlande, H., Colaço, M., and Cotta, R. (2017). *Finite Difference Methods in Heat Transfer*. 2nd edition, CRC Press.

Papachristoudis, G. (2019). The bias-variance tradeoff. Published in Towards Data Science, https://towardsdatascience.com/the-bias-variance-tradeoff-8818f41e39e9. [Online; accessed 2022-06-20].

Park, C. H. and Park, H. (2005). Nonlinear discriminant analysis using kernel functions and the generalized singular value decomposition. *SIAM journal on matrix analysis and applications*, 27(1):87–102.

Pascal, R. (2003). *Modélisation du traitement thermique superficiel par induction*. PhD thesis, Ecole Centrale de Lyon.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Petzold, T. (2014). *Modelling, analysis and simulation of multifrequency induction hardening*. PhD thesis, Technischen Universitat Berlin.

Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, 5:532–538.

Ripepi, M., Verveld, M. J., Karcher, N., Franz, T., Abu-Zurayk, M., Görtz, S., and Kier, T. (2018). Reduced-order models for aerodynamic applications, loads and mdo. *CEAS Aeronautical Journal*, 9(1):171–193.

Rivat, M. (2002). Trempe de contour avec chauffage par induction bi-fréquence d'engrenages à denture droite. *Traitement thermique*, pages 41–45.

Rodigin, N. (1950). Induction heating of steel parts with current of industrial frequency.

Rokicki, P. (2017). Induction hardening of tool steel for heavily loaded aircraft engine components. *Archives of Metallurgy and Materials*, 62(1):315–320.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326.

Rudnev, V., Loveless, D., and Cook, R. L. (2017). *Handbook of Induction Heating*. Manufacturing Engineering and Materials Processing.

Ruhmelhart, D., Hinton, G., and Wiliams, R. (1986). Learning representations by back-propagation errors. *Nature*, 323(533-536):10.

Sancarlos, A., Cameron, M., Abel, A., Cueto, E., Duval, J.-L., and Chinesta, F. (2021). From rom of electrochemistry to ai-based battery digital and hybrid twin. *Archives of Computational Methods in Engineering*, 28(3):979–1015.

Santner, T. J., Williams, B. J., Notz, W. I., and Williams, B. J. (2003). *The design and analysis of computer experiments*. Springer New York, NY, 2 edition.

Saul, L. K. and Roweis, S. T. (2000). An introduction to locally linear embedding. *unpublished. Available at: http://www. cs. toronto. edu/˜ roweis/lle/publications. html*.

Saul, L. K. and Roweis, S. T. (2003). Think globally, fit locally: unsupervised learning of low dimensional manifolds. *Journal of machine learning research*, 4(Jun):119–155.

Schapire, R. E. (2003). *The Boosting Approach to Machine Learning: An Overview*, pages 149–171. Springer New York, New York, NY.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319.

Scikit-learn. Underfitting vs. Overfitting. `https://scikit-learn.org/stable/auto_examples/model_selection/plot_underfitting_overfitting.html`. [Online; accessed 2022-06-15].

Senhaji, A. (2017). Simulation numérique de la chauffe par induction électromagnétique d'un disque en aisi 4340. Master's thesis, École de technologie supérieure.

Shinde, V., Longatte, E., Baj, F., Hoarau, Y., and Braza, M. (2016). A galerkin-free model reduction approach for the navier–stokes equations. *Journal of Computational Physics*, 309:148–163.

Simpson, P. G. (1960). *Induction heating: coil and system design*. McGraw-Hill, New York.

Simpson, T. W., Poplinski, J., Koch, P. N., and Allen, J. K. (2001). Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers*, 17(2):129–150.

Sirovich, L. (1987). Turbulence and the dynamics of coherent structures. i. coherent structures. *Quarterly of applied mathematics*, 45(3):561–571.

Slukhotskii, A. and Ryskin, S. (1974). Inductors for induction heating. *Energy Publ., St. Petersburg, Russia*.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.

Sobester, A., Forrester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons.

Spezzapria, M., Dughiero, F., Forzan, M., and Candeo, A. (2012). Multiphysics tem simulation of contour induction hardening process on aeronautical gears. *Journal of Iron and Steel Research International*, 19(S1):95–98.

Stansel, N. R. (1949). *Induction heating*. McGraw-Hill Book Company.

Stauffer, P. R., Cetas, T. C., and Jones, R. C. (1984). Magnetic induction heating of ferromagnetic implants for inducing localized hyperthermia in deepseated tumors. *IEEE Transactions on Biomedical Engineering*, BME-31, no. 2(8):235—-251.

Stein, M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151.

Stone, J. V. (2002). Independent component analysis: an introduction. *Trends in cognitive sciences*, 6(2):59–64.

Sumithra, P. and Thiripurasundari, D. (2017). Review on computational electromagnetics. *Advanced Electromagnetics*, 6(1):42–55.

Sun, F., Xie, G., Song, J., and Markides, C. N. (2022). Proper orthogonal decomposition and physical field reconstruction with artificial neural networks (ann) for supercritical flow problems. *Engineering Analysis with Boundary Elements*, 140:282–299.

Taguchi, G. (1987). System of experimental design; engineering methods to optimize quality and minimize costs. *UNIPUB/Kraus International Publications New York*, 1.

Tian, Y. and Zhang, Y. (2022). A comprehensive survey on regularization strategies in machine learning. *Information Fusion*, 80:146–166.

TRANSVALOR (2020). Forge® - induction for heating and heat treating. *Training*.

Van den Eynde, G., Belloni, F., Star, S., Degroote, J., and Vierendeels, J. (2020). Reduced order modelling using a pod-based identification method for parametrized pdes. In *Proceedings of the 6th European Conference on Computational Mechanics: Solids, Structures and Coupled Problems, ECCM 2018 and 7th European Conference on Computational Fluid Dynamics, ECFD 2018*, pages 796–805, Spain. CIMNE - International Centre for Numerical Methods in Engineering.

Vanderbei, R. J. (1999). Loqo user's manual—version 3.10. *Optimization methods and software*, 11(1-4):485–514.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer, New York.

Venturi, D. and Karniadakis, G. E. (2004). Gappy data and reconstruction procedures for flow past a cylinder. *Journal of Fluid Mechanics*, 519:315–336.

Verveld, M. J., Kier, T., Karcher, N., Franz, T., Abu-Zurayk, M., Ripepi, M., and Görtz, S. (2016). Reduced order models for aerodynamic applications, loads and mdo.

Vologdin, V. (1939). Surface hardening by the induction method. *Metallurgizdat, Moscow*.

Wanser, S. (1995). *Simulation des phénomènes de chauffage par induction. Application à la trempe superficielle*. PhD thesis, L'Ecole centrale de Lyon.

Wasserman, P. D. (1989). *Neural computing: theory and practice*. Van Nostrand Reinhold Co.

Wendland, H. (2004). *Scattered data approximation*, volume 17. Cambridge university press.

William, J. and Mehl, R. (1939). Reaction kinetics in processes of nucleation and growth. *Transactions of the Metallurgical Society of AIME*, 135:416–442.

Wolfgang, S. and Hans-Joachim, P. (2002). Anwendungen des zweifrequenz-simultan-verfahrens zum induktiven randschichthärten. *Elektrowärme international*, (1):13–18.

Wu, G., Liang, Y., Lin, W., Lee, H., and Lim, S. (2003). A note on equivalence of proper orthogonal decomposition methods. *Journal of Sound and vibration*, 265(5):1103–1110.

Wu, H.-T. and Wu, N. (2018). Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding. *The Annals of Statistics*, 46(6B):3805–3837.

Xiao, D., Fang, F., Buchan, A. G., Pain, C. C., Navon, I. M., and Muggeridge, A. (2015). Non-intrusive reduced order modelling of the navier–stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 293:522–541.

Xiao, D., Yang, P., Fang, F., Xiang, J., and Pain, C. (2017). A non-intrusive reduced-order model for compressible fluid and fractured solid coupling and its application to blasting. *Journal of Computational Physics*, 330:221–244.

Xiao, D., Yang, P., Fang, F., Xiang, J., Pain, C. C., and Navon, I. M. (2016). Non-intrusive reduced order modelling of fluid–structure interactions. *Computer Methods in Applied Mechanics and Engineering*, 303:35–54.

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture*, 1(2):129–142.

Ye, J., Janardan, R., and Li, Q. (2004). Two-dimensional linear discriminant analysis. *Advances in neural information processing systems*, 17:1569–1576.

Yotov, K., Pingali, K., and Stodghill, P. (2005). Think globally, search locally. In *Proceedings of the 19th annual international conference on Supercomputing*, pages 141–150.

Zaman, P. B., Sultana, N., and Dhar, N. R. (2020). Quantifying the effects of cooling condition, tool type and cutting parameters on machinability of turning aisi 4140 steel using full factorial doe. *Journal of Production Systems and Manufacturing Science*, 1(2):23–39.

Zararsiz, G., Elmali, F., and Ozturk, A. (2012). Bagging support vector machines for leukemia classification. *International Journal of Computer Science Issues*, 9(6):355.

Zgraja, J. (2019). Dual-frequency induction heating generator with adjustable impedance matching. *IEEE Transactions on Industrial Electronics*, 66(11):8308–8317.

Zhang, Z. and Zha, H. (2003). Nonlinear dimension reduction via local tangent space alignment. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 477–481. Springer.

Zhao, H. (2021). A reduced order model based on machine learning for numerical analysis: An application to geomechanics. *Engineering Applications of Artificial Intelligence*, 100:104194.

Appendix A

Identification of the electromagnetic material properties

A.1 Electric resistivity

The electrical resistivity of materials (ρ), represents its capacity to oppose the flow of electric current. Based on experimental data provided in the literature (Magnabosco et al., 2006; Akram et al., 2009; Barba et al., 2018), a model for the evolution of the resistivity is implemented in FORGE® as follows:

$$\rho(T) = \begin{cases} \rho_{ref} \left[\alpha \; exp\left(\dfrac{T}{\tau_1}\right) + 1 - \alpha \right] & \text{if } T < T_{trans} \\ \rho_2 \left[\alpha_2 \; exp\left(\dfrac{-T}{\tau_2}\right) + 1 \right] & \text{if } T > T_{trans} \end{cases} \tag{1}$$

with

$$\begin{cases} \rho_2 = \rho_{ref} \; \alpha \; exp\left(\dfrac{T_{trans}}{\tau_1}\right) \left[1 + \dfrac{\tau_2}{\tau_1}\right] + \rho_{ref} \, (1 - \alpha) \\ \alpha_2 = -\dfrac{\rho_{ref} \; \alpha \; \tau_2 \left[\left(\dfrac{1}{\tau_1} + \dfrac{1}{\tau_2} \right) T_{trans} \right]}{\rho_2 \; \tau_1} \end{cases} \tag{2}$$

where T, ρ_2, α_2, T_{trans}, ρ_{ref}, α, τ_1, τ_2 are the temperature, the continuity conditions, and the predefined user parameters, respectively. These parameters are fitted such that ρ evolution matches with the experimental results available in the literature. The identified parameters are listed in Table A.1.

Table A.1: Parameters implemented in the FORGE® model to describe the electrical resistivity evolution

Parameters	Values	Units
T_{trans}	710	°C
ρ_{ref}	1.60×10^{-07}	Ω.m
α	2.8	-
τ_1	700	°C
τ_2	160	°C

The curve showing the electrical resistivity evolution as a function of temperature according to the FORGE® model is shown in Figure 1.

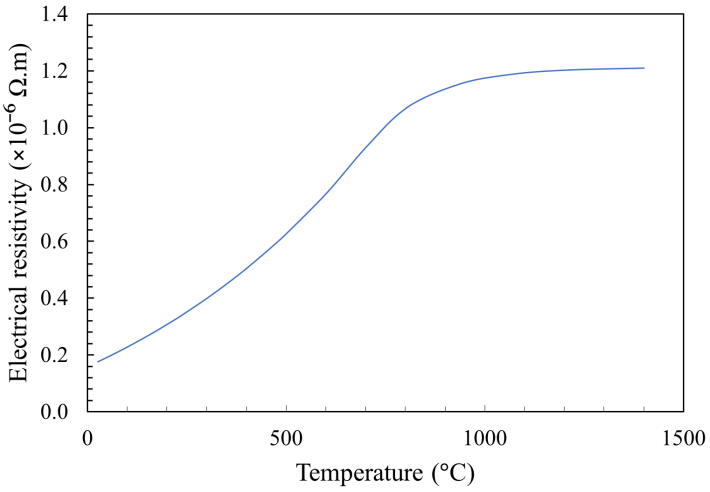


Figure 1: Electrical resistivity of C45 steel according to FORGE® model

A.2 Relative magnetic permeability

The relative magnetic permeability (μ_r) is a dimensionless parameter that indicates the capacity of a material to conduct a magnetic flux. It depends on the temperature and the magnetic field H and is expressed by:

$$\mu_r(T, H) = \frac{B(T, H)}{\mu_0 H} \tag{3}$$

where B is the magnetic induction flux (Tesla), H is the magnetic field (A.m^{-1}) and μ_0 is the permeability of vacuum having a value of $4\pi \times 10^{-7}$ H.m^{-1}.

It is worth noting that the C45 steel is ferromagnetic, hence the relative magnetic permeability is different from unity. In the literature, no experimental data for C45 steel were available, but results for low-alloyed steels with a chemical composition close to ours were provided in (Bristiel, 2001) as well as results from different models such as the Flux 2D model (Wanser, 1995) and the Ansys model (Akram et al., 2009) for the C45 steel. The advantage of FORGE® is that it provides a model that takes into account both the temperature and the magnetic field H (this is not the case for other models). Since μ_r is directly related to the magnetic induction B, FORGE® proposes two different expressions of B as shown in Figure 2. Such evolution enables to take into account the magnetization effect.

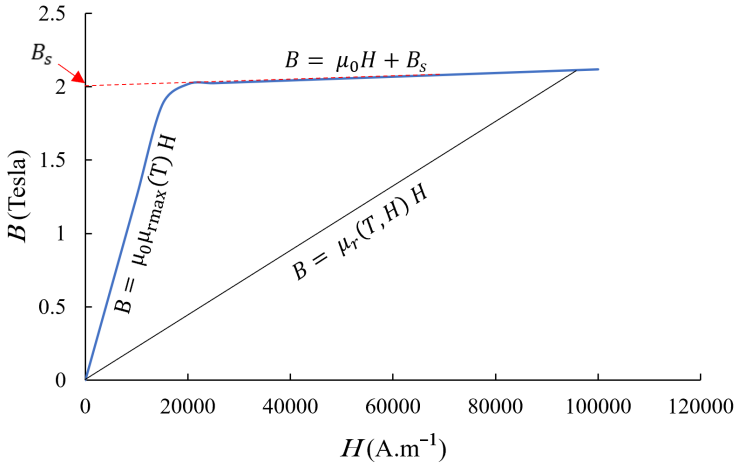


Figure 2: Model of B for ferromagnetic steels

such that B_s is the saturation magnetization (Tesla) and μ_{rmax} is the maximum magnetic permeability (dimensionless) and both depend on the temperature such that:

$$B_s(T) = Max\left[B_{sref}\left(1 - exp\left(\frac{T - T_{curie}}{C}\right)\right), 0\right] \tag{4}$$

$$\mu_{rmax}(T) = Max\left[\mu_{rmaxref}\left(1 - exp\left(\frac{T - T_{curie}}{C}\right)\right), 1\right] \tag{5}$$

Then, to calculate B, we consider two conditions

$$B(T,H) = \begin{cases} \mu_0 \mu_{rmax}(T)\ H & \text{if } B < B_{int} \\ \mu_0\ H + B_s(T) & \text{if } B > B_{int} \end{cases} \tag{6}$$

where B_{int} is defined by Eq.(7), T_{curie} is the Curie temperature beyond which a ferromagnetic steel passes to a non-magnetic steel of relative magnetic permeability equal to unity, and C, B_{sref} and $\mu_{rmaxref}$ are user-defined constants.

$$B_{int} = \frac{B_s(T)\ \mu_{rmax}(T)}{\mu_{rmax}(T) - 1} \tag{7}$$

Table A.2 shows the parameters used for the simulation according to the literature data.

Table A.2: Parameters used in the FORGE® model to describe the relative permeability

Parameters	Values	Units
T_{curie}	760	°C
C	140	°C
B_{sref}	1.6	Tesla
$\mu_{rmaxref}$	100	-

Finally, the evolution of B with respect to H for different temperatures, according to the FORGE® model, for the C45 steel is presented in Figure 3.

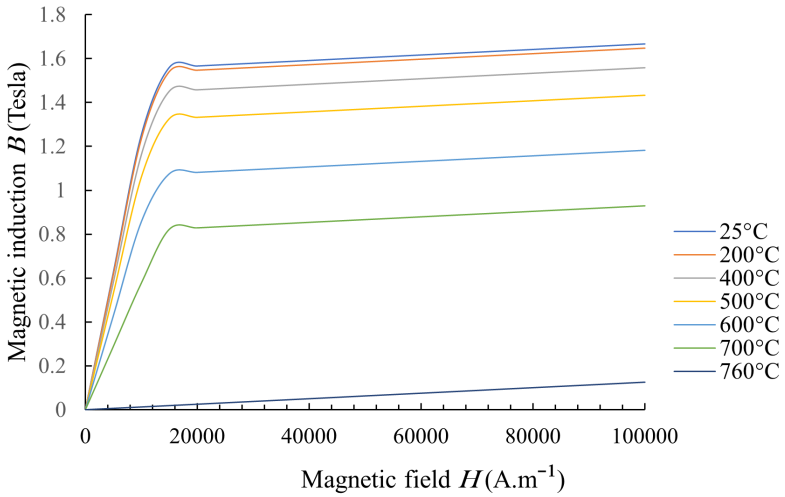


Figure 3: Magnetic induction B of C45 steel

Appendix B

Identification of the thermal material properties

B.1 Density

The density of a substance is its mass per unit volume. Figure 1 shows the evolution of the density as a function of temperature for the involved phases.

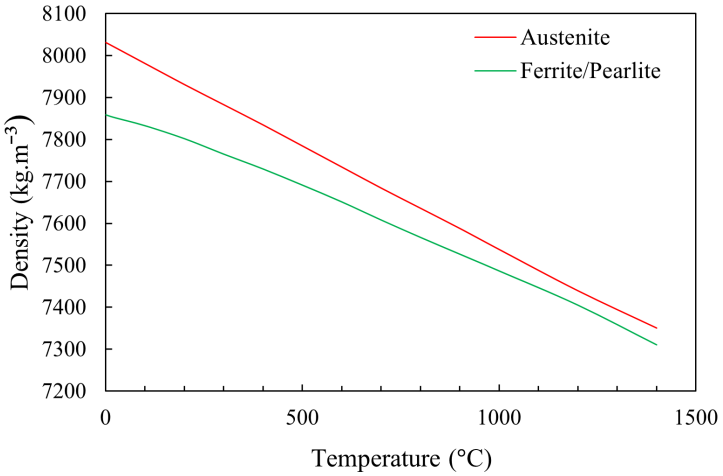


Figure 1: Density evolution per phase of C45 steel

B.2 Thermal conductivity

The thermal conductivity λ represents the capacity of a material to conduct heat. Figure 2 shows the evolution of the thermal conductivity of C45 steel as a function of temperature for the included phases.

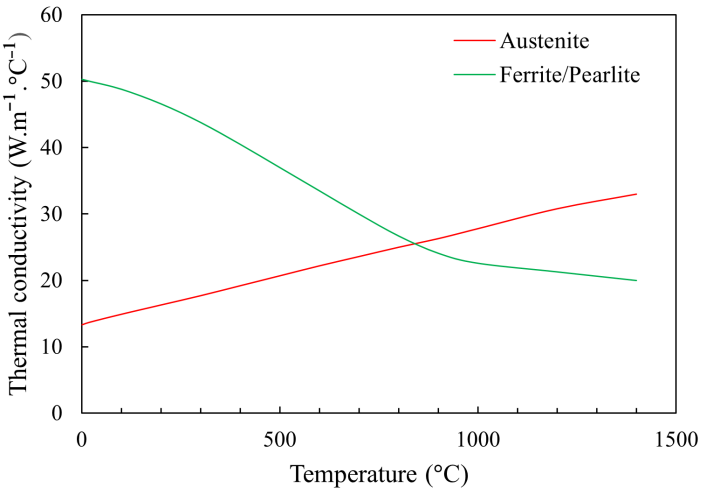


Figure 2: Evolution of the thermal conductivity per phase of C45 steel

B.3 Specific heat

The specific heat C_p is the amount of thermal energy required to raise the temperature of a body of 1 kg mass by 1°C. Figure 3 shows the evolution of this property as a function of temperature for the involved phases.

The specific heat slightly increases with temperature at the beginning. Then, it increases rapidly until it reaches a maximum value which represents the latent heat of transformation. Once the austenite starts to be formed, C_p decreases. In the literature, the maximum value of C_p is often reached between 700 and 730 °C, which represents the austenitic transformation temperature at thermodynamic equilibrium and slow transformations. Since very high heating rates are considered in our study, this value is no longer valid and the ferrite/perlite curve is shifted to higher temperature values.

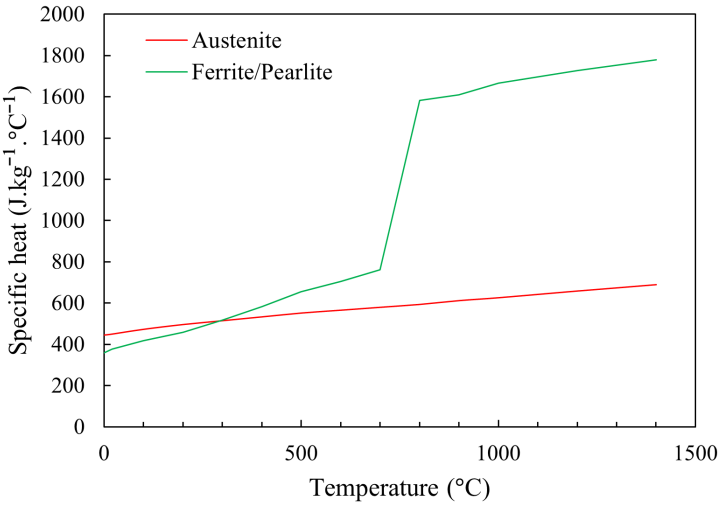


Figure 3: Evolution of the specific heat per phase of C45 steel

Appendix C

Identification of the mechanical material properties

C.1 Young's modulus and Poisson ratio

The Young's modulus E differs little when passing from one phase to another. Therefore, the same evolution curve can be considered for all phases. Based on the references (Laakso, 1959; Lee et al., 2010a), the evolution of Young's modulus is implemented in FORGE® as illustrated in Figure 1.

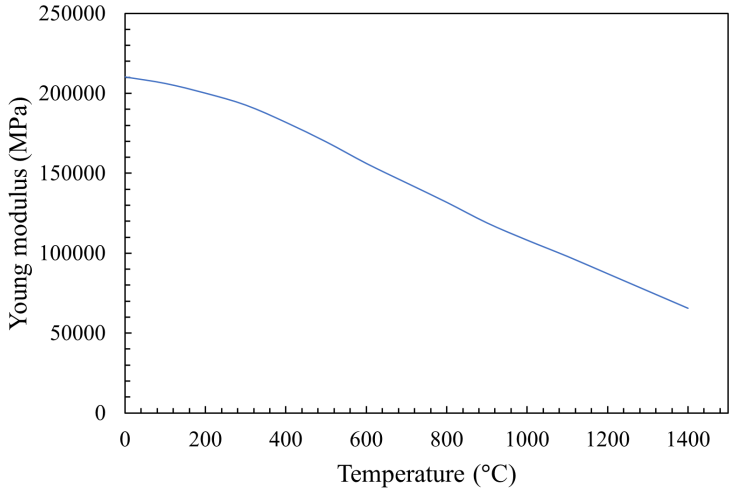


Figure 1: Young's modulus evolution with temperature

The Poisson ratio for C45 was considered identical for all phases and it takes a constant value of 0.3.

C.2 Thermal expansion coefficient

For the quantification of thermal deformation, the expansion coefficient α evolution with temperature and for the involved phases must be defined. Based on data from the literature (Laakso, 1959; Ding and Shin, 2012), the evolution of α for the C45 is illustrated in Figure 2.

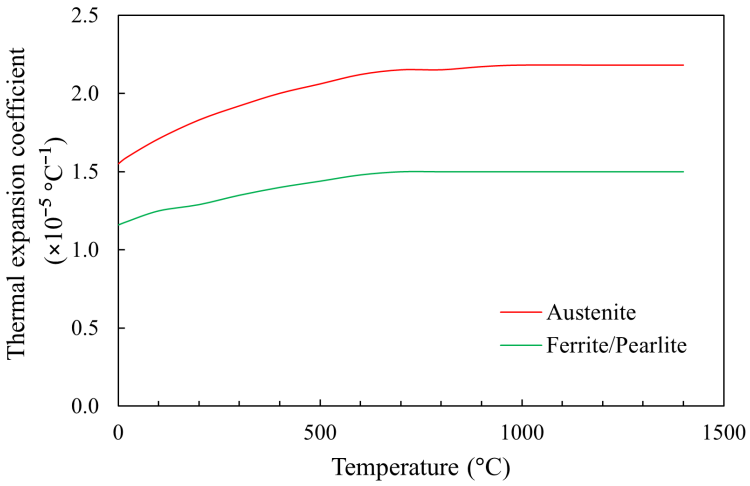


Figure 2: Evolution of the thermal expansion coefficient

C.3 Parameters of the constitutive law

The identification of the rheological law parameters is generally based on numerous tensile tests allowing to describe the irreversible deformation of the steel during the process. The lack of data in the literature led us to retain the default data provided by FORGE® toolbox for C45 using JmatPro datasets. The setting of the parameters (yield strength σ_0, work hardening parameters H and n, viscosity parameters K and m) as a function of the temperature and involved phases is presented in Table (C.1), (C.2) and (C.3) for austenite, ferrite and pearlite, respectively.

Table C.1: Constitutive law parameters for austenite

Temperature (°C)	σ_0 (MPa)	H	n	K	m
20	180	792	0.4	0	0.2
300	138	565	0.4	0	0.2
750	72	200	0.4	45	0.2
900	22	78	0.4	73	0.2
1200	8	78	0.4	73	0.2
1400	2	78	0.4	73	0.2

Table C.2: Constitutive law parameters for ferrite

Temperature (°C)	σ_0 (MPa)	H	n	K	m
20	209	354	0.4	0	0.2
300	107	188	0.4	0	0.2
750	14.4	29.5	0.4	21.2	0.2
900	6.1	6.8	0.4	18.9	0.2
1200	2.1	6.8	0.4	18.9	0.2

Table C.3: Constitutive law parameters for pearlite

Temperature (°C)	σ_0 (MPa)	H	n	K	m
20	363	948	0.3	0	0.2
50	360	945	0.3	0	0.2
100	355	930	0.3	0	0.2
150	350	920	0.3	0	0.2
200	345	910	0.3	0	0.2
250	335	885	0.3	0	0.2
300	325	860	0.3	0	0.2
350	310	820	0.3	48	0.2
400	295	780	0.3	95	0.2
450	270	715	0.3	143	0.2
500	240	645	0.3	190	0.2
550	200	560	0.3	238	0.2
600	150	451	0.3	285	0.2
630	120	360	0.3	300	0.2
680	33	208	0.3	308	0.2
1200	33	208	0.3	308	0.2

C.4 Volume change due to phase transformation

Regarding the complexity of identification of the volume variation at phase change, the data provided by FORGE® toolbox was retained. This evolution as a function of temperature for the phases involved is presented in Figure 3.

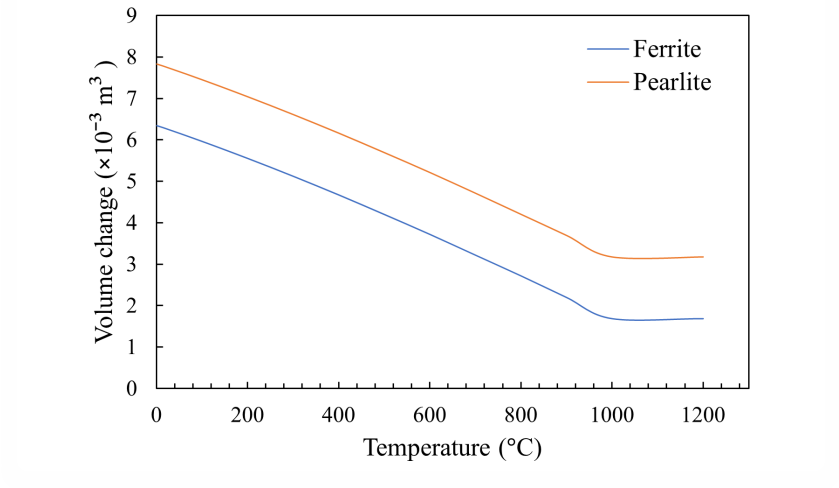


Figure 3: Volume change due to phase transformation

Khouloud DEROUICHE

Réduction de modèle et métamodélisation paramétrique du procédé multiphysique de chauffage par induction

Résumé : La modélisation par éléments finis (MEF) représente aujourd'hui l'outil de calcul le plus attrayant pour prédire et optimiser de nombreux problèmes industriels. Cependant, la MEF devient inefficace en ce qui concerne les problèmes complexes multi-physiques paramétrés, tels que le traitement de chauffage par induction, en raison de son coût de calcul élevé. L'objectif de cette thèse est de définir une méthodologie de réduction de modèles multi-physiques adaptée au procédé de chauffage par induction et de proposer une solution paramétrique pour quelques quantités physiques d'intérêt, notamment l'évolution temporelle de la température et du taux d'austénite sur un pignon droit en acier C45, en utilisant une approche de modélisation non intrusive basée sur les données comme alternative à la MEF pour une prédiction en temps réel. Pour ce faire, un ensemble de solutions synthétiques a été collecté, au niveau de certains capteurs dans la pièce et pour différentes valeurs de paramètres d'entrée (fréquence et puissance), en se basant sur des données de la simulation numérique obtenues via le logiciel de calcul par éléments finis FORGE®. En effet, une étude de faisabilité et de convergence a d'abord été effectuée afin de figer une configuration qui converge et qui suit les bonnes tendances. Les résultats obtenus par simulation selon un échantillonnage de type hypercube latin ont ensuite été traités. Pour le modèle de température, une réduction dimensionnelle par la méthode "proper orthogonal decomposition" (POD) couplée avec trois méthodes de régression non linéaire (sPGD, SVR, et GB) ont été appliquées pour construire une base réduite et créer un modèle représentatif de la solution dans l'espace de faible dimension. Pour le taux d'austénite, deux métamodèles ont été développés pour différents instants qui caractérisent la transformation austénitique. Les résultats ont montré que les méthodes sPGD et SVR donnent comparativement une meilleure prédiction. Par conséquent, une prédiction en temps réel de l'évolution temporelle de la température et du taux d'austénite peut être calculée pour un nouvel ensemble des paramètres d'entrée et au niveau des capteurs considérés. Ensuite, une interpolation spatiale a été réalisée pour étendre la solution partout dans la zone affectée thermiquement. Pour la température, deux approches ont été proposées; la première est basée sur la réduction de dimensionnalité non linéaire par la méthode "locally linear embedding" et la méthode "POD" avec interpolation par fonction de base radiale, tandis que la deuxième est basée sur la "gappy POD". Les deux approches génèrent de bonnes approximations malgré leurs différences. Pour le taux d'austénite, une généralisation de l'approche proposée précédemment a été effectuée en prenant en considération des paramètres géométriques. Une comparaison des trois méthodes de régression a été menée. Enfin, une étude de l'effet d'un changement dimensionnel du pignon sur l'évolution de la température a été effectuée, ceci sans passer par un nouveau plan d'expérience, mais en s'appuyant sur les résultats de la géométrie de référence. Pour ce faire, deux approches ont été proposées pour prédire l'évolution de la température dans des nouvelles géométries. La première approche est basée sur le réseau de neurone en utilisant comme paramètres d'entrée quelques incréments initiaux des courbes temporelles de la température. La deuxième approche est basée sur la "POD" et la régression par sPGD en utilisant la puissance de chauffe comme quantité intermédiaire. Il a été montré que les résultats sont prometteurs, cependant, il est difficile d'approximer des phénomènes non-linéaires dépendant du temps à partir des données partielles extraites au début du procédé.

Mots clés: Modélisation par éléments finis; Modélisation non intrusive; Chauffage par induction; Réduction de dimensionnalité; Métamodèle; Acier C45; régression non linéaires; Interpolation