UNIVERSITÉ PARIS-SUD 11

Faculté des sciences d'Orsay

THÈSE

Présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES DE
L'UNIVERSITÉ PARIS-SUD 11

Spécialité : PHYSIQUE

École Doctorale « Sciences et Technologies de l'Information
des Télécommunications et des Systèmes »

par

Sándor Bilicz

Application of Design-of-Experiment Methods and Surrogate Models in Electromagnetic Nondestructive Evaluation

Soutenue le 30 mai 2011 devant la Commission d'examen:

M.	Bíró	József	(Examinateur)
M.	Dular	Patrick	(Examinateur & Rapporteur)
M.	Idier	Jérôme	(Examinateur & Rapporteur)
M.	Lesselier	Dominique	(Président)

Membres invités :

M.	Gyimóthy	Szabolcs	(Co-directeur de thèse)
M.	Lambert	Marc	(Co-directeur de thèse)
M.	Calmon	Pierre	
M.	Vazquez	Emmanuel	

Rapporteurs :

M.	Dular	Patrick
M.	Harsányi	Gábor
M.	Idier	Jérôme

Application of Design-of-Experiment Methods and Surrogate Models in Electromagnetic Nondestructive Evaluation

PhD Dissertation

prepared under the "co-tutel" joint supervision scheme between the
UNIVERSITY OF PARIS-SUD 11
and the
BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS

Author:
Sándor Bilicz

Advisors:
Marc Lambert
Laboratoire des Signaux et Systèmes UMR8506
(CNRS-SUPELEC-Univ Paris-Sud)
and
Szabolcs Gyimóthy
Department of Broadband Infocommunications and Electromagnetic Theory
(Budapest University of Technology and Economics)

2011

Acknowledgement

I am grateful for the help I have got from my advisors *Marc Lambert* and *Szabolcs Gyimóthy* during the past several years. Not only this PhD Dissertation, but also my MSc Thesis was written under their supervision. I really learned a lot from both of them. Marc Lambert showed me how an enthusiastic and tenacious research looks like. Szabolcs Gyimóthy's accuracy, judiciousness and unbreakable optimism was exemplary, without doubt – not only in research, but also in my teaching activity at the Budapest University of Technology and Economics. Both of them participated in the jury of the PhD defense as invited members.

I say thanks for the careful reviewing of this Dissertation to *Patrick Dular* from the University of Liège, to *Gábor Harsányi* from the Budapest University of Technology and Economics and to *Jérôme Idier* from the Institut de Recherche en Communications et Cybernétique de Nantes. Special thanks go to Patrick Dular and Jérôme Idier for having accepted also to be member of the jury. I am grateful for *József Bíró*, who also participated in the jury, representing the Budapest University of Technology and Economics.

Though I would like to express my gratitude for all people I have been in touch during my work, I must highlight the most impressive ones herein. *József Pávó* taught me to "sight" the electromagnetic field, whatever this means. Without his contribution, I think I would not have started to deal with electromagnetic nondestructive evaluation –and I possibly would not have gotten in touch with the Laboratoire des Signaux et Systèmes. A valuable technical help was also given by József, as he provided his code for the numerical simulation of the illustrative eddy-current testing setup. An unforgettable experience is the cooperation with *Dominique Lesselier*. I have learnt so many tiny things from him (e.g., to start with the bibliography when reading a paper), –it is impossible to explain by words some of them–, and, what impressed me the most: his exceptionally familiar manner. Dominique Lesselier accepted to be a member of the jury of the PhD defense as well –moreover, he has been chosen as the president of the jury–, for what I am thankful for him. For the topic of my PhD work, I must say thanks to *Emmanuel Vazquez*, who also participated in the jury as an invited member. He called our attention to the kriging approach, moreover, he also provided some codes as well at the beginning. Though in this Dissertation, the work we have done in the domain of radar forest observations is only slightly concerned, I would like to thank to *Laetitia Thirion-Lefevre* for this. I believe that this cooperation –basically initiated by her– is fruitful.

In terms of organizations, I would like to thank the Department of Broadband Infocommunications and Electromagnetic Theory at the Budapest University of Technology and Economics and the Laboratoire des Signaux et Systèmes for the milieu I had the privilege to work in during the three years of this co-tutel PhD study. Both contributed to the financial background of my work as well. I am grateful also to DIGITEO for providing me a scholarship, and to the Commissariat à l'Énergie Atomique (CEA) for having taken part in the establishment of the frame of my work. CEA was represented also at the defense by *Pierre Calmon*, to whom I say thanks for having accepted the invitation for being an invited jury member.

The last but not least, I express my sincere gratitude to my family and my friends for their affection.

Abstract

Electromagnetic Nondestructive Evaluation (ENDE) is applied in various industrial domains for the exploration of hidden in-material defects of structural components. The methods rely on the fact that the electromagnetic (EM) constitutive parameters of the material are locally changed in the presence of a defect. Based on the measured EM field (generated by an external source and interacting with the examined specimen), ENDE aims at the characterization of the defects, i.e., the solution of the related *inverse problem*. To this end, one has to be able to determine the EM field corresponding to a known defect, i.e., to solve the *forward problem*. Practically, this is performed via the mathematical modeling (based on the Maxwell's equations) and the numerical simulation (e.g., by the Method of Moments) of the studied ENDE setup. Such simulators can provide fine precision, but at a price of computational cost. However, the solution of an inverse problem often requires several runs of these "expensive-to-evaluate" simulators, making the inversion procedure firmly demanding in terms of runtime and computational resources. To overcome this challenge, *surrogate modeling* (SM) is getting more and more widespread in electromagnetics. A surrogate model imitates the true model, but as a rule, it is much less complex than the latter. A way to construct such surrogates is to perform a couple of simulations and then to approximate the model based on the obtained data. The choice of the "prototype" simulations is usually controlled by a sophisticated strategy, drawn from the tools of *Design-of-Experiments* (DoE).

The goal of the research work presented in this Dissertation is the improvement of ENDE methods by using SM and DoE techniques. The EM simulator –assuming a parametric defect model– is treated as a black-box, i.e., accounting only for the input-output relationship. The input is the set of parameters describing the defect geometry, the output is related to the measurable EM field in function of the receiver position, respectively. Three self-sufficient approaches are then discussed in detail.

First, an optimization-based inversion algorithm is presented. By tuning the input parameters of the simulator, the best similarity (in terms of an appropriate objective function) between the measured and simulated data is to be achieved. Inspired by the need for reducing the number of simulation runs, the objective function is minimized by using the "Efficient Global Optimization" (EGO) algorithm, which is known to converge within a relatively small number of iterations. EGO chooses the inputs to be simulated one-by-one, yielding a sequence converging to the global minimum, i.e., to the solution of the inverse problem. The algorithm is based on the surrogate modeling of the objective function using kriging (modeling by means of a Gaussian process).

We propose two generic surrogate modeling methods as well. Both are based on a sequential sampling strategy for the generation of a problem-specific *database* of corresponding input parameter - output data pairs. This database will subsequently be used as the support of a surrogate model. To improve the performance of this model, the database generation is adapted both to the modeled forward problem and to the goal of the surrogate being built. In the first approach, a functional kriging interpolator is fitted to the samples, and the objective is the reduce the discrepancy between the true and the interpolated output data. To this end, the sampling strategy is driven by the estimated uncertainty of the functional kriging prediction (via a *jack-knife* variance-estimator). The second approach aims at uniformly filling the "output-space" (the domain of the feasible output data) by samples. The main benefit of such output space-filling is the meta-information on the modeled problem provided by the structure of the database. We present two inverse mappings to exploit this meta-information for the quantitative characterization of the related inverse problem.

All approaches are illustrated by examples drawn from Eddy-Current Testing, using 2, 4, 6 and occasionally 8 defect parameters.

Contents

Application des méthodes de plans d'expérience numérique et de modèles de substitution pour le contrôle non destructif électromagnétique

1 Introduction

Le contrôle non destructif électromagnétique (CNDE) est appliqué dans des domaines variés pour l'exploration de défauts cachés affectant des structures. Ces méthodes s'appuient sur le fait que les paramètres électromagnétiques constitutifs des matériaux (permittivité, conductivité, perméabilité, ...) sont localement modifiés par la présence de ce ou de ces défauts. De façon générale, le principe peut se poser en ces termes : un objet inconnu perturbe un milieu hôte donné, connu pour l'essentiel (certains éléments caractéristiques de l'hôte peuvent être connus avec imprécision ou même être non déterminés) et illuminé par un signal électromagnétique connu, et la réponse est mesurée sur un ou plusieurs récepteurs de positions connues. Cette réponse contient des informations sur les paramètres électromagnétiques (permittivité, conductivité, ...) et géométriques (emplacement, taille et forme) des objets recherchés et toute la difficulté du problème traité ici consiste à extraire ces informations du signal obtenu. Plus connu sous le nom de « problèmes inverses », ces travaux s'appuient sur une résolution appropriée des équations de Maxwell.

Au « problème inverse » est souvent associé le « problème direct » complémentaire, qui consiste à déterminer le champ électromagnétique perturbé connaissant l'ensemble des paramètres géométriques et électromagnétiques de la configuration, défaut inclus. En pratique, cela est effectué via une modélisation mathématique et des méthodes numériques (par exemple, la méthode de moments) permettant la résolution numérique de tels problèmes. Les simulateurs correspondants sont capables de fournir une grande précision sur les résultats mais à un coût numérique important. Sachant que la résolution d'un problème inverse exige souvent un grand nombre de résolution de problèmes directs successifs, cela rend l'inversion très exigeante en termes de temps de calcul et de ressources informatiques. Pour surmonter ces challenges, les *modèles de substitution* (« surrogate modeling » (SM) en anglais) qui imitent le modèle exact peuvent être une solution alternative intéressante. Une manière de construire de tels modèles de substitution est d'effectuer un certain nombre de simulations exactes et puis d'approximer le modèle en se basant sur les données obtenues. Le choix des simulations (« prototypes ») est normalement contrôlé par une stratégie tirée des outils de méthodes de *plans d'expérience numérique* (« design of experiments », (DoE) en anglais).

Dans cette thèse, l'utilisation des techniques de modélisation de substitution et de plans d'expérience numérique dans le cadre d'applications en CNDE est examinée. Le simulateur EM – en supposant un modèle de défaut paramétrique[1] – est traité comme une boîte noire, c'est-à-dire, sans connaître explicitement les relations entre l'entrée et la sortie. L'entrée est un ensemble des paramètres descriptifs de la géométrie du défaut, la sortie est l'ensemble des données reliées au champ EM mesuré. Trois approches indépendantes sont examinées en détail : une méthode d'inversion basée sur l'optimisation d'une fonction objectif et deux approches plus générales pour construire des modèles de substitution en utilisant des échantillonnages adaptatifs.

Ce résumé est organisé comme suit. Dans la section 2, nous présentons les exemples illustratifs que nous

[1] C'est à dire décrit par un petit nombre de paramètres géométriques et/ou électromagnétiques

utiliserons tout au long de la thèse. Dans les trois sections 3, 4 et 5, les trois chapitres principaux de la thèse sont résumés. Toutes les sections commencent par un bref sommaire en italique de ma contribution la plus importante dans le domaine, la suite décrivant avec plus de détails les méthodes et les résultats obtenus.

2 Les exemples illustratifs en CND par courants de Foucault

Les approches proposées dans le cadre de cette thèse sont appliquées sur des exemples en CND par courants de Foucault. Une configuration planaire de référence a été choisie pour l'ensemble de la thèse, elle est constituée d'une plaque conductrice d'épaisseur finie et de dimensions latérales infinies affectée par une ou plusieurs fissures minces rectangulaires. Ces fissures sont à caractériser à partir de la mesure de la variation d'impédance d'une bobine alimentée en courant alternatif et déplacée au dessus de la plaque. Dans la plaque, des courants de Foucault sont générés par le courant alimentant la bobine. En présence d'une fissure, la circulation de ces courants est perturbée entraînant une variation de l'impédance de la bobine.

Dans tous les exemples présentés dans le document sont utilisées des données synthétiques obtenues grâce à un simulateur numérique. Ce simulateur (qui n'a pas été développé dans le cadre de la thèse) s'appuie sur une formulation intégrale des champs pour simuler l'interaction des fissures avec le champ électromagnétique. Les équations correspondantes sont résolues en utilisant une discrétisation par méthode de moments, et la variation de l'impédance de la bobine est calculée à l'aide du théorème de réciprocité.

Quatre exemples canoniques sont définis, en utilisant une hypothèse de (i) une fissure (définie par 2 ou 4 paramètres), (ii) deux fissures parallèles (définies par 6 ou 8 paramètres). Les paramètres sont les positions et/ou les dimensions de la/des fissure(s). Ces exemples sont utilisés tout au long de la thèse.

L'ensemble de toutes les entrées possibles est défini comme *l'espace d'entrée*. Les données (la variation d'impédance en fonction de la position de la bobine sur une surface de scan) sont définies via *l'opérateur direct* (réalisé par la simulation numérique). On appelle l'ensemble de ces données fonctionnelles *l'espace de données* ou *l'espace de sortie*.

3 Inversion par optimisation globale de type « Efficient Global Optimization »

*J'ai conçu une méthode d'inversion pour le contrôle non destructif en combinant un simulateur numérique de phénomènes électromagnétiques avec l'algorithme EGO (*Efficient Global Optimization*), lequel a besoin d'un petit nombre de simulations directes pour résoudre un problème inverse même si le nombre de paramètres à reconstruire peut être relativement important.*

Une manière classique d'envisager la résolution d'un problème inverse est de le réécrire sous la forme d'un problème d'optimisation souvent non linéaire d'une fonctionnelle coût (également appelée fonction objectif) mesurant l'écart entre les données mesurées et celles simulées. Le but étant alors de trouver les paramètres d'entrée optimaux de façon à ce qu'une fois introduit dans le simulateur numérique les données de sortie simulées soient aussi proches que possible des données mesurées. Cependant, on doit garder à l'esprit que non seulement le simulateur est coûteux à évaluer mais également que la fonction objectif peut avoir de nombreux minima locaux, il est donc nécessaire, autant que faire se peut, de réduire le nombre de simulations exactes à effectuer tout en conservant l'aptitude de l'algorithme à explorer les zones de l'espace des paramètres encore vierges.

Dans cette thèse, l'utilisation de la méthode EGO est proposée. Cette méthode – destinée à l'optimisation des fonctions coûteuses à évaluer – s'appuie sur un échantillonnage séquentiel judicieux de l'espace d'entrée grâce à l'utilisation d'une interpolation de la fonction d'objectif par krigeage. Le krigeage a été initialement développé dans le cadre de la géostatistique pour la modélisation des distributions spatiales, et est devenu une technique d'approximation très populaire maintenant. C'est une méthode stochastique d'interpolation spatiale qui prévoit la valeur prise par une fonction ou un phénomène naturel en des positions non échantillonnées par une combinaison linéaire sans biais et à variance minimale d'observations du phénomène ou de la fonction en des positions voisines. À cette fin, le krigeage modélise la fonction à interpoler par un processus aléatoire gaussien. Ce processus est prédit à partir d'une forme linéaire (combinaison linéaire des observations qui sont pensées comme les réalisations de variables aléatoires) dans un cadre stochastique : la variance d'erreur de la prédiction est alors minimisée. Une caractéristique très avantageuse du krigeage est également de fournir une estimation de cette variance qui est liée à la confiance à apporter à la prédiction.

L'algorithme EGO a besoin d'un certain nombre d'observations initiales au début de la procédure. Après l'initialisation, les évaluations nouvelles de la fonction objectif sont effectuées une par une. Le prochain échantillon d'entrée est choisi comme le point de l'espace d'entrée le plus prometteur dans le sens de l'optimisation : on choisit le point où on s'attend de trouver le valeur minimale de la fonction objectif en se basant les observations déjà effectuées. Ce choix est formalisé via un problème d'optimisation auxiliaire, qui est aussi défini sur l'espace d'entrée, mais dont la fonction objectif (qui s'appelle « expected improvement » (EI) en anglais) est beaucoup plus simple à évaluer car seulement basée sur la prédiction par krigeage de la fonction objectif originale. On peut penser à la fonction « expected improvement » (EI) comme une combinaison de deux facteurs : la variance de la prédiction et la prédiction elle-même. Ainsi si la variance de la prédiction est élevée (c'est-à-dire, la confiance dans la prédiction est faible) ou si la prédiction de la fonction coût est basse dans une certaine région de l'espace d'entrée (proche de la solution optimale), la valeur de la fonction EI sera élevée. EGO propose donc un équilibre entre une recherche locale dans les « régions prometteuses » et une recherche globale dans les « régions inexplorées ». En effectuant la prochaine observation toujours au point où l'EI est maximal, on obtient une séquence d'observations qui converge vers le minimum global de la fonction objectif (cette convergence peut être démontrée sous certaines conditions restrictives sur la fonction objectif). La vitesse de la convergence n'est pas connue en théorie, mais normalement très peu d'observations sont nécessaires pour obtenir le minimum global.

Nous présentons des exemples numériques illustratifs pour des cas allant de 2 à 8 paramètres d'entrée à optimiser. Les observations initiales sont choisies en utilisant un plan hypercube latin dans l'espace d'entrée. Comme règle d'arrêt, on utilise simplement une limite supérieure du nombre des itérations. L'algorithme d'inversion est efficace et la solution du problème inverse est trouvée dans presque tous les cas. On ne présente pas de comparaison approfondie entre notre méthode et des autres approches, mais on mentionne que certains algorithmes d'inversion stochastiques (par exemple, les algorithmes génétiques) ont besoin de beaucoup plus de simulations directes qu'EGO pour obtenir une solution raisonnable.

4 Modèle de substitution adaptatif par krigeage fonctionnel

Une méthode adaptative de création de modèle de substitution pour le contrôle non destructif a été proposée. Elle utilise le krigeage fonctionnel pour interpoler l'opérateur direct en se basant sur un ensemble d'observations choisies adaptativement afin d'améliorer la précision du modèle de substitution au cours de sa construction.

À la différence de la méthode d'inversion qui est présenté § 3, on peut imaginer des approches dont le but de l'échantillonnage adaptatif n'est plus la solution d'un problème inverse particulier mais la construction d'un modèle de substitution plus général. Une fois obtenu ce modèle de substitution peut alors être utilisé

dans la résolution de problèmes inverses. Pour autant l'objectif principal du travail proposé ici est de fournir une méthode de construction adaptative d'un modèle de substitution ayant une précision aussi grande que possible quels que soient les paramètres d'entrée.

Nous présentons un schéma de construction de modèle de substitution qui se compose de deux étapes : (i) le choix d'un certain nombre d'échantillons d'entrée de façon optimale et le calcul des données correspondantes par le simulateur ; (ii) la connexion d'un interpolateur aux échantillons précédemment calculés. L'avantage d'une telle séparation en deux étapes est la concentration de la difficulté du problème et du coût numérique du calcul dans la première étape normalement effectuée seulement une fois par des « experts », l'utilisateur final n'ayant pas accès au simulateur numérique (qui peut être compliqué et/ou beaucoup plus général que ce dont l'utilisateur a besoin pour son problème spécifique) mais seulement à la « base de données » pre-calculée et à l'interpolateur (modèle de substitution) peu coûteux en temps et suffisamment précis.

Nous proposons l'usage du krigeage fonctionnel comme interpolateur. C'est une extension récente de la théorie originale du krigeage qui sert à l'interpolation des données fonctionnelles. L'idée essentielle du krigeage fonctionnel est de considérer les fonctions de données – au lieu des observations ponctuelles d'une fonction – comme des entités individuelles et de fonder une prédiction sur ces observations fonctionnelles. Il modélise l'opérateur direct par un *processus fonctionnel* aléatoire gaussien. Pour autant la version fonctionnelle est très similaire à celle du krigeage classique : l'interpolation a une forme linéaire dont les coefficients sont choisis dans un cadre stochastique, en minimisant la variance de l'erreur de la prédiction (plus précisément une version intégrée de la variance appelée « trace-variance » en anglais).

Cependant, la précision fournie par un modèle de substitution (par exemple, dans le sens de l'erreur d'interpolation maximal) dépend fortement du choix des échantillons et de l'interpolateur utilisé. Les échantillons doivent donc être choisis conformément au problème physique à modéliser et à l'interpolateur utilisé c'est pourquoi nous avons choisi de développer un algorithme d'échantillonnage adaptatif. La stratégie que nous proposons est notamment contrôlée par l'erreur du modèle de substitution estimée dans le cadre stochastique du krigeage fonctionnel.

Notre algorithme d'échantillonnage est une procédure séquentielle initialisée par un certain nombre d'échantillons auxquels sont rajoutés un par un et de façon itérative de nouveaux échantillons jusqu'à ce qu'une règle d'arrêt soit satisfaite. Le choix de l'échantillon à ajouter implique la résolution d'un problème d'optimisation auxiliaire à chaque itération. La fonction objectif de cette optimisation est définie sur l'espace d'entrée et permet de choisir la position optimale du prochaine échantillon de façon à réduire l'erreur d'interpolation du krigeage fonctionnel autant que faire ce peut. Cette fonction objectif, choisie de façon heuristique, est une combinaison de deux facteurs : (i) l'un force le remplissage de l'espace d'entrée avec les échantillons, et (ii) l'autre mesure la variance de l'erreur de prédiction du processus aléatoire qui peut être considérée comme une indication de l'erreur d'interpolation. Ce choix permet également un équilibre entre favoriser l'ajout de points dans une région plutôt vide et l'ajout de points dans une région qui améliore le modèle de substitution. Cependant, cette variance n'est pas connue exactement et doit donc être estimée numériquement en utilisant l'estimateur de « jackknife » (une méthode statistique proche de la validation croisée pour l'estimation de fiabilité). Le problème d'optimisation est résolu en utilisant un algorithme de « recherches par coordonnées » sur une grille discrète de l'espace d'entrée.

Des exemples numériques (en utilisant de 2 à 6 paramètres d'entrée et différents nombres des échantillons) sont utilisés pour illustrer les avantages et inconvénients de l'approche adaptative proposée par rapport à des approches classiques (krigeage fonctionnel ou interpolation linéaire par morceaux à partir d'un échantillonnage « régulier » de l'espace d'entrée).

5 « Space filling » de l'espace des données (ou de sortie) et méthodes inverses

Une méthode d'échantillonnage adaptative pour la génération des bases de données dont l'objectif est le remplissage homogène (« space filling ») de l'espace de sortie est proposée. Une fois la base de données « optimale » obtenue, deux techniques permettant d'exploiter certaines méta-informations fournie par de telles base en vue de la caractérisation quantitative du problème inverse correspondant ont été développées.

Une base de données qui se compose de paires paramètres d'entrée–paramètres de sortie (données) peut être considérée comme une représentation discrète de l'opérateur direct. Elle peut, de la même manière que présentée § 4, former un cadre pour un modèle de substitution utilisant un interpolateur également simple.

Les plans d'essai classiques fournissent beaucoup d'outils pour la génération des ensembles d'échantillons en satisfaisant certains critères liés à la répartition des échantillons d'entrée (par exemple, le remplissage de l'espace, l'uniformité, …).

Nous présentons ici un algorithme d'échantillonnage dont le but est le remplissage « optimal » dans un certain sens de l'espace de sortie. La répartition des échantillons de données doit satisfaire deux conditions : (i) aucun échantillon n'est « trop proche » d'un autre échantillon ; (ii) pour tous les éléments de l'espace de données (ou de sortie), un échantillon existe « pas trop loin » d'un autre échantillon. L'avantage d'une telle approche est double : (i) en général, l'erreur d'une interpolation simple de type « plus proche voisin » est plus petite qu'en utilisant des bases de données qui remplissent l'espace d'entrée ; (ii) la répartition des échantillons d'entrée fournit certaines *méta-informations* sur le problème direct modélisé. Ces méta-informations sont fondamentalement reliées à la vitesse de variation de l'opérateur direct (qui a un rôle central pour la caractérisation du problème inverse) ; ainsi dans les régions où la densité des échantillons d'entrée est importante l'opérateur direct varie rapidement alors que les zones où il y a peu d'échantillons indique une faible variation.

Notre algorithme fonctionne par insertion et suppression alternées d'échantillons. La méthode s'appuie sur l'optimisation des distances mutuelles entre les échantillons dans l'espace de sortie. Pour contrôler ces distances, nous avons introduit les fonctions de distance, qui sont définies sur l'espace d'entrée mais expriment des distances mesurées dans l'espace de données (ou sortie). En faisant cela, le critère de l'insertion d'un échantillon peut être formalisé comme un problème d'optimisation qui est résolu numériquement en utilisant un algorithme de « recherches par cordonnées » sur une grille discrète sur l'espace d'entrée. Cependant, les fonctions de distance sont « coûteuses-à-évaluer » (parce qu'elles impliquent l'opérateur direct), ainsi nous proposons l'interpolation par krigeage pour approximer ces fonctions et donc réduire le coût de calcul de l'échantillonnage.

En se basant sur les fonctions de distance, nous avons développé deux approches pour la caractérisation quantitative du problème inverse. La première méthode vise à l'application inverse d'un niveau de bruit connu de l'espace de données à l'espace d'entrée. L'image de ce niveau de bruit décrit l'aspect mal-posé du problème inverse dans un certain sens : plus une telle « cellule de bruit » est large dans l'espace d'entrée (à niveau de bruit constant) plus il sera difficile de déterminer précisément les paramètres d'entrée. Il est également possible de considérer cette approche comme l'analyse de l'incertitude sur les résultats du problème inverse. La deuxième approche est aussi une application dans le cadre du problème inverse, mais le but est de trouver les images dans l'espace d'entrée des cellules de Voronoi définies dans l'espace de données (à partir des échantillons de la base de donnée et une norme utilisée pour mesurer les distances). Comme la répartition des échantillons de données est « uniforme », les images de ces cellules aussi caractérisent le problème inverse dans un certain sens. Plus précisément, la taille des images des cellules peut être considérée comme une description de la « précision » que l'on peut attendre de l'inversion en utilisant une telle base de

données. Bien évidement, cette précision dépend des paramètres d'entrée. Ces deux méthodes se ramènent à des inégalités simples en utilisant les fonctions de distance ; une fois de plus, pour en réduire le coût de calcul, le krigeage est utilisé pour évaluer, par interpolation, les fonctions distance.

Les méthodes (la génération des bases de données en remplissant l'espace d'entrée et les applications inverses) sont également illustrées par des exemples numériques en utilisant de 2 à 6 paramètres d'entrée). En plus des caractérisations numériques des bases de données (également comparées aux échantillonnages « régulières » dans l'espace d'entrée), nous présentons l'usage de la méthode de « réduction d'échelle multidimensionelle »[2] pour la visualisation réduite de la répartition des échantillons de données dans un espace bidimensionnel.

Pour illustrer la généralité de l'approche ici développée nous présentons aussi une étude hors du cadre du contrôle non destructif. Notre algorithme d'échantillonnage a été appliqué sur un simulateur numérique modélisant l'interaction d'une onde radar avec un couvert forestier. Les paramètres d'entrée sont alors la fréquence du radar et l'angle d'incidence des ondes, et les données observées sont certaines valeurs caractéristiques dérivées du champ électromagnétique mesuré, à savoir les coefficients de rétrodiffusion pour les différentes polarisations parallèles et croisées, l'atténuation et la hauteur interférométrique du couvert forestier). La structure des bases de données obtenues a fourni des informations intéressantes à la communauté de la télédétection.

[2]Traduction libre de la méthode dite de « Multidimensional Scaling (MDS) ».

Kísérlettervezési módszerek és helyettesítő modellek alkalmazása az elektromágneses roncsolásmentes anyagvizsgálatban

1 Bevezetés

Elektromágneses roncsolásmentes anyagvizsgálati (EMRA) módszereket az ipar számos területén alkalmaznak szerkezeti elemek rejtett anyaghibáinak feltárására. Az eljárások azon alapszanak, hogy az anyaghiba jelenléte lokálisan megváltoztatja az anyag elektromágneses (EM) paramétereit. Az EMRA célja az anyaghiba jellemzése a mért EM tér (amelyet egy külső forrás kelt, majd kölcsönhatásba kerül a vizsgált mintadarabbal) ismeretében, azaz a vonatkozó *inverz probléma* megoldása. Ehhez meg kell tudnunk határozni az ismert anyaghibákhoz tartozó EM teret, azaz képesnek kell lennünk a *direkt probléma* megoldására. A gyakorlatban erre a vizsgált EMRA eljárás matematikai modellezésén (a Maxwell-egyenletek alapján) és numerikus szimulációján (pl. a momentum-módszerrel) keresztül nyílik mód. Az ilyen EM szimulátorok igen pontos eredményeket is szolgáltathatnak, viszont ennek általában magas számításigény az ára. Egy inverz probléma megoldása azonban gyakran számos „költséges” (számítástechnikai értelemben) direkt szimuláció elvégzését igényli, így az inverzió igen idő- és erőforrásigényessé válik. Ennek leküzdésére a „helyettesítő modellezés” egyre inkább teret hódít az elektromágnesesség területén. A helyettesítő modell utánozza a valódi modellt, azonban az utóbbinál rendszerint sokkal egyszerűbb felépítésű. A helyettesítő modellek előállításának egy lehetséges módja, hogy néhány szimulációt elvégzünk, majd az eredmények segítségével közelítjük a valódi modellt. A „prototípus” szimulációk megválasztására többnyire valamilyen különleges stratégiát – egy kísérlettervezési módszert – használunk.

Az értekezésben bemutatott kutatómunka célja az EMRA eljárások hatékonyabbá tétele kísérlettervezési módszerek és helyettesítő modellek segítségével. Az EM szimulátort – egy paraméteres anyaghiba-modellt feltételezve – fekete doboznak tekintjük, azaz csak a bemenet-kimenet kapcsolatra támaszkodunk. A bemenet az anyaghiba geometriáját leíró paraméterek halmaza, a kimenet pedig a mérhető EM térrel (a mérőeszköz pozíciójának függvényében) kapcsolatos. Három önálló megközelítéssel foglalkozunk részletesen: egy optimalizáción alapuló inverziós eljárást és két, általánosabban használható helyettesítő modellek előállítására szolgáló adaptív mintavételezési módszert mutatunk be.

Ez az összefoglaló a következőképpen épül fel. A 2. szakaszban bemutatjuk a teszt példákat, amelyeket az értekezésben illusztrációként használunk. Az ezt követő három szakaszban (3., 4. és 5.) az értekezés három fő fejezetét összegezzük. Az egyes szakaszok az elért tudományos eredmény tézisszerű megfogalmazásával kezdődnek, a szakaszok további része lényegében a tézisek magyarázata.

2 Az illusztratív példák

A módszereket az örvényáramú roncsolásmentes anyagvizsgálatból vett példákkal illusztráljuk. A feltételezés szerint egy végtelen kiterjedésű fémlemezben téglalap alakú, vékony repedések találhatók. A feladat ezeknek a repedéseknek a jellemzése egy váltakozó árammal táplált tekercs impedanciaváltozásának

ismeretében. A tekercs mágneses tere a lemezben örvényáramokat kelt, amelyeknek az eloszlása repedés jelenlétében megváltozik, következésképpen a tekercs impedanciája is változik.

A példáinkban csak szintetikus (nem mérésből származó) adatokat használunk. A rendelkezésünkre álló szimulátor (amelynek a megalkotása nem képezi az értekezésben bemutatott munka részét) a repedés kölcsönhatását az elektromágneses térrel integrálegyenlet alakjában fogalmazza meg. Az integrálegyenlet numerikus megoldása a momentum módszeren alapszik, a tekercs impedanciaváltozásának számítása pedig a reciprocitási elv alapján történik.

Négy kanonikus példát definiálunk, (i) egy repedést feltételezve (2 vagy 4 paraméterrel), ill. (ii) két párhuzamos repedést feltételezve (6 vagy 8 paraméterrel). A paraméterek a repedés(ek) pozícióját ill. méretét adják meg.

A paraméterek együttese alkotja a modell *bemenetét*; az összes lehetséges bemeneti értékek halmaza pedig a *bemeneti tér*. A mért adat – amit a modell *kimenetének* nevezünk – a tekercs impedanciaváltozása a tekercspozíció függvényében, egy téglalap alakú tartományon. A bemenet-kimenet kapcsolatot formálisan a *direkt operátor* segítségével adjuk meg, amelyet a vizsgált elrendezés EM szimulátora realizál.

3 Inverzió a „Hatékony Globális Optimalizáció” algoritmussal

Kidolgoztam egy, roncsolásmentes anyagvizsgálatban használható inverziós módszert egy numerikus elektromágneses szimulátor és az ún. „Hatékony Globális Optimalizáció” algoritmus összekapcsolásával, amely inverziós eljárás kis számú direkt szimuláció elvégzését igényli az inverz probléma megoldásához még a keresett paraméterek viszonylag nagy száma esetén is.

Az inverzió egy klasszikus módja a mért adatok és egy alkalmas direkt szimulátor kimenetének összehasonlítása és az eltérés csökkentése a szimulátor bemenetének hangolásával. A (regularizált) inverz probléma megoldása ezek után a bemeneti paraméterek azon értékkombinációja lesz, amely a „legkisebb” eltérést eredményezi (egy megfelelően választott célfüggvény, pl. az eltérés négyzetes középértéke értelmében). Az ilyen optimalizáció alapú inverzió egyik legfőbb hátulütője az alkalmazott szimulátor számításigénye. Egyfelől az inverzió során végzett szimulációk számát kívánatos kis értéken tartani, másfelől viszont a célfüggvénynek több lokális minimuma is lehet (ami esetleg az inverz probléma gyengén meghatározottságára utal), így egy globális optimalizációs stratégiát kell használnunk.

Az értekezésben bemutatjuk a „hatékony globális optimalizáció” (az angol *Efficient Global Optimization* elnevezés alapján EGO-ként rövidített) módszer alkalmazását. Ez az eljárás – amelyet „költséges” kiértékelésű függvények optimalizálására fejlesztettek ki – egy, a célfüggvény *kriging* interpolációja alapján működő szekvenciális mintavételezési módszer.

A kriging interpolációt eredetileg a geostatisztikusok fejlesztették ki az 1960-as években, de mára általánosan használt függvényapproximációs technikává vált. Egy ismeretlen skalárfüggvényt interpolálhatunk néhány mintája alapján a kriging módszer segítségével. Az eljárás alapgondolata, hogy a függvényt egy gaussi véletlen folyamattal modellezzük, amelyet a megfigyelések egy bizonyos értelemben optimális lineáris kombinációjaként megbecsülünk. A modellező véletlen folyamat kovariancia függvényét többnyire becsülnünk kell a rendelkezésre álló minták alapján. A puszta interpoláció mellett a kriging felvilágosítást ad a közelítés bizonytalanságáról is, mivel megadja a folyamat becslési hibájának a varianciáját is.

Az EGO algoritmus a célfüggvény néhány kezdeti kiértékelése után egyesével hajtja végre a további függvényhívásokat. A soron következő minta beszúrása a bemeneti tér legígéretesebb pontjába történik: a módszer azt a bemeneti pontot választja ki, ahol a célfüggvény globális minimumának megtalálására a

legnagyobb esély mutatkozik az addig elvégzett kiértékelések alapján. Ez a kiválasztás szintén egy (másodlagos) optimalizálási feladatként fogalmazódik meg, azonban a célfüggvény (amelyet „várható javulásnak”, angolul „expected improvement”-nek (EI) neveznek) kiértékelése általában sokkal gyorsabb, mivel nem az eredeti célfüggvényen, hanem annak kriging interpolációján alapul. Az EI tekinthető két tényező együttes hatását kifejező függvénynek: a kriging becslés magas varianciája (azaz az interpoláció bizonytalansága) és a célfüggvény alacsony becsült értéke a bemeneti tér valamely tartományában egyaránt magas EI értéket eredményez. Az EGO algoritmus így „egyensúlyoz” az ígéretes tartományokban történő lokális keresés és a feltérképezetlen tartományok globális felderítése között. A soron következő minta maximális EI alapján történő választása egy olyan kiértékelés-sorozatot eredményez, amely a célfüggvény globális minimumához, azaz a regularizált inverz probléma megoldásához tart (bizonyos enyhe feltételek teljesülése esetén). A konvergencia sebessége általában nem ismert előre, azonban többnyire igen kevés függvényhívás elegendő a globális minimum megtalálásához.

Az inverziós algoritmus megvalósítása a bemeneti tér szabályos ráccsal történő diszkretizációján alapszik. Ez a rács bizonyos értelemben kifejezheti az inverzió elvárt pontosságának felhasználó által történő előírását. A másodlagos optimalizálási feladat (az EI maximalizálása) megoldására egy, ezen a rácson történő koordinátatengelyek mentén történő keresést („coordinatewise search algorithm”) alkalmazunk.

A módszert 2 - 8 bemeneti paramétert használó példákkal illusztráljuk. A kezdeti kiértékeléseket a bemeneti térben generált latin hiperkocka segítségével választjuk; a kezdeti minták száma a bemeneti paraméterek számával lineárisan növekszik. Az algoritmus leállási feltétele egy egyszerű felső korlát az iterációk számára, amely szintén lineárisan növekszik a bemeneti paraméterek számával (a 8 paraméteres példában 320 iterációt engedélyezünk). Az inverziós módszer igen hatékonynak tűnik: az inverz probléma helyes megoldását majdnem minden esetben megtalálta a példáinkban. Bár alapos összehasonlítást nem végeztünk az EGO és más optimalizációs módszerek között, megemlítjük, hogy bizonyos egyéb sztochasztikus inverziós algoritmusok (pl. genetikus algoritmusok használatával) jóval több direkt szimuláció elvégzését igénylik, mint az EGO.

4 Helyettesítő modellezés a „függvény-kriginggel”

Kidolgoztam egy, roncsolásmentes anyagvizsgálatban használható helyettesítő modellezési eljárást, amely „függvény-kriginget” használ a direkt operátor interpolációjára, egy olyan adatbázisra támaszkodva, amelyben a tárolt mintákat az eljárás adaptívan választja meg a helyettesítő modell pontosságának növelése érdekében.

Az előző szakaszban tárgyalt inverziós módszer mellett elképzelhetők olyan adaptív mintavételezési eljárások is, amelyeknek célja nem egy partikuláris inverz probléma megoldása, hanem egy általánosabban alkalmazható helyettesítő modell megalkotása. Természetesen ilyen modelleket alkalmazva is lehet a végső cél az inverzió, azonban a legfontosabb kritérium a helyettesítő modell pontossága a teljes bemeneti tér felett.

Az értekezésben bemutatunk egy kétlépcsős helyettesítő modellezési eljárást: (i) kiválasztunk egy bizonyos számú bemeneti mintát, amelyekre a szimulációt elvégezzük, és az eredményeket egy adatbázisban tároljuk, majd (ii) egy interpolátort illesztünk a kapott mintákhoz. Az eljárás nagy előnye, hogy a számításigényes első lépcsőt „szakértők” hajthatják végre (akik rendelkeznek a bonyolult és/vagy a felhasználó igényeinél sokkal általánosabb célú szimulátorral), továbbá az adatbázis-generálást csak egyszer kell elvégezni. Ezek után a „végfelhasználó” a második lépcsőben csak a tárolt adatokhoz fordul és interpolációt alkalmaz. Ilyen módon lényegében egy problémaspecifikus helyettesítő modellt lehet a végfelhasználó rendelkezésére bocsátani.

Interpolátorként a „függvény-kriginget” alkalmazzuk, amely – az eredeti elmélet egy igen új kiterjesztése révén – lehetővé teszi teljes függvények, mint entitások becslését, szemben a klasszikus kriging céljával, ami egyetlen függvény pontonkénti becslését jelenti. A módszer alapgondolata a vizsgált függvény értékű direkt operátor modellezése egy függvény értékű („functional”) gaussi véletlen folyamattal. Ettől a különbségtől eltekintve, az eljárás igen hasonló a klasszikus kriginghez: az interpoláció a kiértékelések lineáris kombinációja, ahol az együtthatókat egy sztochasztikus formalizmuson keresztül határozzuk meg, a becslés hibavarianciáját minimalizálva (pontosabban ennek egy integrált változatát, amit angolul „trace-variancie”-nek neveznek). Mivel az EM szimulátorok gyakran függvény értékű kimenettel rendelkeznek, ez az eljárás különösen hasznosnak bizonyul ezen a területen.

A helyettesítő modell nyújtotta pontosság (pl. a maximális interpolációs hiba a bemeneti tér felett) azonban erősen függ a minták és az interpolátor megválasztásától. Következésképpen a minták megválasztásánál egyaránt figyelemmel kell lennünk a modellezett direkt problémára és a választott interpolátorra, azaz egy adaptív mintavételező algoritmust kell használnunk. Kidolgoztunk egy ilyen stratégiát, amelyet a helyettesítő modell – a függvény-kriging sztochasztikus keretein belül becsült – interpolációs hibája vezérel.

A mintavételezési algoritmus egy szekvenciális stratégia, amely néhány kezdeti mintával indul, majd az új megfigyelések egymás után adódnak hozzá az adatbázishoz, mígnem egy leállási feltétel teljesül. A soron következő minta megválasztása egy optimalizálási feladatként fogalmazható meg, amelyet minden ciklusban meg kell oldanunk. Ennek a másodlagos optimalizálásnak a célfüggvénye a bemeneti tér felett értelmezett, és azt mutatja meg, hogy hol érné meg a leginkább a következő mintát elhelyezni a helyettesítő modell hibájának lehető legnagyobb csökkentése érdekében. Ez a célfüggvény két tényező heurisztikus kombinációja: (i) a bemeneti tér jó kitöltésének és (ii) a kriging becslés hibavarianciájának – amely az interpolációs hiba bizonyos becslésének is tekinthető – csökkentésének igénye egyaránt szerepet kap benne. A hibavariancia azonban nem ismert zárt alakban, ezért annak becsült értékét használjuk (a „zsebkés” elnevezésű varianciabecslő statisztikai módszer alkalmazásával). Az optimalizálási feladat megoldására egy, a bemeneti teret diszkretizáló szabályos rácson történő koordinátatengelyek mentén történő keresést alkalmazunk.

A bemutatott numerikus példák (2 - 6 bemeneti paramétert alkalmazva) igazolják a javasolt adaptív mintavételezés előnyét bizonyos egyéb módszerek (függvény-kriging ill. szakaszonként lineáris interpoláció szabályos, rács-diszkretizációval történő mintavételezéssel a bemeneti térben) felett az esetek többségében, azonban egy ellenpélda is bemutatásra kerül.

5 Kimeneti teret kitöltő adatbázisok és inverz leképezések

Kidolgoztam egy, a roncsolásmentes anyagvizsgálatban használható adaptív mintavételezési eljárást a kimeneti teret kitöltő anyaghiba-adatbázisok létrehozására, valamint kidolgoztam két módszert az ilyen adatbázisokban tárolt bizonyos meta-információ kinyerésére a kapcsolódó inverz probléma kvantitatív jellemzése céljából.

Egy bemeneti paraméter - kimeneti adat párokból álló adatbázis tekinthető az adott direkt operátor egyfajta diszkrét reprezentációjának. Ilyen adatbázisok képezhetik olcsó helyettesítő modellek alapját, ha a tárolt mintákhoz egy interpolátort illesztünk (ahogyan az előző szakaszban erre rámutattunk).

A klasszikus kísérlettervezési módszerek számos eszközt biztosítanak olyan mintahalmazok generálására, amelyek eleget tesznek bizonyos, a bemeneti minták eloszlására vonatkozó előírásnak (pl. térkitöltés, egyenletesség).

Az értekezésben bemutatunk egy mintavételezési módszert, amelynek célja a kimeneti tér bizonyos értelemben optimális kitöltése. A kimeneti adatok mintáinak eloszlására két feltételt szabunk: (i) egyik minta sincs „túl közel" egy másik mintához; (ii) a kimeneti tér bármely pontjától „nem túl messze" található egy tárolt minta. Az ilyen kimeneti teret kitöltő adatbázisok haszna kettős: (i) a legközelebbi szomszéd interpoláció hibája többnyire kisebb, mint a bemeneti teret kitöltő adatbázisok használata esetén; (ii) a bemeneti minták eloszlása bizonyos meta-információt hordoz a modellezett direkt operátorról (pl. jelzi azokat a tartományokat a bemeneti térben, amelyek fölött a direkt operátor „lapos", vagy éppenséggel gyorsan változik), amely információ a kapcsolódó inverz probléma szempontjából is nagy jelentőséggel bír.

A mintavételezési stratégia felváltva végez egymás utáni mintabeszúrásokat és eltávolításokat, a kimeneti minták páronkénti távolságától vezérelten. Ezeknek a távolságoknak a kézben tartására bevezettük a bemeneti térben definiált, ám a kimeneti térben mért távolságokat leíró távolságfüggvényeket. Ezekkel a mintabeszúrási feladat optimalizálási problémaként fogalmazható meg, amelyet egy, a bemeneti teret diszkretizáló szabályos rácson történő koordinátatengelyek mentén történő keresés alkalmazásával oldunk meg. Mivel azonban a távolságfüggvények kiértékelése költséges (ugyanis tartalmazzák a direkt operátort), a kriging interpolációt alkalmazzuk a közelítésükre, a mintavételezési stratégia számításigényének csökkentése érdekében.

A távolságfüggvényekre és azok kriging interpolációjára támaszkodva kifejlesztettünk két módszert az inverz probléma kvantitatív jellemzésére. Az első eljárás egy, a kimeneti térben adott zajszint bemeneti térbe történő inverz leképezését célozza meg. A zajszint képe bizonyos értelemben leírja az inverz probléma gyengén meghatározottságát: minél nagyobb egy ilyen „zaj-cella" a bemeneti térben (egy adott zajszint esetén), annál gyengébben meghatározott az inverz probléma a vizsgált tartományban. Másképpen szólva, ez az eljárás tekinthető a bizonytalanság-terjedés vizsgálatának az inverz irányban. A második eljárás – amely szintén egy inverz leképezés – a kimeneti térben az adatbázis mintái és a távolságot definiáló norma által megszabott Voronoi cellák bemeneti térbe történő visszavetítésére szolgál. Mivel a kimeneti minták eloszlása „egyenletes", a cellák képei tekinthetők az adott adatbázissal az inverzió során elérhető „pontosság" (az adatbázis nyújtotta „felbontás") bizonyos leírásának is. Mindkét módszer a távolságfüggvényekre felírt egyszerű egyenlőtlenségek megoldását igényli, amelyek számításigényének csökkentésére a távolságfüggvények kiring interpolációját alkalmazzuk.

A módszereket (a kimeneti teret kitöltő adatbázis-generálást ill. az inverz leképezéseket) 2 - 6 bemeneti paramétert alkalmazó példákkal illusztráljuk. Az adatbázisok numerikus jellemzésén (és a bemeneti térben szabályos rács-mintavételezést alkalmazó adatbázisokkal való összehasonlításán) túl bemutatjuk a „többdimenziós skálázás" használatát a kimeneti minták eloszlásának redukált, kétdimenziós ábrázolására.

Bemutatunk továbbá egy érdekes vizsgálatot az EMRA keretein kívül is: az adatbázis-generáló algoritmust alkalmaztuk erdős területek radarral történő megfigyelésének direkt problémájára. A bemeneti paraméterek az alkalmazott frekvencia és az EM hullámok beesési szöge voltak, kimeneti adatoknak pedig a szórt EM térből származtatható bizonyos jellemzőket (reflexió, csillapítás, az erdő interferometrikus magassága) tekintettünk. A kapott adatbázisok szerkezete értékes következtetések levonását teszi lehetővé a távérzékeléssel foglalkozó szakemberek számára.

List of abbreviations

BLUP	Best Linear Unbiased Prediction
CV	Cross Validation
CWS	Coordinatewise Search
DoE	Design-of-Experiments
ECT	Eddy-Current Testing
EGO	Efficient Global Optimization
EI	Expected Improvement
EM	Electromagnetic
ENDE	Electromagnetic Nondestructive Evaluation
FEM	Finite Element Method
FF	Full-Factorial
I/O	Input/Output
LHS	Latin Hypercube Sampling
LOOCV	Leave-One-Out Cross Validation
MDS	Multidimensional Scaling
ML	Maximum Likelihood
NDE	Nondestructive Evaluation
NN	Nearest Neighbour
OSF	Output space-filling
PDF	Probability Density Function
REML	Restricted Maximum Likelihood
SM	Surrogate Model(ing)

List of notations

$\mathbf{p}$	Input vector
p_i	The ith component of the input vector
$\mathbf{p}_i$	The ith input sample
$\mathbb{P}$	Input space
$q(t)$	Output function
$q_{\mathbf{p}_i}(t)$	The ith output sample
$\mathbb{Q}$	Output space
$\mathbb{D}_n$	Database consisting of n samples
$\left(\mathbf{p}_i,\, q_{\mathbf{p}_i}(t)\right)$	The ith sample in $\mathbb{D}_n$
$\mathcal{F}\{\mathbf{p}\}$	Forward operator
$\mathcal{G}^{(n)}\{\mathbf{p}\}$	Approx. fwd. operator based on n samples
$Q(\mathbf{p})$	Objective/auxiliary/distance function
$\xi(\mathbf{p})$	Gaussian process
$\xi_{\mathbf{p}}(t)$	Functional gaussian process
$k(\cdot)$	Covariance function
$k_T(\cdot)$	Trace-covariance function

1

Introduction

The main goal of the work presented in this Dissertation is to contribute to the Electromagnetic Nondestructive Evaluation (ENDE) by introducing Design-of-Experiments (DoE) and Surrogate Modeling (SM) methods. In this chapter, a brief overview of the studied domains is given. First, the basic concepts, goals and tools of NDE are discussed, mostly focusing on the attempts for the solution of the inverse problem. We tried to gather the most important approaches of NDE inversion in a short bibliographic overview. Then, DoE and SM are discussed. We point out that effective SM hardly exists without an appropriate DoE tool. Started from the classical approaches and concluded by the contributions from the recent years, a short review of the related literature is given. Finally, the research objectives of this work are specified and the organization of the Dissertation is presented.

1.1 Nondestructive evaluation

1.1.1 Goals and techniques

NDE is applied to gather information about the inner material structure of solid objects (*specimens* in the following) without damaging them. This occurs for various industrial applications, typically, to evaluate the expected remaining lifetime of a structure. The unrevealed material faults (e.g., cracks, inclusions, corrosion or voids –generally, any deviation from the "normal" estate–, hereafter *defects*) might lead to an unexpected fracture of the structure, possibly causing disaster if this happens in operation. On one hand, the sake of safety would inspire the change for a new component long before the current one reaches its expected lifetime, but on the other hand, economical considerations suggest to push the tenure to the limit. Finding a good trade-off under these concurrent conditions is impossible without an appropriate NDE method.

Without completeness, the ultrasonic, the X-ray and the branch of electromagnetic methods can be mentioned as the most popular ones. Among electromagnetic NDE methods [Blitz, 1991; McIntire, 1986], Eddy-Current Testing (ECT), DC-magnetic field measurements, methods based on flux-leakage and on Barkhausen-noise are the most widespread. All of these electromagnetic approaches are based on the fact that the presence of a defect locally changes the electromagnetic parameters (conductivity, magnetic permeability) of the specimen. A detailed presentation of the above mentioned methods would be out of our scope, herein we confine ourself to ECT, being used as an illustration of the developed approaches in this Dissertation.

The principle of ECT is to generate eddy-currents within the conductive specimen by using a time-varying magnetic field. In the presence of a defect, the eddy-current paths are changed compared to the defect-free case. Thus, the measured magnetic field of the eddy-currents carries some information about the

inner material structure. The exciting magnetic field is usually generated by a coil, fed with time harmonic-current. To receive the field due to the eddy-currents, in the simplest case, the variation of the impedance of the exciting coil is measured at pre-defined coil positions. The ECT signal consists in the set of the measured impedance variations. More complicated ECT configurations (separated exciting and receiving coil, coil-arrays, fluxset-sensors, current-impulse excitation, etc.) are also available, however, we do not deal with them herein. Typical ECT applications are the control of riveted bindings of airplanes, the control of turbine windings, or the evaluation of steam-generator pipes in nuclear reactors. Needless to say that all of these tasks must be treated with special attention.

The problems related to nondestructive evaluation can traditionally be classified as *forward* (also called *direct*) problems and *inverse* problems.

The forward problem of NDE consists in the determination of some observable quantities (e.g., the impedance variation of the coil in ECT, generally: the output (also called response) signal) based on the exact knowledge of the configuration, including the defect. Theoretically, this can be achieved in two ways: by performing a measurement on the configuration of interest, or by constructing an appropriate mathematical model and carrying out a numerical simulation. When the forward problem occurs as a part of an inversion scheme, one always recourses to modeling and simulation, whereas the measurement usually serves for the preliminary validation of the simulator. In electromagnetic NDE, the models are based on the Maxwell's equations. For the numerical simulation, numerous methods are used (typical and well-known ones are the Finite Element Methods (FEM), various discretization techniques and Method-of-Moments schemes for integral equations, Finite Difference (FD), Finite Difference Time Domain (FDTD) methods, Finite Integration Techniques (FIT), . . .). These numerical simulators are implemented in commercial softwares as well. Some of the most widely-used ones are the COMSOL Multiphysics, FLUX 3D (FEM softwares), XFdtd (an FDTD simulator), CST (a FIT software), CIVA (integral equation-based code), see the references: [*CIVA*; *COMSOL*; *CST*; *FLUX 3D*; *XFdtd*].

As a rule, these numerical simulators are able to provide accurate results, but at the price of a high computational cost. The computer resource requirements and time consumption of these simulators make them "expensive-to-run".

The inverse problem is the counterpart of the forward one: the configuration is to be determined based on the knowledge of some observable quantities [Anger, 1990; Tarantola, 2005]. In nondestructive evaluation, the inverse problem means the characterization of the defect using some observed data. The goal is then to transform the information carried by the observed data into knowledge on the defect. The "knowledge on the defect" is represented by a certain defect description, which is a key point of the inversion. For instance, this description can be a *contrast function* defined over a volume within the specimen. The contrast function expresses the difference (due to the defect) between the actual and the normal value of the Electromagnetic (EM) constitutive parameters. When a magnetic material is considered, both electric and magnetic contrast functions can be introduced. Another way to describe defects is to introduce a parametrized model, then a finite set of scalar parameters uniquely defines a defect.

The available information is usually not sufficient for the complete characterization of the defect. The inverse problems are often *ill-posed*: the solution might not exist, might not be unique, and might be unstable [Hadamard, 1952]. These attributes do even hold when one has an exact knowledge on the observable output –which is not the practical case as some noise always corrupts the measurements. The lack of information might be coped with (i.e., the ill-posedness can be reduced) by involving some *a priori* knowledge on the defect –in mathematical terms, the problem can be *regularized*. The regularization might be, for instance, an assumption on the contrast function (e.g., it can only have two values: the defect is homogeneous) or on the set of parameters (e.g., lower and upper bounds for each parameter). The choice of a parametric model is in fact also a regularizing assumption. The first step of regularization is very often the specification of a volume domain (the so-called region of interest), and the defect is assumed to lie within this region.

Beyond the theoretical pitfall of being ill-posed, the inverse problems give rise to computational challenges as well. NDE inverse problems are solved via forward problems in most of the cases. Due to the high computational cost of the forward simulation, the whole inversion process might be extremely time-consuming –which is a central inspiration of our work being presented in this Dissertation.

1.1.2 Some results in NDE inversion

Some significant efforts related to NDE inverse problems (with special emphasis on ECT) are collected in this section. The following sub-categories are not disconnected and obviously do not cover the whole domain –it is far much wider than even a rough taxonomy could be depicted within such limited section. Depending on how the defect is described, which a priori information is available and how it is complicited to regularize the problem, a huge variety of approaches exists. The referred works can rather be considered as arbitrarily chosen "capstones" of the research field, to underline and illustrate its diversity. Though many inversion schemes are closely related to the forward model they are based on, we focus on the main idea of the reconstruction method, often ignoring modeling issues.

Optimization-based approaches

The central idea of optimization-based inversion is to compare the measured output signal to the simulated one (via an objective function), and try to achieve the best similarity between them by systematically varying the assumed (currently simulated) defect configuration. The inversion then boils down to the optimization of one (or more) objective function(s). Depending upon many factors (e.g., the forward model, the description of the defect, the regularization enforced, the optimization strategy, etc.) several methods exist. The common point of them is the iterative nature of the inversion scheme. The optimization-based approaches form two main classes: methods accounting for the gradient of the objective function, and others being gradient-free techniques.

The gradient-based schemes account for the gradient of the objective function. This obviously implies additional computational cost, since besides the forward solution, this gradient is also to be computed. In [Norton and Bowler, 1993], a classical ECT inversion scheme is presented. The defect is described by a single scalar contrast function (related to the electric conductivity of the medium). The goal is to minimize an objective functional (defined over a set of possible contrast functions, expressing the mean-squared discrepancy between the measured and simulated output data). The gradient (the Gateaux derivative) of the objective functional with respect to the contrast function can be analytically expressed by solving two forward problems: the ordinary one and its adjoint. Then, a parametrized defect model is chosen, i.e., the objective functional is reduced to a function of the N parameters of the chosen model. The gradient of the objective function with respect to the parameters can also be expressed, making possible the use of a gradient-based iterative optimization scheme for the minimum search (the steepest descent, the conjugate gradient and the Levenberg-Marquardt algorithms are proposed in [Norton and Bowler, 1993]). Main benefits of such adjoint problem-based gradient computation are that (i) only two forward solutions are needed at each iteration (instead of N, which would be the case when computing the gradient numerically) and (ii) the computed gradient is exact. The approach of [Norton and Bowler, 1993] can be extended to the characterization of an ideally thin crack, i.e., which is represented by a mathematical surface on which the normal component of the electric field is zero [Bowler, Norton, and Harrison, 1994].

By assuming that the EM field within a volumetric defect is equal to the field in the defect-free specimen (this is the Born-approximation [Nair and Rose, 1990]), the inverse problem can be linearized, thus, the gradient of the objective functional becomes straightforward to compute. In [Sabbagh et al., 1990], a linearized model and an inversion scheme using the linear conjugate gradient algorithm are discussed.

Besides the classical gradient-based approaches (such as the steepest decent, conjugate gradient or Levenberg-Marquardt methods), more sophisticated tools have also been used to solve NDE inverse problems. The "modified gradient" methods (i) do not require a full forward solution in each iteration, and (ii) do not successively linearize the differential equations. The paper [Souriau et al., 1996] investigates the use of a modified gradient algorithm, where, in addition, a binary constraint is specified (i.e., the defect is assumed to be homogeneous with known constitutive parameters). Also in [Lambert, Lesselier, and Kooij, 1998], a modified gradient approach is used for the retrieval of a cylindrical obstacle entirely buried in a known homogeneous half-space.

In many approaches, the contrast function and the total EM field are considered as unknowns in the inversion formalisms. However, by choosing the contrast function and the distribution of equivalent sources (representing the defect in, e.g., an integral equation-based model) as unknowns in the inversion method, the convergence of the iterative scheme can significantly be accelerated, beyond the advantages already mentioned in the case of the modified gradient approaches. This is the main principle of the contrast-source inversion (CSI) [van den Berg and Kleinman, 1997]. The retrieval of a dielectric object is discussed in [Abubakar and Berg, 2002], whereas an eddy-current application is presented in [Dos Reis, Lambert, and Lesselier, 2002], using the CSI method under a binary constraint (i.e., the assumption of homogeneity of the defect). More recent versions of CSI are dealt with, e.g., in [Abubakar et al., 2008; Trillon et al., 2010] –in both contributions, a finite-difference method is used in the inversion scheme.

Three gradient-based approaches are compared in [Monebhurrun, Duchêne, and Lesselier, 1998] via an eddy-current steam generator inspection study: a modified gradient method, a kind of contrast-source method and an approach using the so-called localized nonlinear approximation (not discussed here).

An interesting inversion method is proposed in [Santosa, 1996]: the shape of the object to be retrieved is modeled by level sets. The controlled evolution of level sets for the solution of inverse problems is reviewed in detail in [Dorn and Lesselier, 2006]. An ECT application (metal tube inspection) can be found in [Abascal et al., 2008].

Inversion schemes based on Finite Element Method (FEM) forward modeling are discussed in [Badics et al., 1998a; Badics et al., 1998b]: a pre-calculated "reaction" data set (related to the finite element discretization) is used to speed-up both the forward simulation and the computation of the gradient of the objective function. A similar approach is presented in [Fukutomi et al., 1998], combined with a gradient-free (trust region) optimization scheme.

The gradient-free schemes are preferable in the cases when the gradient of the objective function cannot easily be computed. This might happen when the defect is assumed to be a polygon-shaped crack or a cuboid-shaped void, with the edge lengths as parameters. A crack-shape reconstruction algorithm is proposed in [Pávó and Miya, 1994]: a thin-crack model (using a surface integral equation) is combined with a simple, zeroth-order optimization strategy. In [Fukutomi et al., 1998], an inversion scheme using precomputed finite element databases (similar to the method in [Badics et al., 1998a; Badics et al., 1998b]), combined with a gradient-free (trust region) optimization approach.

The stochastic approaches are typical gradient-free optimization methods. In electromagnetic NDE, such approaches have also been widely used for inversion. For instance, a genetical algorithm –an evolutionary strategy– is used in [Zaoui, Marchand, and Pávó, 2001] to characterize thin cracks in a metal plate. Particle Swarm Optimization (PSO) –a population based stochastic optimization technique– is applied in [Cacciola et al., 2007] for imaging of corrosion based on ECT measurements. An example for the combined use of stochastic and deterministic methods is presented in [Pávó, 1994]: the stochastic simulated annealing is applied first to search for some "rough" solutions, then, taking them as initial configurations, the final solution is found by means of the deterministic downhill simplex strategy.

As a rule, the stochastic approaches need many more objective function evaluations (i.e., forward simu-

lations) to obtain the solution of the inverse problem than deterministic optimization strategies. On the other hand, however, the stochastic methods usually perform a global search over the whole feasible domain of input parameters, whereas the gradient-based deterministic tools are inherently local-search algorithms (i.e., possibly stalling at a local minimum of a multimodal objective function).

Signal processing approaches

A number of defect reconstruction methods is based on signal processing tools applied on the measured response signal. In [Simone and Morabito, 2000], the authors propose a method for the separation of multiple cracks; in [Taniguchi et al., 2001], a wavelet-transform approach is dealt with, whereas image processing methods are used to determine the orientation of a crack in [Taniguchi et al., 2002].

A MUSIC (MUltiple SIgnal Classification)-type algorithm is proposed in [Ammari et al., 2007] for the localization of a collection of small inclusions buried in a homogeneous medium. MUSIC is applied on the response signal, more precisely, on the multistatic response matrix, whose singular value pattern is related to the defect configuration. The imaging of a thin inclusion is investigated in [Park, 2010; Park and Lesselier, 2009], also using MUSIC-based algorithms. An ECT application for the MUSIC approach is presented by [Henriksson, Lambert, and Lesselier, 2010].

The so-called monotonicity method can also be considered as a signal processing tool in a certain sense. Based on the monotonicity method, a non-iterative inversion scheme is presented in [Tamburrino and Rubinacci, 2002]. The approach has been extended to eddy-current testing cases, as recently discussed in [Tamburrino, Ventre, and Rubinacci, 2010].

Neural networks

In [Ramuhalli, Upda, and Upda, 2001], neural networks are trained and applied within an iterative inversion scheme in a magnetic flux leakage NDE technique. In [Bensetti et al., 2004], the authors also use a neural network in an ECT application, whereas the combined use of a signal processing tool (principal component analysis) and a neural network is presented in [Le Bihan, Pávó, and Marchand, 2008]. An optimally chosen dataset can also be used to train the neural network for defect reconstruction [Gyimóthy, Le Bihan, and Pávó, 2007].

Statistical approaches

The regularized inverse problems can be interpreted as Bayesian inference problems, thus, they can be handled in a statistical framework. These approaches are also out of the scope of this Dissertation; we only mention that, e.g., [Idier, 2008] is an up-to-date textbook in the domain. This statistical viewpoint can be applied in the framework of ECT, as presented in [Prémel and Baussard, 2002].

Surrogate model-based approaches

An approach closely related to the work presented later on in this Dissertation for NDE inversion is the use of surrogate models. The latter is a cheap approximation of the usually time-consuming forward simulations (see § 1.2). Meshing methods –inspired by meshes of Finite Element Methods– have been experienced to be effective for the generation of surrogate models in ECT. Basic concepts are discussed in [Pávó and Gyimóthy, 2007], and more advanced, recent developments are addressed in [Gyimóthy, Kiss, and Pávó,

2009]. In [Franceschini, Lambert, and Lesselier, 2009], the authors also deal with the use of such mesh-databases in ECT. An ECT inversion technique based on a mesh-supported metamodel and a particle swarm optimization approach is presented in [Douvenot, Lambert, and Lesselier, 2010a; Douvenot, Lambert, and Lesselier, 2010b]. Beyond the mere solution of the inverse problem, in the latter two papers and also in [Gyimóthy and Pávó, 2005], also some manners to characterize the inverse problem are proposed.

1.2 Design of experiments and surrogate modeling

1.2.1 Basic concepts and terms

Design-of-experiments (DoE) is the collective name of information-gathering techniques related to observations of a given phenomenon. At the beginning of DoE's long history, the word "experiment" referred to classical (e.g., physical, chemical) experiments, whereas nowadays running a computer code of a simulation is also considered as a "computer experiment". The experiment is controlled by a set of *input* variables (or factors in DoE terms) and a set of *output* (or *response*) variables (or functions) can be observed. The set of all feasible input combinations spans the *input space* (also called experimental domain). One execution of the experiment (with a specified value for all inputs) is called an experimental run or trial. The main difference between real and computer experiments is the fully deterministic nature of the latter: multiple runs at the same input values result in exactly the same output values in the case of computer experiments.

Experimental runs are performed to draw conclusions about the studied phenomenon. To optimize the gain of information from a set of runs, one must execute these runs according to an appropriate DoE method. Depending on what the term *information* means and on the sense the optimality is defined in, various DoE approaches are available. The main branches are briefly summarized in § 1.2.2.

DoE is usually applied for complex experiments, whose inner structure cannot exactly be defined (real experiments) or is not in the focus of interest at the level of execution (computer experiments). Ignoring the knowledge of the inner functionality leads us to black-box models: only the relation of the input-output variables, the so-called Input/Output (I/O) function is dealt with. In our approaches only such black-box models are considered.

DoE methods are often involved in building a *surrogate model* (also called *metamodel*) for the cheap approximation of an expensive-to-run experiment. Surrogate modeling –being almost inseparable from design-of-experiments (thus, the notation DoE/SM is often used hereafter)– is an exhaustively studied research domain nowadays. Surrogate models are usually based on the interpolation of the I/O function of the black-box to be modeled. In this case, DoE is used to select those input values at which experimental runs are then performed. The I/O function is interpolated based on these observations –which interpolation can be done at a much less computational cost than performing an experimental run. When DoE is used to build a surrogate model, the input values at which an experimental run is performed are called *input samples*, the related observations are the *output samples* –then, a DoE method is indeed a sort of *sampling strategy*.

1.2.2 Some results of DoE and surrogate modeling

To illustrate the efforts that have been –and are still being– made in the research of DoE/SM, a brief overview of the literature is given below. However, this is neither a complete nor a taxonomical presentation, it rather aims at revealing the bibliographical background of our work.

Classical DoE

Classical approaches (see, e.g., [Fang, Li, and Sudjianto, 2006; Ryan, 2007; Santner, Williams, and Notz, 2003] as recent textbooks) are based only on geometrical considerations related to the distribution of input samples. These methods do not take the model (to be fitted to the samples, e.g., an interpolator) into account, i.e., the sampling is independent from the subsequent use of the samples.

For instance, a full-factorial (FF) design sets the input samples at the nodes of a regular grid spanned in the input space. A drawback of the FF design is the exponential dependence of the sample number on the number of the input parameters. Other variants of the factorial design also exist (e.g., central composite design, or fractional factorial design). Latin Hypercube Sampling (LHS) was introduced to ensure the uniform representation of the whole range of all input parameters in the design, but at a much smaller sample number than using a FF design [McKay, Beckman, and Conover, 1979]. Several additional criteria, like the maximin property [Biedermann and Dette, 2001] or the orthogonality [McKay, Beckman, and Conover, 1979] of the sampling can also be specified for LHS designs. Both factorial and Latin hypercube designs aim at providing a uniform sample distribution in the input space (space-filling design). However, LHS has better space-filling properties than factorial designs.

The fundamental contribution [Sacks et al., 1989] provides a thorough review of the first attempts about computer experiments, moreover, it proposes a method for the interpolation based on the samples. More recently, classical DoE approaches are discussed in [Giunta, Wojtkiewicz, and Eldred, 2003; Pronzato, 2008] (the latter with applications in control theory, citing almost 200 references). Let us note that in [Giunta, Wojtkiewicz, and Eldred, 2003], the authors use the expression “classical DoE” related to the physical experiments, whereas all computer-related approaches are denoted as “modern”.

Model-oriented DoE

The model-oriented (sometimes called quasi-adaptive) DoE approaches formalize a criterion related to the model (to be fitted to the samples). In other words, the model has to be chosen first, then the sampling is fitted to that specific model by optimizing a certain criterion.

Without going into details, we only mention the traditional A- and D-optimality criteria [Khuri and Cornell, 1996], both attempting to reduce the uncertainty of the estimated model parameters (e.g., the coefficients of linear regression) due to noisy observations. In computer experiments, these criteria are rarely used, since no random error occurs and in case of more sophisticated models (like kriging[1]), no simple relation exists between the uncertainty of model parameters and the confidence of the resulted model.

A sampling scheme for kriging models, aiming at reducing the averaged uncertainty of the kriging prediction, is proposed in [Sacks, Schiller, and Welch, 1989]. The sampling problem is transformed into an optimization task which is then solved by a quasi-Newton method. Another kriging-related design method –using simulated annealing for the optimization– is presented in [Brus and Heuvelink, 2007], via a geostatistical application.

Adaptive DoE

Adaptive DoE methods account not only for the model but for the output samples actually observed as well. Thus, most of the adaptive methods are sequential sampling techniques, i.e., the observation is made (an

[1] Kriging –a stochastic technique for function approximation– plays a main role in this Dissertation. Details and references can be found in § B.

experimental run is performed) right after an input sample has been chosen. The choice of the subsequent sample depends on all the previous observations.

The development of such adaptive strategies is one of the main research domain of DoE. A very up-to-date overview of the state-of-art, with more than 100 references, is given in [Viana, Gogu, and Haftka, 2010]. Another thorough review can be found in [Jin, Chen, and Sudjianto, 2002], focusing on kriging-based surrogate models and presenting two sampling methods. The first one aims to achieve a maximin sample pattern in the input domain (i.e., maximizing the minimal pairwise distance between the input samples), using a weighted distance, based on the observations. The second one is designed to reduce the uncertainty of the kriging prediction. Both methods take into account the observations made during the sampling procedure. More recent contributions are, e.g., [Kleijnen and Beers, 2004; van Beers and Kleijnen, 2008] in which a step-by-step uncertainty reduction of kriging interpolators is dealt with. They use jackknifing and bootstrapping, respectively –both being well-known statistical methods for variance estimation.

A sequential sampling method is proposed in [Bettinger et al., 2008], to improve the inverse interpolation (i.e., interpolation of the inverse of the I/O function to solve the inverse problem) capability of the resulted database by spreading out the samples in the output space, i.e., a sort of output space-filling is used. The thesis [Picheny, 2009] is devoted to the error and uncertainty reduction of surrogate models.

An interesting goal occurs in [Picheny et al., 2010]: a specified "target region" in the output space is aimed to be correctly described by the surrogate model, the precision at other regions is of less importance.

A recent manner of adaptive sampling is based on meshing methods (already mentioned in § 1.1.2). Some basic concepts are discussed in [Pávó and Gyimóthy, 2007], and a more advanced and recent mesh-based sampling method is addressed in [Gyimóthy, Kiss, and Pávó, 2009]. By appropriate mesh refinement, smoothing and deformation, the resulted sample-set can fully be adapted to the forward model and to the applied interpolator. The authors of [Franceschini, Lambert, and Lesselier, 2009] also deal with mesh-databases used in ECT. In [Gyimóthy et al., 2009], the sensitivity information on the modeled I/O function is taken into account in a mesh-based sampling method.

DoE and SM for optimization: response surface methods

The above discussed approaches aim at generic surrogate models, being good approximations all over the input space. However, DoE is often used to build surrogate models for the solution of a specific optimization problem. In this context, the surrogate-based approximation of the objective function (to be optimized) is called *response surface*. In the following, some contributions from the domain of such response surface methods are mentioned.

An exhaustive overview of surrogate-based optimization approaches is given in [Queipo et al., 2005], illustrated by a practical example drawn from aerospace industry. A more recent review paper is [Barton, 2009]. An electromagnetics-related presentation of the recent research activity on the field can be found in [Sykulski, 2007].

Many SM approaches use kriging as interpolator. Kriging is a stochastic technique for function approximation. The main features are summarized in § B, only some references are given herein. An essential textbook from its original domain, geostatistics, is [Chilès and Delfiner, 1999]. By now, kriging is widely used in other fields as well: the textbook [Stein, 1999] presents kriging in a general framework; a recent review paper is, e.g., [Kleijnen, 2009], respectively. A thorough presentation with technical aspects is given in [Martin and Simpson, 2005]. Some applications of kriging in response surface optimization are presented in, e.g., [Lebensztajn et al., 2004] (magnetostatic design problem) and [Dalla'Rosa, Raizer, and Pichon, 2008] (optimal localization of antennas).

A taxonomy of response surface-based global optimization tools is provided in [Jones, 2001]. The simplest approaches –according to Jones– fit a polynomial (often quadratic) surface to the samples (e.g., in a least-square sense), then find the extremum of the approximating surface (see, e.g., [Canova, Gruosso, and Repetto, 2003], presenting an electromagnetic design problem). More sophisticated methods use interpolation, e.g., by splines or kriging, then the extremum of the interpolating surface is to be found. An even more sophisticated technique is to take into account the estimated uncertainty of the kriging response surface, then the convergence to the global extremum is guaranteed under some mild assumptions on the objective function (this is the so-called Efficient Global Optimization (EGO) algorithm, discussed below). Some one-stage kriging-based optimization approaches are also mentioned in [Jones, 2001].

The EGO algorithm (e.g., an early appearance in [Mockus, Tiesis, and Zilinskas, 1978], a solid presentation in [Jones, Schonlau, and Welch, 1998] or a recent work [Villemonteix, Vazquez, and Walter, 2008]) is designed for the global optimization of expensive-to-evaluate functions. In this Dissertation, the use of EGO for the solution of eddy-current testing inverse problems is presented (§ 3). Some other engineering applications of EGO can be found, e.g., in [Morgans et al., 2008] (the shape of a horn-loaded loudspeaker is optimized), in [Gao and Wang, 2009], dealing with the optimization of an injection-molding process or in [Couckuyt et al., 2010], presenting other global optimization methods but focusing on EGO (illustrations are drawn from microwave filter and antenna design problems). The original EGO algorithm is extended in [Knowles, 2006] to multiobjective optimization (via a Pareto-optimality criterion, yielding the "ParEGO" algorithm).

When the gradient of the objective function with respect to the input parameters is easily available, the use of derivative-based surrogate models might be considered (e.g., [Leary, Bhaskar, and Keane, 2004], based on the method in [Morris, Mitchell, and Ylvisaker, 1993]).

1.3 Research objectives and organization of the dissertation

As a conclusion of the above overview of NDE inversion techniques and DoE/SM approaches, we can state that the research activity in both domains is still intensive. Many attempts are being made for the solution of inverse problems of NDE, however, the high computational cost of the applied numerical simulator is often a serious bottleneck. On the other hand, SMs –built by an appropriate DoE method– have already shown good performance in coping with such expensive-to-evaluate simulations in a wide range of engineering problems. Thus, we feel entitled to expect success from the application of DoE/SM tools in electromagnetic nondestructive evaluation as well.

The general objective of the research presented in this Dissertation is to improve the performance of ENDE methods by involving DoE and SM approaches. To this end, both existing DoE/SM methods are applied to ENDE and new DoE/SM tools are developed as well. Although the methods are presented in the framework of ENDE and illustrated by a specific ECT setup, we believe that a more general use is also possible in other cases (see, e.g., § E).

The terms and notations used in this thesis are summarized, and the illustrative ECT example (referred throughout the Dissertation) is presented in § 2. Thereafter, our research activity is discussed in the following order.

Optimization-based inversion

The inverse problem of NDE is addressed via an optimization-based inversion method in § 3. By connecting a numerical ECT simulator (based on an integral equation scheme) to the EGO algorithm, a new effective tool of inversion follows, providing a way to involve the discretization applied in the numerical simulator

and the user's precision requirements as well. The EGO –already concerned in the bibliographic overview (§ 1.2.2)– is an up-to-date surrogate-based DoE tool, which has not been applied to ECT inverse problems so far, according to our best knowledge.

Surrogate modeling by functional kriging

§ 4 is devoted to a surrogate modeling technique using functional kriging. The latter is a recent development of mathematics, making possible the prediction of a whole function as a single entity (in contrast with the classical kriging approach of pointwise prediction of a function). Besides the use of the functional kriging model for NDE problems, we have developed an adaptive sampling strategy to build such surrogate models in a certain sense optimal way (based on the estimated uncertainty of the kriging prediction).

Output space-filling databases and inverse mappings

We have developed an adaptive sampling strategy, aiming at the generation of a special set of corresponding input-output samples (of the studied forward problem), i.e., a special database. By using this DoE method, the output samples in the resulted database have a more or less uniform distribution in the output space (domain of the response signals). In other words, a *space-filling* in the output space is realized. In § 5, this new DoE method is discussed, and the properties of the resulting databases are examined. The structure of such databases might provide some meta-information on the related inverse problem which is also in the focus. Two techniques of inverse mapping are also presented in § 5, aiming at a quantitative characterization of the inverse problem.

These three main chapters are self-consistent (only the canonical examples of the numerical illustrations are defined in a separate chapter), thus, they are basically independent from each other, though a few cross-references can also be found. The main new scientific achievements of each chapter are summarized as "thesis points" in the final chapter, § 6, along with the references to the papers in which we have published the results. Finally, the Appendix considers some issues (either being classical textbook knowledge, or not closely related to the main message) referred in the Dissertation.

2

The illustrative eddy-current testing example

In the following chapters, the same eddy-current testing example will be dealt with. This chapter is devoted to this canonical example. First, the physical configuration is discussed. Certain details of the numerical field computation –being slightly out of our scope– are ignored here and presented in § A. However, the applied discretization in the numerical simulation must be taken into account in the DoE/SM approaches, thus, this is also discussed in this chapter. Then, the basic terms and notations –mainly from the vocabulary of design-of-experiments and surrogate modeling– are defined and links to the context of eddy-current testing are established. Finally, four specific cases of the ECT example are defined, involving 2, 4, 6 and 8 defect parameters.

2.1 Eddy-current testing configuration

We consider a classical ECT setup which has commonly been addressed in the literature –as a typical benchmark configuration– because of its relative simplicity.

A non-ferromagnetic, homogeneous metal plate is assumed to be affected by one single or a pair of infinitesimally thin cracks. The plate is of a finite thickness d, but it is assumed to be infinite in its other dimensions[1]. The cracks are assumed to be rectangular-shaped, planar cracks, being perpendicular to the plate surface. In case of two cracks, their planes are parallel. Eddy-currents are generated in the metal plate by a pancake-type air-cored coil, driven by time-harmonic current. The numerical description of the configuration is given in Table 2.1.

A sketch of the configurations with one and two cracks is shown in Figs. 2.1a and 2.1b, respectively. The cracks open to the plate surface opposite with the coil, i.e., they are OD-type (“outer defect”) flaws in the terms of NDT.

The single-crack configuration provides 4 parameters to describe the defect: A and B are the coordinates of the crack center in the xy plane, L is the length and D is the depth of the crack, respectively. In the double-crack configuration, 8 parameters are available: A is the coordinate of centre of the double-crack along the axis x, h is the distance between the cracks, B_1 and B_2 are the center coordinates along y, whereas the lengths L_1, L_2 and the depths D_1, D_2 describe the crack sizes, respectively. Let us note that the crack depths are usually given in percentages of the plate thickness.

[1]Obviously, in the reality, neither is the plate infinite, nor is the crack infinitesimally thin. However, both approximations are valid if the coil is farther from the plate edges than several skin-depths (also called penetration depth, a characteristic distance related to the diffusion of the EM field in the specimen, determined by the frequency and the constitutive parameters), and the aperture of the crack is considerably smaller than its length, depth and the skin-depth.

Table 2.1: Parameters of the studied ECT configuration. Both the plate and the coil are the same as in the JSAEM-benchmarks [Takagi, Uesaka, and Miya, 1997].

Coil	Inner radius (r_1)	0.6 mm
	Outer radius (r_2)	1.6 mm
	Height (a)	0.8 mm
	Distance to the plate (lift-off) (b)	0.5 mm
	Number of turns	140
	Frequency (f)	150 kHz
Plate	Thickness (d)	1.25 mm
	Conductivity (σ)	10^6 S/m
	Relative permeability (μ_r)	1
	Skin-depth (δ) at 150 kHz	1.30 mm

During the evaluation, the coil is scanning above the damaged zone, and the variation of its impedance is measured. As a result of the interaction between the electromagnetic field and the cracks, the impedance of the coil changes compared to its value in case of flawless plate. In both configurations, the scanned surface is the same rectangular region. An impedance measurement is performed at each node of the following grid (i.e., the center of the coil fits to the following coordinates):

$$\begin{aligned} x_c &= k \times 0.5\,\text{mm}, \quad k = 0, \pm 1, \ldots, \pm 4, \\ y_c &= l \times 0.5\,\text{mm}, \quad l = 0, \pm 1, \ldots, \pm 16. \end{aligned} \tag{2.1}$$

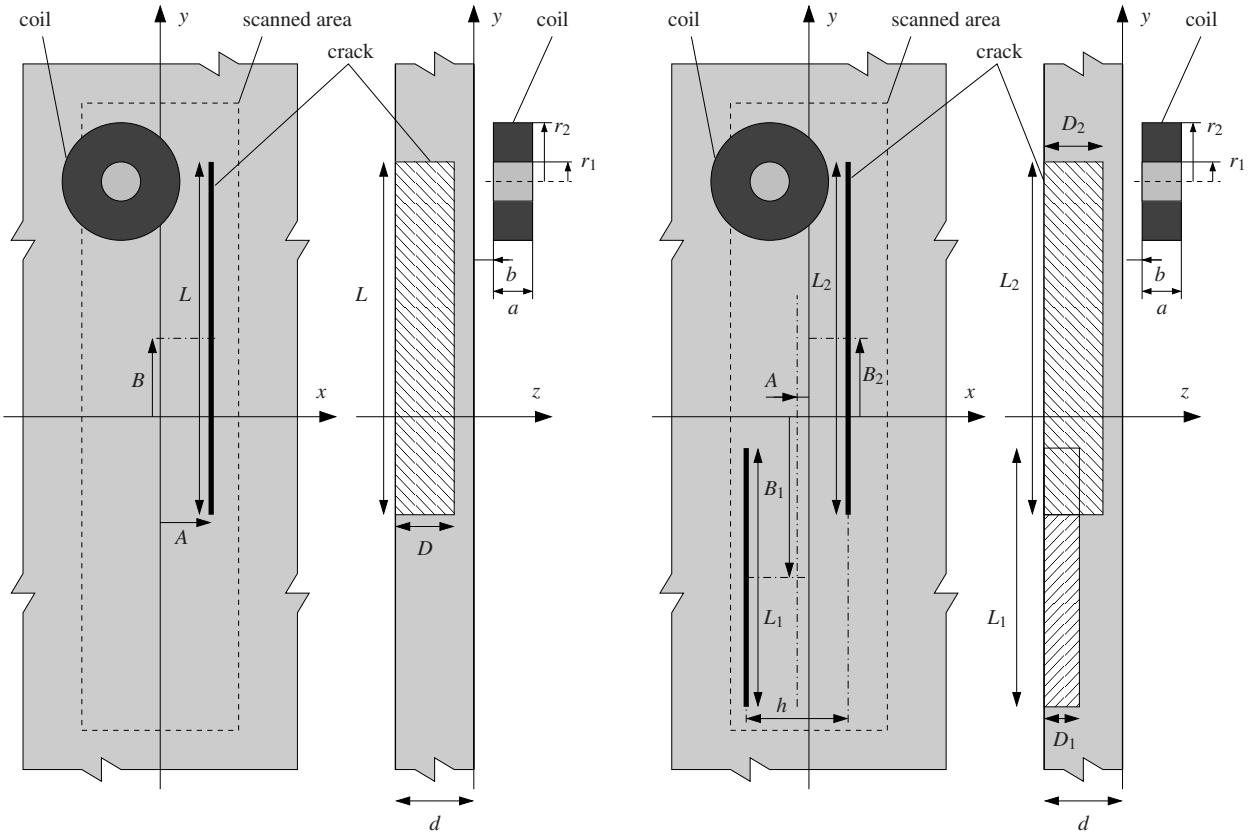
Consequently, the coil impedance is sampled at $9 \times 33 = 297$ positions over a 4 mm $\times$ 16 mm region.

2.2 Numerical simulation and effect of the discretization

A more detailed discussion of the mathematical model and the numerical simulation of the involved electromagnetic phenomena is found in § A. Herein we only mention that a surface integral method is used to describe the interaction of the crack with the electromagnetic field. This integral equation (or a coupled system of two integral equations) is defined over the surface of the crack(s). The unknown in the equation(s) is the current-dipole density –being a "secondary source"– which represents the crack in the model. In the numerical simulation, the surface of the crack(s) is discretized, and the current-dipole density is assumed to vary piecewise linearly (bilinear dependence on the y and z coordinates). By using the Method of Moments, the original integral equation is reduced to to a system of linear equations. The coefficients in the system matrix are related to the self- and mutual interaction between the electromagnetic fields produced by the current-dipole densities over the surface elements of the crack (i.e., they are related to the Green's function). One can considerably speed-up the simulation if these coefficients –being independent from the position and the total size of the crack– are pre-computed and stored in a database. Our simulator at hand actually uses this technique, however, the drawback of doing so is that the shape and the size of the surface elements are fixed, obviously. The surface elements are rectangles, the edge length along y is 0.125 mm, and 5 % of d along z, respectively.

Similar pre-computation is carried out for the electromagnetic field produced by the coil in flawless plate. Taking into account the chosen coil positions (see the previous section) and the discretization of the crack surface, as a consequence, the following constraints are specified to the position and the sizes of the crack(s):

- the planes are in the $x = k \times 0.125$ mm (k integer) planes,
- the center along the y axis has to be $y = l \times 0.125$ mm (l integer),



(a) Single crack (4-parameters).

(b) Double crack (8-parameters).

Figure 2.1: The studied canonical problems with one single and a pair of cracks, respectively.

- the length is equal to $m \times 0.25$ mm (m positive integer),
- the depth is equal to $n \times 5$ % of d (n positive integer).

These conditions are considered in the implementation of the approaches presented later on in this Dissertation. However, the usage of the developed DoE/SM tools is not restricted to such discrete-input simulators: as it will be pointed out, a continuous-input simulator could also be used.

2.3 Terms and notations

Herein we formalize certain basic concepts (some of them have already been mentioned in the previous sections) via the ECT example discussed above.

The input of the model is a set of N input parameters, collected into the vector $\mathbf{p}$. For instance, in the single-crack example, if all parameters are enabled to vary, the input vector is a four-dimensional vector, which could naturally be

$$\mathbf{p} = [A, B, L, D]^{\mathsf{T}}. \tag{2.2}$$

However, for the sake of simplicity (e.g., to avoid problems arising from the different units of the elements of $\mathbf{p}$), the scaled input vectors are used in the followings. The scaling is based on the lower and upper bounds of each parameter (also defined for the studied problem):

$$J_{\min} \leq J \leq J_{\max} \tag{2.3}$$

holds where J stands for an arbitrary parameter. In the four-parameter example, the scaled input vector is then:

$$\mathbf{p} = \left[\frac{A - A_{\min}}{A_{\max} - A_{\min}}, \frac{B - B_{\min}}{B_{\max} - B_{\min}}, \frac{L - L_{\min}}{L_{\max} - L_{\min}}, \frac{D - D_{\min}}{D_{\max} - D_{\min}} \right]^{\mathsf{T}}. \tag{2.4}$$

In the following, the term "input vector" will refer to a such scaled vector, i.e.,

$$\mathbf{p} \in [0, 1]^N \tag{2.5}$$

always holds. In the examples presented in this Dissertation, each parameter can independently vary from the others. However, there is no theoretical difficulty in defining the bounds as functions of the other input parameters.

The set of all feasible input vectors is called the *input space*[2] $\mathbb{P}$. The simulator which we actually use can only handle discrete values of the input parameters (§ 2.2), consequently, the input space is a finite set of input vectors. Let us define the discretization of the ith input parameter as

$$p_i \in \left\{ k_i \Delta p_i \;\middle|\; \forall k_i \in \{0,\ 1,\ \ldots,\ K_i\} \right\}, \tag{2.6}$$

where K_i is a positive integer and Δp_i is the step-size along the ith parameter:

$$\Delta p_i = \frac{1}{K_i}. \tag{2.7}$$

Let $\mathbf{e}_i$ be the ith unit-step vector:

$$\mathbf{e}_i = \left[\overset{(1)}{0},\ \overset{(2)}{0},\ \ldots,\ \overset{(i-1)}{0},\ \overset{(i)}{\Delta p_i},\ \overset{(i+1)}{0},\ \ldots,\ \overset{(N)}{0} \right], \quad \forall i = \{1,\ 2,\ \ldots,\ N\}. \tag{2.8}$$

The input space $\mathbb{P}$ is then formally defined as

$$\mathbb{P} = \left\{ \sum_{i=1}^{N} k_i \mathbf{e}_i \;\middle|\; \forall [k_1, k_2, \ldots, k_N] \in \mathbb{K} \right\} \tag{2.9}$$

with the index set $\mathbb{K}$

$$\mathbb{K} = \{0,\ 1,\ \ldots,\ K_1\} \times \{0,\ 1,\ \ldots,\ K_2\} \times \cdots \times \{0,\ 1,\ \ldots,\ K_N\}. \tag{2.10}$$

The output of the model is a function $q(t)$ which is –in our examples– equal to the ΔZ variation of the complex impedance of the probe coil at the position t. The notation q is used for the sake of generality, i.e., to emphasize that other models with other kind of output can also be considered. According to (2.1), the position t is within the rectangular region T:

$$t \in T = [-2\,\mathrm{mm}, 2\,\mathrm{mm}] \times [-8\,\mathrm{mm}, 8\,\mathrm{mm}] \tag{2.11}$$

in the xy plane. However, the impedance measurement is performed only at $M = 297$ discrete locations, i.e., only samples of $q(t)$ at certain $t = t_1, t_2, \ldots, t_M$ positions are observed.

The relation between the input vector and the corresponding output function is formalized via the *forward operator* $\mathcal{F}$:

$$q(t) = \mathcal{F}\{\mathbf{p}\}, \quad \forall \mathbf{p} \in \mathbb{P}. \tag{2.12}$$

[2]Let us note that herein we do not assign any rigorous mathematical definition to the term "space". In some works (e.g., [Tarantola, 2005]), the term "manifold" also appears as a hint for a more formal approach. Considering that we have a finite number of elements in the "space", we face a more challenging situation. In spite of all that, we believe that the following discussions in this Dissertation are understandable even without recoursing to the formal tools.

$\mathcal{F}$ is a one-valued operator. In our discrete-input examples, the only assumption on $\mathcal{F}$ is its boundedness. However, in the case of continuous-input models, the continuity and smoothness of $\mathcal{F}$ is usually also provided by the underlying physical phenomena. Note that the theoretical forward operator corresponding to a certain model and the one which is realized by a numerical simulation of the model are different. These issues (e.g., estimation of the modeling and simulation errors) are not considered hereafter, the symbol $\mathcal{F}$ stands for the realization of the forward operator by means of the numerical simulator at hand.

The *output space*[3] $\mathbb{Q}$ is defined as the codomain of $\mathcal{F}$ over $\mathbb{P}$:

$$\mathbb{Q} = \left\{ q(t) = \mathcal{F}\{\mathbf{p}\} \;\middle|\; \forall \mathbf{p} \in \mathbb{P} \right\}. \tag{2.13}$$

$\mathbb{Q}$ is a subset of all square integrable functions on the domain T:

$$\mathbb{Q} \subset L^2(T). \tag{2.14}$$

An important assumption has to be made on the adequacy of the discretization of the input space (formula (2.9)) with respect to the forward operator $\mathcal{F}$. Without giving a rigorous definition, we assume that the output space $\mathbb{Q}$ gives a "good description" of the studied problem, i.e., if one could model defects that are not represented in $\mathbb{P}$, the resulted output signal would be "similar" to an appropriate element of $\mathbb{Q}$ (see also the related part of § 3.4.2, discussing a practical interpretation of these issues). This can obviously be provided by choosing the discretization –being fine enough– with respect to $\mathcal{F}$. In our examples, this assumption holds[4]. However, this might become counterproductive when extending the approach to other applications.

Both the input and output spaces are equipped with norms, thus, the distance is defined on them. The Euclidean vector norm is used in $\mathbb{P}$:

$$\|\mathbf{p}_a - \mathbf{p}_b\| = \sqrt{\sum_{i=1}^{N} (p_{a,i} - p_{b,i})^2}, \tag{2.15}$$

where $p_{a,i}$ and $p_{b,i}$ are the ith element of the vector $\mathbf{p}_a$ and $\mathbf{p}_b$, respectively. In the output space, the root mean square norm is used:

$$\|q_a(t) - q_b(t)\| = \sqrt{\frac{1}{T} \int_T |q_a(t) - q_b(t)|^2 \mathrm{d}t}, \tag{2.16}$$

where the factor T^{-1} ensures that $q(t)$ and $\|q(t)\|$ are of the same dimension (impedance in our examples). We use the same symbol to denote the vector and function norms, however, the argument always tells which one is referred. Since $q(t)$ is observed only at $M = 297$ discrete points, the functional norm is approximately evaluated:

$$\|q_a(t) - q_b(t)\| \approx \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left| q_a(t_i) - q_b(t_i) \right|^2}. \tag{2.17}$$

Let us note that more sophisticated integral schemes could also be used, e.g., by using a piecewise second-order polynomial interpolation (Simpson-rule), or by reproducing the function $q(t)$ in terms of linear combination of basis functions (the Radial Basis Function method). However, the involved output signals appear to be smooth enough (see § A.3) to let us use a such simple approximation of the integral.

This formal description of the model of the nondestructive evaluation setup as an operator between the input and output spaces makes possible the easy generalization of the approaches presented later on: a wide range of problems can be addressed in a similar way.

[3] See the footnote 2 on the usage of the term "space".

[4] One can obviously simulate cracks of arbitrary sizes even by using the discretization scheme presented in § A. Furthermore, simulations by other schemes (e.g., [Pávó and Lesselier, 2006]) –not based on the spatial discretization of the crack– also prove that the considered forward problem is "smooth enough" to be appropriately described by such finite input and output spaces.

Table 2.2: Numerical description of the single-crack examples (1 and 2): Bounds and discretization of the input parameters (in the original, unscaled domain), and the cardinality $|\mathbb{P}|$ of the input space, respectively.

	Min	Max	Unit-step (Δp)	No. of grid nodes (K)
L (mm)	1	10	0.5	19
D (%)	5	90	5	18
				$\lvert\mathbb{P}\rvert = 342$

(a) Example 1.

	Min	Max	Unit-step (Δp)	No. of grid nodes (K)
A (mm)	−2	2	0.5	9
B (mm)	−3	3	0.5	13
L (mm)	1	10	0.5	19
D (%)	5	90	5	18
				$\lvert\mathbb{P}\rvert = 40{,}014$

(b) Example 2.

2.4 The canonical examples

Four canonical examples (based on the two configurations presented in § 2.1) are defined for the illustration of the methods discussed later on. The description of the examples is as follows, numerical details are given in Tables 2.2 and 2.3.

1. **The 2-parameter example:** The single-crack configuration is considered (Fig. 2.1a). The position of the crack is assumed to be known ($A = B = 0$), thus only the length (L) and depth (D) are allowed to vary (Table 2.3a).
2. **The 4-parameter example:** All the four parameters involved in the single-crack crack configuration (Fig. 2.1a) are allowed to vary (Table 2.3b).
3. **The 6-parameter example:** The double-crack configuration is considered (Fig. 2.1b). All parameters except the depths –being assumed $D_1 = D_2 = 40\,\%$– are allowed to vary (Table 2.4a).
4. **The 8-parameter example:** All the eight parameters involved in the double-crack configuration (Fig. 2.1b) are allowed to vary (Table 2.4b).

Let us note that the discretization of the parameters respects the constraints arisen by the numerical simulator at hand (§ 2.2). Due to these constraints, a finer discretization cannot be defined for the depth, whereas this could be done for the other parameters.

Obviously, the configurations chosen could be parametrized in other ways as well. In the double-crack configuration, the presented parametrization provides the advantage of ensuring the unicity of the input parameter vector corresponding to a given configuration (i.e., by alternating the location of two cracks of equal size, the input parameter vector remains the same). Moreover, the non-realistic configurations (e.g., overlapping cracks) are ignored. A disadvantage, however, is that cracks are not allowed to be in the same plane even if they would not overlap.

The cardinality of the input spaces (i.e., the number of the feasible input vectors) is also given in Tables 2.2 and 2.3. This cardinality is growing rapidly with the N number of dimensions, as it is shown in Fig. 2.2. This phenomenon and its disadvantageous consequences in certain applications are denoted as the "curse of dimensionality".

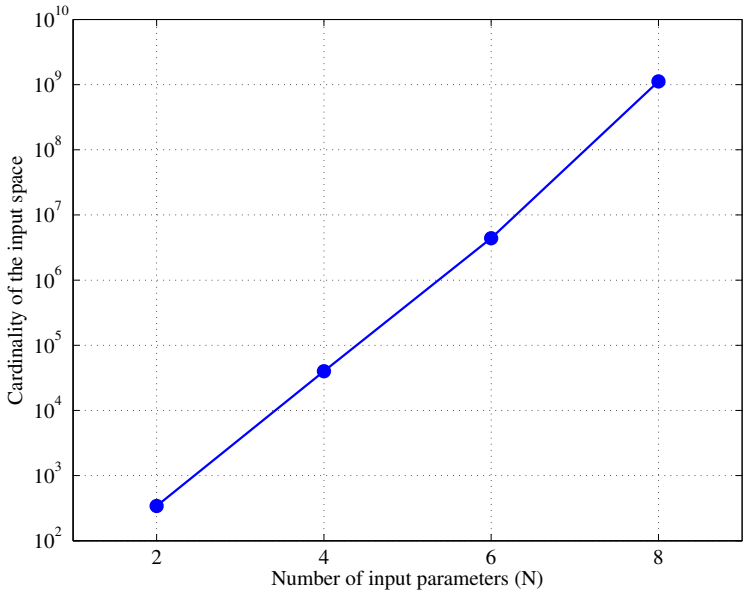


Figure 2.2: The cardinality of the input spaces appears to grow exponentially with the number of input parameters.

Table 2.3: Numerical description of the double-crack examples (3 and 4): Bounds and discretization of the input parameters (in the original, unscaled domain), and the cardinality $|\mathbb{P}|$ of the input space, respectively.

	Min	Max	Unit-step (Δp)	No. of grid nodes (K)
A (mm)	-1	1	0.25	9
h (mm)	0.25	2	0.25	8
B_1, B_2 (mm)	-3	3	0.5	13
L_1, L_2 (mm)	1	10	0.5	19
				$\lvert\mathbb{P}\rvert \approx 4.39 \times 10^6$

(a) Example 3.

	Min	Max	Unit-step (Δp)	No. of grid nodes (K)
A (mm)	-1	1	0.25	9
h (mm)	0.25	2	0.25	8
B_1, B_2 (mm)	-3	3	0.5	13
L_1, L_2 (mm)	1	10	0.5	19
D_1, D_2 (%)	5	80	5	16
				$\lvert\mathbb{P}\rvert \approx 1.12 \times 10^9$

(b) Example 4.

3

Inversion by the “Efficient Global Optimization” algorithm

In this chapter, an inversion scheme is presented and illustrated by ECT examples. Parametric defect models are chosen and the inversion is traced back to the optimization of an objective function expressing the discrepancy between the measured and simulated output signals. Since the objective function is expensive-to-evaluate, the EGO algorithm –being based on the kriging model of the objective function– is used for the optimization. According to the numerical studies (with 2, 4, 6 and 8 input parameters), the presented approach is indeed effective, however, a thorough comparison to other inversion techniques has not been carried out yet.

The research work presented in this chapter forms the basis of the thesis point 1 (in § 6). We have published our results related to the presented approach in the following papers: [Bilicz et al., 2008; Bilicz et al., 2009a; Bilicz et al., 2009b; Bilicz et al., 2009d].

3.1 Inverse problem and regularization

The inverse problem of NDE consists in the characterization of the defect based on some observed output data. The first step of a possible inversion approach is to choose *a priori* a parametric defect model with N scalar parameters, along with their lower and upper bounds. The central assumption of the method is that this defect to be characterized can be sufficiently described by the parametric model chosen. The latter imposes a restriction on the solution of the inverse problem, i.e., the parametric model can be considered as a regularization tool.

The solution of the regularized inverse problem is then the defect (from the parametric class) which yields the “most similar” output data (via the forward model) to the actually measured data. In other words, the best similarity between the measured and simulated output data is sought, as this is believed to imply the best similarity between the true defect and the one we choose from the parametrized class.

For a formal description –partially recalling some terms and notations from § 2.3 – let us denote the vector of N parameters by the input vector $\mathbf{p}$:

$$\mathbf{p} = [p_1,\, p_2,\, \ldots,\, p_N]^{\mathsf{T}}, \tag{3.1}$$

and the set of all feasible input parameters –the input space– by $\mathbb{P}$. The functional output data $q(t)$ is related to the input vector via the forward operator $\mathcal{F}$:

$$q(t) = \mathcal{F}\{\mathbf{p}\},\ \forall \mathbf{p} \in \mathbb{P}. \tag{3.2}$$

In practice, $\mathcal{F}$ is realized by a numerical simulator of the studied NDE configuration. The measured output signal is denoted by $\tilde{q}(t)$. The discrepancy between the simulated and measured output signal is formally expressed by the objective function $Q(\mathbf{p})$:

$$Q(\mathbf{p}) = \frac{\|\tilde{q}(t) - \mathcal{F}\{\mathbf{p}\}\|}{\|\tilde{q}(t)\|}. \tag{3.3}$$

Any valid function norm can be used, however, the behaviour of the objective function (also the location of its minima) depends on the choice of the norm. In our examples, the classical norm (2.17) is used. By normalizing with $\|\tilde{q}(t)\|$, a dimension-free objective function is obtained; moreover, the performance of the inversion algorithm on different examples can easily be compared via the values of $Q(\mathbf{p})$. The general scheme of such optimization-based inversion methods is shown in Fig. 3.1.

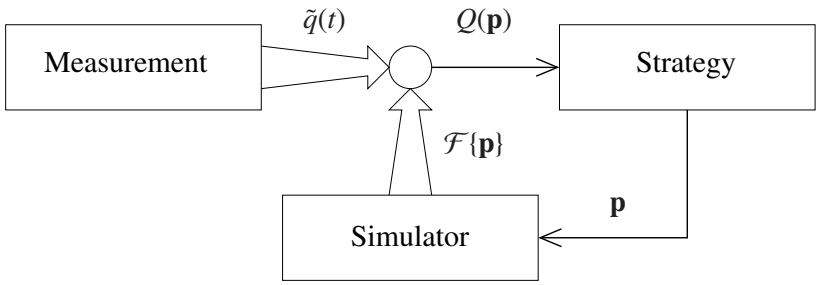


Figure 3.1: General scheme of optimization-based inversion by using a parametric model.

Via the objective function $Q(\mathbf{p})$, the inverse problem is then spelled as the following optimization task:

$$\mathbf{p}^\star = \arg\min_{\mathbf{p}\in\mathbb{P}} Q(\mathbf{p}). \tag{3.4}$$

As a rule, this minimum-search problem is computationally expensive, since the objective function $Q(\mathbf{p})$ involves the usually expensive-to-evaluate forward operator $\mathcal{F}$. Let us recall that in our examples, $\mathbb{P}$ is a finite set of input vectors, defined by a rectangular grid (§ 2.3). A natural idea would then be to perform a so-called exhaustive search over $\mathbb{P}$, i.e., evaluate $Q(\mathbf{p})$ for each $\mathbf{p} \in \mathbb{P}$. However, for the above reason, this would be numerically intractable. One has to consider a more sophisticated optimization strategy to solve (3.4). The required method has to fulfill two main criteria:

1. Due to the computational expense of $Q(\mathbf{p})$, only a small number of objective function evaluations are allowed in order to keep the runtime tractable.
2. A global minimum-search is needed, since $Q(\mathbf{p})$ might have several local minima[1] in $\mathbb{P}$.

We have chosen the EGO algorithm, which appears to simultaneously fulfill both above requirements. In the next section, EGO is presented in detail.

3.2 Efficient Global Optimization

As already indicated in the Introduction, the EGO algorithm is designed for the optimization of expensive-to-evaluate objective functions, since usually a relatively small number of function evaluations are needed to find the global minimum. The method is widely used in other engineering domains as well. Biographical references have already been given in section § 1.2.2; herein we only recall the solid presentation found in [Jones, Schonlau, and Welch, 1998].

[1] In the "grid-sense", defined for the discrete input spaces in § C.2.

The EGO algorithm is a sequential sampling performed on the objective function, controlled by a criterion based on all previous observations of $Q(\mathbf{p})$. Thus, the method can be considered as an adaptive DoE approach. Furthermore, the heart of EGO is the surrogate modeling of the objective function by kriging. The algorithm is built up as follows.

0. *Measurement:* The $\tilde{q}(t)$ measured output is obtained and passed to the inversion algorithm. We do not consider measurement issues herein.

1. *Initialization:* A set of n initial input vectors

$$\mathbf{p}_1,\ \mathbf{p}_2,\ \ldots,\ \mathbf{p}_n$$

are arbitrarily chosen in $\mathbb{P}$. Practically, a space-filling classical DoE approach (e.g., Latin Hypercube Sampling (LHS)) is used. The corresponding objective function values

$$Q(\mathbf{p}_1),\ Q(\mathbf{p}_2),\ \ldots,\ Q(\mathbf{p}_n)$$

are computed via the forward operator.

2. *Kriging prediction:* The kriging model of the objective function is constructed, based on the observations. A detailed discussion on the kriging prediction can be found in § B, only the key points are outlined herein. Kriging provides an approximation $\hat{Q}(\mathbf{p})$ for the objective function, based on the n observations. The approximation is an interpolation, i.e.,

$$\hat{Q}(\mathbf{p}) \approx Q(\mathbf{p}),\quad \hat{Q}(\mathbf{p}_i) = Q(\mathbf{p}_i),\ \forall i = 1,\ 2,\ \ldots,\ n. \tag{3.5}$$

Beyond the approximate objective function $\hat{Q}(\mathbf{p})$, its uncertainty is also provided via the estimated variance $\hat{\sigma}^2(\mathbf{p})$. The higher this variance, the less our "confidence" in $\hat{Q}(\mathbf{p})$ at the given point $\mathbf{p}$.

3. *Sample insertion:* (Recursive definition.) The next, $(n+1)$th observation of $Q(\mathbf{p})$ should be performed at a certain $\mathbf{p}$, promising the highest "improvement" in the sense of minimum-search, i.e., the lowest value of $Q(\mathbf{p})$. Since the objective function is known only at the already observed n points, one cannot know the location of the global minimizer, obviously. However, based on the kriging prediction $\hat{Q}(\mathbf{p})$ and its estimated variance $\hat{\sigma}^2(\mathbf{p})$, a good guess can be made. Let us define the "expected improvement" (EI) as

$$\eta(\mathbf{p}) = \left(Q_{\min} - \hat{Q}(\mathbf{p})\right)\Phi\left(\frac{Q_{\min} - \hat{Q}(\mathbf{p})}{\hat{\sigma}(\mathbf{p})}\right) + \hat{\sigma}(\mathbf{p})\varphi\left(\frac{Q_{\min} - \hat{Q}(\mathbf{p})}{\hat{\sigma}(\mathbf{p})}\right), \tag{3.6}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ is the normal cumulative distribution and normal probability density function, respectively (both for zero mean and unit variance). The value $Q_{\min}$ is the current best objective function observation:

$$Q_{\min} = \min\left\{Q(\mathbf{p}_1),\ Q(\mathbf{p}_2),\ \ldots,\ Q(\mathbf{p}_n)\right\}. \tag{3.7}$$

The derivation and interpretation of the form (3.6) in the stochastic framework is given in § B.3. The $(n+1)$th observation is then performed at the most promising point (having maximal expected improvement). Formally, the following optimization problem is to be solved:

$$\mathbf{p}_{n+1} = \arg\max_{\mathbf{p}\in\mathbb{P}} \eta(\mathbf{p}). \tag{3.8}$$

This is similar to the original optimization task (3.4), however, the expected improvement $\eta(\mathbf{p})$ can usually be evaluated at a much lower computational cost than $Q(\mathbf{p})$. Thus, despite of the need for performing (3.8) in each iteration, EGO indeed appears to be effective.

The objective function is evaluated at the new point $\mathbf{p}_{n+1}$, the value $Q(\mathbf{p}_{n+1})$ is added to the set of observations and the number of samples is increased: $n \coloneqq n+1$.

4. *Stopping criterion:* The incremental strategy goes on by jumping to step 2, to build an updated prediction, yet involving the new observation, until a stopping criterion is met. This criterion can be related to the n number of observations, to the maximal expected improvement ($\max_{\mathbf{p}\in\mathbb{P}} \eta(\mathbf{p})$), or more complex stopping conditions can also be specified.

This optimization scheme –also presented by a flowchart in Fig. 3.2– is proven to find the global minima of the objective function, under some mild assumptions[2]. A discussion on the convergence of the EGO algorithm would be out of scope of this Dissertation. Some recent results are given, e.g., in [Vazquez and Bect, 2007; Vazquez and Bect, 2010], some other questions are still open. Though no rigorous statements can be made on the speed of the convergence, a relatively few objective function evaluations are usually enough to find the global minimum, according to the literature (e.g., [Jones, Schonlau, and Welch, 1998]) and to our own experience in the numerical examples (§ 3.5), respectively.

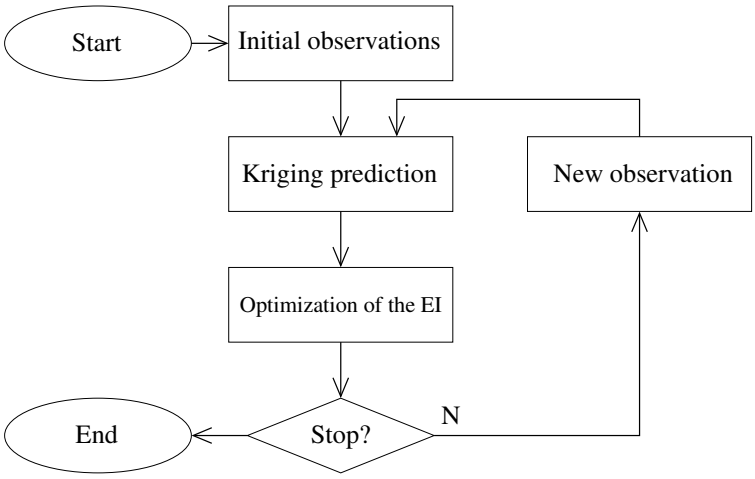


Figure 3.2: Flowchart of the EGO algorithm.

The Expected Improvement (EI) plays a central role in the EGO algorithm, thus, the latter is sometimes referred as "EI-algorithm". A heuristic interpretation of the EI function (3.6) is the following. The first term

$$\left(Q_{\min} - \hat{Q}(\mathbf{p})\right) \Phi\left(\frac{Q_{\min} - \hat{Q}(\mathbf{p})}{\hat{\sigma}(\mathbf{p})}\right) \tag{3.9}$$

is high where the prediction $\hat{Q}(\mathbf{p})$ is far below the current best value $Q_{\min}$ (let us note that $\Phi(x) \to 1$ if $x \to \infty$). Thus, this term represents the "promising" regions (i.e., where small objective value is predicted by kriging) of $\mathbb{P}$. The second term

$$\hat{\sigma}(\mathbf{p})\varphi\left(\frac{Q_{\min} - \hat{Q}(\mathbf{p})}{\hat{\sigma}(\mathbf{p})}\right) \tag{3.10}$$

is more related to the uncertainty of the prediction. The second term is high where $\hat{\sigma}(\mathbf{p})$ is high, which is the case in the so-called "unexplored" regions of $\mathbb{P}$ (let us note that $\varphi(0) = 1/\sqrt{2\pi}$). The formula (3.6) as a sum of the these two terms, then indeed balances the local search in promising regions and global exploration of the whole space $\mathbb{P}$.

[2]This makes sense only in continuous search domains. In our case, as a consequence of the finite cardinality of $\mathbb{P}$, the global minimizer is surely found (in the worst case, the algorithm boils down to an exhaustive search).

3.3 Illustrative one-dimensional analytical example

In this section, the performance of the EGO algorithm is illustrated on a simple analytical example. Let the input space be 1-dimensional: $\mathbb{P} \in [0, 1]$; and consider the following objective function[3] defined on $\mathbb{P}$:

$$Q(p) = 579p^4 - 1110p^3 + 684.5p^2 - 141.071p + 15. \tag{3.11}$$

The discrete input space is defined by a grid: $\Delta p = 0.005$, thus, the number of grid nodes is equal to 201. Three initial observations ($p = 0,\ 0.5,\ 1$) are carried out, then, 7 subsequent iterations are performed. The auxiliary optimization problem (3.8) is solved by an exhaustive search on the grid. The evolution is presented in Fig. 3.3 and a numerical description is also given in Table 3.1.

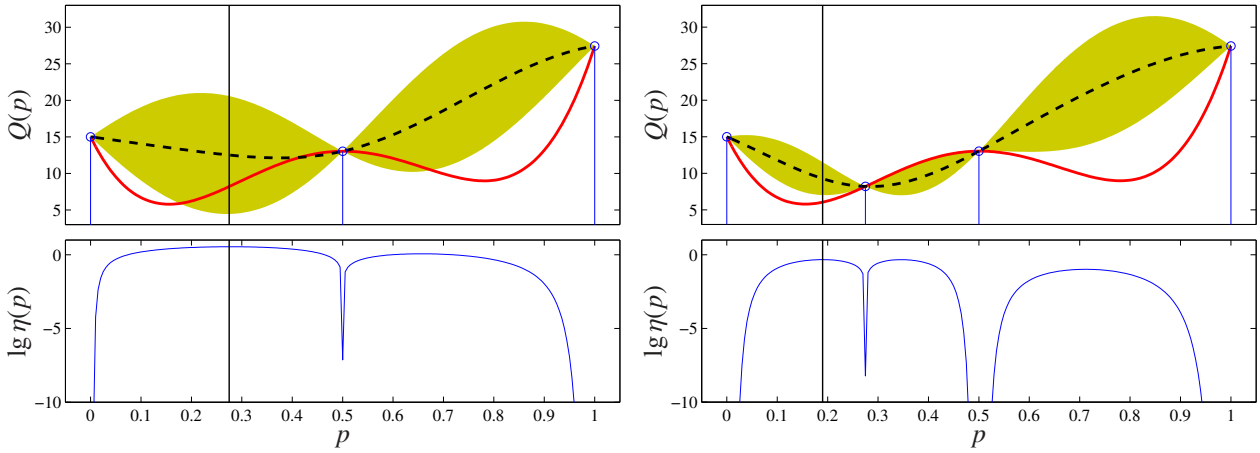
In Fig. 3.3, a marker line shows the current maximizer of the expected improvement, where the next observation is performed in the next iteration. Let us note that $Q(p)$ has two local minima, and both of them are explored by the EGO algorithm. The observations are mainly performed around these local minimizers. During the iterations, the maximal expected improvement tends to decrease, however, this decrease is not monotonous. The estimated variance (uncertainty) of the prediction is also decreasing during the iterations: our confidence increases with the number of observations. The true global minimizer of $Q(p)$ is $p^\star = 0.1555$, the minimum is $Q(p^\star) = 5.7797$. A very good approximate minimizer (the best achievable on the applied grid) is found in the 6th iteration (Table 3.1). However, the fact that the sought minimizer is found does not cause an immediate and significant decrease in the value of the maximal expected improvement (Fig. 3.3g). Generally, the EI starts to significantly decrease when all promising and apparently unexplored (with predictions of high uncertainty) regions of the input space have been sampled; in our example, this seems to happen in the 7th iteration.

The stopping criterion of the EGO algorithm is still an open question. Several heuristic rules (based on the $\max \eta(p)$, $Q_{\min}$, n or on their combination) are proposed in the literature, but unfortunately no general rule has been established yet. In this analytical example, the iteration is simply stopped after 7 iterations.

Table 3.1: Evolution of the EGO algorithm in the analytical example. The expected improvement is computed first only after all initial observations.

Iter.	p	$Q(p)$	$\max \lg \eta(p)$
	0	15.0000	–
	0.500	13.0270	–
	1.000	27.4290	0.5418
1	0.275	8.1976	−0.3345
2	0.190	6.0480	−0.9806
3	0.710	9.7485	−1.1380
4	0.160	5.7847	−2.4289
5	0.150	5.7875	−2.2555
6	0.155	5.7798	−3.1965
7	0.810	9.1739	−7.1344

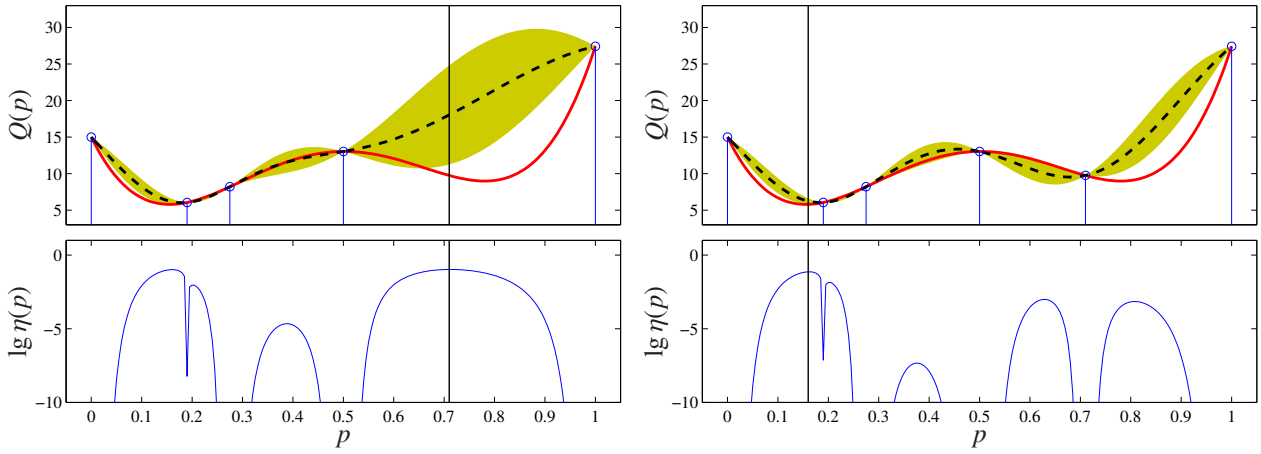
[3] Inspired by the test function used in [Kleijnen and Beers, 2004].



(a) Initialization.

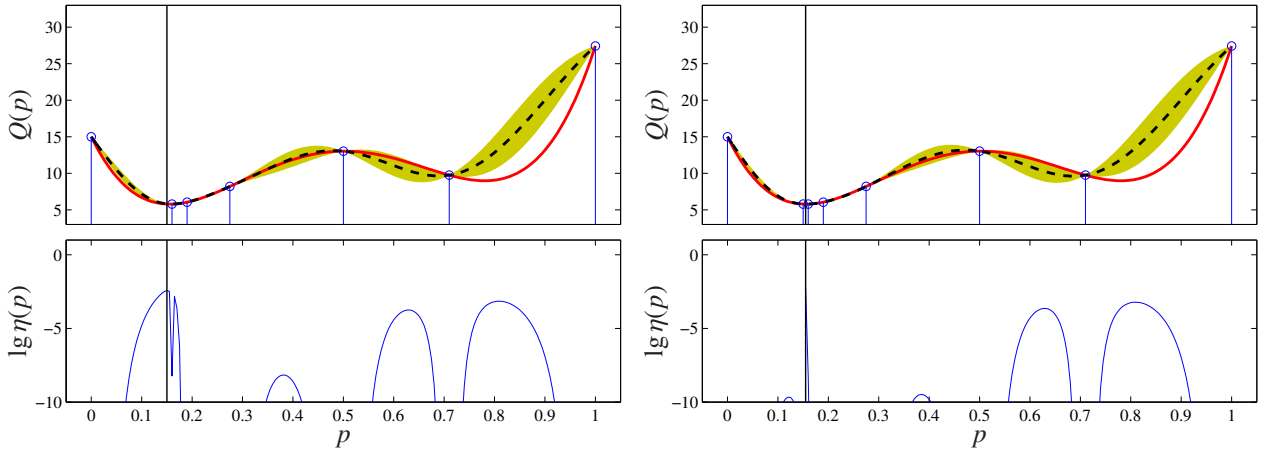
(b) Iteration 1.

Figure 3.3: Performance of the EGO algorithm in the analytical example. Top figures: true objective function $Q(p)$ (red solid line), kriging prediction $\hat{Q}(p)$ (black dashed line), observations (dots) and the estimated standard deviation $\hat{\sigma}(p)$ (represented by the filled area). Bottom figures: logarithm to base 10 of the expected improvement. (To be continued on the next page.)



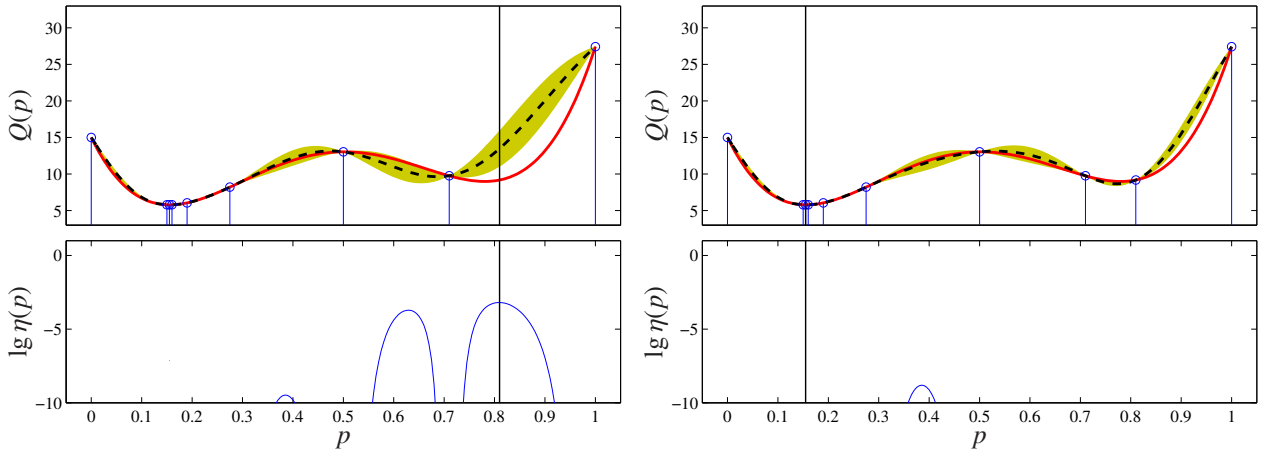
(c) Iteration 2.

(d) Iteration 3.



(e) Iteration 4.

(f) Iteration 5.



(g) Iteration 6.

(h) Iteration 7.

Figure 3.3: (Continued.) Performance of the EGO algorithm in the analytical example. Top figures: true objective function $Q(p)$ (red solid line), kriging prediction $\hat{Q}(p)$ (black dashed line), observations (dots) and the estimated standard deviation $\hat{\sigma}(p)$ (represented by the filled area). Bottom figures: logarithm to base 10 of the expected improvement.

3.4 Considerations on the implementation

3.4.1 Optimization of the Expected Improvement function

Although the EI function $\eta(\mathbf{p})$ is computationally relatively cheap, the solution of the involved auxiliary optimization task (3.8) might be complicated. $\eta(\mathbf{p})$ probably have several local maxima[4] in $\mathbb{P}$, since $\eta(\mathbf{p})$ is always non-negative and exactly zero at the observed points (see Fig. 3.3 for a simple illustration). This leads us to consider sophisticated optimization tools. If the input space was the continuous unit hypercube $\mathbb{U} = [0, 1]^N$, one could use, e.g., the semi-analytic method proposed in [Jones, Schonlau, and Welch, 1998], being a sort of "branch and bound" algorithm. However, any other global optimization technique –such as the wide range of stochastic methods (simulated annealing, swarm or evolutionary approaches, etc.) could be used to evaluate the point insertion criterion

$$\mathbf{p}_{n+1} = \arg\max_{\mathbf{p}\in\mathbb{U}} \eta(\mathbf{p}). \tag{3.12}$$

Let us note again that the expected improvement $\eta(\mathbf{p})$ can be evaluated at a relatively low computational cost, thus, the number of evaluations during the solution of (3.12) is not of main importance.

In our examples, however, the input space $\mathbb{P}$ is a finite set, defined by a rectangular grid (§ 2.3). A natural way to solve (3.8) is then performing an exhaustive search over $\mathbb{P}$. Though doing so, the global maximizer would surely be found, in the case of high number of input parameters and/or fine discretization, the growing cardinality of $\mathbb{P}$ might make the exhaustive search intractable. In this Dissertation, a Coordinatewise Search (CWS) algorithm is used instead. The CWS operates on the rectangular grid which defines $\mathbb{P}$. Since the CWS is a local optimization technique, we use it in a multistart scheme: the search is started subsequently from each observation points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$. A detailed description of the applied CWS variant can be found in the § C.3.

3.4.2 Benefits of the proposed approach

In this Dissertation, the discrete nature of the input space is due to the applied simulator which can only handle discrete values of input parameters. However, in return for this constraint, a considerable acceleration of the numerical EM simulation is provided (§ A).

The CWS algorithm for the solution of (3.8) fits to the grid-defined input space in a natural way. Thus, involving CWS –taking into account the simulator at hand– can be considered as an application-oriented implementation of the EGO algorithm. This scheme indeed appears to be an efficient tool of inversion in NDE, as illustrated by the ECT examples of § 3.5.

A further interpretation of the grid-defined discrete input space can be related to the *required precision* of the solution of the regularized inverse problem. Let us assume a case with a continuous input space $\mathbb{U} = [0, 1]^N$, and let the forward operator be defined on $\mathbb{U}$. The solution of the regularized inverse problem is then

$$\mathbf{p}^\star_{\text{cont}} = \arg\min_{\mathbf{p}\in\mathbb{U}} Q(\mathbf{p}). \tag{3.13}$$

Let us assume that one is able to define a required precision of the solution of the inverse problem: the obtained solution $\mathbf{p}^\star$ is acceptably precise if

$$|p^\star_{\text{cont},i} - p^\star_i| < \Delta p_i \tag{3.14}$$

[4] In the "grid-sense", defined for the discrete input spaces in § C.2.

holds for all components of the input vectors ($i = 1, 2, \ldots, N$). Then one can heuristically believe that a such satisfying approximate solution $\mathbf{p}^\star$ can be found by solving

$$\mathbf{p}^\star = \arg\min_{\mathbf{p}\in\mathbb{P}} Q(\mathbf{p}), \tag{3.15}$$

with the discrete input space $\mathbb{P}$ being a rectangular grid of a unit-step size Δp_i along the ith input parameter, according to (2.9). Tracing back a continuous optimization problem to a discrete one in such a way certainly does not ensure that the condition (3.14) holds (some further hints can be found in § C.1). However, it is believed to hold if the discretization of the input space is fine enough with respect to Q. Consequently, defining the input spaces in terms of a rectangular grid can be viewed as a manner of specifying the required precision of the result of the inversion to a certain extent.

3.5 Numerical examples

3.5.1 Generally about the examples

Five groups of ECT crack-reconstruction examples are presented. The first four cases are directly related to the canonical examples defined in § 2.4: a defect belonging to the given input space is to be reconstructed based on the corresponding output data. The fifth case is a mix of two canonical examples: a defect belonging to a double-crack example is to be characterized under the assumption of a single crack.

Each group of examples includes three different runs –denoted by a), b) and c)– in which different defects are assumed. In these numerical studies, only synthetic output data are used (obtained by the simulator at hand), i.e., no measurement has been performed. Using the same simulator in the inversion procedure and to generate synthetic reference data is called "inverse crime". Generally one must avoid such situations, however, the demonstration and the evaluation of the EGO algorithm's performance is convenient in so doing (as one knows the exact solution of the regularized inverse problem).

The number of initial samples is generally equal to $2N$ (N is the number of input parameters) except the first example with 2 parameters, where 10 initial samples are chosen (as our experience showed that the involved objective functions are quite simple in this case). The initial observations are chosen by a random LHS (using a maximin criterion) over the input space[5]. The linear dependence of the number of initial samples on the number of input parameters is proposed in [Jones, Schonlau, and Welch, 1998]. This choice also appears to be appropriate according to our numerical studies presented herein.

In each iteration, the hyperparameters of the covariance function (involved in the kriging model) are re-estimated by the Restricted Maximum Likelihood (REML). These classical issues are not our own contributions, thus, we put them to the Appendix. For details on the kriging model and the choice of the covariance, see §§ B.1 and B.2.

The stopping criterion is simply an upper bound on the number of iterations, equal to four times the number of initial samples. Almost in all cases, the EGO algorithm found the global minimizer of the objective function.

To present the performance of the EGO algorithm, two characteristic properties are plotted in function of the iteration number. Let us assume that in the kth iteration, the algorithm intends to add the nth observation (thus, $(n - k)$ initial observations were performed). The maximal expected improvement

$$\eta_{\max} = \max_{\mathbf{p}\in\mathbb{P}} \eta(\mathbf{p}) \tag{3.16}$$

[5]Derived from a continuous LHS (§ D.3).

is defined based on the previous kriging model, based on $(n-1)$ observations. On the contrary, the current best observed objective function value is the minimum of all previous observations (according to (3.7)), including the nth one as well. This convention emphasises the fact that the improvement in the observation of Q is predicted by η in a certain sense.

Three further characteristics help us to evaluate the performance of the inversion method. The number of observations needed to find the smallest value of the objective function is $n^\star$ (given as a sum of the number of initial and the subsequent observations). Let us note that in most of our examples, the smallest found value is exactly zero, i.e., the global minimizer has surely been obtained. In some other examples, however, there is no guarantee for having found the global minimizer since the smallest observed value of Q –within the limited number of iterations having been performed– is not zero. The time-consumption of the algorithm is assessed via two essential time values:

- the average time of hyperparameter estimation (of the covariance, by REML), $\bar{t}_{\text{est}}$;
- the average time of the optimization of the EI by the CWS algorithm, $\bar{t}_{\text{opt}}$.

Both of them are computed as the arithmetic mean of the corresponding times of all iterations. The third main component of the total time-consumption of the inversion scheme is obviously the runtime of the numerical EM simulator. However, this time is not considered herein, since this is not related to the performance of the EGO algorithm.

For the computations, a PC with a 64-bit CPU with 4 cores at 2.33 GHz and 16 GB memory has been used. All algorithms are implemented in MATLAB®.

3.5.2 Examples with 2 parameters

These examples are based on the canonical example defined as the point 1 in § 2.4. Let us recall that the length (L) and the depth (D) of a single crack are enabled to vary, all other parameters are assumed to be fixed. The input space consists of 342 feasible input vectors.

The numerical summary of the 2-parameter test cases is found in Table 3.2. Cracks of a wide range of sizes are involved in the runs. In all three runs, the assumed defect has exactly been found, i.e., the $Q(\mathbf{p}) = 0$ value is obtained by the optimization algorithm. Let us recall that the algorithm was started by 10 initial observations, then 40 subsequent iterations have been made. The average time of covariance estimation is approximately twice higher than the CWS optimization of the expected improvement function. However, the total time spent on the EGO-related computations during the 40 iterations is less than 1 min in all presented cases.

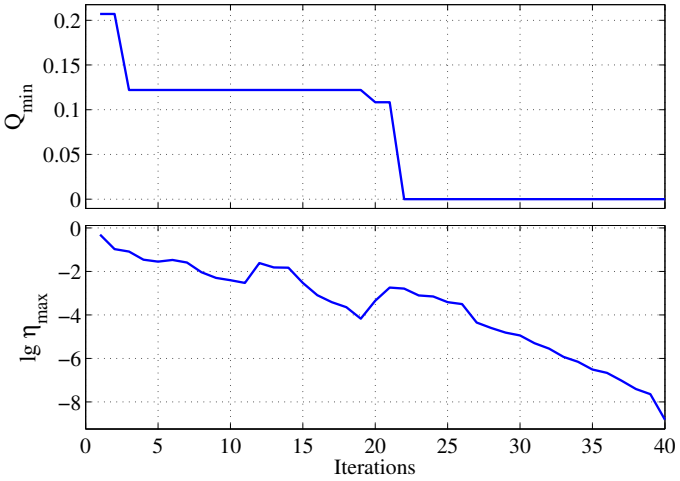
The evolution of the EGO algorithm is plotted in Fig. 3.4. As the number of observations is increasing, $\eta_{\max}$ appears to have a decreasing trend. However, this decrease is neither monotonous, nor of similar order of magnitude in the different cases. Within 40 iterations, $\eta_{\max}$ reaches approximately 10^{-9} in the case 1/a, and 10^{-18} in the case 1/c. Moreover, no simple relation can be revealed between the value of $\eta_{\max}$ and the fact of obtaining the global minimizer. These properties make difficult to set an appropriate stopping criterion which (i) ensures the termination of the iterations shortly after reaching the global minimizer and (ii) prevents premature stopping.

In Fig. 3.5, the performance of the EGO algorithm is illustrated. The sequence of the observations is traceable as the iteration numbers are also plotted. At the same time, the objective function is also displayed. Let us notice that no simple rule seems to appear in the sequence of the observations (e.g., the observations are not clustering around the minimizer). Furthermore, the promising regions of the input space (where Q is small) are well-explored during the search. Let us also notice that these promising regions are quite spacious (espacially the small-crack regions in the cases 1/b and 1/c). This flatness of the objective

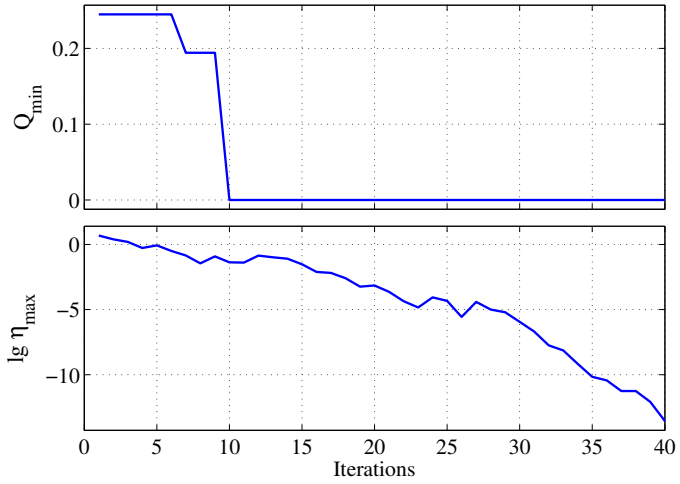
Table 3.2: Numerical description of the 2-parameter examples.

Example		1/a	1/b	1/c
Assumed defect	L (mm)	2.5	2.5	8.5
	D (%)	75	20	20
Performance	$n^\star$	10 + 22	10 + 10	10 + 10
	$\bar{t}_{\text{est}}$ (s)	0.85	0.71	0.80
	$\bar{t}_{\text{opt}}$ (s)	0.36	0.31	0.36

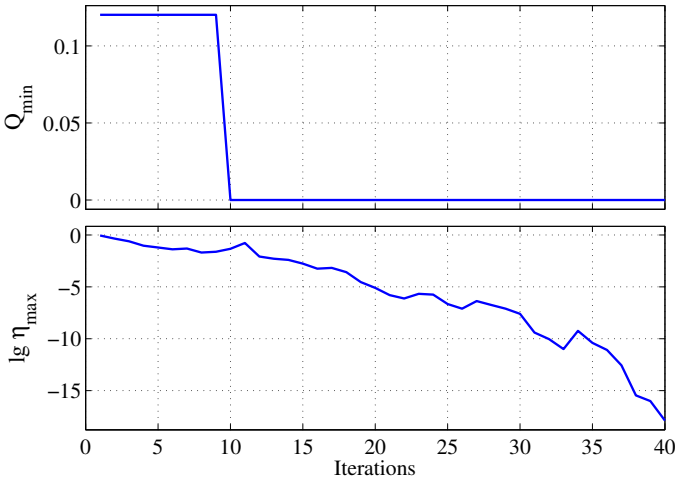
function –the global minimizer is almost invisible in the figures– reflects the ill-posedness of the studied inverse problem, as small changes of the output data result into large variations of the corresponding input parameter values. Consequently, a relatively small noise –which always corrupts the measurements– can cause high uncertainty in the solution of the inverse problem when a small crack is to be characterized. Let us also refer to § 5.5.2, where similar conclusions are drawn by means of other tools being designed for the characterization of the inverse problem.



(a) Case 1/a.

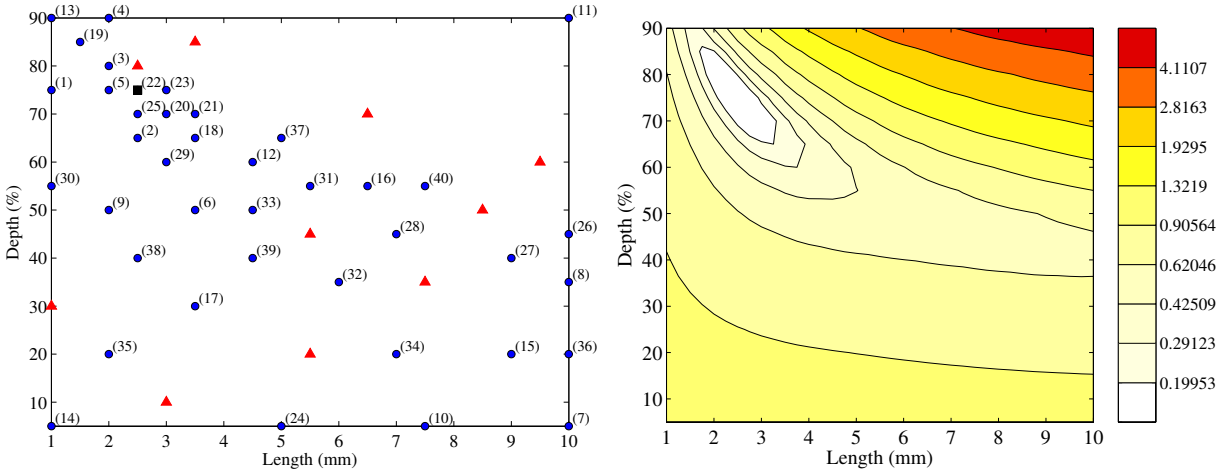


(b) Case 1/b.

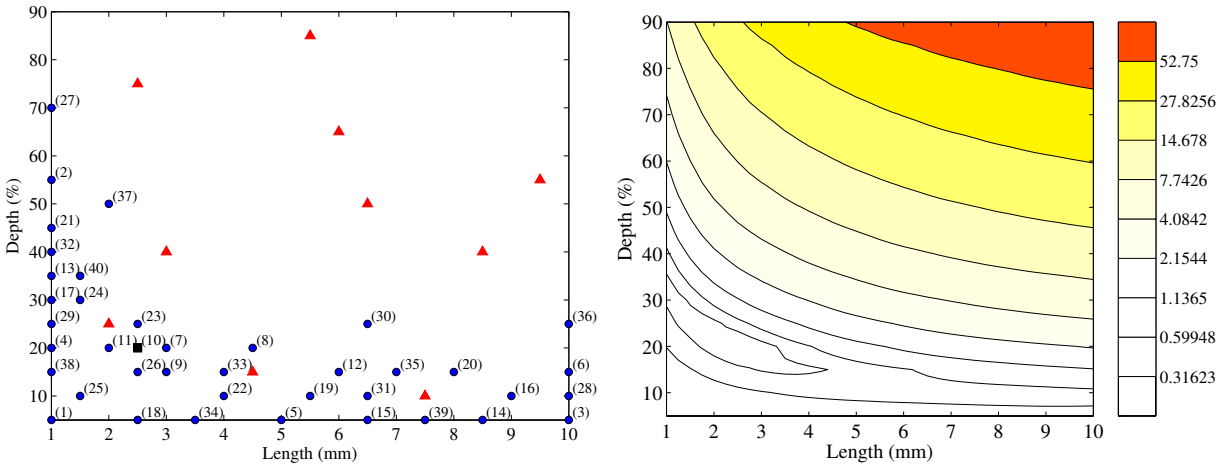


(c) Case 1/c.

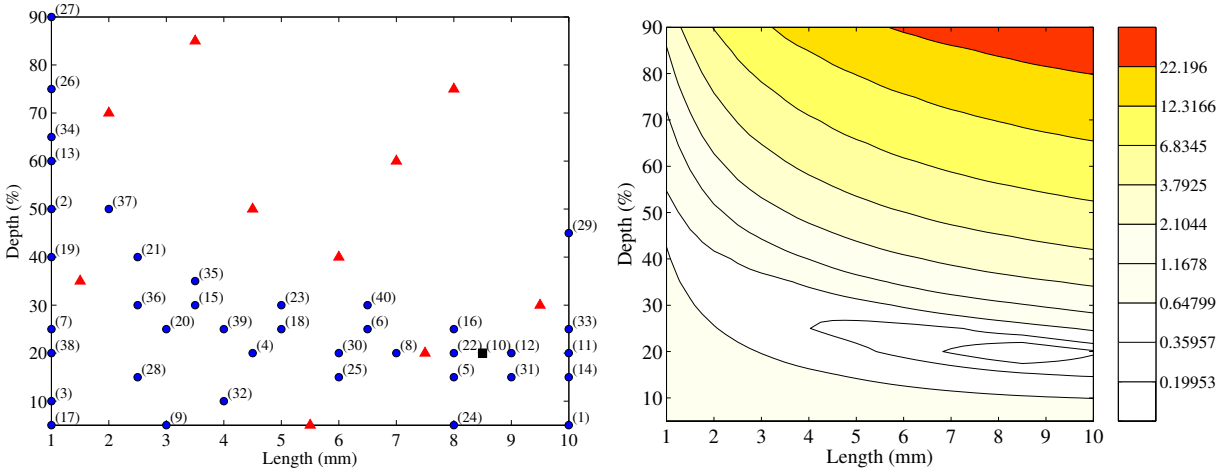
Figure 3.4: Evolution of the inversion method in the 2-parameter cases. Top: current minimum of the previous observations. Bottom: logarithm to base 10 of the maximum of the expected improvement.



(a) Case 1/a.



(b) Case 1/b.



(c) Case 1/c.

Figure 3.5: Sequence of the observations (*left*) and the objective functions (*right*, contour plots with logarithmic scale) in the 2-parameter cases. Red triangles: initial samples. Blue dots: sequentially added samples (with the number of iteration in brackets). Black square: the global minimizer.

3.5.3 Examples with 4 parameters

The 4-parameter examples involve the canonical configuration defined as the point 2 in § 2.4. Let us recall that the position (A, B) in the plane of the plate, the length (L) and the depth (D) of a single crack are enabled to vary, the input space consists of approximately 4×10^4 feasible input vectors. After 40 initial observations, 160 iterations have been performed.

The numerical summary of the three runs (with three different defects to be retrieved) is given in Table 3.3. In all three cases, the assumed defect has exactly been found. The average time of covariance estimation and CWS optimization is considerably higher that in the previous example.

The performance of the algorithm is shown in Fig. 3.6. Similar comments can be made here as in the previous example: the maximal EI tendentiously decreases during the iterations, however, its range is quite different in the three examples. The strongly non-monotonous nature of $\eta_{\max}$ in the case 2/b well illustrates how the EGO algorithm discovers newer and newer regions of the input space, and the expected improvement can rapidly increases even from one iteration to the next one. Unfortunately, this phenomenon shows again that the stopping criterion cannot be based only on the current value of $\eta_{\max}$.

Table 3.3: Numerical description of the 4-parameter examples.

Example		2/a	2/b	2/c
Assumed defect	A (mm)	1.5	1.5	1.5
	B (mm)	2.0	2.0	2.0
	L (mm)	2.5	2.5	8.5
	D (%)	75	20	20
Performance	$n^\star$	40 + 76	40 + 126	40 + 41
	$\bar{t}_{\text{est}}$ (s)	18.0	31.3	19.8
	$\bar{t}_{\text{opt}}$ (s)	11.0	7.86	8.81

3.5.4 Examples with 6 parameters

The double-crack configuration –defined in the point 3 of § 2.4– is considered in the 6-parameter examples. Let us recall that the positions (A, h, B_1 and B_2) and the lengths (L_1, L_2) of two parallel cracks are to be retrieved, whereas the depth of both cracks is assumed to be 40 % of the plate thickness. The input space consists of approximately 4.39×10^6 feasible input vectors.

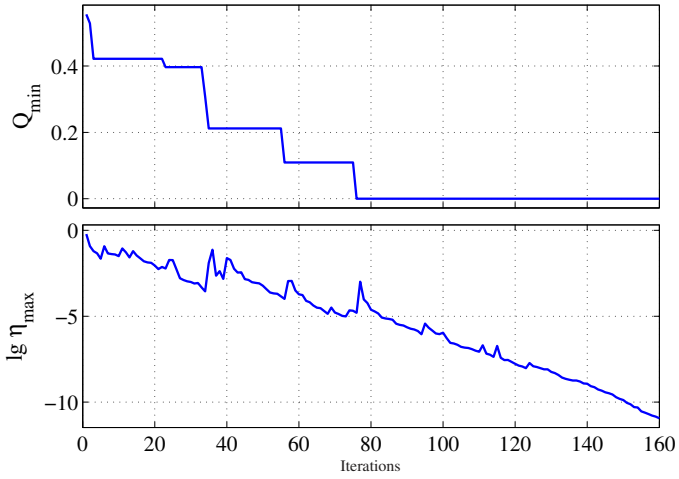
60 initial observations had been performed before the iterative procedure started. The latter was stopped after 240 cycles in all runs. Similarly to the previous examples, the assumed defects have exactly been found by the inversion algorithm. The numerical summary of the runs is given in Table 3.4. In this 6-parameter case, the average time of covariance estimation and of CWS search appear to be similar, and both have considerably increased in comparison with the previous 4-parameter examples. This practically means that the computation time related to the EGO algorithm in the whole procedure increased up to approximately 8 hours.

Let us notice that the number of observations needed to find the global minimizer ($n^\star$) is surprisingly small: it does not exceed 200 even in the worst case (among the three presented ones), however, the global search is performed over an input space whose cardinality is over 4 million.

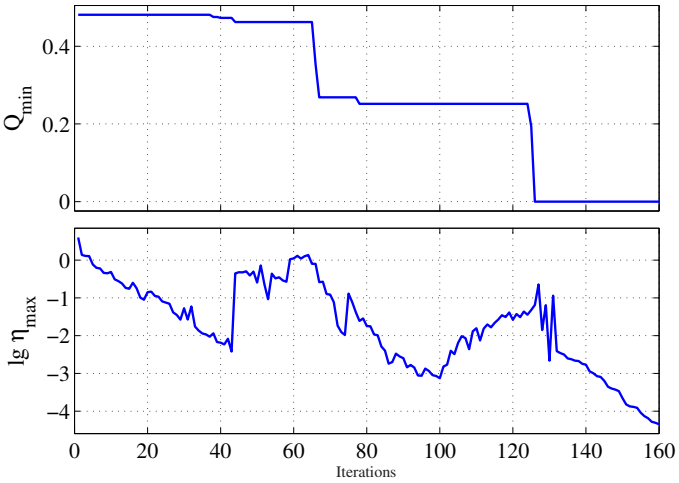
The graphical evaluation of the performance is shown in the plots of Fig. 3.7. Although the defect configurations to be retrieved are far from being similar in the cases 3/a and 3/b, these cases look quite similar in the sense of both $n^\star$ and the pretty linear decrease of the log-scaled $\eta_{\max}$.

Table 3.4: Numerical description of the 6-parameter examples.

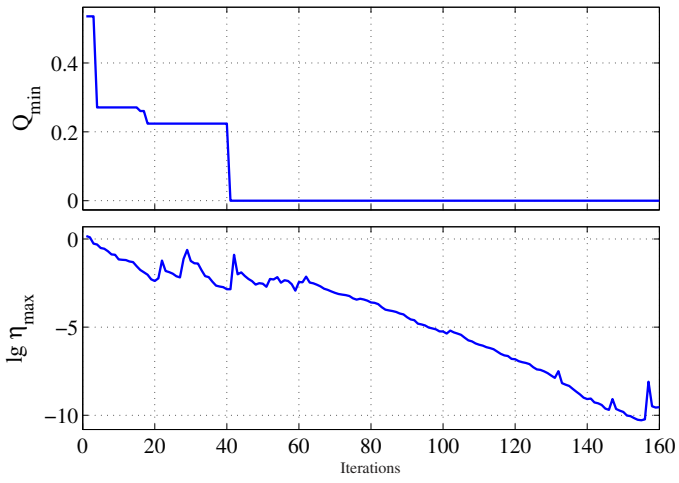
Example		3/a	3/b	3/c
Assumed defect	A (mm)	0.75	0.75	0.75
	h (mm)	0.5	1.75	0.5
	B_1 (mm)	−2.0	−2.0	2.0
	B_2 (mm)	2.0	2.0	2.0
	L_1 (mm)	8.5	2.5	2.5
	L_2 (mm)	8.5	2.5	8.5
Performance	$n^\star$	60 + 17	60 + 30	60 + 79
	$\bar{t}_{\text{est}}$ (s)	53.7	54.7	72.7
	$\bar{t}_{\text{opt}}$ (s)	55.4	49.7	54.7



(a) Case 2/a.

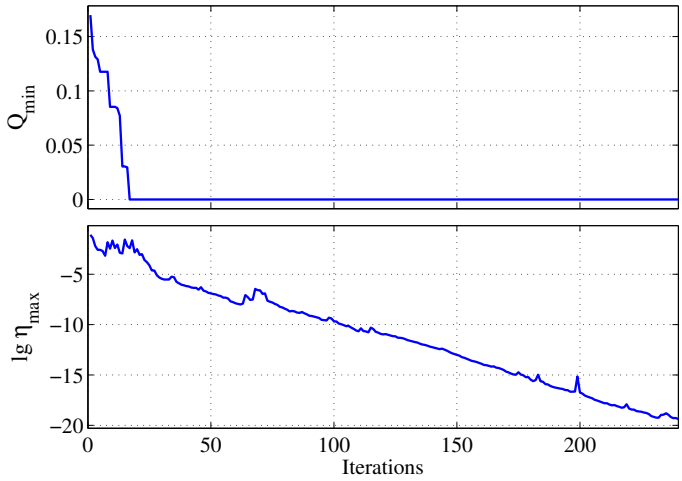


(b) Case 2/b.

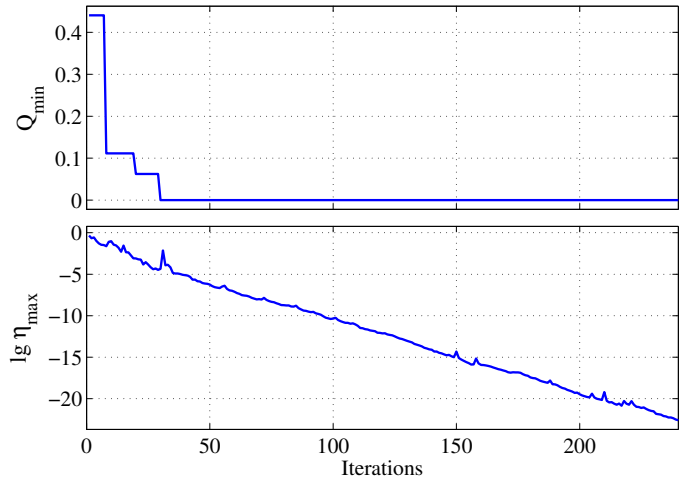


(c) Case 2/c.

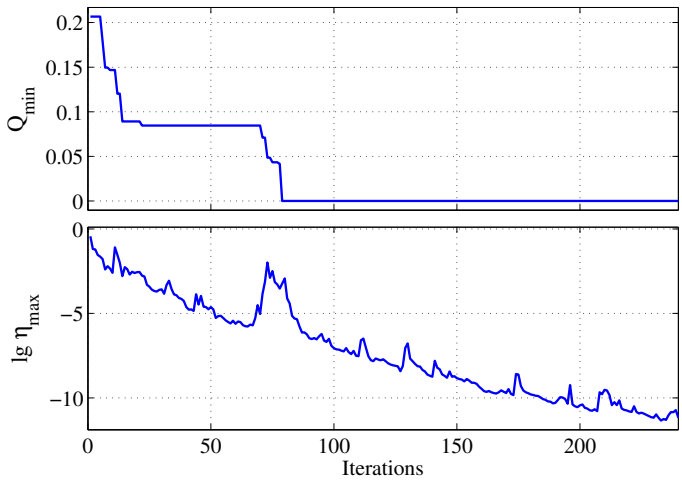
Figure 3.6: Performance of the inversion method in the 4-parameter cases. Top: current minimum of the previous observations. Bottom: logarithm to base 10 of the maximum of the expected improvement.



(a) Case 3/a.



(b) Case 3/b.



(c) Case 3/c.

Figure 3.7: Performance of the inversion method in the 6-parameter cases. Top: current minimum of the previous observations. Bottom: logarithm to base 10 of the maximum of the expected improvement.

3.5.5 Examples with 8 parameters

The 8-parameter examples are based on the double-crack configuration defined in the point 4 of § 2.4. Let us recall that the positions (A, h, B_1 and B_2), the lengths (L_1, L_2) and the depths (D_1, D_2) of two parallel cracks are to be retrieved. The input space consists of approximately 1.12×10^9 feasible input vectors.

Table 3.5: Numerical description of the 8-parameter examples.

Example		4/a	4/b	4/c
Assumed/found defect	A (mm)	0.75/0.75	0.75/0.75	0.75/0.5
	h (mm)	0.5/0.5	0.5/0.5	0.5/0.25
	B_1 (mm)	−2.0/−2.0	−2.0/−2.0	2.0/2.0
	B_2 (mm)	2.0/2.0	2.0/1.5	2.0/2.0
	L_1 (mm)	8.5/8.5	8.5/8.5	2.5/4.0
	L_2 (mm)	8.5/8.5	8.5/9.0	8.5/2.0
	D_1 (%)	75/75	75/75	75/45
	D_2 (%)	75/75	20/20	20/75
Performance	$n^\star$	80 + 134	80 + 205	80 + 178
	$\bar{t}_{\text{est}}$ (s)	162.9	171.6	145.6
	$\bar{t}_{\text{opt}}$ (s)	231.0	263.1	283.5

The initialization of the EGO algorithm has involved 80 initial observations. The number of iterations that the algorithm made is 320. Among the three runs of this example, the assumed defect could only be reconstructed in one case (4/a) within 320 iterations. The numerical summary in Table 3.5 presents both the assumed and the retrieved defects.

The time-consumption of the EGO algorithm obviously increased compared to the 6-parameter example. In this case, the total EGO-related time of a run (with 320 iterations) can exceed 40 hours.

Let us have a closer look at the unsuccessful cases 4/b and 4/c. In the case 4/b, the retrieved configuration appears to be quite similar to the assumed configuration and the corresponding value of the objective function is indeed small (Fig. 3.9b). A few further iterations would probably be enough to obtain the true solution, however, in a real case when noise corrupts the measurement, one would possibly be satisfied with such precision of the defect characterization. As a simple trial, we applied the CWS algorithm directly to the optimization of the objective function, started by the input vector found by the EGO algorithm. The CWS algorithm then rapidly converged to the true solution.

In the case 4/c, the search of EGO has been stalled at a local minimum of the objective function, the found defect is quite different from the assumed one, and the corresponding value of the objective function is still high. To visualize the objective function, 8 cross-sections of the input space are chosen, all including the found solution (see in Table 3.5). In all of these cross-sections, 7 of the input parameters are kept constant and only the remaining one varies. In Fig. 3.8, the objective function is plotted in the 8 cross-sections, and the found solution is indeed a local minimizer, as it can clearly be seen from the figures.

The evolution of the optimization in all the three cases is presented in Fig. 3.9. Let us notice that in the case of 4/a, the plot of $\eta_{\max}$ might suggest correctly that the global minimizer has been found, whereas the oscillation of $\eta_{\max}$ in the case 4/c might warn the user that still spacious unexplored regions are present in the input space.

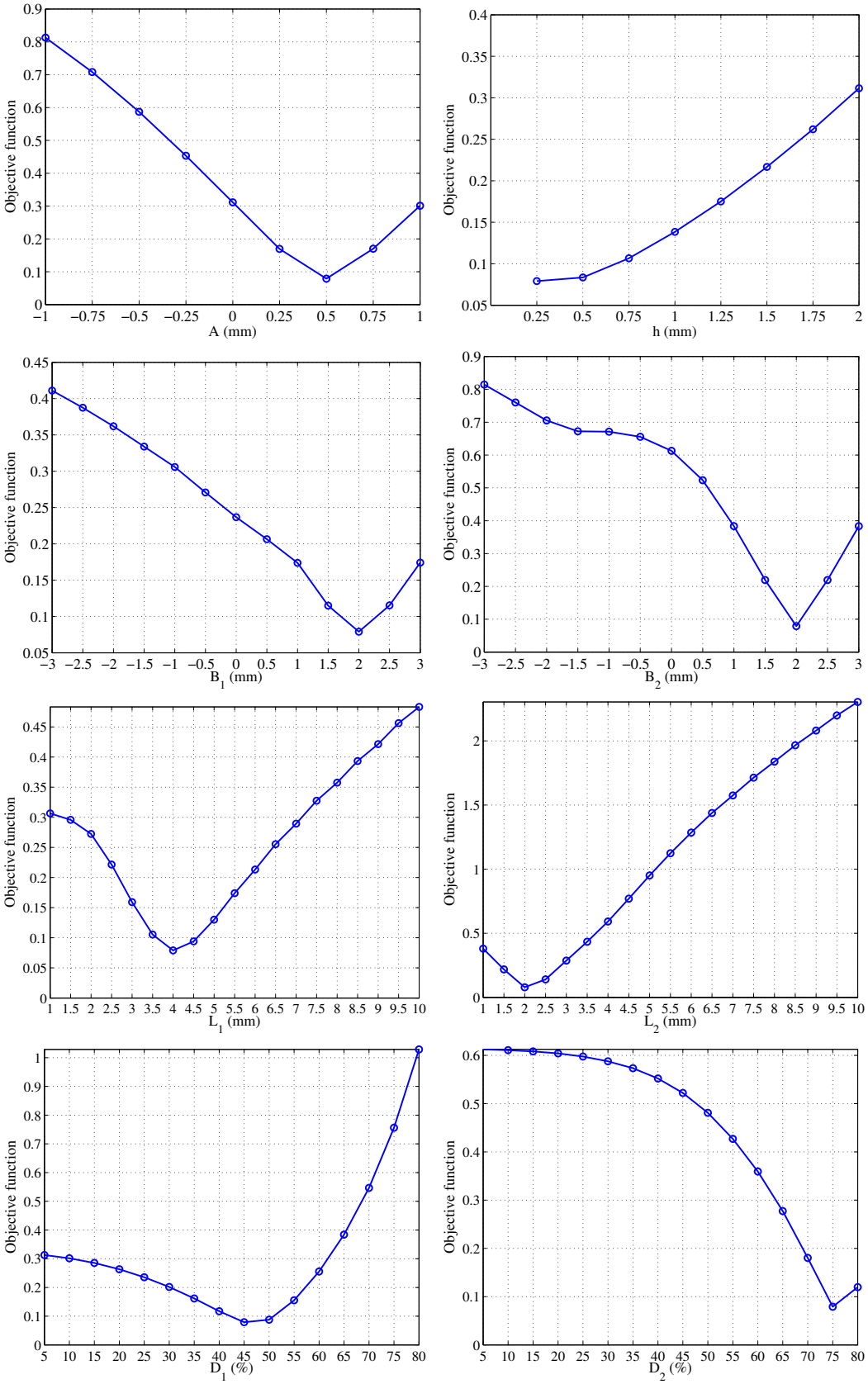
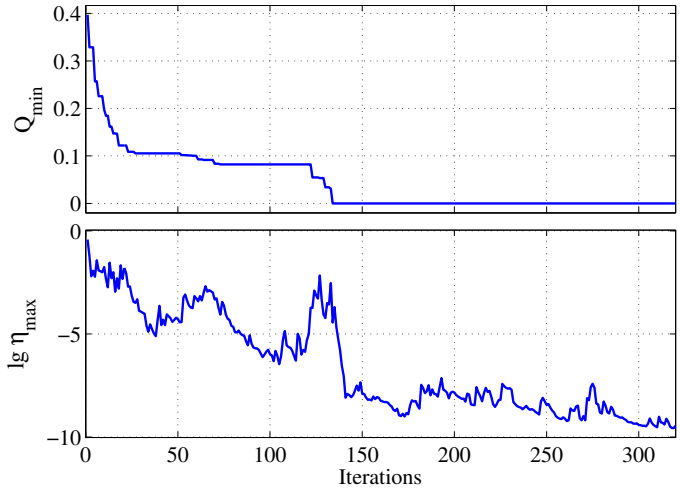
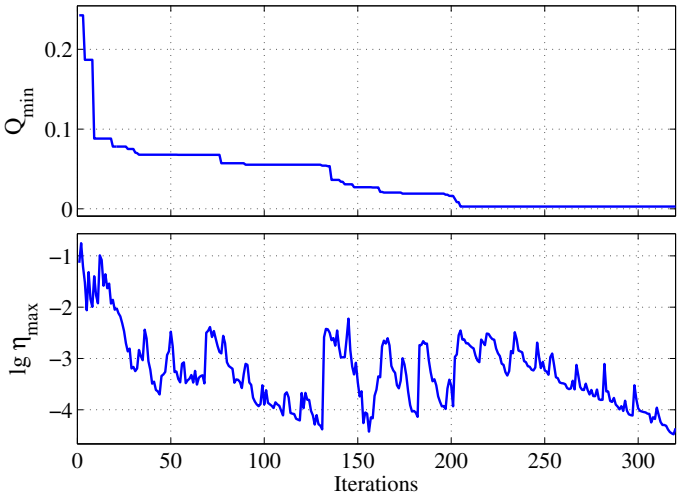


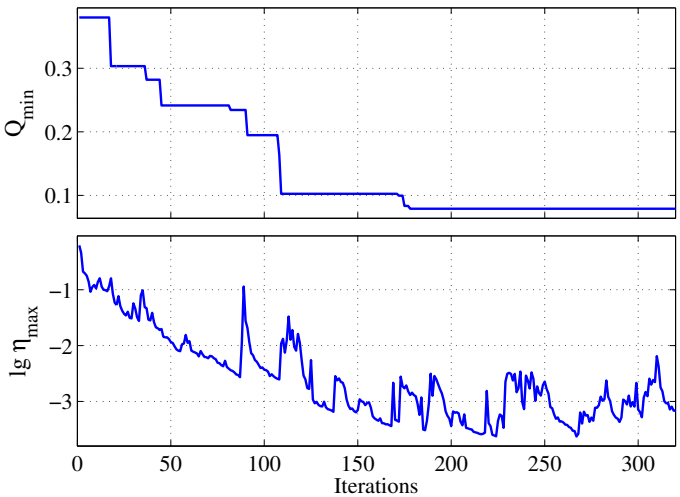
Figure 3.8: Cross-section plots of the objective function of 8 parameters in the case 4/c around the (local) minimum found by the EGO algorithm.



(a) Case 4/a.



(b) Case 4/b.



(c) Case 4/c.

Figure 3.9: Performance of the inversion method in the 8-parameter cases. Top: current minimum of the previous observations. Bottom: logarithm to base 10 of the maximum of the expected improvement.

3.5.6 "Hybrid" examples: double-crack reconstruction with a single-crack model

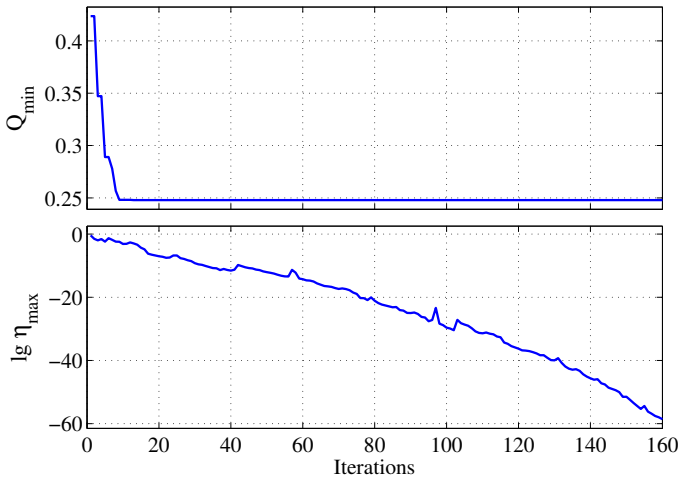
To include an example where the "inverse crime" does not occur any more, thus, to test the proposed inversion method under different circumstances, the following test case is presented. The assumed defects of the previous 8-parameter example (Table 3.5) are aimed to be characterized using the 4-parameter single crack model (the same as in § 3.5.3). Since 4 input parameters are involved, the number of initial samples was 40, and 160 subsequent iterations were performed.

The numerical summary is given in Table 3.6. The retrieved defects are as expected based on our knowledge on the physical background of the studied ECT setup: as a rule of thumb, we can say that in such cases, the *union* of the crack-surfaces (projected to the yz plane) should be approximated by the retrieved single crack. This –in good agreement with the results– also says that the effect of a shallow crack in the presence of a much deeper one does not really appear in the result of the single-crack reconstruction. Let us also note that in the case 5/a, the "union of the crack surfaces" would imply a single crack of 12.5 mm length, however, the upper limit of L in the 4-parameter model is 10 mm.

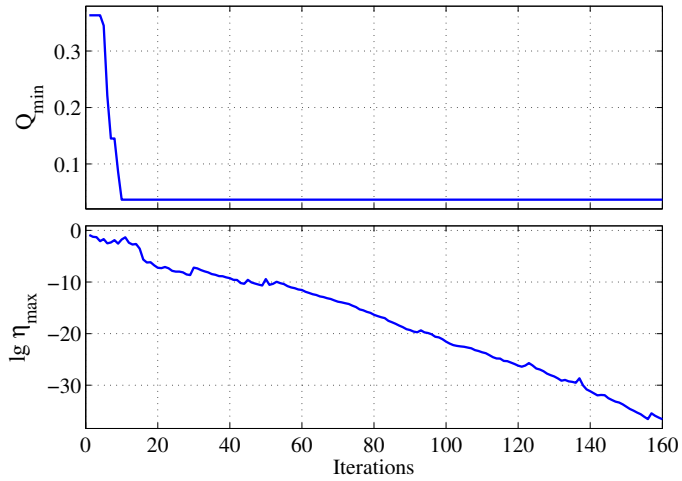
The related plots of the evolution of the EGO algorithm are shown in Fig. 3.10. The global minimizer has (supposedly) been found within a small number of iterations, especially in the cases 5/a and 5/b. Moreover, $\eta_{\max}$ decreases to an extremely small value by the end of the optimization, making one believe that the global minimizer has indeed been found. Obviously, the minimum of the objective function cannot be zero in this example, but not more than this can be known on $Q_{\min}$ a priori.

Table 3.6: Numerical description of the "hybrid" examples. The assumed defects are defined in Table 3.5.

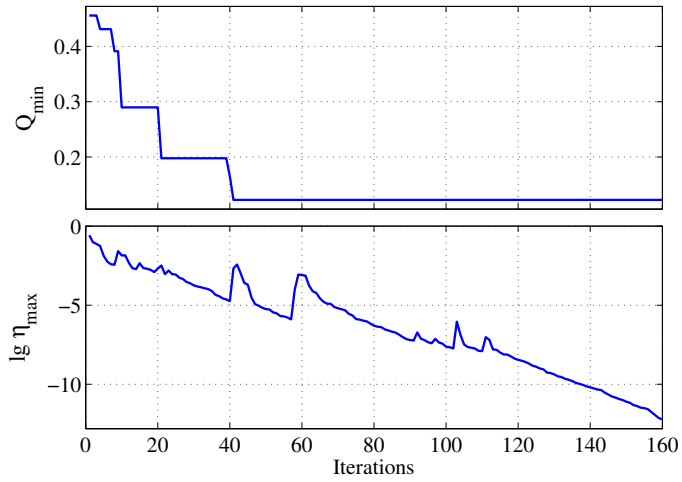
Example		5/a	5/b	5/c
Found defect	A (mm)	0.5	0.5	0.5
	B (mm)	0.0	−2.0	2.0
	L (mm)	10	8.5	3.0
	D (%)	80	75	70
Performance	n^*	40 + 53	40 + 10	40 + 41
	$\bar{t}_{\text{est}}$ (s)	42.4	44.0	48.2
	$\bar{t}_{\text{opt}}$ (s)	10.9	11.3	11.7



(a) Ex. 5/a.



(b) Ex. 5/b.



(c) Ex. 5/c.

Figure 3.10: Performance of the inversion method in the 4-parameter "hybrid" cases (double-crack characterization using a single-crack model). Top: current minimum of the previous observations. Bottom: logarithm to base 10 of the maximum of the expected improvement.

3.6 Conclusion and perspectives

The EGO algorithm –with our application-oriented implementation– appears to be a useful tool for the solution of inverse problems in the presented ENDE configuration, based on our numerical studies. Let us highlight the main advantages of the method:

- The number of forward simulations needed to solve the regularized inverse problem is small. The size of the initial sample set depends linearly on the N number of input parameters in the presented examples. Similarly, the maximal number of subsequently added samples increases linearly with N. This property makes the method particularly favourable in the case of expensive-to-run forward simulators (the one we actually use is relatively cheap-to-run, in comparison with, e.g., a simulator based on FEM). Let us note that certain global optimization methods need much higher number of forward simulations in an inversion scheme. For instance, in [Zaoui, Marchand, and Pávó, 2001], the solution of inverse problems with 5 and 10-parameter defect models is presented, using a genetic algorithm: the number of forward simulations in that approach is of several thousand (up to 20 thousand).

- The EGO algorithm treats the forward model as a black-box. Thus, other EM simulators can easily be used in the inversion scheme. Moreover, the EGO does not account for the derivatives of the objective function, contrarily with the wide family of gradient-based methods. Since the derivatives (with respect to the input parameters) might not be expressed analytically and/or might be expensive-to-compute, the derivative-free property of EGO is a considerable advantage.

- Defining the input space as a discrete grid is a way to trace back the general continuous optimization problem to a discrete search task. If this discretization is appropriate with respect to the objective function at hand, a physical interpretation can also be given to this mathematical trick: the required precision of inversion can be controlled to a certain extent. Moreover, the simulator we actually use fits naturally to the discrete grid of the input space.

Though the above advantages are believed to be impressive, some disadvantages of the proposed method must also be noted:

- The speed of the convergence of EGO is not known theoretically. Although we experienced a fast convergence, there is no general guarantee for that.

- No reliable stopping criterion has been found so far. The maximal EI ($\eta_{\max}$) could be used to have a guess how much improvement is expected form further iterations. However, according to our examples, a lower bound on $\eta_{\max}$ of its own is not a reliable stopping criterion.

- In most of our examples, the initialization appears to be appropriate. However, in other cases (when the objective function is more complicated, e.g., probably in the last 8-parameter case 4/b), more initial observations might be needed. When little is known on the behaviour of the objective function a priori, finding an appropriate initial sample set is rather a trial-and-error task.

- The auxiliary computations –estimation of the covariance and the CWS optimization– are getting more and more time-consuming as the number of input parameters is increasing. In Fig. 3.11, the average times (of the cases a), b) and c) in all of the examples) are presented. In spite of the fast increase of the computation time, EGO can still be overall effective in the case of expensive-to-run forward simulations.

- Even if the algorithm seems to be easy-to-automatize, certain issues (e.g., the covariance estimation or the stopping rule) must be treated with special attention which cannot be expected from an "end-user". Consequently, this inversion scheme is more appropriate for professional use (i.e., controlled by "experts"), rather than for "end-users".

- Finally, a strange contradiction must be mentioned. The objective function $Q(\mathbf{p})$ is modeled by a Gaussian random process, however, $0 \leq Q(\mathbf{p})$ always holds. This positivity-constraint is not involved in the random process model. When computing the expected improvement, in fact, the expectation of the truncated normal distribution of $\hat{\xi}(\mathbf{p})$ (upper bounded by $Q_{\min}$) is considered (see § B.3). One might miss also a lower bound of zero to involve the knowledge of $0 \leq Q(\mathbf{p})$ in the model. However, we did not study this issue (believed to be far from straightforward) in detail, but let us only note that the present expression of the expected improvement is a "weak" form in the above sense. A plausible remedy would be to model the logarithm of the objective function to eliminate the positivity-constraint. In so doing, the regularity of the modeled function would highly be violated for small $Q(\mathbf{p})$ values, which is to be avoided. Let us also mention, however, that some attempts to perform "positive kriging" to model non-negative functions have been discussed in the literature (e.g., [Chilès and Delfiner, 1999]).

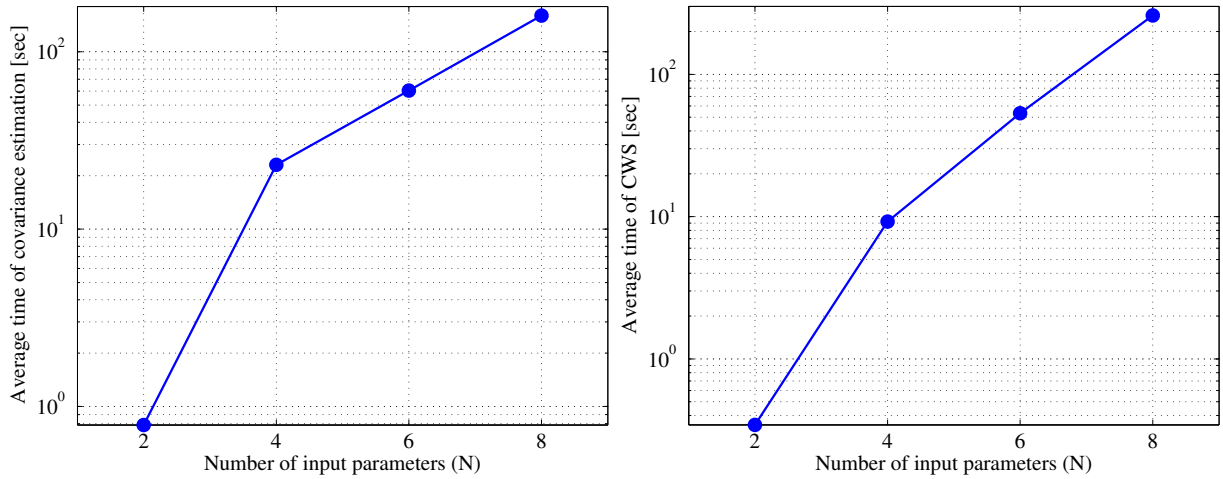


Figure 3.11: Average time of the EGO-related computations (per iteration) in function of the number of input parameters.

According to the presented examples, the developed scheme can be used to solve inverse problems of ENDE. However, the method can obviously be improved. For the future, the following main directions of research and development can be outlined:

- The fast convergence of the EGO algorithm could be more highlighted by carrying out an exhaustive comparison with other global optimization techniques.
- The effect of the initialization (number and repartition of the input samples) influences the performance of the algorithm in the subsequent iterations. Some heuristic rules might then be established concerning the initialization and the performance during the iterations.
- Though EGO is a derivative-free method, one might obtain the derivatives of the objective function at a low computational cost. In this case, it might be reasonable to consider some extensions of kriging, accounting for the derivatives (e.g., [Mardia et al., 1996]) to obtain more precise predictions for the objective function.
- Some heuristic stopping rule might be established based on a combination of $\eta_{\max}$, the number of iterations, $Q_{\min}$ (perhaps on these values within a given number of subsequent iterations). If the derivatives of the objective function are also available (or predicted by an appropriate kriging model, e.g., [Vazquez and Walter, 2005]), then the stopping criterion might further be improved.
- The EGO algorithm could be used as a part of a multi-stage optimization scheme. For instance, an

approximate solution is sought on a coarse grid, then a second search (might also be EGO on a finer grid or e.g., a gradient-method) is performed around the found approximate solution.

4

Surrogate modeling by functional kriging

A plausible idea in surrogate modeling is to evaluate the given forward operator at certain values of the input parameters, then to approximate the studied model knowing the observed data without performing any further forward simulation. This approximation often consists in interpolating the forward operator based on the observations. The performance of the yielded surrogate model obviously depends on both the choice of the input values (where observations are performed) and the applied interpolator. Consequently, to obtain a high-precision interpolation-based surrogate model, one must consider adaptive DoE approaches to generate the sample-set, which take into account both the interpolator (subsequently fitted to the samples) and the behaviour of the modeled forward operator as well.

In this chapter, the use of the so-called functional kriging technique for surrogate modeling is presented. Functional kriging –a recent extension of the classical theory– makes possible the direct prediction of functional data. As an adaptive DoE method, we use a sequential sampling strategy for the generation of the database. This algorithm is driven by the estimated uncertainty of the kriging model being built. The approach is illustrated and the yielded surrogate models are evaluated via numerical studies in the domain of ECT, with defect models using 2, 4 and 6 input parameters.

This work forms the basis of the thesis point 2 (in § 6). We have published our results related to the presented approach in the following papers: [Bilicz et al., 2009c; Bilicz et al., 2010b; Bilicz et al., 2010e]. We have recently submitted the following papers concerning the topic of this chapter: [Bilicz et al., 2011a; Bilicz et al., 2011b; Vaskó et al., 2011].

4.1 Interpolation-based surrogate modeling

Let us note first that some terms and notations used in the following have been defined in § 2.3. A cheap surrogate model of the usually expensive-to-evaluate forward operator $\mathcal{F}$ can be obtained by the following two-stage approach:

1. Let us evaluate the forward operator at n different input vectors

$$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n \in \mathbb{P}, \tag{4.1}$$

i.e., generate the *database*

$$\mathbb{D}_n = \left\{ \left(\mathbf{p}_1, q_{\mathbf{p}_1}(t)\right), \left(\mathbf{p}_2, q_{\mathbf{p}_2}(t)\right), \ldots, \left(\mathbf{p}_n, q_{\mathbf{p}_n}(t)\right) \right\}, \tag{4.2}$$

where

$$q_{\mathbf{p}_i}(t) = \mathcal{F}\{\mathbf{p}_i\}, \quad \forall i = 1, 2, \ldots, n. \tag{4.3}$$

The database $\mathbb{D}_n$ (also called sample-set or training set) of n corresponding input parameter - output data pairs can be considered as a sort of discrete representation of the forward operator $\mathcal{F}$.

2. Let us approximate $\mathcal{F}$ based on the samples in $\mathbb{D}_n$. The $\mathcal{G}^{(n)}$ approximate forward operator, i.e., the surrogate model is required to interpolate $\mathcal{F}$ and to be "similar" to $\mathcal{F}$ in the whole input space:

$$\begin{aligned} \mathcal{G}^{(n)}\{\mathbf{p}_i\} &= \mathcal{F}\{\mathbf{p}_i\}, \quad \forall i = 1, 2, \ldots, n, \text{ and} \\ \mathcal{G}^{(n)}\{\mathbf{p}\} &\simeq \mathcal{F}\{\mathbf{p}\}, \quad \forall \mathbf{p} \in \mathbb{P}. \end{aligned} \tag{4.4}$$

The upper index (n) emphasizes the sample-based nature of the surrogate model.

The main advantage of separating the sampling and the interpolation steps consists in concentrating the computational burden (the forward simulations) to the first stage. The generation of the database is needed to be performed only once, practically by "experts", equipped with the numerical simulator (which might be an expensive commercial software and/or more complex than an "end-user" can be expected to know its correct usage). In the second stage, the "end-user" recourses only to the pre-computed database and the interpolator, respectively. In other words, such interpolation-based surrogate modeling can be considered as an "off-line" simulation of the forward problem.

The choice of the input samples $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ needs an appropriate DoE tool. Besides the classical approaches (e.g., factorial or LHS designs), more sophisticated tools also exist. Similarly, a wide range of interpolation methods is available. However, both sampling and applied interpolator strongly influence the properties of the yielded surrogate model. Herein we propose the use of functional kriging (§ 4.2), being a recent extension of the traditional kriging approach of pointwise prediction of scalar functions (§ B) to the modeling of functional data. For a recent overview of functional data analysis, see, e.g. [Delicado et al., 2009]. We use the so-called ordinary kriging approach, discussed, e.g., in [Giraldo, Delicado, and Mateu, 2007; Giraldo, 2009]. To improve the precision of the kriging-based surrogate model, an adaptive sampling strategy is involved in the database generation, driven by the estimated uncertainty of the kriging prediction (§ 4.3).

4.2 Functional kriging interpolation

The basic idea of functional kriging is the prediction of a whole function as a single entity, instead of pointwise prediction of scalar values of an unknown function as performed in the traditional framework. Let us consider the output $q_{\mathbf{p}}(t)$, being a function of $t \in T$ (t is actually a two-element vector, referring to a point of the surface T, see (2.11)) for all $\mathbf{p} \in \mathbb{P}$. The functions $q_{\mathbf{p}}(t)$ are observed at n points, i.e., the database (4.2) has been generated. Our goal is then to predict the $q_{\mathbf{p}}(t)$ functional output for an unobserved $\mathbf{p} \in \mathbb{P}$. To this end, $q_{\mathbf{p}}(t)$ is modeled by a functional Gaussian random process

$$\xi_{\mathbf{p}}(t), \quad \mathbf{p} \in \mathbb{P}, \quad t \in T. \tag{4.5}$$

A such functional process can intuitively imagined –without a rigorous formalization– as a set of scalar processes $\xi_p(t_0)$, for each $t_0 \in T$. These scalar processes (over $\mathbb{P}$) model the output $q_{\mathbf{p}}(t)$ at a certain fixed $t = t_0$. We call the functional process Gaussian, if the scalar processes for each $t_0 \in T$ are Gaussian processes, i.e., every finite combination

$$\left[\xi_{p_1}(t_0), \xi_{p_2}(t_0), \ldots, \xi_{p_K}(t_0)\right] \tag{4.6}$$

has a multivariate normal distribution. The modeling functional process is characterized by its mean and covariance, similarly to § B.1. Since we usually have no prior information on $q_{\mathbf{p}}(t)$, we assume a constant but unknown mean function –this is an important property of the ordinary kriging approach used–, i.e.,

$$\mathbb{E}\left[\xi_{\mathbf{p}}(t)\right] = m(t), \quad \forall \mathbf{p} \in \mathbb{P}. \tag{4.7}$$

The covariance of the process over $\mathbb{P}$ depends on $\mathbf{p}$ and on t as well:

$$k(\mathbf{p}_i, \mathbf{p}_j, t) = \mathbb{E}\left[\left(\xi_{\mathbf{p}_i}(t) - m(t)\right)\left(\xi_{\mathbf{p}_j}(t) - m(t)\right)\right]. \tag{4.8}$$

Functional kriging provides a linear prediction of the array of modeling processes $\xi_{\mathbf{p}}(t)$, based on the observations of the modeling process, collected in a vector

$$\boldsymbol{\xi}(t) = [\xi_{\mathbf{p}_1}(t), \xi_{\mathbf{p}_2}(t), \ldots, \xi_{\mathbf{p}_n}(t)]^{\mathsf{T}}. \tag{4.9}$$

We have chosen and applied a simple approach, the prediction is written in the form[1]

$$\hat{\xi}_{\mathbf{p}}(t) = \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}}\boldsymbol{\xi}(t). \tag{4.11}$$

Similarly to the scalar function case (§ B.1), our requirements on the prediction are:

- The unbiasedness of the prediction, i.e.,

$$\mathbb{E}\left[\hat{\xi}_{\mathbf{p}}(t)\right] = \mathbb{E}\left[\xi_{\mathbf{p}}(t)\right], \quad \forall t \in T \tag{4.12}$$

  must hold. Assuming a constant mean (4.7), the unbiasedness constraint reads as the equality

$$\sum_{i=1}^{n} \lambda_i(\mathbf{p}) = 1. \tag{4.13}$$

- The prediction must be optimal in the sense of the integrated variance of prediction error, i.e., the following minimization problem is to be solved:

$$\boldsymbol{\lambda}(\mathbf{p}) = \arg\min_{\boldsymbol{\chi}(\mathbf{p})\in\mathbb{R}^n} \int_T \mathbb{E}\left[\left(\xi_{\mathbf{p}}(t) - \boldsymbol{\chi}(\mathbf{p})^{\mathsf{T}}\boldsymbol{\xi}(t)\right)^2\right] \mathrm{d}t, \tag{4.14}$$

  where the unbiasedness constraint is taken into account.

The prediction (4.11) is called the Best Linear Unbiased Prediction (BLUP) if it satisfies (4.12) and the coefficients are chosen optimally in the sense of (4.14). Once the coefficients $\boldsymbol{\lambda}(\mathbf{p})$ are determined, the prediction of the output $q_{\mathbf{p}}(t)$ at an unobserved point $\mathbf{p} \in \mathbb{P}$ is

$$\hat{q}_{\mathbf{p}}(t) = \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}}\mathbf{q}(t), \tag{4.15}$$

where the vector $\mathbf{q}(t)$ consists of the observed functional outputs

$$\mathbf{q}(t) = \left[q_{\mathbf{p}_1}(t), q_{\mathbf{p}_2}(t), \ldots, q_{\mathbf{p}_n}(t)\right]^{\mathsf{T}}. \tag{4.16}$$

To obtain the optimal coefficients $\boldsymbol{\lambda}(\mathbf{p})$, yielding the BLUP, let us introduce the so-called trace-covariance:

$$k_T(\mathbf{p}_i, \mathbf{p}_j) \coloneqq \int_T k(\mathbf{p}_i, \mathbf{p}_j, t)\mathrm{d}t. \tag{4.17}$$

[1] Let us note that more complicated approaches also exist. For instance, the "total model":

$$\hat{\xi}_{\mathbf{p}}(t) = \int_T \boldsymbol{\lambda}(\mathbf{p}, t, s)^{\mathsf{T}}\boldsymbol{\xi}(s)\mathrm{d}s. \tag{4.10}$$

This scheme (4.10) appropriately takes into account the relation between the ith observation of the process at the point s ($\xi_{\mathbf{p}_i}(s)$) and the predicted process at the point t ($\hat{\xi}_{\mathbf{p}}(t)$). However, the determination of the functional coefficients $\lambda_i(\mathbf{p}, t, s)$ is a rather complicated task. In this Dissertation, we consider only the (4.11) functional kriging prediction.

The trace-covariance $k_T(\mathbf{p}_i, \mathbf{p}_j)$ can be considered as an integrated characteristic of the statistical behaviour of the functional process over $\mathbb{P}$. One can prove that the optimization problem (4.14) under the constraint (4.13) is equivalent to solving the following linear system of equations via a Lagrangian formalization:

$$\begin{bmatrix} \mathbf{K}_T & \mathbf{1} \\ \mathbf{1}^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}(\mathbf{p}) \\ \mu(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} \mathbf{k}_T(\mathbf{p}) \\ 1 \end{bmatrix}, \tag{4.18}$$

where the trace-covariance matrix $\mathbf{K}_T$ and vector $\mathbf{k}_T(\mathbf{p})$ are defined as

$$\mathbf{K}_T = \Big[k_T(\mathbf{p}_i, \mathbf{p}_j)\Big]_{ij} \quad \text{and} \quad \mathbf{k}_T(\mathbf{p}) = \big[k_T(\mathbf{p}, \mathbf{p}_i)\big]_i, \quad \forall i, j = 1, 2, \ldots, n \tag{4.19}$$

and $\mu(\mathbf{p})$ is the Lagrange-multiplier (for a detailed derivation, see [Giraldo, Delicado, and Mateu, 2007]).

Neither the covariance, nor the trace-covariance is known explicitly, but one must predict them from the observed data. In our approach, only the trace-covariance is involved and to be estimated. To this end, we assume that the trace-covariance is stationary (invariant to shifting in $\mathbb{P}$), i.e., it can be re-defined as a one-variate function

$$k_T(\mathbf{p}_a - \mathbf{p}_b) := k_T(\mathbf{p}_a, \mathbf{p}_b). \tag{4.20}$$

The assumption of stationarity is usually not based on physical considerations, but often a necessary simplification, since the use of non-stationary covariance models, as a rule, is much more challenging. The function $k_T(\cdot)$ is then assumed to be an element of a parametrized function class, e.g., exponential, rational or polynomial. In our numerical studies, the so-called Matérn function is used as a covariance model. In § B.2, the Matérn function is described in detail, along with an interpretation of its parameters (called the "hyperparameters"). To fit the covariance model to the observed data, the hyperparamerters are estimated based on the samples (4.2) by using a functional cross-validation technique (§ B.2.2).

4.3 Jackknife variance estimation

The quality of the kriging interpolator can easily been quantified via the variance of the interpolation error. More precisely, we use the trace-variance, being an integrated quantity:

$$\sigma_T^2(\mathbf{p}) = \int_T \mathbb{E}\left[\left(\hat{\xi}_\mathbf{p}(t) - \xi_\mathbf{p}(t)\right)^2\right] \mathrm{d}t. \tag{4.21}$$

Let us note that the unbiasedness constraint (4.12) is used in the above formula. This variance $\sigma^2(\mathbf{p})$ has a straightforward interpretation in the stochastic framework: the higher the variance at a point $\mathbf{p}$, the less certain the prediction $\hat{q}_\mathbf{p}(t)$. This expression has already occurred in (4.14) where the goal was to determine the coefficient vector $\boldsymbol{\lambda}(\mathbf{p})$ for a given sample-set $\mathbb{D}_n$. However, now we will use the trace-variance to optimize the sample pattern, i.e., to obtain a database $\mathbb{D}_n$ yielding small variance, thus, reduced uncertainty in the kriging prediction.

To estimate the trace-variance (4.21), we use jackknifing, a popular technique of statistics. Jackknifing was originally described in [Quenouille, 1949; Quenouille, 1956; Tukey, 1958]. By now, jackknifing has become classical, see, e.g., the textbooks [Efron and Tibshirani, 1993; Shao and Tu, 1995]. The jackknifing technique is similar to cross-validation (§ B.2.2). The central idea of the method is to define the $\tilde{q}_\mathbf{p}^{(-j)}(t)$ *pseudo-values*:

$$\tilde{q}_\mathbf{p}^{(-j)}(t) = n\hat{q}_\mathbf{p}(t) - (n-1)\hat{q}_\mathbf{p}^{(-j)}(t), \quad j = 1, 2, \ldots, n, \tag{4.22}$$

where $\hat{q}_\mathbf{p}(t)$ is the kriging prediction according to (4.15), and $\hat{q}_\mathbf{p}^{(-j)}(t)$ is the kriging predicition based on a reduced database: the jth sample of (4.2) is ignored, and $\hat{q}_\mathbf{p}^{(-j)}(t)$ is obtained by using the remaining $n-1$ samples, via (4.18) and (4.15). Obviously, the coefficients in the linear prediction using $n-1$ samples cannot

be derived by simply deleting the jth coefficient of the full-sample case, but one has to re-solve the reduced version of (4.18). The pseudo-values make possible a modified prediction of $q_{\mathbf{p}}(t)$:

$$q_{\mathbf{p}}(t) \approx \hat{q}_{\mathbf{p}}^{\text{Jack}}(t) = \frac{1}{n}\sum_{j=1}^{n} \tilde{q}_{\mathbf{p}}^{(-j)}(t). \tag{4.23}$$

We do not directly use this prediction in the following, but let us only mention that due to its attractive statistical properties (not discussed herein), it is often used as a predictor in some applications. Let us focus on the expression of $\hat{q}_{\mathbf{p}}^{\text{Jack}}(t)$: it is the mean of the n pseudo-values. Consequently, the empirical variance of the pseudo-values reflects some knowledge on "how difficult" the prediction is at a certain $\mathbf{p}$ and t, i.e., how much uncertainty can be assigned to the prediction. Formally, the jackknife estimation of the variance is equal to the corrected empirical variance of the mean of the pseudo-values:

$$\hat{\sigma}^2_{\text{Jack}}(\mathbf{p}, t) = \frac{1}{n(n-1)}\sum_{j=1}^{n}\left(\tilde{q}_{\mathbf{p}}^{(-j)}(t) - \hat{q}_{\mathbf{p}}^{\text{Jack}}(t)\right)^2. \tag{4.24}$$

The trace-variance estimation is easily obtained by integrating both sides of the equation (4.24) over the domain T:

$$\hat{\sigma}^2_{\text{T, Jack}}(\mathbf{p}) \equiv \int_T \hat{\sigma}^2_{\text{Jack}}(\mathbf{p}, t)\mathrm{d}t = \frac{1}{n(n-1)}\sum_{j=1}^{n}\int_T \left(\tilde{q}_{\mathbf{p}}^{(-j)}(t) - \frac{1}{n}\sum_{k=1}^{n}\tilde{q}_{\mathbf{p}}^{(-k)}(t)\right)^2 \mathrm{d}t. \tag{4.25}$$

A considerable drawback of jackknifing can be noticed: the related computational load is relatively high and rapidly growing with n, since the variance estimation in a kriging model using n samples requires to perform $n + 1$ kriging predictions (one based on all n samples and further n based on $n - 1$ samples). The computational cost will also be discussed via the numerical examples (§ 4.8).

4.4 Adaptive sampling algorithm

The surrogate model –represented by $\mathcal{G}^{(n)}$– is obviously aimed to be as precise as possible, i.e., a good approximation is required in (4.4). Formally, the interpolation error

$$\varepsilon^{(n)}(\mathbf{p}) = \left\|\mathcal{G}^{(n)}\{\mathbf{p}\} - \mathcal{F}\{\mathbf{p}\}\right\| \tag{4.26}$$

must be as small as possible for a given number of samples n. Though the interpolation error $\varepsilon(\mathbf{p})$ cannot be exactly determined without performing a forward simulation at the input point $\mathbf{p}$, one can have a guess on the quality of the prediction based on the trace-variance (4.25) estimated by jackknifing. One can heuristically assume that high $\hat{\sigma}^2_{\text{T, Jack}}(\mathbf{p})$ indicates poor prediction, i.e., high interpolation error.

We propose a sequential adaptive sampling strategy –inspired by the work presented in [Kleijnen and Beers, 2004]– to build the database (4.2) so that the functional kriging interpolator has small interpolation error. The algorithm inserts new samples one-by-one, as illustrated by the flowchart shown in Fig. 4.1. The main stages can be described as follows.

1. *Initialization:* The algorithm starts by choosing a set of initial input samples (e.g., using a classical space-filling method, such as LHS). Then, the corresponding outputs are computed by evaluating $\mathcal{F}$, i.e., an initial database $\mathbb{D}_n$ is generated.

2. *Sample insertion:* (Recursive definition.) The next input sample should be chosen so that the quality of the interpolation improves as much as possible. The maximal improvement of the kriging interpolation –in the sense of maximal interpolation error– is then obtained if the next sample is chosen as

$$\mathbf{p}_{n+1} = \arg\min_{\mathbf{p}\in\mathbb{P}}\left(\max_{\mathbf{p}'\in\mathbb{P}}\varepsilon^{(n+1)}(\mathbf{p}')\right). \tag{4.27}$$

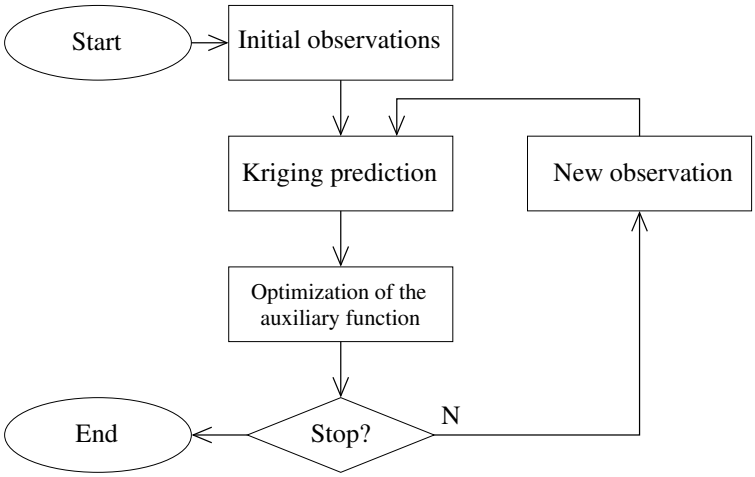


Figure 4.1: Flowchart of the adaptive sampling strategy for the generation of functional kriging surrogate models.

This optimization problem is obviously hard to solve. Instead of the above implicit definition of $\mathbf{p}_{n+1}$, one rather needs an explicit approximation of the best choice for the next sample. Based on the jackknife estimation of the trace-variance, a heuristic approximation is available. A natural idea could be to maximize $\hat{\sigma}^2_{\mathrm{T,\,Jack}}(\mathbf{p})$, i.e.,

$$\mathbf{p}_{n+1} = \arg\max_{\mathbf{p}\in\mathbb{P}} \hat{\sigma}^2_{\mathrm{T,\,Jack}}(\mathbf{p}). \tag{4.28}$$

However, as it will be illustrated in § 4.7, this approach would fail, since $\hat{\sigma}^2_{\mathrm{T,\,Jack}}(\mathbf{p})$ is not necessarily zero at the sampled locations $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ but can be rather high around these points. Consequently, by using (4.28), the input samples might cluster in some regions of $\mathbb{P}$, which is to be avoided.

We propose to use the following heuristic auxiliary function:

$$Q^{(n)}(\mathbf{p}) = \left(\min_{i=1,2,\ldots,n} \|\mathbf{p} - \mathbf{p}_i\| \right) \times \hat{\sigma}^2_{\mathrm{T,\,Jack}}(\mathbf{p}). \tag{4.29}$$

The first factor of $Q^{(n)}(\mathbf{p})$ is the distance between $\mathbf{p}$ and the nearest input sample, whereas the second factor is related to the estimated uncertainty at $\mathbf{p}$. Then the criterion of adding the next sample is

$$\mathbf{p}_{n+1} = \arg\max_{\mathbf{p}\in\mathbb{P}} Q^{(n)}(\mathbf{p}). \tag{4.30}$$

In so doing, one can intuitively believe that the strategy balances between the classical input-space filling (represented by the first factor of $Q^{(n)}(\mathbf{p})$) and the adaptive, variance-driven sampling (via the second factor of $Q^{(n)}(\mathbf{p})$), respectively[2].

Once $\mathbf{p}_{n+1}$ is determined by solving the optimization problem (4.30), the corresponding output is computed:

$$q_{\mathbf{p}_{n+1}}(t) = \mathcal{F}\{\mathbf{p}_{n+1}\}. \tag{4.31}$$

The new sample is appended to $\mathbb{D}_n$, and n is increased by 1.

[2]Other goals of the sampling can also be addressed in a similar way (i.e., via the optimization of an appropriate auxiliary function). For instance, in [Picheny et al., 2010], a sequential algorithm (similar in spirit to ours) is presented. Therein the goal is to give good approximations by the surrogate model *around a certain output level*. They consider the prediction of scalar functions (instead of functional data), thus, a scalar target value can be specified, around which one requires small interpolation errors. Roughly, their approach is also a heuristic combination of a factor being related to the "target detection", and an other for the estimation of the interpolation error (they use the classical kriging variance for this purpose).

3. *Stopping criterion:* The sample insertion goes on until a stopping criterion is met, then the final database is ready, the algorithm is terminated. The stopping rule might be related to the number of samples, the maximal estimated trace-variance, etc.

The above algorithm is believed to yield a database $\mathbb{D}_n$ which is adapted both to the applied functional kriging interpolator and to the forward operator $\mathcal{F}$ under study. Consequently, one expects to obtain a precise surrogate model by fitting a functional kriging interpolator to the samples of $\mathbb{D}_n$.

We have to highlight that the idea of using jackknifing to improve the performance of a kriging-based surrogate model is ascribed to the authors of [Kleijnen and Beers, 2004]. However, extending the approach to functional kriging, formalizing the sampling method as an optimization problem (4.30) and the definition of the modified objective function (4.29) are our own contributions.

4.5 Some further comments on the variance estimation

The traditional kriging framework (§ B) provides a way to estimate the variance of the prediction error (see (B.12)). This formula plays a central role in certain applications, e.g., in the EGO algorithm (§ 3), and it can be extended to the functional kriging case as well, by using the trace-covariance. The estimated trace-variance is

$$\hat{\sigma}_T^2(\mathbf{p}) = k_T(\mathbf{p}, \mathbf{p}) - \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}} \mathbf{k}_T(\mathbf{p}) - \mu(\mathbf{p}), \tag{4.32}$$

where $\mu(\mathbf{p})$ is the Lagrange-multiplier from (4.18) (the derivation can be found, e.g., in [Giraldo, Delicado, and Mateu, 2007]).

Though the classical form (B.12) is widely used, suggesting also the use of (4.32) in the functional case, its validity is questionable. The objection of several mathematicians (see, e.g., [Hertog, Kleijnen, and Siem, 2006; Kleijnen and Beers, 2004]) is that the coefficients $\boldsymbol{\lambda}(\mathbf{p})$ in (4.11) and (4.15) *do* depend on the observed data since the trace-covariance is predicted using the observed data. Consequently, the prediction (4.15) is in fact not linear, however, linearity is assumed when deriving the formulae (4.32) and (B.12). These formulae are supposed to underestimate the variance. The main motivation of the authors of [Kleijnen and Beers, 2004] to use jackknifing in an adaptive sampling algorithm is basically the aim of correcting the classical variance formula.

Beyond this theoretical doubt, an intuitively explainable drawback also arises concerning (4.32). Let us notice that once the trace-covariance is chosen (e.g., by Cross Validation (CV) estimation), $\hat{\sigma}_T^2(\mathbf{p})$ according to (4.32) depends only on the repartition of the input samples $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, whatever the currently observed output data are. Consequently, an adaptive sampling algorithm cannot be based on the formula (4.32). On the contrary, the jackknife estimation of the trace-variance is computed *directly* from the observed data. According to the brief classification of DoE tools in § 1.2.2, a sampling method based on the classical formula (4.32) would be counted as a *model-oriented* strategy, such as the work in [Brus and Heuvelink, 2007], whereas jackknifing makes possible the construction of *adaptive* schemes.

We have to emphasize again the heuristic nature of the presented jackknifing-based sampling method. Both the classical kriging theory (including its variance formula) and the jackknifing technique are well established on a solid theoretical basis. However, there is no more theoretical background when combining these tools in an adaptive sampling algorithm as presented. For instance, one of the main points where one might have doubts is the fact that the jackknife estimation of the trace-variance at the sampled input points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ is not necessarily zero (as it is illustrated in § 4.7). On the other hand, however, kriging interpolates the observed data, i.e., zero variance should be expected at the sampled input points.

4.6 Practical considerations

4.6.1 Corner points of the input space

As a rule, kriging shows a bad performance when it is used for extrapolation, i.e., in the cases where prediction is performed at an input point being outside the convex hull of the already observed input points [Kleijnen and Beers, 2004; van Beers, 2005; Walter and Pronzato, 1997]. This pitfall should be taken into account when constructing a kriging-based surrogate model. To this end, the initial sample set we choose always includes the corner points of the input space (let us note that in all of our examples, $\mathbb{P} \subset [0,1]^N$), i.e., the 2^N vertices of the unit-hypercube. Furthermore, the jackknife estimation of the trace-variance is computed by sequentially leaving out only the observations not located at the corner points. Thus, assuming that n samples are included in the database, jackknifing leaves out $n-2^N$ samples. In so doing, the estimated trace-variance at the corner points is exactly zero.

4.6.2 Optimization of the auxiliary function

A key point of the presented adaptive sampling scheme is the solution of the optimization problem (4.30). The complexity of this task is twofold: the auxiliary function $Q^{(n)}(\mathbf{p})$ is probably multimodal (in the "grid-sense" defined in § C.2, as $Q^{(n)}(\mathbf{p}_i) \equiv 0$ at the sampled points) and expensive-to-evaluate (since jackknifing is a costly tool). Consequently, an effective global optimization method is needed to solve (4.30).

Only in the simplest cases (like the one-dimensional toy example in § 4.7) can one consider the method of exhaustive search (i.e., evaluating $Q^{(n)}(\mathbf{p})$ all over $\mathbb{P}$) to find the global maximum of the auxiliary function. The advantage of this approach is to ensure that the global maximum is found on a pre-defined grid. If the grid is appropriately fine, the found maximizer approximates well the true maximizer. A drawback of the exhaustive search is the extremely high computational cost which rapidly increases with the number of input parameters.

For more complicated cases (like the numerical examples presented in § 4.8) we propose the use of a variant of CWS algorithm on a pre-defined grid in the input space. If the input space was continuous, the definition of the search grid could reflect some prior knowledge of the modeled forward problem (i.e., finer discretization along the parameter axes along which $\mathcal{F}$ is assumed to vary more rapidly). In our examples, however, the input spaces are *ab ovo* discrete grids, defined according to the constraints arisen by the simulator at hand (see § 2.4).

The applied CWS algorithm is presented in § C.3. Since CWS is a local optimization tool, we either use it in a multistart scheme or combine it with a preliminary exhaustive search to solve (4.30). Thus, we define two implementations for the adaptive sampling, referred as I and II:

I: The starting points of CWS in the multistart scheme are all n input samples. The search is sequentially started from all starting points. The CWS can thus converge to different local maxima, the final solution is the highest among these local maxima.

II: An LHS with a maximin criterion (§ D.2) is generated in the input space[3]. The auxiliary function is then evaluated at all points of the LHS design, i.e., a sort of reduced exhaustive search is performed in $\mathbb{P}$. The CWS algorithm is started only once from the point of the LH design where the maximal $Q^{(n)}(\mathbf{p})$ was found by the exhaustive search.

By using these heuristic optimization methods, the true global maximum of $Q^{(n)}(\mathbf{p})$ might be missed, however, the computational cost of the optimization becomes tractable.

[3]This LHS is derived from a continuous LHS (§ D.3).

4.7 An illustrative analytical example

To illustrate the presented surrogate modeling approach, a polynomial function $f(p)$ is chosen[4], defined on the $p \in [0, 1]$ interval:

$$f(p) = -579p^4 + 1110p^3 - 684.5p^2 + 141.071p + 17. \tag{4.33}$$

This function is considered as a simple input-output function (i.e., an explicit form of a forward operator). The input and output spaces are both one-dimensional, thus, not the functional kriging but the traditional approach (§ B) is involved, however, the variance of the interpolation error is estimated by using jackknifing.

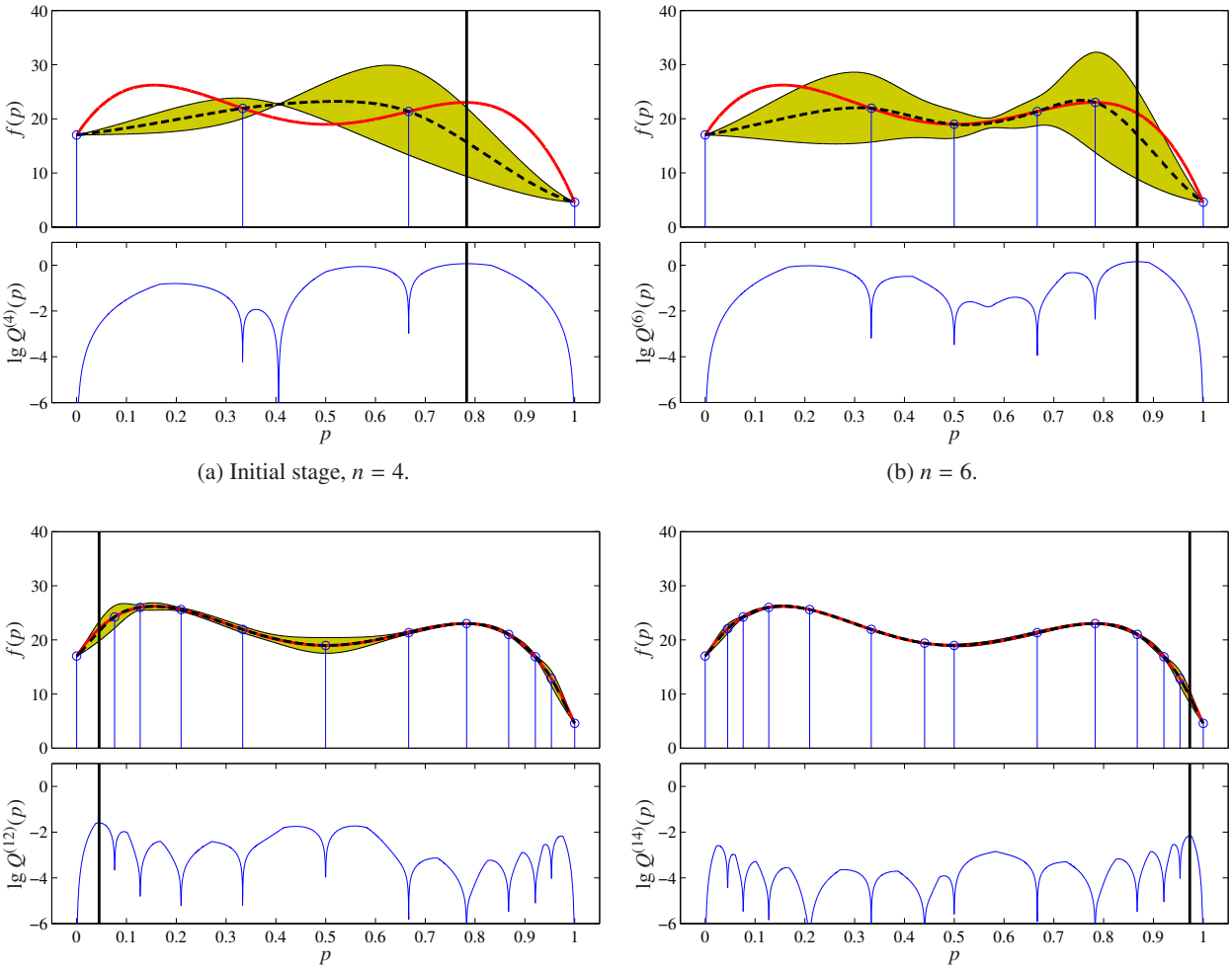
The Matérn function is used to approximate the covariance of the modeling process and the hyperparameters are estimated via cross-validation (see § B.2 and § B.2.2 in particular).

The performance of the algorithm is presented in Fig. 4.2. Prior to starting the adaptive sampling loop, 4 initial observations are made (the input samples are equally spaced in the interval $[0, 1]$). The optimization problem (4.30) is solved in each cycle by performing an exhaustive search on a fine grid consisting of 5000 equally spaced nodes.

The stopping criterion was simply an upper bound on n: totally 14 samples are included in the final database. As conclusions of this toy example, let us notice the followings:

- The objective function $Q^{(n)}(p)$ is multimodal (making its optimization challenging).
- The estimated variance appears to have a decreasing trend during the iterations, however, the decrease is not monotonous.
- The estimated variance is not necessarily zero at the sampled locations. Furthermore, $\hat{\sigma}^2_{\text{Jack}}(p)$ appears to be maximal around an already sampled location when $n = 6$ (at $p \approx 0.78$ in Fig. 4.2b). This justifies why one should avoid the form (4.28) and use (4.29) as objective function instead.
- The estimated variance appears to indeed be in a direct relation with the observations. For instance, in Fig. 4.2a, one can see an asymmetric profile of $\hat{\sigma}^2_{\text{Jack}}(p)$ in spite of the equal spacing of the input samples in the interval $[0, 1]$. If the traditional variance formula (B.12) of kriging was used, $\hat{\sigma}^2(p) = \hat{\sigma}^2(1 - p)$ would be found in the initial, $n = 4$ case.
- Since the samples at the bounds of the input space ($p = 0$ and $p = 1$) are not left out when computing the jackknife variance estimation, $\hat{\sigma}^2_{\text{Jack}}(p) \equiv 0$ holds at these points.
- The repartition of the input samples in the final stage (Fig. 4.2d) is not equidistant but adapted to the function $f(p)$ to be modeled. More samples seem to be inserted to the regions where the $|\mathrm{d}f/\mathrm{d}p|$ is high, however, there is no direct relation between the sampling and the derivative of $f(p)$.

[4]Inspired by the test function used in [Kleijnen and Beers, 2004].



(a) Initial stage, $n = 4$.

(b) $n = 6$.

(c) $n = 12$.

(d) Final stage, $n = 14$.

Figure 4.2: Performance of the adaptive sampling in the case of an analytical toy function to be modeled. Only a couple of stages of the sequential sample insertion are plotted. Top figures: solid red line: true function $f(p)$; dashed black line: kriging prediction $\hat{f}(p)$; stems: samples; filled area: $\hat{f}(p) \pm 2\sqrt{\sigma^2_{\text{Jack}}(p)}$. Bottom figures: logarithm to base 10 of the criterion function $Q^{(n)}(p)$. Black sticks: current maximum of the criterion function (where the next sample is added).

4.8 Numerical examples

4.8.1 Generally about the examples

The proposed surrogate modeling method is tested via the the numerical examples presented in this section. The two canonical configurations of § 2.4 are involved in the tests, using 2, 4 and 6 input parameters, respectively.

To initialize the sampling algorithms, $2^N + 1$ initial observations are performed (N is the number of input parameters): the 2^N corner points of the input space $\mathbb{P}$ (i.e., the input vectors given by the coordinates $\{0, 1\}^N$) and the midpoint of the input space (i.e., the input vector in $\mathbb{P}$ being the closest to the point $[0.5, 0.5, \ldots, 0.5]$).

Let us recall that the Matérn function is used to approximate the trace-covariance of the modeling process and the hyperparameters are estimated via cross-validation (see § B.2 and § B.2.2). According to our experience (based on some further numerical studies not presented here), the estimated hyperparameters do not vary significantly when the number of samples increases, moreover, the ρ_k ranges of the different input parameters are all approximately equal to 0.5. Thus, in the presented numerical examples, the trace-covariance is actually not estimated but the following fixed hyperparameters are used:

$$\nu = 1.5N,\ \rho_k = 0.5,\ \sigma_0^2 = 1. \tag{4.34}$$

Let us note that neither the optimal coefficients $\boldsymbol{\lambda}(\mathbf{p})$ of the predictor (4.15) nor the objective function of the cross-validation (B.29) depend on the hyperparameter σ_0^2, thus, the latter is arbitrarily set to 1. The choice of ν and ρ_k is according to the usual "rule of thumbs". However, if this surrogate modeling method is applied to other problems, one must ensure the correct choice of the trace-covariance hyperparameters, e.g., by performing a CV estimation in each cycle of the sampling.

In the first cycle, only one sample (the midpoint) could be removed when estimating the trace-variance by jackknifing (see § 4.6.1). However, in this case jackknifing is not defined due to the division by $n - 1$, actually being zero, thus, in the very first cycle not the objective function (4.29) is considered, but only its first factor (the distance of the nearest input sample). In the subsequent cycles, the algorithm is as discussed in § 4.4.

The stopping criterion of the sequential sampling is simply an upper bound on the n number of samples. The proposed surrogate models are compared to other surrogate models generated by different approaches:

- Functional kriging interpolation based on a sample-set obtained by a full-factorial design (see § D.1);
- Multilinear interpolation (§ D.4) based on a sample-set obtained by a full-factorial design.

The upper limit of n in the adaptive sampling strategies is appropriately chosen so that it fits the sample number of the corresponding full-factorial design to be compared with. The full-factorial designs have approximately equispaced levels[5] (not necessarily with the same number of levels along the different input parameters) and the lower and upper bounds (0 and 1) of each input parameter are always represented by a level, respectively.

The surrogate model is required to approximate the forward operator with a small discrepancy, formally, with a small interpolation error $\varepsilon(\mathbf{p})$ as defined in (4.26). To characterize the overall quality of the surrogate model, one can consider (among others) the root mean square interpolation error $\varepsilon_{\mathrm{RMS}}$ or the maximal interpolation error $\varepsilon_{\max}$. Both quantities are traditionally used to characterize predictors, interpolators, etc. In our examples, the input space $\mathbb{P}$ is a finite set, thus, the root mean square error is expressed as

$$\varepsilon_{\mathrm{RMS}} = \sqrt{\frac{1}{|\mathbb{P}|} \sum_{\mathbf{p} \in \mathbb{P}} \varepsilon^2(\mathbf{p})}, \tag{4.35}$$

[5] As far as it is possible in the discretized input space. See § D.3.

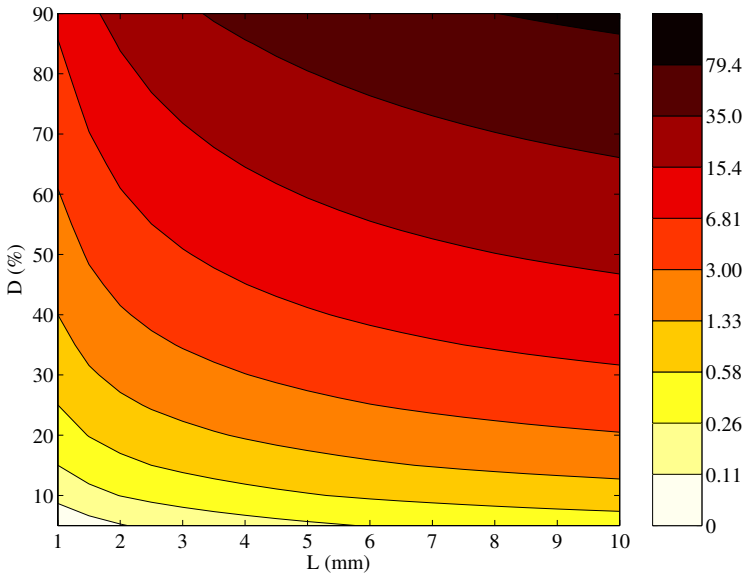


Figure 4.3: Norm of the impedance signal ($\left\|q_{\mathbf{p}}(t)\right\|$ in mΩ) in function of the crack parameters in the first canonical configuration (point 1 in section § 2.4). Let us note that a logarithmic colorscale is applied since the plotted values vary in a large scale.

whereas the maximal error is simply

$$\varepsilon_{\max} = \max_{\mathbf{p}\in\mathbb{P}} \varepsilon(\mathbf{p}). \tag{4.36}$$

Let us note that to compute these "true" errors of the final surrogate model, one usually has to face intractable computational load, since the forward operator must be evaluated at each point of $\mathbb{P}$ to obtain the true output data. This is why in the 4 and 6-parameter cases, we only approximate these errors as

$$\varepsilon_{\mathrm{RMS}} = \sqrt{\frac{1}{|\mathbb{S}|}\sum_{\mathbf{p}\in\mathbb{S}} \varepsilon^2(\mathbf{p})}, \quad \text{and} \quad \varepsilon_{\max} = \max_{\mathbf{p}\in\mathbb{S}} \varepsilon(\mathbf{p}), \tag{4.37}$$

where $\mathbb{S}$ is a subset of $\mathbb{P}$ with a much smaller cardinality. We choose $\mathbb{S}$ as an approximate Latin hypercube design with maximin criterion (see § D.2 and § D.3).

The interpolation errors are not normalized but given in mΩ. The comparison of the different surrogate models is possible in so doing, however, one might need an idea how the interpolation error is related to the norm of the true output data, i.e., how much the relative error is. To this end, we have plotted the norm of the output data (the impedance signal) in function of the crack sizes in the case of the single-crack configuration (point 1 in § 2.4) in Fig. 4.3.

To optimize the sample-insertion criterion function, i.e., to solve (4.30), the two approaches discussed in § 4.6.2 are applied. The scheme I is applied for the 2 and 4-parameter examples, whereas the approach II is involved in also the 4-parameter and in the 6-parameter examples.

For the computations, a PC with a 64-bit CPU with 4 cores at 2.33 GHz and 16 GB memory has been used. All algorithms are implemented in MATLAB®.

4.8.2 Examples with 2 parameters

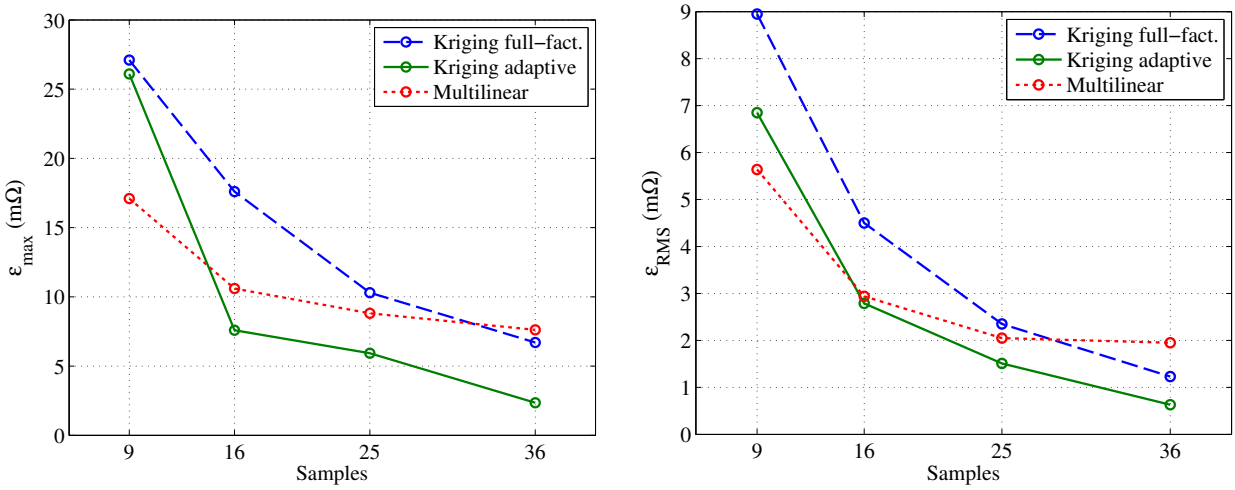
These examples involve the canonical configuration defined as the point 1 in § 2.4: the length (L) and the depth (D) of a single crack are enabled to vary, all other parameters are assumed to be fixed. The input space

consists of 342 feasible input vectors. The adaptive surrogate model is compared to a functional kriging and a multilinear interpolator, both based on a full-factorial design including the same number of samples. The maximal and RMS interpolation errors are computed as (4.35) and (4.36), i.e., by evaluating $\varepsilon(\mathbf{p})$ at all 19×18 points of the input space.

The numerical results are summarized in Table 4.1 and plotted in Fig. 4.4. The sample number n is given as a product of the number of the levels of the related full-factorial design. One can see that both the maximal and the RMS errors monotonously decrease when the number of samples increases. The kriging interpolator gives better approximations if the samples are chosen adaptively (by implementation I) than by using a full factorial design. The proposed adaptive kriging approach also outperforms the multilinear interpolation except for the $n = 3 \times 3$ case.

Table 4.1: Maximal and RMS interpolation error in the 2-parameter example (mΩ).

Samples (n)	Kriging				Multilinear	
	Full-factorial		Adaptive I.		Full-factorial	
	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$
3×3	27.1	8.95	26.1	6.85	17.1	5.64
4×4	17.6	4.50	7.59	2.79	10.6	2.94
5×5	10.3	2.35	5.93	1.51	8.81	2.05
6×6	6.70	1.23	2.35	0.63	7.62	1.95



(a) Maximal error.

(b) RMS error.

Figure 4.4: Maximal and RMS interpolation error in function of the number of samples in the 2-parameter example.

Since only two input parameters are involved in this numerical study, the distribution of the interpolation error can easily be visualized in a two-dimensional colormap. The kriging interpolators based on the full-factorial and the adaptive sampling are compared in Fig. 4.5. Let us notice that the interpolation error is concentrated mainly in the large-crack region of the input space in case of full-factorial sampling. The adaptive sampling can reduce the interpolation error by inserting more samples in this region. The inhomogeneous repartition of the adaptively added samples refers to the properties of the forward operator to be modeled at hand. It is also worth noting that the levels of the full-factorial designs are not exactly equispaced due to the discrete input space (§ D.3).

The most precise interpolator (in the sense of RMS error) is achieved by the adaptive kriging approach,

with $n = 36$ samples. The corresponding $\varepsilon_{\mathrm{RMS}}$ is approximately equal to the norm of the impedance signal of a single (centered) crack with length $L \approx 1$ mm and depth $D \approx 25\,\%$ according to Fig. 4.3.

4.8.3 Examples with 4 parameters

The canonical configuration defined as the point 2 in § 2.4 is involved in the presented examples with 4 parameters. The position (A, B), the length (L) and the depth (D) of a single crack are enabled to vary. The input space consists of approximately 4×10^4 feasible input vectors. Two adaptive surrogate models are generated (by using implementations I and II, respectively). Their performance is compared to a functional kriging and a multilinear interpolator, both are based on a full-factorial design including the same number of samples. The maximal and RMS interpolation errors are approximately computed by using the formulae (4.37) where the test sample-set $\mathbb{S}$ is chosen according to a random Latin hypercube design using maximin criterion, including 840 samples[6]. A similar random LH design with 256 points are used (and re-generated in each cycle) for the exhaustive search in implementation II of the adaptive sampling.

The results are given numerically in Table 4.2 and plotted in Fig. 4.6. Both adaptive kriging approaches outperform the kriging method based on a full-factorial design (except in the case of the smallest database in the sense of $\varepsilon_{\max}$). The adaptive implementation II performs similarly to implementation I, except in the case of 3^4 samples. This phenomenon can be explained by the fact that the CWS optimization of the criterion function might be more successful by using implementation II in the case of small sample number (since implementation II always performs first an exhaustive search using 256 points). As soon as the number of samples gets comparable to the 256 search points of implementation II, the multistart scheme (impelementation I) appears to give similar results. A possible consequence might be that implementation I would outperform II if n was much higher than 256. This case is not examined, since the time consumption of the multistart scheme is rapidly growing with n (§ 4.8.5).

The multilinear interpolation performs more-or-less similarly to the kriging interpolator being based on the same full-factorial design, the RMS errors are almost equal in the case of $n = 144$ and $n = 256$.

The most precise interpolators among all examined ones in the sense of RMS error are the two adaptive kriging approaches with $n = 256$ samples. They provide an $\varepsilon_{\mathrm{RMS}}$ being approximately equal to the norm of the impedance signal corresponding to a single (centered) crack with length $L \approx 5$ mm and depth $D \approx 20, \ldots, 30\,\%$ according to Fig. 4.3.

Table 4.2: Maximal and RMS interpolation error in the 4-parameter example (mΩ).

Samples (n)	Kriging						Multilinear	
	Full-factorial		Adaptive I.		Adaptive II.		Full-factorial	
	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$
$3 \times 3 \times 3 \times 3$	30.8	9.36	33.6	8.47	18.2	5.95	29.2	6.35
$3 \times 3 \times 4 \times 4$	21.7	5.95	11.6	3.83	11.6	3.71	28.9	5.89
$4 \times 4 \times 4 \times 4$	17.6	4.59	7.66	2.36	9.69	2.36	17.5	4.00

4.8.4 Examples with 6 parameters

The 6-parameter examples are based on the double-crack canonical configuration defined as the point 3 in § 2.4. Let us recall that the positions (A, h, B_1 and B_2) and the lengths (L_1, L_2) of two parallel cracks are enabled to vary, whereas the depth of both cracks is assumed to be 40 % of the plate thickness. The input space consists of approximately 4.39×10^6 feasible input vectors.

[6] Derived from a continuous LHS with $50 \times (2^N + 1)$ points (§ D.3).

An adaptive surrogate model is generated (using implementation II) and compared to a functional kriging and a multilinear interpolator, both based on a full-factorial design including the same number of samples. The maximal and RMS interpolation errors are approximately computed by using the formulae (4.37) where the test sample-set $\mathbb{S}$ is chosen according to a random Latin hypercube design using maximin criterion, including 3249 samples[7]. A similar random LH design with 729 points is used (and re-generated in each cycle) for the exhaustive search in the adaptive sampling.

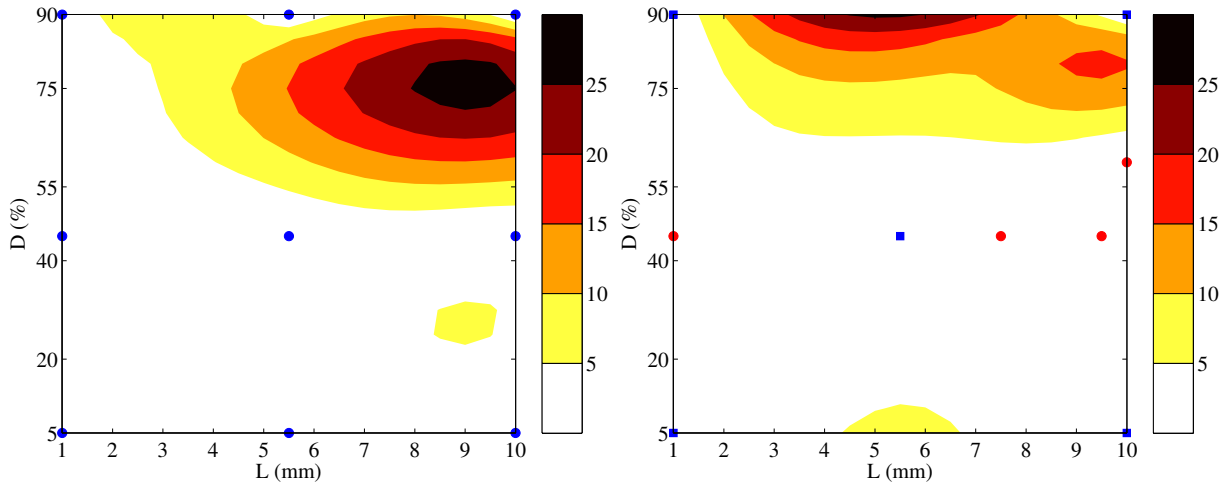
The numerical results are given in Table 4.3. The maximal number of samples in the adaptive metamodel is $n = 729$ (the adaptive approach is in fact limited by the rapidly growing computation time, see § 4.8.5). At a higher number of samples ($n = 3600$) the kriging model based on a full-factorial design is compared to a piecewise multilinear interpolator. Surprisingly, the multilinear interpolation outperforms even the adaptive kriging approach in the sense of both $\varepsilon_{\max}$ and $\varepsilon_{\mathrm{RMS}}$ which is not experienced in the cases with 2 and 4 parameters. Let us also note that the $\varepsilon_{\max}$ corresponding to the multilinear interpolation is higher when n is higher (though this is strange, obviously not impossible).

In spite of the fact that the adaptive surrogate modeling strategy is outperformed by the classical full-factorial design combined with the multilinear interpolation, the 6-parameter example has positive conclusions as well. Let us compare the adaptive metamodels based on 3^N samples in all examples ($N = 2, 4, 6$). This comparison is "fair" in the sense that the number of samples divided by the cardinality of the input space is approximately the same in all cases. One can see that the precision of the 6-parameter surrogate model is finer than either the 2-parameter or the 4-parameter surrogate models in the sense of both $\varepsilon_{\max}$ and $\varepsilon_{\mathrm{RMS}}$. Consequently, a basically precise adaptive surrogate model has been obtained in the presented 6-parameter case, and has been outperformed by an even better multilinear interpolator. Let us refer also to the experiences of the output space-filling sampling strategy in this 6-parameter example (§ 5.5.4). The conclusions therein appear to confirm that the studied forward operator is close to linear, thus, the good performance of the piecewise multilinear interpolation here is understandable.

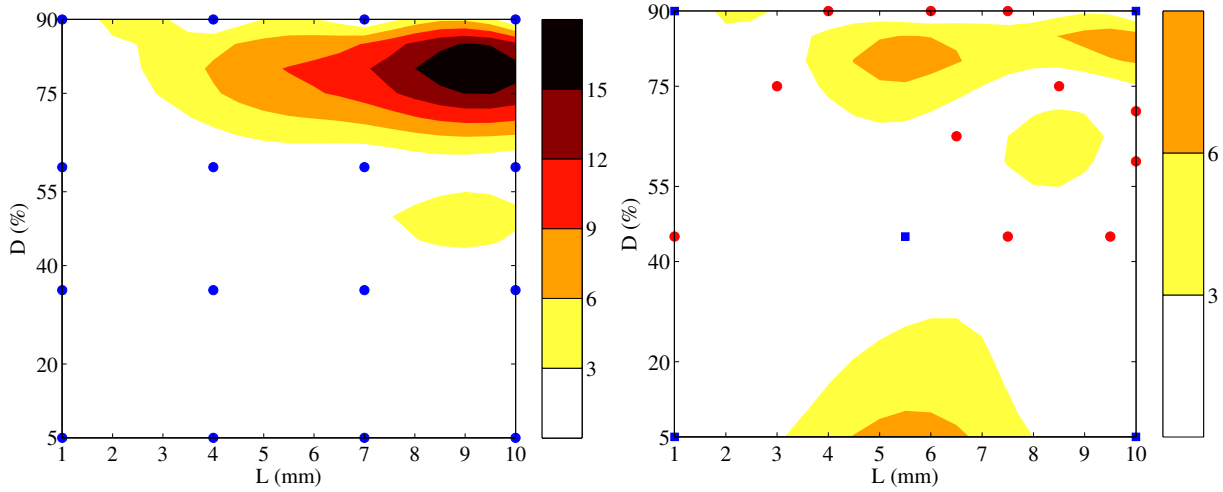
Table 4.3: Maximal and RMS interpolation error in the 6-parameter example (mΩ).

Samples (n)	Kriging				Multilinear	
	Full-factorial		Adaptive II.		Full-factorial	
	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$	$\varepsilon_{\max}$	$\varepsilon_{\mathrm{RMS}}$
$3 \times 3 \times 3 \times 3 \times 3 \times 3$	7.90	3.67	5.43	3.06	3.98	2.19
$3 \times 3 \times 4 \times 4 \times 5 \times 5$	6.02	3.04	n/a	n/a	4.25	2.06

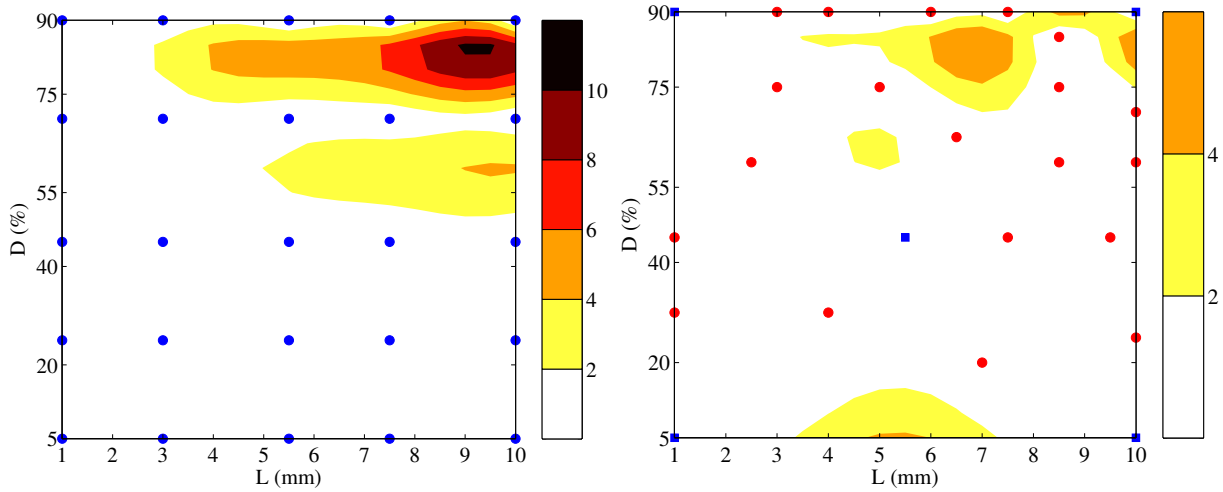
[7] Derived from a continuous LHS with $50 \times (2^N + 1)$ points (§ D.3).



(a) $n = 9$.

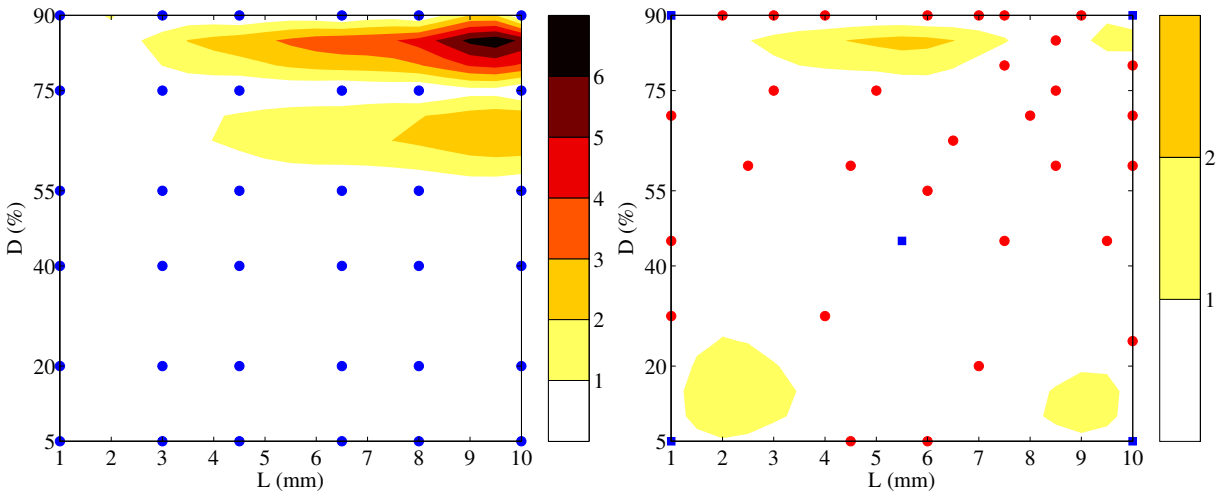


(b) $n = 16$.



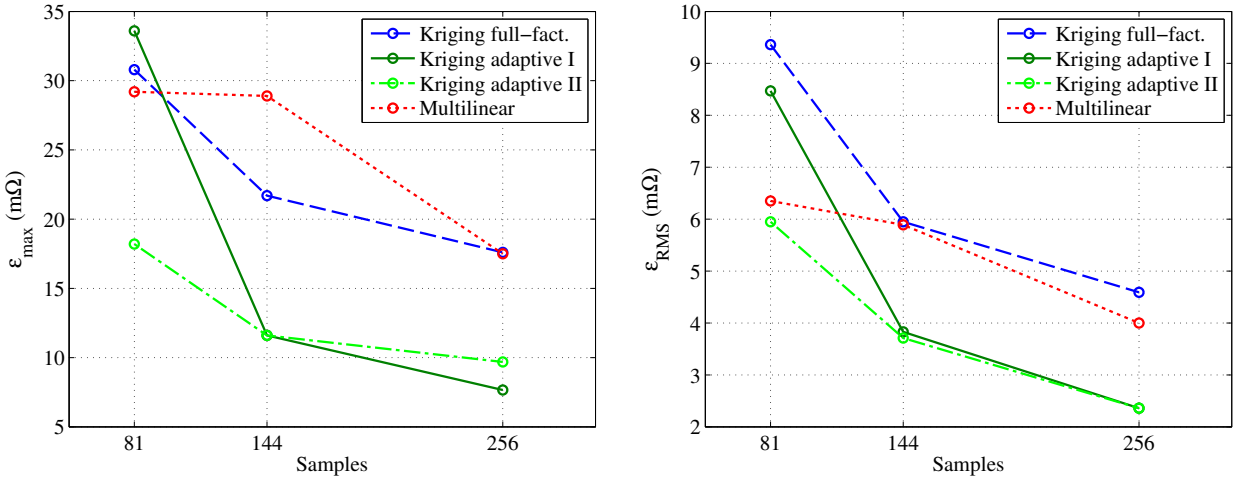
(c) $n = 25$.

Figure 4.5: Interpolation error $\varepsilon(\mathbf{p})$ of the kriging interpolators over the input space in the 2-parameter example with different number of samples (colormap in mΩ). Left: full-factorial sampling. Right: adaptive sampling. The $2^2 + 1$ initial samples are marked by blue squares, whereas the adaptive sampling has inserted the observation points marked by red dots. (To be continued on the next page.)



(d) $n = 36$.

Figure 4.5: (Continued.) Interpolation error $\varepsilon(\mathbf{p})$ of the kriging interpolators over the input space in the 2-parameter example with different number of samples (colormap in mΩ). Left: full-factorial sampling. Right: adaptive sampling.



(a) Maximal error.

(b) RMS error.

Figure 4.6: Maximal and RMS interpolation error in function of the number of samples in the 4-parameter example.

4.8.5 Discussion on the time consumption

Though the computational burden and the time consumption of the database generation (i.e., the adaptive sampling algorithm) is not a central question –since the database needs to be generated only once– we present some details on the time consumption experienced in the studied numerical examples.

In § 4.6.2, two implementations of the adaptive sampling are discussed. The first approach (I) uses a multistart CWS to optimize the objective function, the search is started from all n samples already included in the database. This scheme is applied to the 2- and 4-parameter examples. The time consumption of the approach is shown in Fig. 4.7. In the 2-parameter case (Fig. 4.7a), the time consumption is only a few seconds per cycle, but the increasing tendency (the sampling algorithm slows down with increasing n) can clearly be seen. In the 4-parameter case, however, a much more challenging behaviour is experienced (Fig. 4.7b). The time-consumption appears to grow fast and reaches almost 2 hours per iteration by the end of the sampling ($n = 256$). A slight oscillation around a polynom-like curve is seen in the plot. The oscillations are due to the varying number of steps of the CWS runs till the convergence.

The growth of the runtime of the CWS when the number of input parameters N increases (2, 4, 6) is easily explainable by the facts that

- the number of points in the input space increases exponentially, thus, the search presumably has to roam longer paths in $\mathbb{P}$ until converging to a local optimum, and
- the number of neighbours –where the objective function is to be evaluated in each step of the search– of a point in $\mathbb{P}$ is generally[8] $2N$.

For these reasons, the runtime of the CWS depends exponentially on N. For a fixed N, the time-consumption (in the multistart scheme) is increasing with n due to the following factors:

- the search is sequentially started from all of the n samples;
- to compute the jackknife estimation of the trace-variance, $n - 2^N$ kriging predictions are needed;
- all of these predictions are based on $n - 1$ samples.

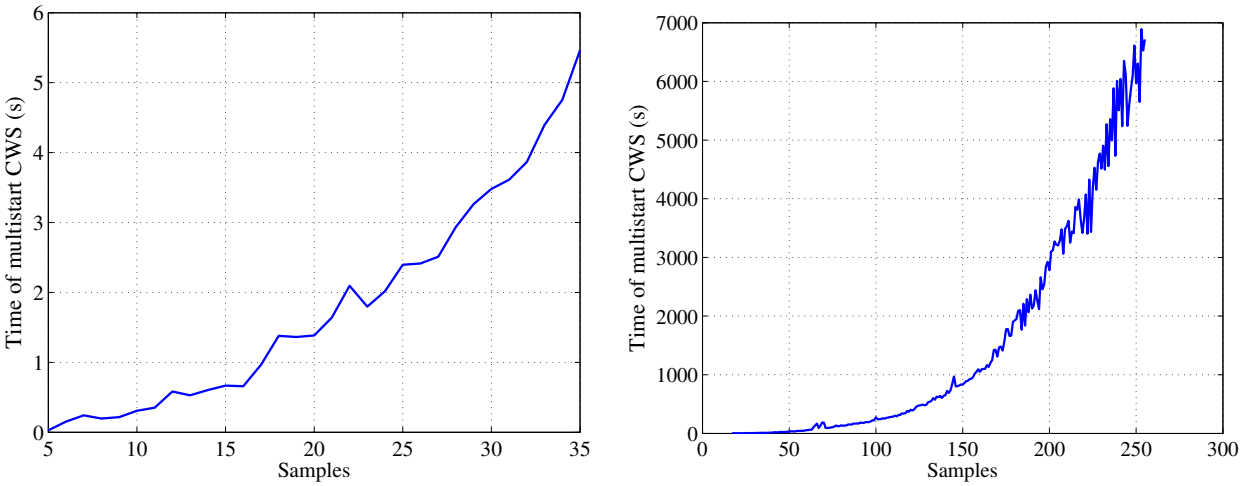
Consequently, the runtime increases with n in a polynomial order.

The second implementation of the adaptive sampling (II) is inspired by the fast growth of the CWS runtime. By decomposing the optimization problem into a preliminary exhaustive search and a single run of CWS, the degree of the polynomial dependence of the time on n can be reduced. Implementation II is applied to the 4- and 6-parameter cases. The time consumption of the exhaustive search is plotted in Fig. 4.8. Let us notice that though the dependence on n is still polynomial, it is of much lower degree than in the case of multistart CWS. The runtime of the subsequent single-start CWS is plotted in Fig. 4.9. In comparison with the multistart scheme in the case of 4 parameters (Fig. 4.7b), one can see that the "mean" of the oscillating curve appears to increase much slower than by using implementation I. The number of steps within a local optimum is found by CWS is random[9], which explains the oscillation. The total time consumption of the sampling method (per iteration) is simply the sum of the runtime of the exhaustive and the coordinatewise search. In the 4-parameter case, one can compare implementations I and II. The scheme II is clearly much faster (sum of the times in Fig. 4.8a and Fig. 4.9a) than the scheme I (Fig. 4.7b).

Due to the enormous computational cost, implementation I was not applied to the 6-parameter case, moreover, even the scheme II is used to generate a database only with a small number of samples (3^6).

[8] Not considering the points on the boundary of the input space.

[9] Obviously, CWS is deterministic, however, as long as the objective function is not known, the length of the roamed path from the starting point to a local maximum is unknown.



(a) 2-parameter case.

(b) 4-parameter case.

Figure 4.7: Time consumption of the multistart CWS (adaptive I approach) per iteration in function of the n number of samples.

Examples with more input parameters (e.g., the 8-parameter canonical configuration, § 2.4) are not considered for the same reason. The computational cost of the presented adaptive surrogate modeling approach is indeed a bottleneck. However, as long as the forward operator to be modeled is much more expensive-to-evaluate than the sampling-related computations (such situations can easily be imagined), our approach is still of good interest.

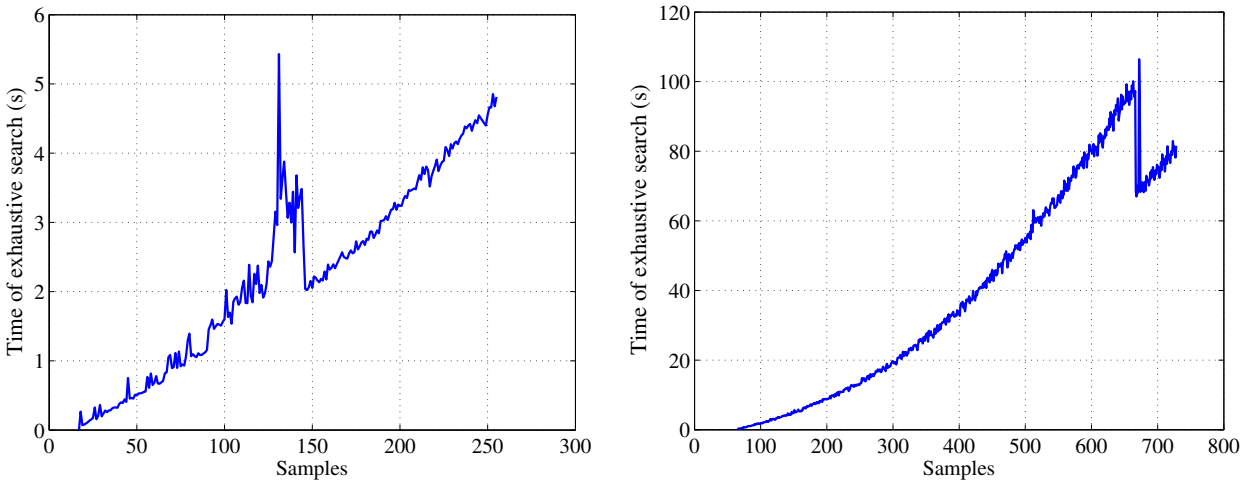
Once a database is generated, the interpolation by functional kriging is not time consuming any more. We assume that the hyperparameters of the trace-covariance model are provided along with the sample set. Then, one just has to use the prediction formula (4.15), and the coefficients $\boldsymbol{\lambda}(\mathbf{p})$ are determined by solving the equation system (4.18). First the right-side has to be assembled by evaluating the trace-covariance function at a given $\mathbf{p}$. If the system matrix is inverted or decomposed in advance, the kriging prediction can further be speeded-up. According to our experience, as long as n is of several hundreds, the kriging interpolation can be performed within the order of some milliseconds.

4.9 Conclusions and perspectives

The presented functional kriging framework can effectively be used to build interpolation-based surrogate models. Though in some of the studied numerical examples the traditional multilinear interpolation shows better performance, functional kriging still yields good approximations. The adaptive sampling strategy –aiming at improving the surrogate model– indeed results in functional kriging interpolators of higher precision than by using full-factorial designs.

Let us recapitulate the main advantages of the proposed approach.

- Due to the sampling strategy which directly accounts for the forward operator to be modeled, the surrogate model is adaptive, i.e., it is fitted to the physical problem at hand. Thus, the approach belongs to the family of adaptive DoE tools which are nowadays the most up-to-date and still exhaustively studied DoE techniques.
- The functional kriging interpolator itself can also be considered to be adaptive to some extent: via the appropriate choice of the hyperparameters of the trace-covariance model, the interpolator can be



(a) 4-parameter case.

(b) 6-parameter case.

Figure 4.8: Time consumption of the exhaustive search (adaptive II approach) per iteration in function of the n number of samples. Let us disregard the abrupt changes in the speed of the algorithm. The computations have been performed on a shared computer with multi-core CPU and the current state of the simultaneous processes of other users actually influenced the runtime of our algorithm.

fitted to the observed data.

- The generation of the metamodel (the sampling) and its use for the approximation of the forward operator are well-separated steps. The time-consuming first part is plausibly done by "experts", whereas the second stage can be performed by the "end-user" (not necessarily equipped even by the numerical simulator of the EM phenomena) in a really short time. This practically provides the potential for establishing real-time applications.

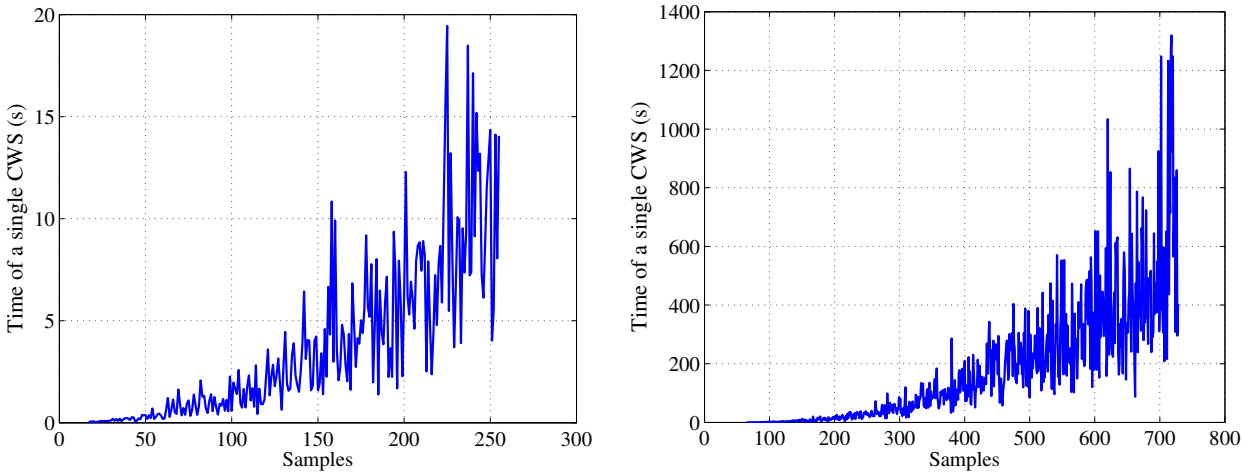
- The method treats the forward operator as a black-box: the inner structure of the model is not involved in the construction of the surrogate model, but only the input-output relationship function is needed. This makes the approach easily applicable to other problems as well.

Besides these impressive advantages, some drawbacks of the approach have to be noted as well.

- One can query the physical meaning of such interpolation-based surrogate modeling approaches. Taking the linear combination of certain pre-computed results might not seem to be a plausible idea in all cases. As long as the input parameters are only the sizes of a crack (or cracks), interpolation is a straightforward idea, since the output signal corresponding to a "middle-sized" crack can naturally be expected "between" the output signals of a "small" and a "large" crack, i.e., their linear combination does make sense. However, when the input parameters are positions, one had better to use a sort of "shifting" in the domain T to derive a prediction for an unobserved $q_{\mathbf{p}}(t)$, instead of using interpolation in $\mathbb{P}$. This cannot be done under the black-box assumption we use.

- The computational cost of both the estimation of the trace-covariance hyperparameters by cross-validation and the estimation of the trace-variance by jackknifing is considerable. The latter depends exponentially on the number of input parameters and polynomially on the number of already added samples in our implementations.

- There is no direct relation between the interpolation error $\varepsilon(\mathbf{p})$ and the estimated trace-variance $\hat{\sigma}^2_{\mathrm{T,Jack}}(\mathbf{p})$. Thus, a reliable stopping criterion based only on the interpolation error cannot be set



(a) 4-parameter case.

(b) 6-parameter case.

Figure 4.9: Time consumption of the single-start CWS (adaptive II approach) per iteration in function of the n number of samples.

for the adaptive sampling.

We think that the presented surrogate modeling method well deserves further study. The possible perspectives of improvement and application are the followings.

- To reduce the computational cost of both the cross-validation and the jackknife estimation, the well-known blocking technique [Shao and Tu, 1995] might be used. In so doing, the n samples are partitioned into k disjunct groups (blocks), and instead of leaving *one* sample out (in the CV and jackknifing), all samples of one block are left out at the same time. In this way, the number of kriging predictions are reduced from n to k.

- The first and second factors of the auxiliary function (4.29) can obviously be raised to any positive power. We have not studied the effect of such modification on the performance of the sampling, but we think this would be worth to consider in the future.

- Our choice of the parametrization and the norm we use to measure the "distance" of different defect configurations in the input space is obviously not the only possible solution for these issues. For instance, one can imagine a parametrization where the cracks are described by their area –which is believed to be a more relevant parameter than the length and the depth in their own– and the ratio of the edge lengths.

- We assumed a constant mean $m(t)$ for the modeling random processes. However, more complex mean functions can also be defined. An interesting issue would be to extend the polynomial mean assumption of the traditional kriging framework (§ B) to the functional kriging case and to study its effect on the yielded interpolators.

- Using the surrogate model, classical inversion algorithms (e.g., the optimization-based approach discussed in § 3) can be run to obtain an approximate solution of the inverse problem. Depending on the user's requirements, this can be considered as the final solution, or as the initial guess for a second stage of inversion operating on the true model, aiming to obtain a more precise final solution.

- The surrogate model makes possible the acceleration of originally time-consuming issues, such as Monte Carlo simulations (let us mention only the early textbook [Hammersley and Handscomb, 1975], since then, the domain has become much more diversified and applied). Monte Carlo sim-

ulations might be involved, e.g., in the analysis of uncertainty propagation, i.e., the study of how the uncertainty of the input parameters influences the output. One might also consider to include some of the model parameters (e.g., the conductivity of the specimen, the shape and position of the probe coil in ECT) in the uncertainty analysis by specifying them as further input parameters of the surrogate model.

- A completely different use of the sample-based surrogate model would be the *inverse interpolation*, i.e., to consider the (measured) output data as the input of an interpolator, aiming at predicting the corresponding values of the input parameters. This setup would be a plausible one-stage solver of the inverse problem to a certain extent. However, the performance of such inverse interpolations would obviously be influenced by the choice of the samples and the applied interpolator. The latter can be, e.g., a sort of kriging method. An adaptive sampling strategy for the generation of a database optimized for the inverse interpolation can be outlined as follows.

  1. Initial database using some classical DoE tool.
  2. Forward interpolation based on the samples, using functional kriging.
  3. Using the interpolated output data as input, perform the inverse interpolation with the chosen interpolator.
  4. Compute the error of the inverse interpolation (the discrepancy of the interpolated and the true input parameters) all over the input space.
  5. Insert a new input sample where the error is maximal, compute the corresponding output and add them to the database. Go to step 2 until a stopping criterion is met.

  We have obtained only some preliminary results [Bilicz et al., 2011a] in this domain so far, however. The proposed scheme arises several theoretical and practical issues, such as the regularization (the inverse interpolation does not necessarily yield an input vector within the input space), or effect of the uncertainty of forward interpolation on the overall sampling strategy. The outlined idea seems to be challenging and it deserves further study.

5

Output space-filling databases and inverse mappings

As it is discussed in § 4, the interpolation-based surrogate models have the potential of outperforming certain classical approaches if the sample-set (on which the interpolation is based) is generated adaptively. To this end, a sequential sampling strategy aiming at improving the functional kriging interpolator being built has been presented in § 4.

In the present chapter, a new sampling strategy is dealt with. The goal of this approach is mainly not the improvement of an interpolator's quality, but rather to fulfill certain criteria related to the distribution of the output samples in the output space. A certain "output space-filling" (OSF) is required from the yielded databases. Whereas classical DoE provides various approaches to obtain an "optimal" repartition of the input samples, the properties of the output sample distribution have not exhaustively been studied so far. Our sampling strategy aims at filling the output space uniformly (in a certain sense).

The main benefit of the methodology discussed in this chapter is the "meta-information" on the studied problem, represented by the structure of such OSF databases, and the tools facilitating the retrieval of this meta-information, respectively. Important conclusions concerning the inverse problem can be drawn both in qualitative and quantitative terms by means of the approaches devised.

The concepts are illustrated by numerical examples (with 2, 4 and 6 input parameters). Our results on this domain have been published in the following papers: [Bilicz, Lambert, and Gyimóthy, 2010a; Bilicz, Lambert, and Gyimóthy, 2010b; Bilicz et al., 2010a; Bilicz et al., 2010c; Bilicz et al., 2010d; Thirion et al., 2010].

5.1 Output space-filling

5.1.1 Motivations and goals

In the framework of classical DoE, the central question is the "optimal" repartition of the input samples in the input space. Various criteria of optimality have been proposed and by now, widely used. The textbooks [Fang, Li, and Sudjianto, 2006; Ryan, 2007; Santner, Williams, and Notz, 2003] give an exhaustive overview. Without completeness, let us mention the family of criterion-based designs (e.g., the maximal entropy design) and the space-filling designs. Within the latter, several classes can be defined based on the sense in which the "space-filling" is defined. The most popular space-filling designs are the LHS, the uniform designs and the distance-based approaches (e.g., maximin, minimax). Hybrid approaches (driven by a combination of several criteria) are also used. The common point –considered as a limitation– of all these

classical methods is that they are based only on criteria related to the input sample repartition (and perhaps to the manner of approximation of the forward operator in the subsequent step).

An essentially different setup appears when one formalizes criteria concerning the repartition of the output samples. However, in certain cases, it might be a natural desire to generate a sample set (corresponding input-output data pairs) whose output samples possess a sort of space-filling property in the output space. Such OSF designs can be useful when the inverse problem is addressed. Intuitively, when performing inversion, the "role" of the input and output space changes to the reverse in a certain sense, then one might need a DoE approach which is also inverted: acting in the output space like if it was the input space (formal definitions will also be given in the following). Let us mention in advance that this situation is inherently more difficult than the classical space-filling, since the forward operator can still be controlled from the input space, whereas the criteria are specified in the output space.

Let us recall the basic concepts of the sample-based surrogate modeling, already discussed in § 4.1. The database of n samples is defined as

$$\mathbb{D}_n = \left\{\left(\mathbf{p}_1,\, q_{\mathbf{p}_1}(t)\right),\, \left(\mathbf{p}_2,\, q_{\mathbf{p}_2}(t)\right),\, \ldots,\, \left(\mathbf{p}_n,\, q_{\mathbf{p}_n}(t)\right)\right\}, \tag{5.1}$$

where the input-output data pairs are linked via the given forward operator (for the basic terms and notations, see § 2.3):

$$q_{\mathbf{p}_i}(t) = \mathcal{F}\{\mathbf{p}_i\}, \quad \forall i \in I, \tag{5.2}$$

with the index-set $I = \{1, 2, \ldots, n\}$ to simplify further notations. Our main goal is to generate a database $\mathbb{D}_n$ with the property of OSF. As mentioned above, space-filling can be defined in various senses. Herein we consider the following two distance-based criteria:

- For all outputs $q(t) \in \mathbb{Q}$, an output sample $q_{\mathbf{p}_i}(t)$ exist which is "not too far" from $q(t)$. Let us define $\gamma(\mathbf{p})$ as the distance to the nearest sample:

  $$\gamma(\mathbf{p}) = \min_{i \in I} \left\| \mathcal{F}\{\mathbf{p}\} - q_{\mathbf{p}_i}(t) \right\|, \tag{5.3}$$

  then the first criterion is formalized as

  $$\max_{\mathbf{p} \in \mathbb{P}} \gamma(\mathbf{p}) < \Delta, \tag{5.4}$$

  with the positive constant Δ.

- The output samples are evenly distributed in the sense that none of the output samples is "too close" to an other one, and every output sample has its nearest neighbour "not too far". For a formal definition, let us introduce the minimal and maximal distances between the nearest neighbours:

  $$d_{\min} = \min_{i \in I} \left(\min_{j \in I \setminus \{i\}} \left\| q_{\mathbf{p}_i}(t) - q_{\mathbf{p}_j}(t) \right\| \right), \quad d_{\max} = \max_{i \in I} \left(\min_{j \in I \setminus \{i\}} \left\| q_{\mathbf{p}_i}(t) - q_{\mathbf{p}_j}(t) \right\| \right). \tag{5.5}$$

  Then we require –as second criterion– the equivalence of $d_{\min}$ and $d_{\max}$:

  $$\frac{d_{\max}}{d_{\min}} = 1. \tag{5.6}$$

If $\mathbb{D}_n$ simultaneously fulfills the criteria (5.4) and (5.6), we call it an OSF database. Let us note in advance that the rigorous validation of (5.4) is possible only by evaluating $\mathcal{F}$ over the whole input space. In the case of discrete input and output spaces, the exact fulfillment of (5.6) is not possible in general (however, such exact equispacing is obviously not required in practice). Thus, the term OSF will be used when the criteria (5.4) and (5.6) are approximately fulfilled.

A similar approach is discussed in [Bettinger et al., 2008], calling their method as "design for response diversity". However, they neither consider the (5.6) criterion of uniformity nor they formalize the approach for functional outputs. The sampling strategy (discussed in the next section) is also different from the one proposed in [Bettinger et al., 2008].

5.1.2 A simple analytical illustration

To better illustrate the concept of OSF, a simple analytical example is given herein. Let us assume that the input space is the continuous unit-square $\mathbb{U} = [0, 1]^2$. Since the functional output data cannot be easily visualized, we define a fictious forward operator as a vector function:

$$\mathbf{f} \ : \ \mathbb{U} \mapsto \mathbb{R}^2. \tag{5.7}$$

Let the components of $\mathbf{f}$ be

$$\begin{aligned} f_1(\mathbf{p}) &= \cosh \frac{\pi p_1}{2} \cos \frac{\pi p_2}{2}, \\ f_2(\mathbf{p}) &= \sinh \frac{\pi p_1}{2} \sin \frac{\pi p_2}{2}. \end{aligned} \tag{5.8}$$

The output space is then 2-dimensional:

$$\mathbb{Q} = \left\{ \mathbf{f}(\mathbf{p}) \;\middle|\; \forall \mathbf{p} \in \mathbb{U} \right\}. \tag{5.9}$$

In Fig. 5.1, two sampling techniques are illustrated. A Full-Factorial (FF) design with 10 levels along both axes of the input space is used to generate the repartition of the input samples in the first case (Fig. 5.1a). Due to the strong nonlinearity of the function $\mathbf{f}$, the corresponding repartition of the 100 output samples is strongly distorted: a certain region of $\mathbb{Q}$ is oversampled, whereas an other region appears to be badly filled by samples. This can be avoided, and a "homogeneous" output sample distribution can be achieved by using an OSF strategy (Fig. 5.1b). The same number of input samples are spaced in $\mathbb{U}$ in an adaptive way, taking into account the given function $\mathbf{f}$. In so doing, an about uniform output sampling is obtained.

Let us note that in this simple illustration, the sampling strategy (it will be presented in § 5.2) is applied to generate the OSF design; however, due to the small computational cost of $\mathbf{f}$, the exact distance functions are used in the sampling algorithm.

5.2 Adaptive sampling based on the distance functions

For the generation of the above discussed OSF databases, we have devised a sequential sampling algorithm which is presented in the following. Let us first introduce the concept of distance functions: $Q_i(\mathbf{p})$ is the distance function corresponding to the ith sample, defined as

$$Q_i(\mathbf{p}) = \left\| \mathcal{F}\{\mathbf{p}\} - q_{\mathbf{p}_i}(t) \right\|. \tag{5.10}$$

Thus, Q is defined over the input space $\mathbb{P}$, but expresses a distance measured in the output space $\mathbb{Q}$. Let us note that the criteria (5.4) and (5.6) can be recast in a simpler form using the distance functions. The minimal distance $\gamma(\mathbf{p})$ can be rewritten as

$$\gamma(\mathbf{p}) = \min_{i \in I} Q_i(\mathbf{p}), \tag{5.11}$$

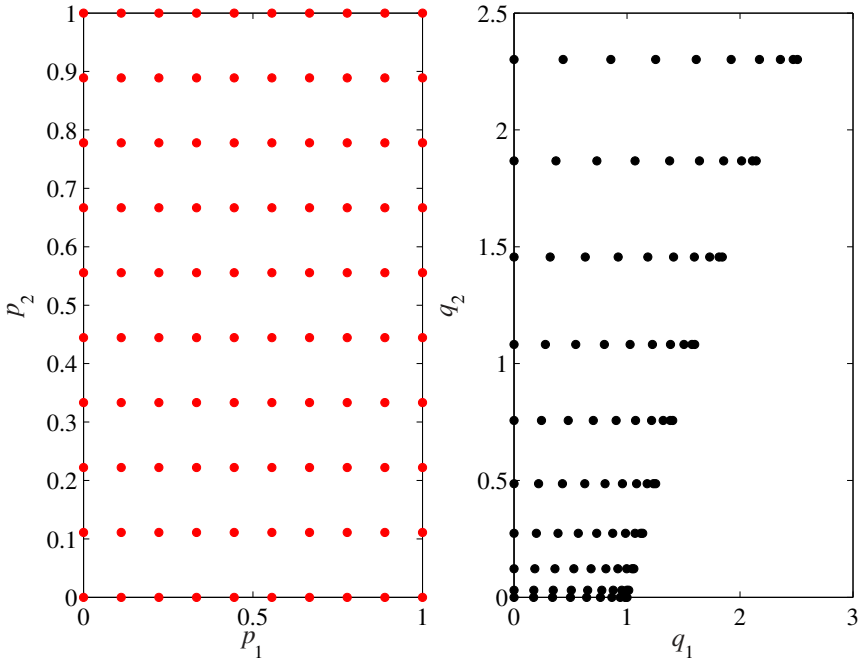
thus, for the space-filling criterion, we have

$$\max_{\mathbf{p} \in \mathbb{P}} \min_{i \in I} Q_i(\mathbf{p}) < \Delta. \tag{5.12}$$

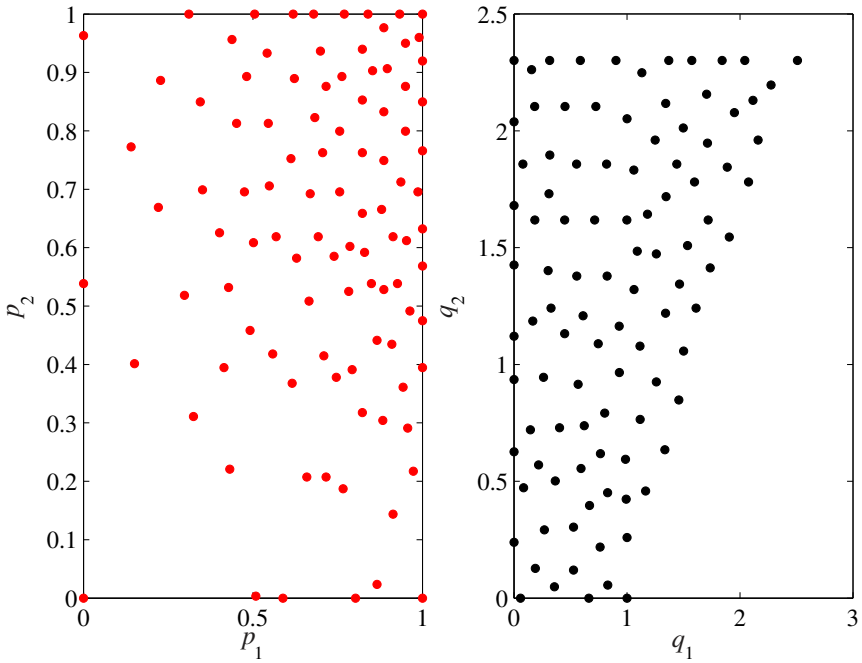
The criterion of equi-spacing has the form

$$\frac{d_{\max}}{d_{\min}} = \frac{\max\limits_{i \in I} \min\limits_{j \in I \setminus \{i\}} Q_i(\mathbf{p}_j)}{\min\limits_{i \in I} \min\limits_{j \in I \setminus \{i\}} Q_i(\mathbf{p}_j)} = 1. \tag{5.13}$$

Let us note that the evaluation of a distance function is generally as expensive as the evaluation of the forward operator, unless the evaluation is carried out at an input $\mathbf{p}_j$ included in the database (then $Q_i(\mathbf{p}_j)$ is computed using the stored output data).



(a) Full-factorial design in the input space.



(b) Output space-filling.

Figure 5.1: Illustrative analytical example. *Left:* input space, *right:* output space.

5.2.1 Sampling as an optimization problem

The sampling strategy is an incremental scheme: after an initialization, the samples are added one-by-one to the database, until a stopping criterion is met. The method is directly based on the criteria of optimality (reformalized as (5.12) and (5.13)). The stages of the strategy can be summarized as follows.

1. *Initialization:* Choose one initial sample $\mathbf{p}_1 \in \mathbb{P}$ (e.g., randomly) and compute the corresponding output. The initial database is then
$$\mathbb{D}_1 = \left\{\left(\mathbf{p}_1, q_{\mathbf{p}_1}(t)\right)\right\}. \tag{5.14}$$

2. *Sample insertion:* (Recursive definition.) Let us assume that n samples have already been added to the database $\mathbb{D}_n$. The $(n+1)$th sample has to be inserted so that $q_{\mathbf{p}_{n+1}}(t)$ will be as far as possible from all previously inserted output samples. This idea comes intuitively from the requirement of OSF and it is justified by the criterion (5.12). Formally, the choice of the next sample comes from the optimization problem
$$\mathbf{p}_{n+1} = \arg\max_{\mathbf{p}\in\mathbb{P}} \left(\min_{i\in I} Q_i(\mathbf{p})\right). \tag{5.15}$$
The search of the next sample is indeed a sort of maximin approach: the minimal distance between the next and all previously added output samples is to be maximized. However, this differs from the classical maximin problems in the sense that herein two spaces are considered: the input sample is sought whereas the criterion is related to the corresponding output sample.

   Once $\mathbf{p}_{n+1}$ has been found, compute the corresponding output and assign the new sample to the database:
$$\mathbb{D}_{n+1} = \mathbb{D}_n \cup \left\{\left(\mathbf{p}_{n+1}, q_{\mathbf{p}_{n+1}}(t)\right)\right\}, \tag{5.16}$$
and n is increased by 1. At this point, a natural idea would be to repeat step 2 again and again. However, the proposed algorithm not only inserts but also removes samples, thus, the procedure goes on with step 3.

3. *Sample removal:* If a certain criterion is fulfilled, a sample is removed from the database. We have arbitrarily set our algorithm to execute a sample removal after every two sample insertions. Other rules can also be imagined, e.g., one removal after every s insertions ($s > 2$), or a criterion based on $d_{\min}$, $d_{\max}/d_{\min}$, etc. The choice of the sample to be removed is heuristically motivated by the equi-spacing criterion (5.13). Let $d_{\min}^{(-r)}$ the distance between the pair of the nearest output samples if the rth sample has been removed:
$$d_{\min}^{(-r)} = \min_{i\in I\setminus\{r\}} \left(\min_{j\in I\setminus\{r,i\}} \left\|q_{\mathbf{p}_i}(t) - q_{\mathbf{p}_j}(t)\right\|\right). \tag{5.17}$$
Then, the index of the sample to be indeed removed is chosen as
$$r^\star = \arg\max_{r\in I} d_{\min}^{(-r)}, \tag{5.18}$$
i.e., the removal must increase the minimal distance between neighbour output samples as much as possible[1]. The reduced database after the removal is
$$\mathbb{D}_{n-1} = \mathbb{D}_n \setminus \left\{\left(\mathbf{p}_{r^\star}, q_{\mathbf{p}_{r^\star}}(t)\right)\right\}, \tag{5.19}$$
and the sample number n is decreased by 1.

[1] Since the distances $d_{\min}$ and $d_{\max}$ are both cheap-to-compute (similarly $d_{\min}^{(-r)}$ and $d_{\max}^{(-r)}$ as well), there is no difficulty in basing the removal rule directly on $d_{\max}/d_{\min}$, i.e., decreasing this ratio instead of increasing $d_{\min}$. We have not considered this variant yet, however.

4. *Stopping criterion:* The iterative-incremental sampling stops when certain stopping rules are met, otherwise the algorithm jumps to step 2. In our numerical studies, this was simply an upper limit of n. However, more sophisticated criteria can also be used: one might rely on the minimal inter-sample distance $d_{\min}$, the maximal distance of unsampled locations $\max_{\mathbf{p}\in\mathbb{P}}\left(\min_{i\in I} Q_i(\mathbf{p})\right)$, their values in a couple of subsequent iterations, etc.

This heuristic algorithm is believed to yield approximately OSF databases. As the input space is a finite set of input vectors, the limit $n \to \infty$ makes no sense. However, if the input space was the continuous unit-hypercube $\mathbb{U} = [0,1]^N$ consisting of input vectors $\mathbf{u}$, the convergence of the algorithm could be discussed. We believe that by continuing the above sampling strategy up to infinity, we would find

$$\frac{d_{\max}}{d_{\min}} \to 1 \quad \text{and} \quad \max_{\mathbf{u}\in\mathbb{U}}\left(\min_{i\in I}\right) Q_i(\mathbf{u}) \to 0, \tag{5.20}$$

under some mild assumptions on $\mathcal{F}$. However, we have not addressed this question; its importance is marginal in the practice.

5.2.2 Realization by kriging

Since the distance functions involve the forward operator directly, they are usually expensive-to-evaluate. Thus, the above method can be numerically intractable, since the optimization problem (5.15) (the core of the sampling) involves the distance functions. To reduce the computational cost of the algorithm to a treatable level, we use the kriging interpolation (see § B) to approximate the distance functions.

The kriging interpolation is based on a set of observed values. However, until n is small, and/or the input samples do not cover the input space sufficiently, the precision of the kriging prediction is usually poor; moreover, when the sampling starts by one single sample, a kriging interpolator cannot be built at all. In addition, the sample removal in the sampling strategy is also questionable. Once an output is computed via the expensive-to-evaluate operator $\mathcal{F}$, one should not delete the information represented by the sample (actually sentenced to removal) but should keep it in the kriging model. For these reasons, not only the current n samples in the database $\mathbb{D}_n$ are used, but also a number of "hidden" samples are stored and involved in the kriging model.

At the initialization stage, n_{init} initial input samples are selected and the corresponding outputs are computed. This initial sampling can be, e.g., a classical space-filling design in the input space. The initial samples and the n_{rem} samples deleted from $\mathbb{D}$ during the sampling procedure are together the hidden samples. The kriging interpolation of the ith distance function can then be written as

$$\hat{Q}_i(\mathbf{p}) = \sum_{j=1}^{n} \lambda_j(\mathbf{p}) Q_i(\mathbf{p}_j) + \sum_{j=n+1}^{n+n_{\text{rem}}} \lambda_j(\mathbf{p}) Q_i(\mathbf{p}_j) + \sum_{j=n+n_{\text{rem}}+1}^{n+n_{\text{rem}}+n_{\text{init}}} \lambda_j(\mathbf{p}) Q_i(\mathbf{p}_j), \quad i \in I. \tag{5.21}$$

The $(n + n_{\text{rem}} + n_{\text{init}})$ $\lambda_j(\mathbf{p})$ coefficients are computed as presented in § B.

In spite of the fact that all n distance functions (the distance functions corresponding to the hidden samples are irrelevant in the sampling) are to be predicted, one kriging model is enough to achieve all predictions; providing us a considerable advantage in computational terms. As a similar behaviour of the functions $Q_i(\mathbf{p})$ can be assumed, the same random process can be used to model all of them, consequently, we have the convenient matrix form for the $\lambda_j(\mathbf{p})$ coefficients:

$$\begin{bmatrix} \hat{Q}_1(\mathbf{p}) \\ \hat{Q}_2(\mathbf{p}) \\ \vdots \\ \hat{Q}_n(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} Q_1(\mathbf{p}_1) & Q_1(\mathbf{p}_2) & \cdots & Q_1(\mathbf{p}_{n+n_{\text{init}}+n_{\text{rem}}}) \\ Q_2(\mathbf{p}_1) & Q_2(\mathbf{p}_2) & \cdots & Q_2(\mathbf{p}_{n+n_{\text{init}}+n_{\text{rem}}}) \\ \vdots & & \ddots & \vdots \\ Q_n(\mathbf{p}_1) & Q_n(\mathbf{p}_2) & \cdots & Q_n(\mathbf{p}_{n+n_{\text{init}}+n_{\text{rem}}}) \end{bmatrix} \begin{bmatrix} \lambda_1(\mathbf{p}) \\ \lambda_2(\mathbf{p}) \\ \vdots \\ \lambda_{n+n_{\text{init}}+n_{\text{rem}}}(\mathbf{p}) \end{bmatrix}. \tag{5.22}$$

Let us recall that the values of the distance function $Q_i(\mathbf{p}_j)$ involved in the n-by-$(n + n_{\text{init}} + n_{\text{rem}})$ distance matrix in (5.22) are easy-to-compute (no forward simulation but only the stored data are needed).

Instead of the (5.15) original optimization problem of the sample insertion, the following task is performed in the realization:

$$\mathbf{p}_{n+1} = \arg\max_{\mathbf{p}\in\mathbb{P}} \left(\min_{i\in I} \hat{Q}_i(\mathbf{p}) \right). \tag{5.23}$$

Some more details on the implementation are given in § 5.2.3

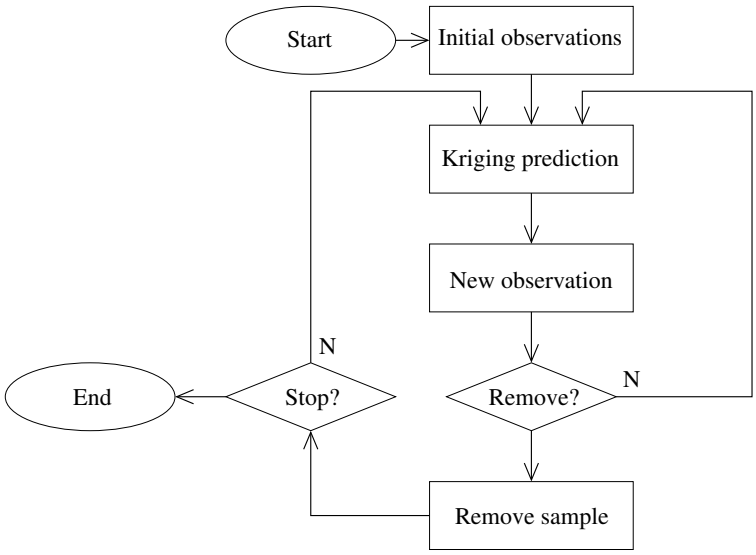


Figure 5.2: Flowchart of the adaptive sampling for the generation of OSF databases.

The sampling algorithm –already including the block of the kriging prediction– is summarized by the flowchart in Fig. 5.2. Having presented the whole realization, one can better justify why letting the sampling algorithm also to remove samples:

- The imperfection of the kriging prediction can mislead the algorithm. When the true distance relations turn out after evaluating the forward operator (at the input previously selected by using kriging), the algorithm still can improve the equi-spacing of the sampling (i.e., decrease the ratio $d_{\max}/d_{\min}$) by removing the unexpectedly "wrong" sample.
- Since the algorithm looks only one step forward, even with the perfect knowledge of all distance functions, the best repartition of the samples is not necessarily found by only adding samples. Though this point is rather heuristic, one can easily explain the need for sample removal with the choice of the first sample: there is no guarantee that the "best" repartition of the samples would include that arbitrarily chosen sample.

5.2.3 Optimization of the criterion function

Even by involving the relatively cheap-to-evaluate approximate distance functions $\hat{Q}_i(\mathbf{p})$, the optimization problem (5.23) can still be computationally demanding when the number of input parameters (thus, the cardinality of $\mathbb{P}$) is high. In our implementation, two variants for the numerical solution of (5.23) are proposed. In both of them, one has to keep in mind that the objective function (subjected to maximization) $\min_{i\in I} \hat{Q}_i(\mathbf{p})$ is probably multimodal (in the "grid-sense" defined in § C.2). The choice of the implementation basically depends on the cardinality of the input space:

I: The CWS algorithm (§ C.3) is used to solve the optimization problem. In order to reduce the probability of stalling at a local maximum, the search is started from all previous input samples (including the n_{init} initial, the n_{rem} removed and the n normal samples, respectively). The global maximum is then chosen among the found local maxima. This variant is proposed in the case of *small* and *middle-sized* input spaces (the 2 and 4-parameter examples).

II. Similarly to the version I, a multistart CWS scheme is used, however, to reduce the computational cost, only the n normal input samples are used as starting points. According to our experience, this approach becomes needed when the cardinality of the input space is in the order of 10^6 (in the 6-parameter example).

Let us note that in the 2-parameter example, the cardinality of the input space is only several hundred, thus, also a simultaneous evaluation of the objective function at all points of $\mathbb{P}$ would have been possible.

5.3 Measures of the database quality

To highlight the benefits and disadvantages of the databases generated by the proposed output-space filling sampling, they are compared to other databases built by using classical DoE approaches. For the quantitative and qualitative assessment, numerous measures can be defined. We believe that by the tools proposed below, a thorough characterization can be carried out.

5.3.1 Space-filling and uniformity

A plausible way of database characterization is to evaluate the criteria (5.4) and (5.6) *a posteriori* on the final database. Let us measure the output-space filling by the distance

$$\gamma_{\max} = \max_{\mathbf{p}\in\mathbb{P}} \left(\min_{i\in I} \left\| \mathcal{F}\{\mathbf{p}\} - q_{\mathbf{p}_i}(t) \right\| \right), \tag{5.24}$$

where $q_{\mathbf{p}_i}(t)$ is the ith output sample in the given database. The expressive meaning of $\gamma_{\max}$: if all output samples were encircled with a "sphere" of radius $\gamma_{\max}$, the union of the spheres would cover the whole output space. Equivalently: a "point" exists in the output space so that if it was encircled by a sphere of radius equal to $\gamma_{\max}$, none of the output samples would be within the sphere[2].

The exact determination of $\gamma_{\max}$ requires the evaluation of the forward operator at all input vectors in $\mathbb{P}$. Since this would be computationally intractable, we approximate $\gamma_{\max}$ as

$$\gamma_{\max} = \max_{\mathbf{p}\in\mathbb{S}} \left(\min_{i\in I} \left\| \mathcal{F}\{\mathbf{p}\} - q_{\mathbf{p}_i}(t) \right\| \right), \tag{5.25}$$

where $\mathbb{S}$ is an appropriate subset of $\mathbb{P}$, with smaller cardinality.

The uniformity of the database is measured via the equi-spacing ratio $d_{\max}/d_{\min}$ (already defined in (5.5) and (5.6))[3].

[2]Herein the terms "sphere" and "point" are obviously meant to be the generalization of the classical Euclidean concepts to the function space $\mathbb{Q}$.

[3]Let us note that in the classical DoE framework, the uniformity of a design is characterized via the empirical distribution function. However, since the output space is a function space, this is not possible (at least, not straightforward to define) in our case.

5.3.2 Forward interpolation error using nearest neighbour interpolation

A naturally arising question is how the OSF databases can be used to establish an interpolation-based surrogate model. In § 4, a functional kriging interpolator has been fitted to a database, having been adaptively generated so that the precision of the yielded surrogate model has small approximation error. In the case of the OSF technique, we have no specific interpolation technique being directly linked to the sampling strategy. However, one can believe by intuition that the error of the nearest neighbour interpolation must be reduced when the output space is evenly filled by samples.

To conveniently define the classical nearest neighbour (NN) interpolation, let us introduce n window functions based on the database (5.1):

$$W_i(\mathbf{p}) = \begin{cases} 1 & \text{if} \quad i = \arg\min\limits_{j \in I} \left\| \mathbf{p} - \mathbf{p}_j \right\|, \\ 0 & \text{otherwise.} \end{cases} \tag{5.26}$$

This definition still does not provide the window functions unambiguously, since the arg min operator might yield more than one i index at a fixed $\mathbf{p}$. The value of such $W_i(\mathbf{p})$ is randomly chosen to one or zero, so that

$$\sum_{i=1}^{n} W_i(\mathbf{p}) \equiv 1 \tag{5.27}$$

holds for all $\mathbf{p} \in \mathbb{P}$. Let us note that this indefiniteness is important in the case of our discrete input space, but it would be irrelevant if the input space was continuous[4].

The nearest neighbour interpolator is then defined as

$$\mathcal{G}^{(n)}\{\mathbf{p}\} = \sum_{i=1}^{n} W_i(\mathbf{p}) q_{\mathbf{p}_i}(t), \tag{5.28}$$

where the upper index (n) emphasizes the sample-based nature of the surrogate operator $\mathcal{G}$. The error of the interpolation is

$$\varepsilon(\mathbf{p}) = \left\| \mathcal{G}^{(n)}\{\mathbf{p}\} - \mathcal{F}\{\mathbf{p}\} \right\| = \sum_{i=1}^{n} W_i(\mathbf{p}) \left\| q_{\mathbf{p}_i}(t) - \mathcal{F}\{\mathbf{p}\} \right\| = \sum_{i=1}^{n} W_i(\mathbf{p}) Q_i(\mathbf{p}). \tag{5.29}$$

Now it is clear that the OSF does not result directly in minimizing the interpolation error $\varepsilon(\mathbf{p})$. However, one can also see that the OSF indirectly reduces $\varepsilon(\mathbf{p})$ as a high value of $Q_i(\mathbf{p})$ over the support of $W_i(\mathbf{p})$ indicates an "empty" region in the output space which is to be filled by the sampling strategy.

Consequently, the nearest neighbour interpolation would naturally fit to the OSF sampling[5]. Since the forward interpolation is a central idea in surrogate modeling, we will also study the maximal interpolation error

$$\varepsilon_{\max} = \max_{\mathbf{p} \in \mathbb{P}} \varepsilon(\mathbf{p}) \tag{5.30}$$

related to the databases. However, bad performance in the sense of $\varepsilon_{\max}$ obviously does not mean the failure of the OSF. Let us further note that $\gamma_{\max}$ is always a lower bound of $\varepsilon_{\max}$.

Similarly to $\gamma_{\max}$, the maximal interpolation error is also expensive-to-compute due to the need for evaluating $\mathcal{F}$ at all points of $\mathbb{P}$. Thus, $\varepsilon_{\max}$ is also approximated as

$$\varepsilon_{\max} = \max_{\mathbf{p} \in \mathbb{S}} \varepsilon(\mathbf{p}), \tag{5.31}$$

with $\mathbb{S}$ being an appropriate and small subset of $\mathbb{P}$.

[4] Since the input vectors at which $W_i(\mathbf{p})$ is indefinite would form a zero-measure subset of the continuous input space.

[5] However, by slightly modifying the sampling strategy, the scheme which directly aims at minimizing the interpolation error can also be constructed, see § 5.6.

5.3.3 Multidimensional scaling

Whereas the repartition of the input samples is easy to visualize in the case of 2 or 3 input parameters, the output space of functional data cannot be displayed in straightforward manner. However, the yielded OSF properties of the databases would be desirable to be visually assessed as well, beyond the quantitative measures we defined above.

Multidimensional Scaling (MDS) provides a framework to visualize datasets lying in high-dimensional spaces (for a solid basis, see, e.g., the textbook [Kruskal and Wish, 1986]). Let us consider the n output samples in the database (5.1). Let $d(i, j)$ denote the distance between the ith and jth output sample:

$$d(i, j) = \left\| q_{\mathbf{p}_i}(t) - q_{\mathbf{p}_j}(t) \right\|. \tag{5.32}$$

The main idea of MDS is to set n points in a k-dimensional Euclidean space so that the pairwise distances between the new points are as similar as possible to the original distances $d(i, j)$. If $k = 2$ or $k = 3$, the repartition obtained can easily be plotted. This plot can then be considered as the low-dimensional representation of the original dataset.

Let us denote the new points by $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_n \in \mathbb{R}^k$. Let $\tilde{d}(i, j)$ be the distance between the ith and jth sample:

$$\tilde{d}(i, j) = \left\| \boldsymbol{\pi}_i - \boldsymbol{\pi}_j \right\| \tag{5.33}$$

in the sense of the Euclidean norm. The misfit of the distances in the original and in the k-dimensional spaces is formalized via the so-called normalized stress:

$$S(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_n) = \sqrt{\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(d(i, j) - \tilde{d}(i, j) \right)^2}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d(i, j)^2}}. \tag{5.34}$$

The points $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_n$ are then chosen so as the stress is minimal:

$$\{\boldsymbol{\pi}_1, \boldsymbol{\pi}_n, \ldots, \boldsymbol{\pi}_n\} = \arg\min_{\boldsymbol{\chi}_i \in \mathbb{R}^k} S(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \ldots, \boldsymbol{\chi}_n). \tag{5.35}$$

The stress can be formalized in other ways as well, and the above optimization problem can be solved by different approaches. In our numerical studies, we have chosen $k = 2$ and used the `mdscale` routine of MATLAB®.

5.4 Characterization of the inverse problem

As mentioned in the introduction of this chapter, the most important goal of the OSF is to learn certain meta-information on the studied forward operator. This meta-information can help us to characterize the related inverse problem, namely, by assessing its ill-posedness.

5.4.1 The regularized inverse problem

Let us recall that the inverse problem of nondestructive evaluation is usually ill-posed [Hadamard, 1952], which can be handled –the problem can be regularized– by involving some *a priori* knowledge on the defect. As it was discussed in § 3.1 (and applied throughout the chapter § 3), the assumption of a parametric defect model is a straightforward way of regularization.

Being based on a parametric defect model, the database of n input-output data pairs can also be considered as a regularization tool. The simplest method to invert the measured data (denoted by $\tilde{q}(t)$) by means of a database is to apply the Nearest Neighbour (NN) interpolation in reverse: let the solution $\hat{\mathbf{p}}$ of the regularized inverse problem be the input sample $\mathbf{p}_i$ corresponding to the output signal being the closest to the measured data:

$$\hat{\mathbf{p}} = \mathbf{p}_i, \quad i = \arg\min_{j \in I} \left\| \tilde{q}(t) - q_{\mathbf{p}_j}(t) \right\|. \tag{5.36}$$

In this dissertation, we do not use the database to indeed solve inverse problems. However, the above scheme is a good illustration of inverse interpolation[6], in addition, it will also be invoked in the next subsections, where the tools of inverse problem characterization are presented.

5.4.2 Voronoi cell mapping

The output space can be partitioned into n disjunct Voronoi cells [Okabe et al., 2000; Voronoi, 1908], whose union completely covers the output space. The Voronoi cells are uniquely determined by the n output samples and the norm (defining the distance) in the output space. The ith Voronoi cell is defined as

$$\Omega_i = \left\{ q(t) \in \mathbb{Q} \;\middle|\; i = \arg\min_{j \in I} \left\| q(t) - q_{\mathbf{p}_j}(t) \right\| \right\}. \tag{5.37}$$

Let us note that the operator "arg min" can theoretically yield more than one index. In the practice, however, this is "almost impossible" when one has an output space $\mathbb{Q}$ being a finite set of functions[7]. The union of all Voronoi cells is equal to the whole output space:

$$\bigcup_{i=1}^{n} \Omega_i = \mathbb{Q}. \tag{5.38}$$

Let us define the inverse-mapped Voronoi cells as the "projection" of the output Voronoi cells into the input space. Let the ith inverse-mapped Voronoi cell be defined as

$$\Psi_i = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \mathcal{F}\{\mathbf{p}\} \in \Omega_i \right\}. \tag{5.39}$$

Since the forward operator $\mathcal{F}$ is assumed to be univalent, the sets Ψ_i are also uniquely determined, furthermore

$$\bigcup_{i=1}^{n} \Psi_i = \mathbb{P} \tag{5.40}$$

holds. The knowledge of the inverse-mapped Voronoi cells is an essential tool for the characterization of the inverse problem. Let us assume that the database is used for inverse interpolation with the nearest neighbour interpolator (according to the formula (5.36)), and for a given measurement $\hat{q}(t)$, one finds the ith output sample as the closest one to $\tilde{q}(t)$. The inverse interpolation then yields $\mathbf{p}_i$ as the solution of the regularized inverse problem. The "reliability" of the solution is in fact closely related to the Ψ_i inverse-mapped Voronoi cell: any of the $\mathbf{p} \in \Psi_i$ input vectors must be considered as an equivalent solution in this simple NN inverse interpolation scheme. In this sense, the inverse-mapped Voronoi cells represent the "resolution", i.e., the precision within reach of the database. Obviously, this precision usually depends on i.

In practical terms, one might imagine that not only the database is given to the end-user, but also a quantitative description of the precision one might reach for each sample. This description can be, e.g., the distance of the farthest input samples corresponding to the same inverse-mapped Voronoi cell Ψ_i. The precision can separately be determined for the different input parameters as well.

[6] More complicated interpolation methods can also be imagined (see the perspectives in § 5.6).

[7] Because in our case, $\mathbb{Q}$ is in fact a zero-measure subset of $L^2(T)$.

A convenient expression of the inverse-mapped Voronoi cells can be obtained by using the distance functions (5.10). The set Ψ_i is defined as the subset of the input space on which $Q_i(\mathbf{p})$ is smaller than any other distance function:

$$\Psi_i = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; i = \arg\min_{j \in I} Q_j(\mathbf{p}) \right\}, \tag{5.41}$$

as directly follows from (5.37) and (5.39).

However, the numerical determination of the inverse-mapped Voronoi cells is computationally expensive. One must perform an exhaustive search in the whole input space in the sense that all elements of $\mathbb{P}$ must be examined and classified into one of the n subsets Ψ_i, via the evaluation of the forward operator $\mathcal{F}$, which might be expensive-to-run. To overcome this computational challenge, we propose the use of the kriging interpolation to predict the distance functions based on the observations (made at the input samples of the studied database, as discussed in § 5.2.2). In this way, not the exact but the approximate inverse-mapped Voronoi cells can be determined as

$$\hat{\Psi}_i = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; i = \arg\min_{j \in I} \hat{Q}_j(\mathbf{p}) \right\}, \tag{5.42}$$

where $\hat{Q}_j(\mathbf{p})$ denotes the kriging prediction of the jth distance function.

5.4.3 Noise level mapping

A measured output $\tilde{q}(t)$ is always corrupted by noise. Let us assume that $\tilde{q}(t)$ is still an element of the $L^2(T)$ space, but not necessarily an element of $\mathbb{Q}$. In a very simple case, the noise can be modeled with a uniform distribution within a "sphere" in $L^2(T)$, with a radius δ and center $\tilde{q}(t)$. Formally, we assume that the "true" (noise-free) output satisfies

$$\|q_{\mathrm{True}}(t) - \tilde{q}(t)\| \leq \delta, \tag{5.43}$$

where δ is the "noise level" (assumed to be known). Using our parametrized defect model as a regularization tool, the set Ξ of the solutions of the regularized inverse problem can be defined as[8]

$$\Xi = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \|\mathcal{F}\{\mathbf{p}\} - \tilde{q}(t)\| \leq \delta \right\}. \tag{5.44}$$

In words, the "sphere" with radius δ excises a certain region from the output space, then the solution of the regularized inverse problem can be defined as the image of this intersection in the input space, where the mapping is given by the inverse of $\mathcal{F}$.

In a special case, the distance functions can conveniently be used to study the inverse mapping of a noise level. Let us assume that we have a database of n samples, furthermore, the measured output $\tilde{q}(t)$ is equal to the ith output sample $q_{\mathbf{p}_i}(t)$. Then, for a given noise level δ, the set of solutions of the regularized inverse problem is

$$\Phi_i^\delta = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \left\|\mathcal{F}\{\mathbf{p}\} - q_{\mathbf{p}_i}(t)\right\| \leq \delta \right\}, \tag{5.45}$$

or equivalently with the distance functions:

$$\Phi_i^\delta = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; Q_i(\mathbf{p}) \leq \delta \right\}. \tag{5.46}$$

In spite of the fact that the measured output is "almost never" equal to any of the output samples of a database, the inverse-mapped "noise patches" Φ_i^δ appear to be valuable tools of the characterization of the

[8] Other definitions are also possible. For instance, in many cases (like in the EGO-based inversion approach in § 3 or in the above mentioned NN formula of inverse interpolation (5.36)), we simply define the solution as the input corresponding to the output being the closest to the measured data. However, herein our goal is to make the concept of "noise level mapping" more understandable.

inverse problem. The extension of these patches in some sense (e.g., the distance of the farthest input samples corresponding to the same cell) indicates how much uncertainty is induced in the solution by the uncertainty of the output data. Let us emphasize that, however, there is no direct link to the cases in which $\tilde{q}(t)$ is not equal to any of the output samples.

Similarly to the inverse-mapped Voronoi cells, the noise patches are also expensive to compute, for the same reasons. By involving the kriging prediction of the distance functions, the computational cost can considerably be reduced, and the approximate noise patches can be written as

$$\hat{\Phi}_i^{\delta} = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \hat{Q}_i(\mathbf{p}) \leq \delta \right\}, \tag{5.47}$$

where $\hat{Q}_i(\mathbf{p})$ denotes the kriging prediction of the ith distance function.

5.5 Numerical examples

5.5.1 Generally about the examples

To illustrate the performance of the proposed OSF sampling strategy and the inverse mappings, some numerical studies are presented in this section, involving the canonical configurations of § 2.4 with 2, 4 and 6 input parameters, respectively. The adaptive sampling is compared to the classical full-factorial design (§ D).

The initial samples of the adaptive sampling are chosen according to 3-level factorial designs (i.e., using 3^N samples in case of N input parameters). The Matérn function (§ B.2) is used as a covariance model. The hyperparameters are estimated by using the REML method (§ B.2.1), based on the initial samples. According to our experience, the estimated covariance hyperparameters are appropriate in the further stages of the sampling algorithm, i.e., one does not need to perform re-estimation during the sequential sampling. However, when the method is applied to other forward problems, updating of the hyperparameters might be needed in each cycle. The data used in the REML scheme is the distance function corresponding to the first sample. The latter is chosen as $\mathbf{p}_1 = [1, 1, \ldots, 1]$. We assume a normal prior distribution on the log-scaled hyperparameters with the means (according to the usual rule-of-thumbs)

$$\sigma^2_{0,\text{prior}} = \text{Var}\left[\text{all observations}\right], \quad \nu_{\text{prior}} = 1.5N, \quad \rho_{i,\text{prior}} = 0.5, \tag{5.48}$$

and variances

$$\sigma_{\sigma^2_0,\text{prior}} = 1, \quad \sigma_{\nu,\text{prior}} = 0.12^2, \quad \sigma_{\rho_i,\text{prior}} = 0.25^2, \tag{5.49}$$

furthermore, no correlation between the hyperparameters is assumed.

The sampling algorithm performs a sample removal after every two insertions. In Table 5.1, the number of samples n and the total number of forward simulations needed to generate the databases ($n + n_{\text{init}} + n_{\text{rem}}$) are summarized. Let us note that to build a database of n samples, the number of forward simulations is at most $n_{\text{init}} + 2(n-1)$ (since two insertions are needed to increase n by 1 and one of the initial samples is used as the first sample). This number can decrease when an initial or a previously "removed" sample is selected again for insertion (this happened only in the 6-parameter example). The stopping criterion is simply an upper limit for n, chosen such that it fits the sample number of the corresponding full-factorial design to which the OSF database is compared.

The full-factorial designs have approximately equispaced levels[9] (with the same number of levels along the different input parameters) and the lower and upper bounds (0 and 1) of each input parameter are always represented by a level.

[9] As far as it is possible in the discretized input space. See § D.3.

Table 5.1: Number of samples (n) and total number of forward simulations needed to generate the database ($n + n_{\text{init}} + n_{\text{rem}}$).

Example	n	$n + n_{\text{init}} + n_{\text{rem}}$
2-parameter	3^2	25
	4^2	39
	5^2	57
	6^2	79
4-parameter	3^4	241
	4^4	591
	5^4	1329
6-parameter	3^6	2141

The numerical characteristics (such as the interpolation error or the minimal distance between the output samples) are not normalized but given in resistance dimension. To give an idea to the reader about the relation between the crack sizes and the norm of the corresponding output signal, let us refer to Fig. 4.3, where we have plotted the norm of the output in function of the crack sizes in the case of the single-crack configuration (point 1 in § 2.4).

As mentioned in § 5.2.3, two approaches are applied to solve the (5.23) optimization problem of the sample insertion. The scheme I is used in the 2 and 4-parameter example, whereas the implementation II is called for the 6-parameter case, respectively.

All computations have been performed on a PC with a 64-bit CPU with 4 cores at 2.33 GHz and 16 GB memory. The algorithms are implemented in MATLAB®.

5.5.2 Examples with 2 parameters

The canonical configuration defined as the point 1 in § 2.4 is involved in the 2-parameter examples. Let us recall that the length (L) and the depth (D) of a single crack are enabled to vary, all other parameters are assumed to be fixed. The input space consists of 342 feasible input vectors.

Numerical characteristics

The minimal distance between an arbitrary point of the output space and an output sample, $\gamma(\mathbf{p})$ (5.11) and the error of nearest neighbour interpolation, $\varepsilon(\mathbf{p})$ (5.29) are evaluated at all $\mathbf{p} \in \mathbb{P}$. The characteristics $\gamma_{\max}$ and $\varepsilon_{\max}$ (defined in (5.24) and (5.30), respectively) are given in Table 5.2, along with the minimal and maximal distance between the nearest neighbours (5.13). Both $\gamma_{\max}$ and $\varepsilon_{\max}$ are smaller for the OSF sampling, for all studied sample numbers. The ratio $d_{\max}/d_{\min}$ also appears to be much smaller for the OSF sampling. The high ratios of the FF sampling are caused by the extremely small $d_{\min}$ distances. One can clearly see that the modeled forward operator is quite "flat" at certain regions, thus, the input-equidistant FF sampling necessarily fails in uniformly filling the output space.

The minimal distance $\gamma(\mathbf{p})$ and the interpolation error $\varepsilon(\mathbf{p})$ are also plotted by colormaps in Figs. 5.3 and 5.4, respectively. The repartition of the input samples in the OSF databases explains why the FF samplings are outperformed. In the region of large cracks, many more input samples are inserted by the adaptive sampling strategy than in the regions of small cracks. This is because the forward operator varies much faster in the large-crack region, whereas it appears to be flat for small cracks. Consequently, the input-equidistant FF design oversamples the small-crack region, at the expense of undersampling the domain of large cracks. Let us also notice that $\gamma(\mathbf{p})$ varies "smoothly" (however, continuity obviously makes

Table 5.2: Numerical characteristics of the databases in the 2-parameter example. (FF: full-factorial design, OSF: output space-filling.)

Samples (n)	$\gamma_{\max}$ (mΩ)		$\varepsilon_{\max}$ (mΩ)		$d_{\min}$ (mΩ)		$d_{\max}$ (mΩ)		$d_{\max}/d_{\min}$	
	FF	OSF	FF	OSF	FF	OSF	FF	OSF	FF	OSF
3×3	25.8	14.5	60.3	33.3	0.17	14.2	44.1	20.3	266.5	1.44
4×4	22.4	13.7	38.7	25.8	0.11	7.50	30.3	19.2	264.6	2.56
5×5	19.9	11.8	26.1	22.2	0.082	5.48	25.6	12.4	311.1	2.26
6×6	15.1	11.8	22.8	17.9	0.061	4.08	20.7	10.6	339.5	2.60

no sense in a discrete input space), whereas $\varepsilon(\mathbf{p})$ appears to abruptly change at the border of the neighbouring window functions (5.26) (input Voronoi cells). Both quantities are exactly zero at the sampled points, and $\gamma_{\max}$ is indeed a lower bound of $\varepsilon_{\max}$.

Visualization by multidimensional scaling

Multidimensional scaling (MDS, § 5.3.3) makes possible the low-dimensional visualization of the high dimensional output data. In Fig. 5.5, the 2-dimensional representations of the output samples are plotted. Let us note that in the (π_1, π_2) image plane, the Euclidean distances still have resistance dimensions, similarly to the distances in the original (output) space. However, we must recall that the absolute position of the points in (π_1, π_2) plane has no meaning, only the pairwise distances carry information. One can easily see the difference between the output sample repartition of the adaptive and the full-factorial samplings. The latter yield quite distorted output sample distribution, many samples appearing to cluster, whereas some others lie much farther. On the contrary, the OSF sampling more-or-less equidistantly spaces the output samples, as it has already been shown by the ratio $d_{\max}/d_{\min}$.

In Table 5.3, the minimized stress (5.34) is given for all cases. The stress appears to be small, meaning that the output sample distribution can accurately be represented (with a small loss of information) in the 2-dimensional plane.

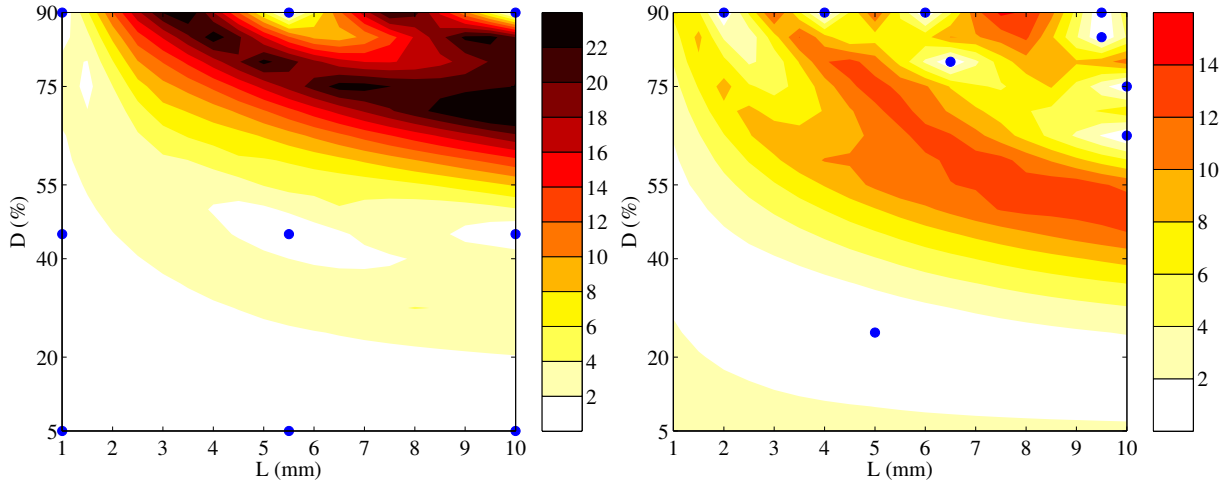
Table 5.3: MDS stress (S) in the 2-parameter example with different number of samples in case of full-factorial and output space-filling sampling.

Samples (n)	FF	OSF
3×3	0.0060	0.0212
4×4	0.0093	0.0204
5×5	0.0101	0.0194
6×6	0.0103	0.0165

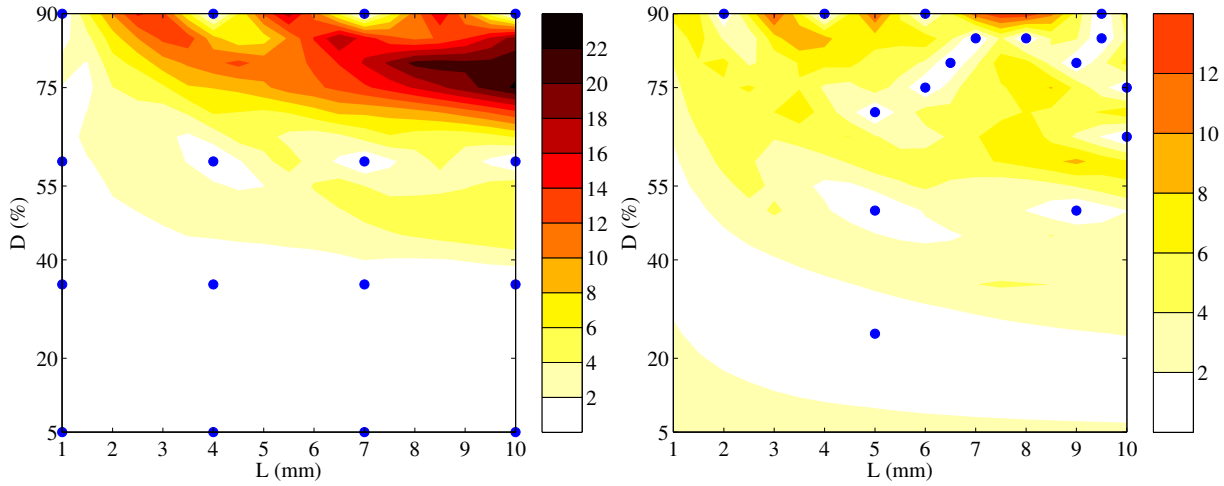
Inverse mappings

In §§ 5.4.2 and 5.4.3, the techniques of Voronoi cell and noise level inverse mapping have been presented. Herein we illustrate these approaches on the 2-parameter example.

The inverse-mapped (approximate) Voronoi cells $\hat{\Psi}_i$ are subsets of the input space (as defined in (5.42)), i.e., in our cases, they are finite sets. For the sake of a more expressive visualization, we do not project the Voronoi cells to the discrete input space $\mathbb{P}$, but onto the continuous domain $\mathbb{U} = [0, 1] \times [0, 1]$. In so doing, the yielded "patches" in $\mathbb{U}$ can be considered as continuous extensions of the cells $\hat{\Psi}_i$, and the "true" $\hat{\Psi}_i$ cells



(a) $n = 9$.



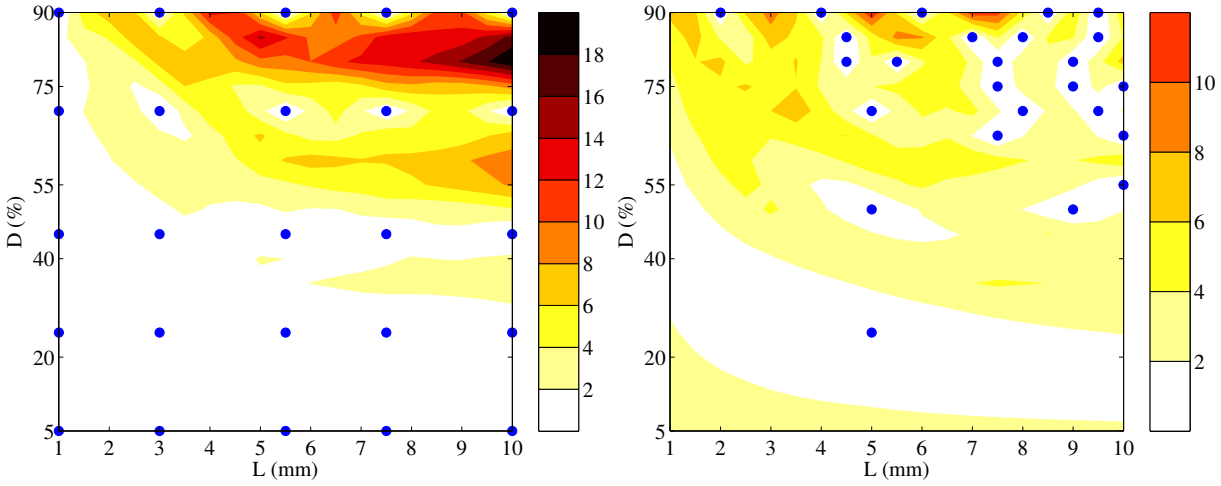
(b) $n = 16$.

Figure 5.3: Minimal distance $\gamma(\mathbf{p})$ over the input space in the 2-parameter example with different number of samples (colormap in mΩ). Left: full-factorial sampling. Right: output space-filling sampling. (To be continued on the next page.)

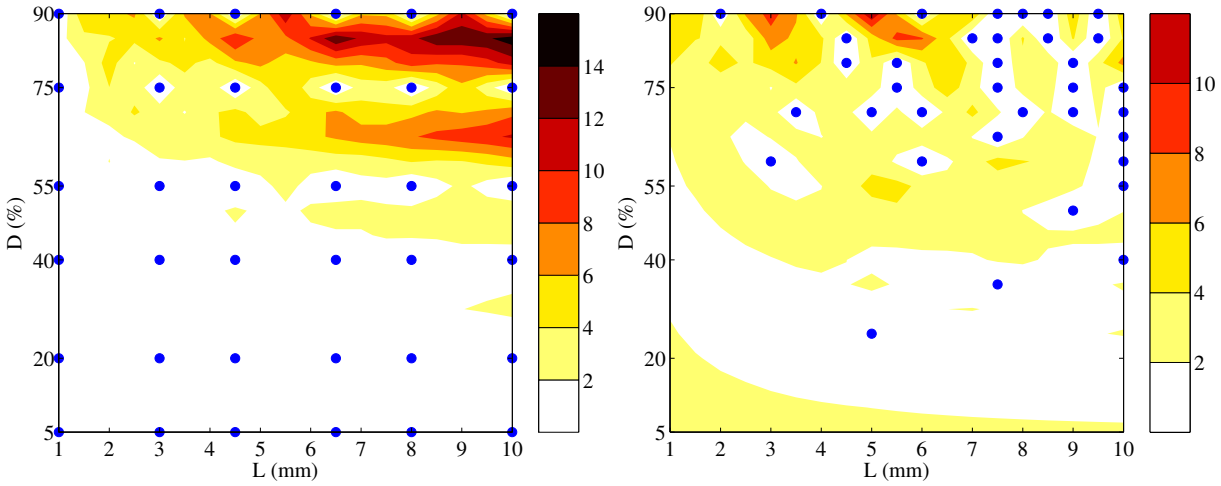
are in fact the sets of input samples within the ith patch[10]. In practice, $\mathbb{U}$ is discretized by a very fine grid (500×500), and the cells $\hat{\Psi}_i$ are numerically determined by evaluating the criterion (5.42) on this fine grid, instead of $\mathbb{P}$.

One can easily see the quite various shapes and dimensions of the inverse-mapped patches in Fig. 5.6. In the first case (Fig. 5.6a), more than half of the input space is covered by one patch, whereas the patch corresponding to the largest crack is extremely tiny, in all four cases. Consequently, in the studied setup, small cracks can be characterized with much higher uncertainty than large cracks. As the space $\mathbb{U}$ has a topology, it makes sense to discuss the connectedness of the patches. A very interesting phenomenon can be seen in Figs. 5.6b and 5.6c: the inverse mapping of one of the output Voronoi cells yields two disconnected patches in $\mathbb{U}$.

[10]One must be careful at this point. The input vectors out of $\mathbb{P}$ have absolutely no physical meaning. In our case, however, it would be a natural idea to assume that $\mathcal{F}$ is indeed defined on $\mathbb{U}$, and the corresponding $\mathbf{p} \in \mathbb{P}$, $q(t) \in \mathbb{Q}$ data pairs can be considered as a sort of discrete representation of $\mathcal{F}$. For consistency and mathematical rigour, we do not apply this approach in this Dissertation.



(c) $n = 25$.

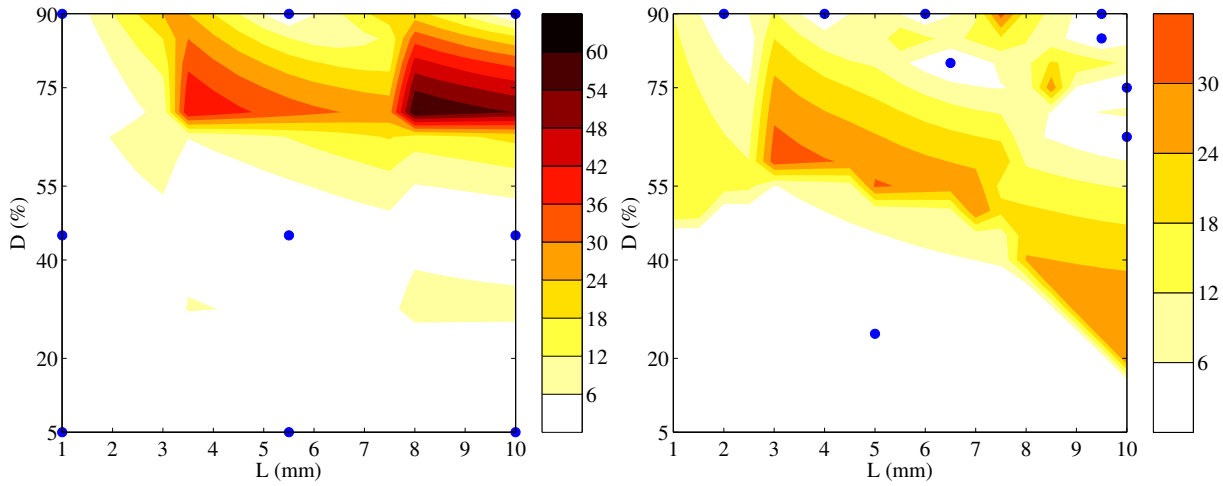


(d) $n = 36$.

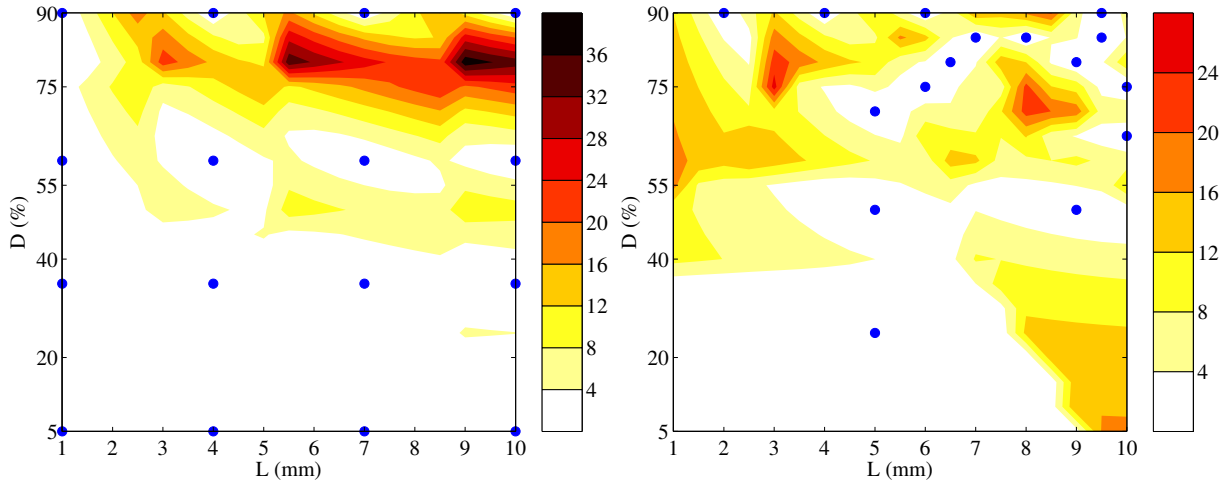
Figure 5.3: (Continued.) Minimal distance $\gamma(\mathbf{p})$ over the input space in the 2-parameter example with different number of samples (colormap in mΩ). Left: full-factorial sampling. Right: output space-filling sampling.

In Fig. 5.7, three different noise levels (§ 5.4.3) are projected onto the input space, using two OSF databases. Similarly to the Voronoi cell mapping, the continuous $\mathbb{U}$ domain is introduced in order to obtain expressive plots. The various shapes and dimensions of the "noise-patches" (encircling the "true" Φ_i^δ sets (5.47)) are salient again. The same level of noise on the output data causes much higher uncertainty in the characterization of small cracks than of large ones. The extensive patches corresponding to small cracks explains why it would not be necessary to sample the small-crack region finer: when using the database for inverse interpolation, even a relatively small noise makes distinguishing small cracks impossible. In other words, small cracks can be "hidden" by the noise, whereas large cracks can still precisely be characterized.

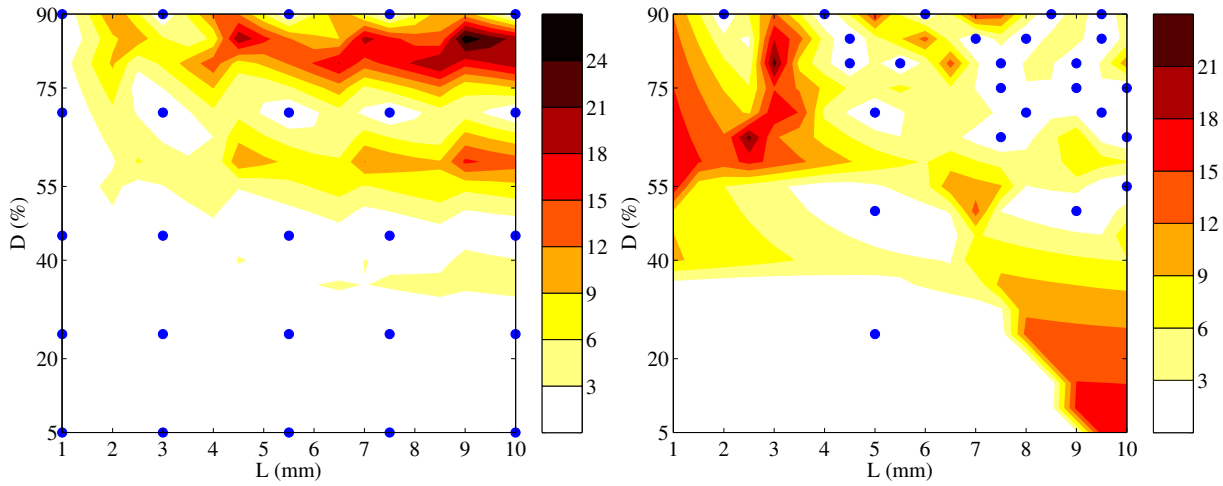
Let us note that the same disconnectedness which we experienced in certain patches of the Voronoi cells might occur in the case of noise patches as well. Though there is no splitting patch among the presented cases, one can assume that in Fig. 5.7b, the patch of the sample $L = 5\,\text{mm}$, $D = 25\,\%$ might become disconnected when δ decreases.



(a) $n = 9$.

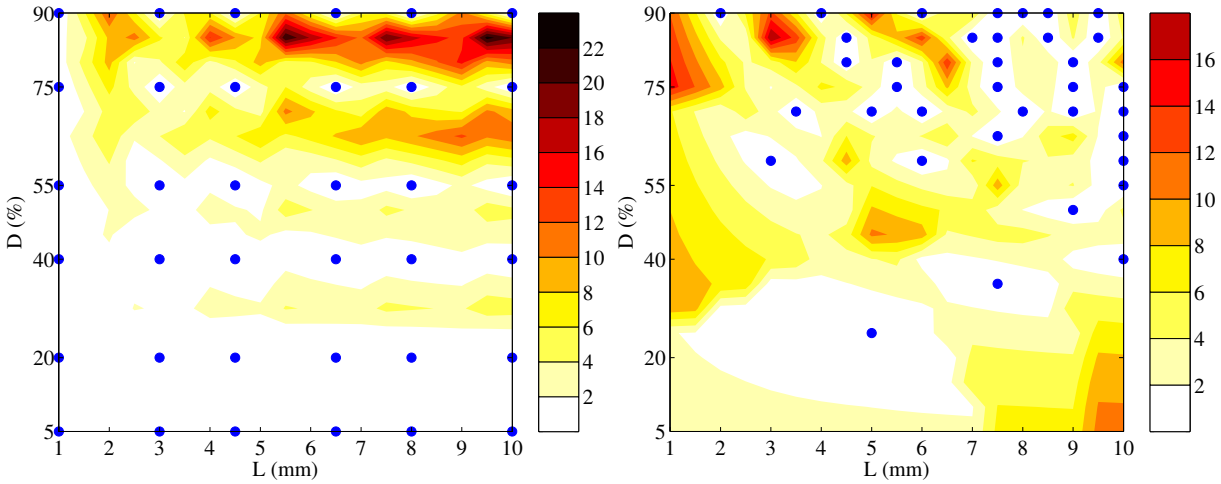


(b) $n = 16$.



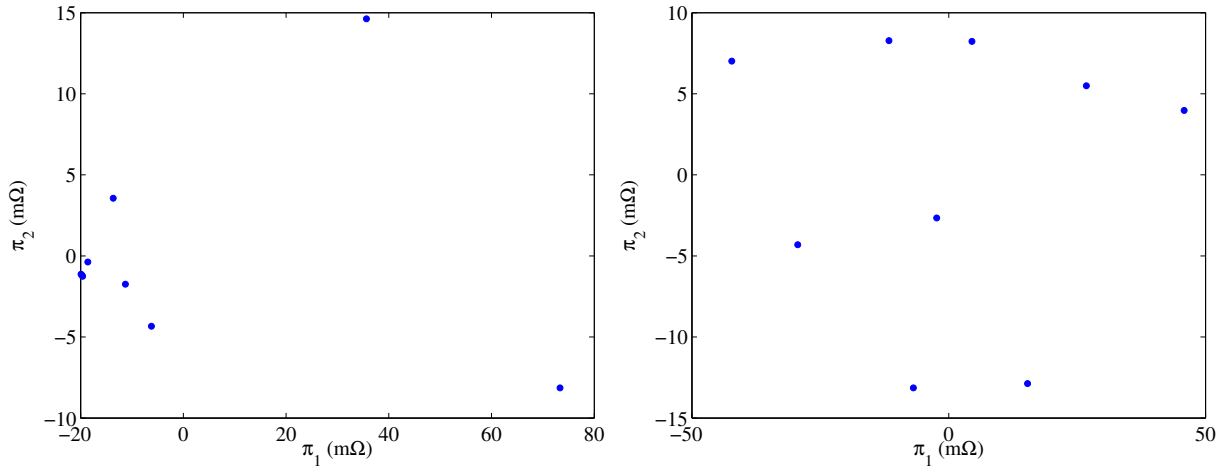
(c) $n = 25$.

Figure 5.4: Interpolation error $\varepsilon(\mathbf{p})$ of the NN interpolators over the input space in the 2-parameter example with different number of samples (colormap in mΩ). Left: full-factorial sampling. Right: output space-filling sampling. (To be continued on the next page.)



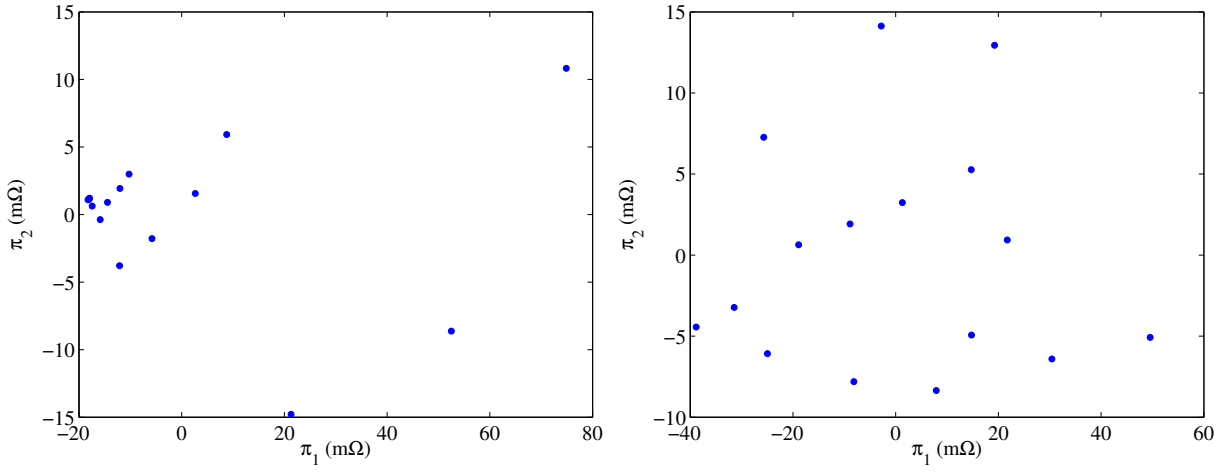
(d) $n = 36$.

Figure 5.4: (Continued.) Interpolation error $\varepsilon(\mathbf{p})$ of the NN interpolators over the input space in the 2-parameter example with different number of samples (colormap in mΩ). Left: full-factorial sampling. Right: output space-filling sampling.

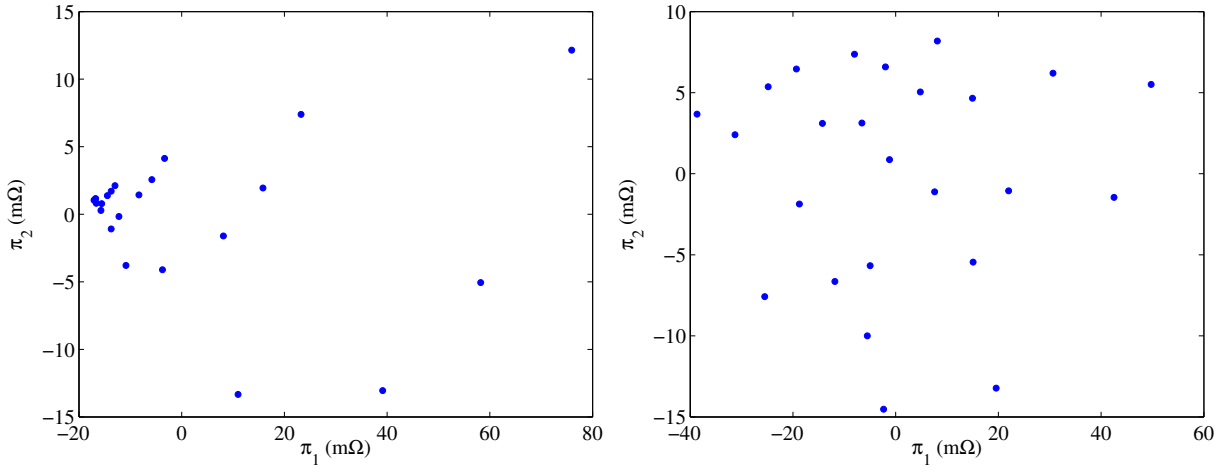


(a) $n = 3 \times 3$.

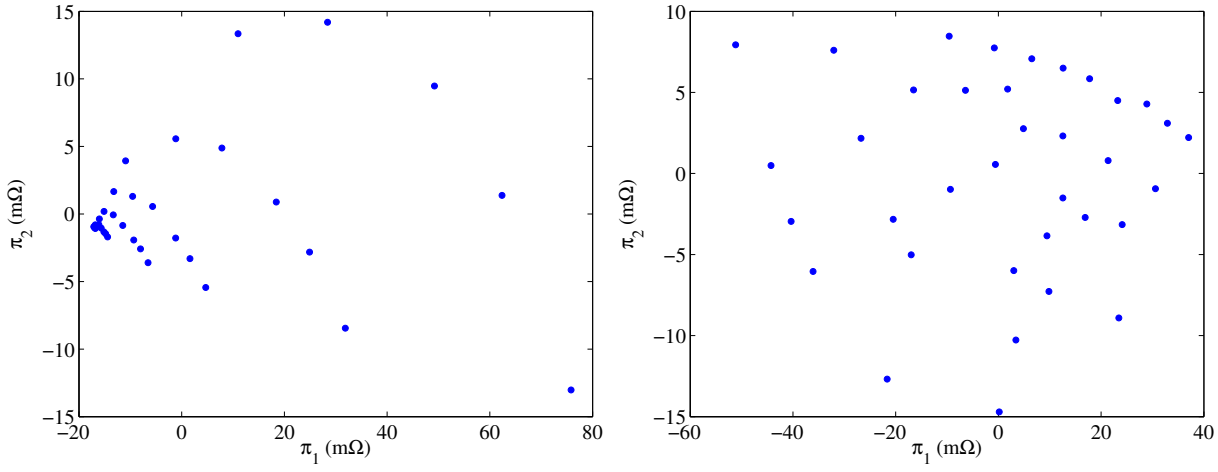
Figure 5.5: 2-dimensional representation of the output samples by multidimensional scaling in the 2-parameter example with different number of samples. Left: full-factorial sampling. Right: output space-filling sampling. (To be continued on the next page.)



(b) $n = 4 \times 4$.

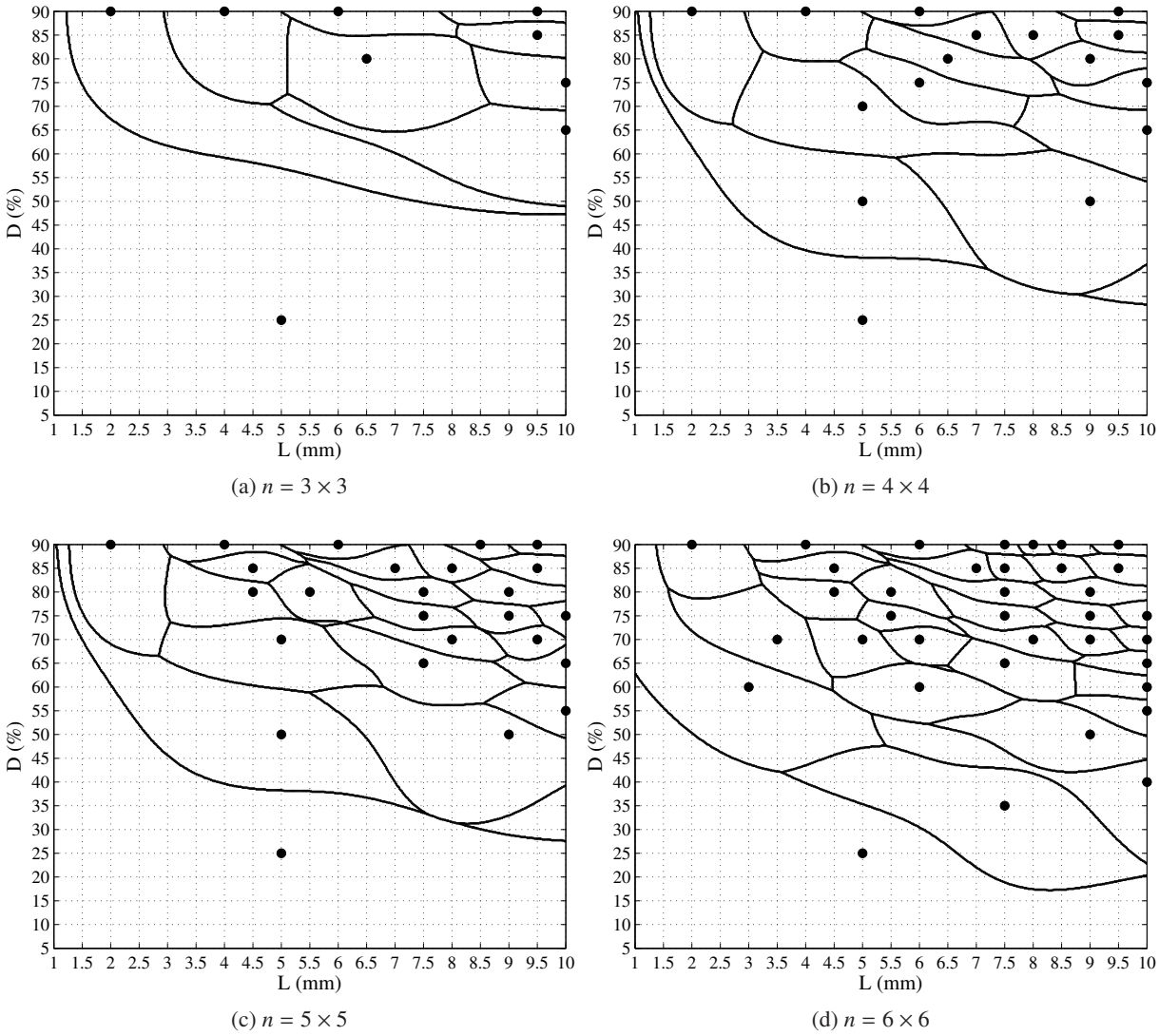


(c) $n = 5 \times 5$.



(d) $n = 6 \times 6$.

Figure 5.5: (Continued.) 2-dimensional representation of the output samples by multidimensional scaling in the 2-parameter example with different number of samples. Left: full-factorial sampling. Right: output space-filling sampling.



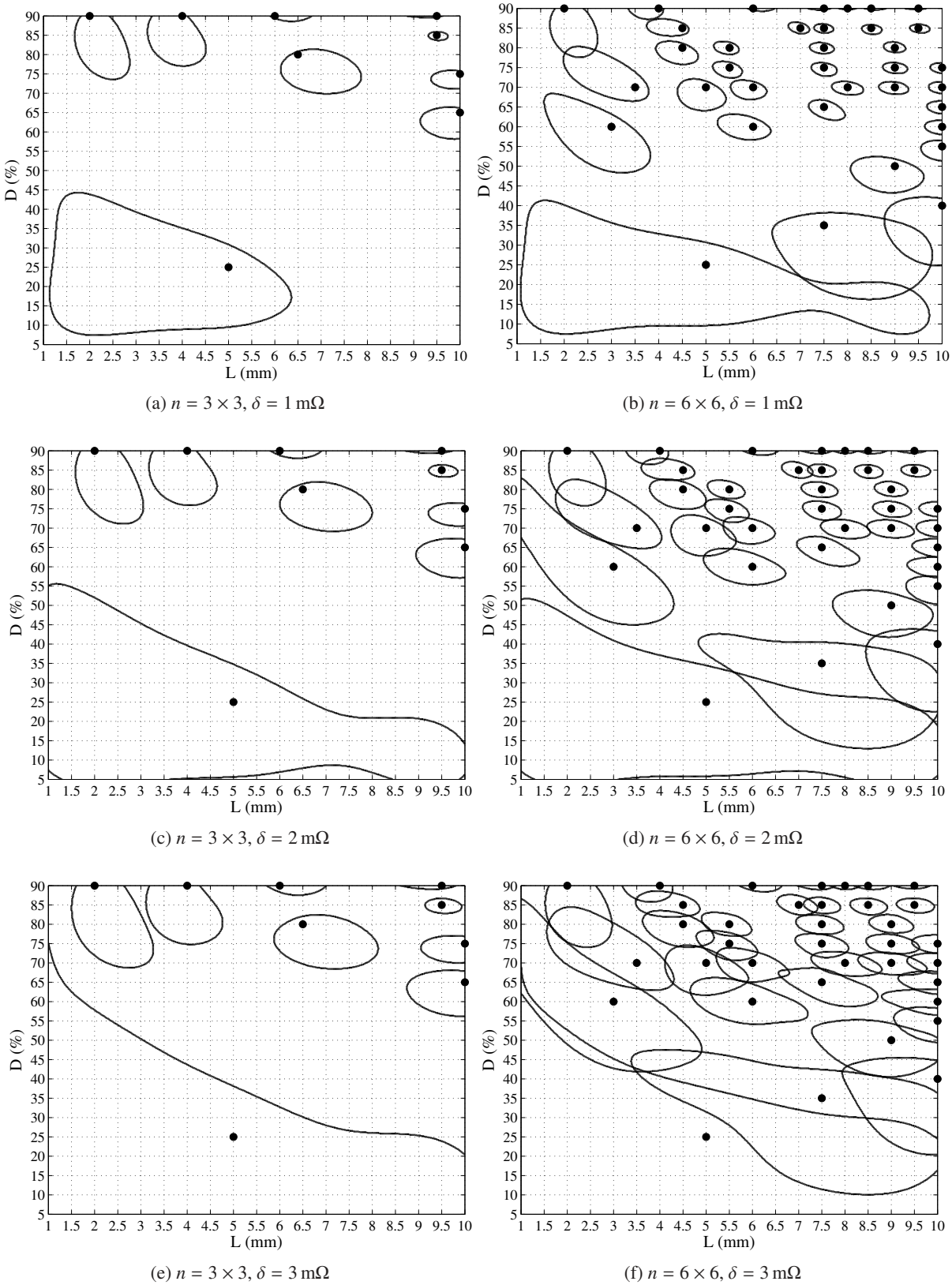
(a) $n = 3 \times 3$

(b) $n = 4 \times 4$

(c) $n = 5 \times 5$

(d) $n = 6 \times 6$

Figure 5.6: Inverse-mapped approximate Voronoi cells in the 2-parameter example. Let us notice the splitting cells of the $L = 5\,\text{mm}$, $D = 50\,\%$ input sample in the cases $n = 4 \times 4$ and $n = 5 \times 5$. The parts at $L = 10\,\text{mm}$, $D = 30\,\%$ correspond to this cell. The input space $\mathbb{P}$ is represented by the nodes of the pale grid lines, however, the inverse projection is performed on a 500×500 fine grid.



(a) $n = 3 \times 3$, $\delta = 1\,\mathrm{m}\Omega$

(b) $n = 6 \times 6$, $\delta = 1\,\mathrm{m}\Omega$

(c) $n = 3 \times 3$, $\delta = 2\,\mathrm{m}\Omega$

(d) $n = 6 \times 6$, $\delta = 2\,\mathrm{m}\Omega$

(e) $n = 3 \times 3$, $\delta = 3\,\mathrm{m}\Omega$

(f) $n = 6 \times 6$, $\delta = 3\,\mathrm{m}\Omega$

Figure 5.7: Inverse-mapped approximate noise levels in the 2-parameter example. The input space $\mathbb{P}$ is represented by the nodes of the pale grid lines, however, the inverse projection is performed on a 500×500 fine grid.

5.5.3 Examples with 4 parameters

The canonical configuration defined as the point 2 in § 2.4 is involved in the presented 4-parameter examples. Let us recall that the position (A, B), the length (L) and the depth (D) of a single crack are enabled to vary. The input space consists of approximately 4×10^4 feasible input vectors.

Numerical characteristics

Databases with three different sample numbers have been generated. The numerical comparison is presented in Table 5.4. Let us note that to compute $\gamma_{\max}$ and $\varepsilon_{\max}$, the approximate formulas (5.25) and (5.31) have been used, with the test set $\mathbb{S}$ being a random Latin hypercube design with 840 points, using maximin criterion[11]. In the sense of $\gamma_{\max}$, the OSF sampling highly outperforms the FF design. However, the interpolation error $\varepsilon_{\max}$ appears to be approximately the same for both samplings. Let us recall, however, this does not mean the failure of the OSF sampling. The reduction of the NN interpolation error might heuristically be expected from the OSF strategy, but cannot be considered as a measure of the performance of the sampling algorithm (§ 5.3.2). In the sense of the ratio $d_{\max}/d_{\min}$ (see formula (5.13)), the OSF sampling highly outperforms the FF databases. The difference is caused by the extremely small $d_{\min}$ distances in the FF design.

Table 5.4: Numerical characteristics of the databases in the 4-parameter example (mΩ). (FF: full-factorial design, OSF: output space-filling.)

Samples (n)	$\gamma_{\max}$ (mΩ)		$\varepsilon_{\max}$ (mΩ)		$d_{\min}$ (mΩ)		$d_{\max}$ (mΩ)		$d_{\max}/d_{\min}$	
	FF	OSF	FF	OSF	FF	OSF	FF	OSF	FF	OSF
$3 \times 3 \times 3 \times 3$	45.1	25.8	69.2	72.5	0.049	20.4	44.1	31.2	903.4	1.53
$4 \times 4 \times 4 \times 4$	29.1	18.5	44.9	44.9	0.032	12.7	30.1	21.8	942.4	1.71
$5 \times 5 \times 5 \times 5$	28.6	15.4	37.1	31.5	0.025	8.05	25.6	20.8	1010.4	2.59

Visualization by multidimensional scaling

The MDS visualization of the output samples has been carried out similarly to the 2-parameter examples. In Fig. 5.8, one can clearly see the clustering output samples in the case of FF sampling. The small values of $d_{\min}$ in Table 5.4 have already referred to this phenomenon, but the MDS plots make clear that indeed many samples cluster in a certain region of the output space.

Interesting regular repartitions of the output samples can also be noticed in the FF-case. Groups with k samples, and aggregations of k groups appear, where k denotes the number of levels along each input parameter in the FF design. This must be caused by the small sensitivity of $\mathcal{F}$ to certain input parameters (in certain regions of the input space). Let us imagine that by changing the ith input parameter from the lower to its upper bound in k steps (the other $N-1$ parameters are fixed), the forward operator yields "similar" outputs: this is how the k output samples form a small group. By fixing the jth parameter to an other value, and changing the ith parameter as before, the k-groups appear somewhere else but still in a regular manner.

Another kind of clustering of the output samples can be noticed in the OSF databases as well. However, this causes much less distortion in the repartition of the samples. In the right-side plot of Fig. 5.8c, 9 "branches" appear, induced by the discretization of the input space $\mathbb{P}$. Let us recall that along the parameter A, 9 levels are defined. The branches on the MDS plot might refer to the fact that a finer discretization along the parameter A would result in output samples being relatively far from the ones in the current output space.

[11] Derived from a continuous LHS (§ D.3).

However, the MDS stress (Table 5.5) is much higher in the 4-parameter case than in the previous example. Consequently, the repartition of the output samples can less properly be represented in the 2-dimensional plane.

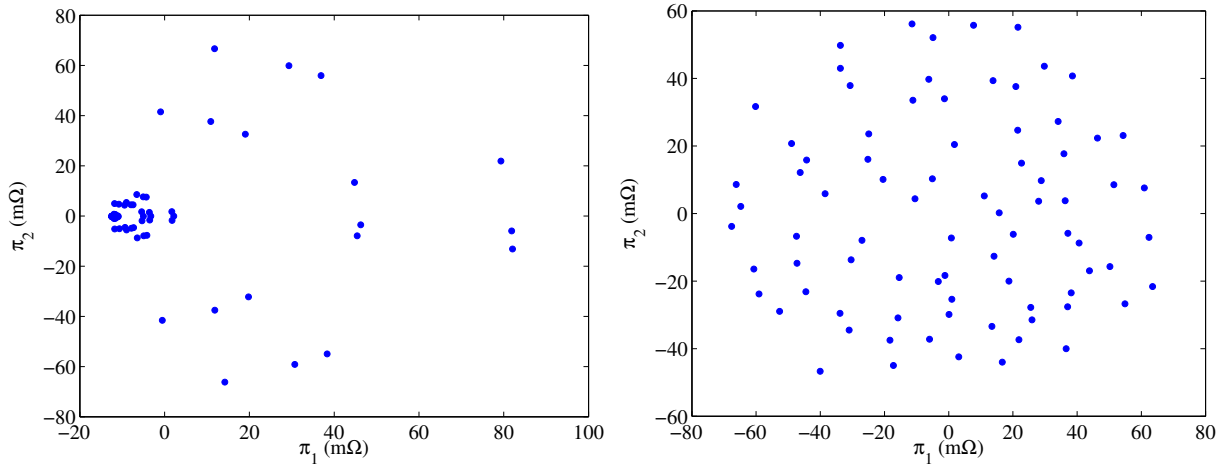
Table 5.5: MDS stress (S) in the 4-parameter example with different number of samples in case of full-factorial and output space-filling sampling.

Samples (n)	FF	OSF
$3 \times 3 \times 3 \times 3$	0.1114	0.2171
$4 \times 4 \times 4 \times 4$	0.1028	0.2133
$5 \times 5 \times 5 \times 5$	0.1002	0.2060

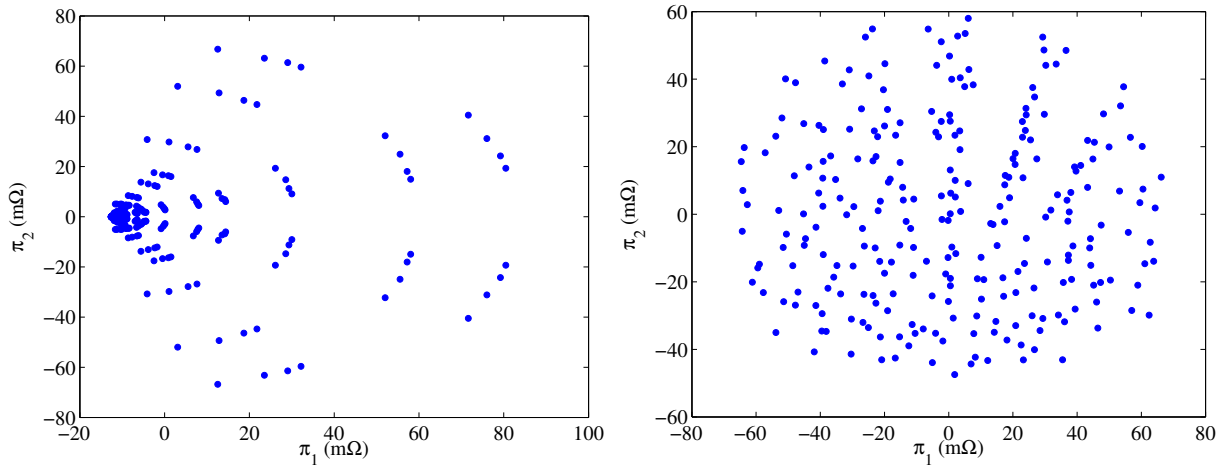
Inverse mappings

In this 4-parameter case, the results of the inverse mappings are difficult to visualize. We present a way only for the noise-patch visualization. In Fig. 5.9, one sample of the OSF database (with 256 samples) is shown, and the approximate noise patches in all the 6 possible cross-sections of the 4-dimensional input space are presented. Let us note that the continuous $\mathbb{U} = [0, 1]^4$ space is introduced again (similarly to the 2-parameter case), the cross-sections are equal to $[0, 1]^2$, and the patches are numerically determined on 200×200 grids. The extensions of the patches along the different input parameter axes are related to the sensitivity of the forward operator to the different input parameters. For instance, the depth D appears to be an influential parameter in all of the 3 cross-sections where it varies. On the contrary, the length L seems to have less effect on the output compared to the 3 other parameters. Let us note that obviously the above conclusions are valid for the chosen particular sample. For other cracks, the experiences might be completely different.

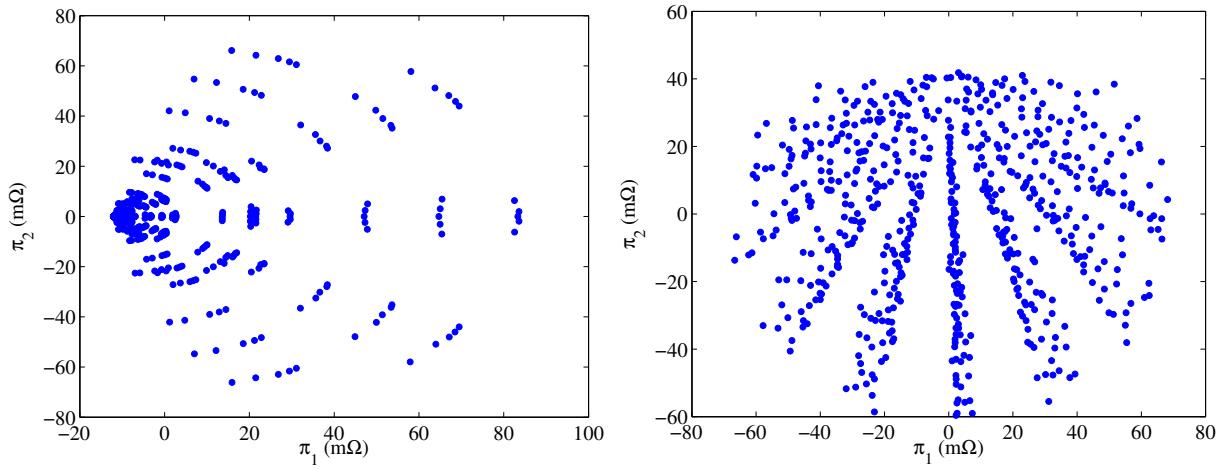
Though there is no limitation concerning the N number of input parameters, let us also note that the determination of the inverse mapped-sets (in either Voronoi or noise mapping) becomes computationally demanding with increasing N, since an exhaustive search has to be carried out over $\mathbb{P}$. A possible remedy could be to explore the "borders" of the inverse-mapped sets (i.e., to find some of the elements being far from the related input sample, or from each other) by using a sort of local search strategy. In so doing, however, the possibly disconnected parts cannot be explored. We do not consider these numerical issues in this Dissertation.



(a) $n = 3 \times 3 \times 3 \times 3$.



(b) $n = 4 \times 4 \times 4 \times 4$.



(c) $n = 5 \times 5 \times 5 \times 5$.

Figure 5.8: 2-dimensional representation of the output samples by multidimensional scaling in the 4-parameter example with different number of samples. Left: full-factorial sampling. Right: output space-filling sampling.

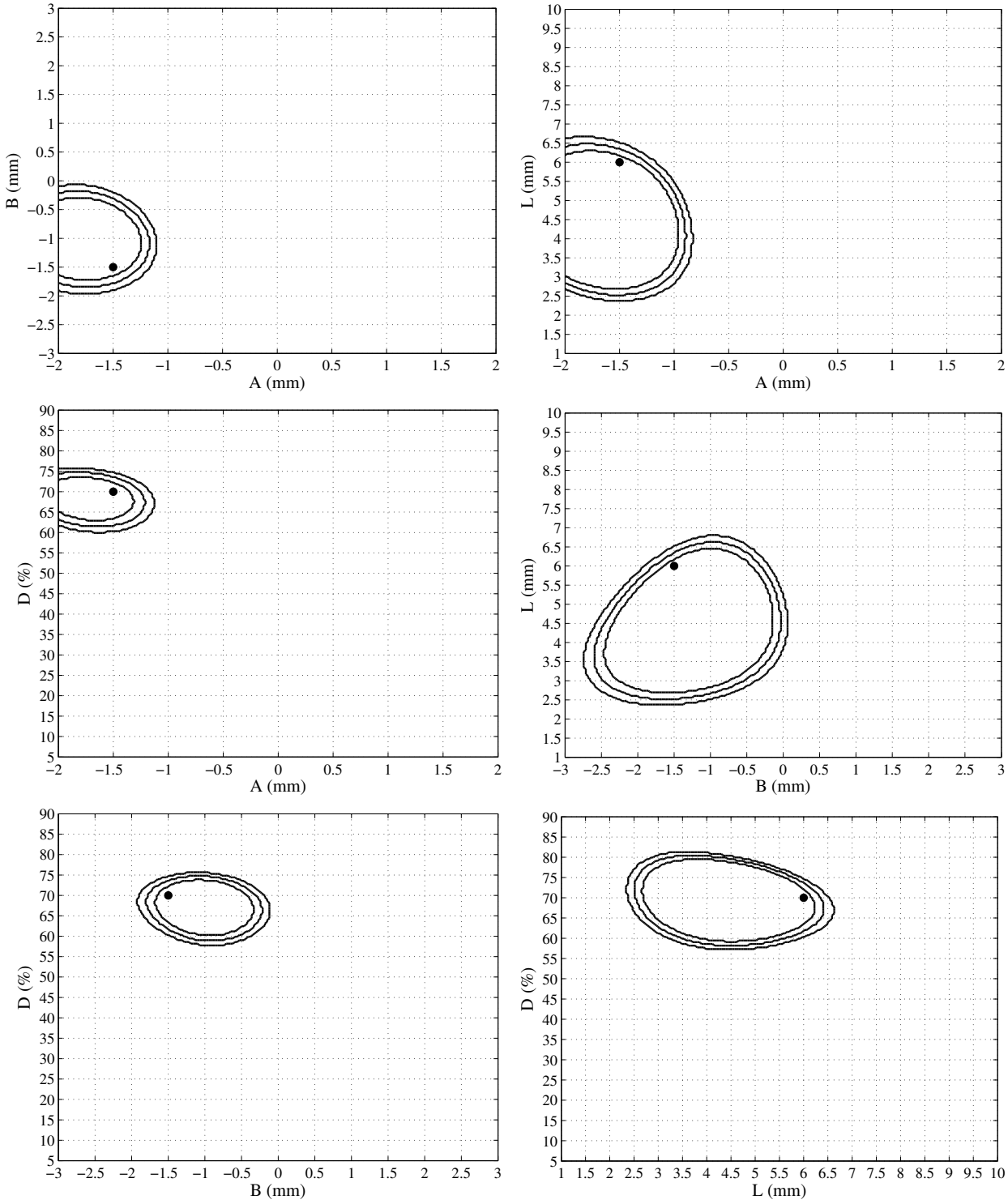


Figure 5.9: Cross-sections of the inverse-mapped approximate noise levels in the 4-parameter example with three different noise levels ($\delta = 1, 2, 3\,\mathrm{m}\Omega$) for one single sample ($A = B = -1.5\,\mathrm{mm}$, $L = 6\,\mathrm{mm}$, $D = 70\,\%$). The input space $\mathbb{P}$ is represented by the nodes of the pale grid lines, however, the inverse projection is performed on 200×200 fine grids in each cross-section.

5.5.4 Examples with 6 parameters

In the 6-parameter examples, the double-crack canonical configuration is involved (defined as the point 3 in § 2.4). Let us recall that the positions (A, h, B_1 and B_2) and the lengths (L_1, L_2) of two parallel cracks are enabled to vary, whereas the depth of both cracks is assumed to be 40 % of the plate thickness. The input space consists of approximately 4.39×10^6 feasible input vectors. The inverse mappings are not presented in this example, due to the difficulty of expressive visualization with 6 input parameters.

Numerical characteristics

In this example, only one database size is considered: an OSF database of 729 samples is compared to the FF sampling with the same number of points. The numerical results are given in Table 5.6. Similarly to the previous example, for the computation of $\gamma_{\max}$ and $\varepsilon_{\max}$, the approximate formulas (5.25) and (5.31) have been used, with the test set $\mathbb{S}$ being a random Latin hypercube design with 3249 points, using maximin criterion[12]. Both $\gamma_{\max}$ and $\varepsilon_{\max}$ are smaller for the OSF sampling, however, in the sense of $\gamma_{\max}$ (being the more expressive characteristic), the difference is not as significant as in the previous examples. The difference in terms of $d_{\max}/d_{\min}$ is still considerable, for the good of the OSF strategy.

Visualization by multidimensional scaling

The MDS plots of the output sample distribution are shown in Fig. 5.10. Though the FF design still yields many output samples clustering in a certain region of $\mathbb{Q}$, the space-filling in terms of $\gamma_{\max}$ does not appear as bad as in the previous examples. Indeed, the numerical results in Table 5.6 confirms that (i) the clustering is much stronger in the FF database ($d_{\max}/d_{\min}$); (ii) the covering of the output space with samples is not much worse by using FF design than OSF sampling ($\gamma_{\max}$). The MDS stress (Table 5.7) is even higher than in the previous 4-parameter example, meaning that the 2-dimensional visualization of the sample-set cannot be done without a considerable loss of information.

The reason why FF yields better results than in the previous cases is the behaviour of the forward operator: it must behave more uniformly all over $\mathbb{P}$, i.e., its sensitivity to the different input parameters does not change much over $\mathbb{P}$. In the 6-parameter example, the depths of the cracks do not vary. However, in the previous examples, the depth appeared to be the most influential parameter (i.e., $\mathcal{F}$ itself and its sensitivity for the other input parameters were the most sensitive for D). Let us recall that in the previous chapter, when the interpolation based on the database was in the focus, we have found that the best interpolator in this 6-parameter example was the piecewise multilinear scheme (§ 4.8.4). Consequently, $\mathcal{F}$ appears to be close to a linear operator over $\mathbb{P}$, which also confirms the good performance of the FF design in the present study.

Table 5.6: Numerical characteristics of the databases in the 6-parameter example (mΩ). (FF: full-factorial design, OSF: output space-filling.)

Samples (n)	$\gamma_{\max}$ (mΩ)		$\varepsilon_{\max}$ (mΩ)		$d_{\min}$ (mΩ)		$d_{\max}$ (mΩ)		$d_{\max}/d_{\min}$	
	FF	OSF	FF	OSF	FF	OSF	FF	OSF	FF	OSF
$3 \times 3 \times 3 \times 3 \times 3 \times 3$	3.56	3.01	8.58	8.54	0.18	2.65	2.96	3.80	16.5	1.43

[12] Derived from a continuous LHS (§ D.3).

Table 5.7: MDS stress (S) in the 6-parameter example with different number of samples in case of full-factorial and output space-filling sampling.

Samples (n)	FF	OSF
$3\times3\times3\times3\times3\times3$	0.2238	0.2504

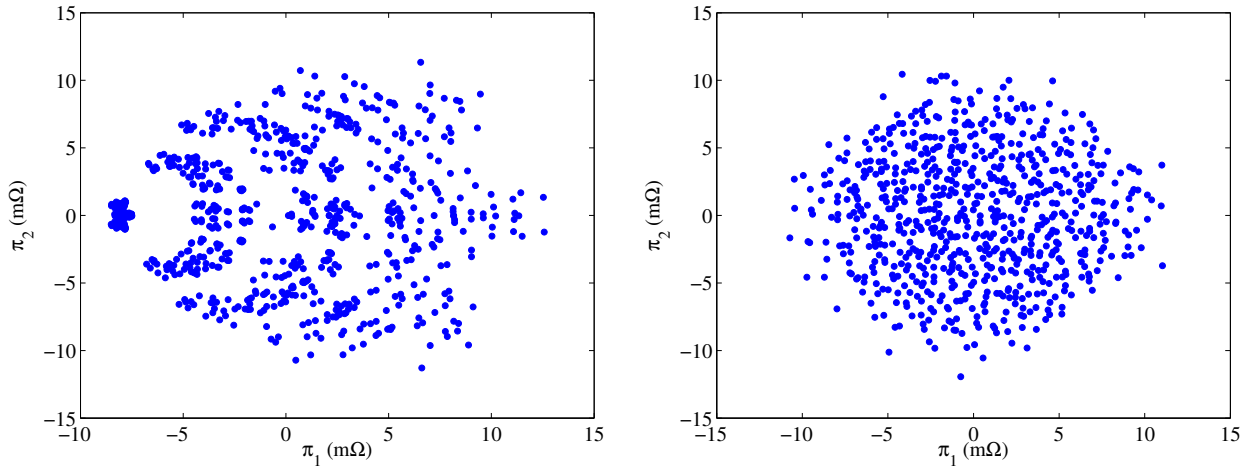


Figure 5.10: 2-dimensional representation of the output samples by multidimensional scaling in the 6-parameter example using 729 samples in both databases. Left: full-factorial sampling. Right: output space-filling sampling.

5.5.5 Discussion on the time consumption

In the presented numerical examples, the CWS-based evaluation of the sample-insertion criterion is by far the most time-consuming issue among all tasks related to the sampling (the evaluation of the forward operator is independent from the sampling algorithm, its time consumption is out of our scope herein).

In Fig. 5.11, the time consumption of the CWS optimization is plotted in function of the iteration number. Let us note that in Figs. 5.11a and 5.11c, the last iteration yields the final database with $n = 36$ and $n = 729$, respectively. However, in Fig. 5.11b, the performance is shown up to a sample number of $n = 788$, on trial.

In the 2-parameter case, the time consumption is not considerable but it appears to grow fast with the iterations. The stochastic behaviour of the time is obviously due to the different number of steps until the convergence of the CWS procedures.

In Fig. 5.11b, a rapidly growing time consumption can be seen. By the end of the iterations, the runtime of the multistart CWS scheme exceeded the 25 min. Let us note, however, that in the last cycles, more than 1000 points were used as starting points for the optimization. The total time spent on the CWS was approximately 13 days. This extreme time consumption inspired the construction of the implementation II (§ 5.2.3), which indeed resulted in a faster database generation procedure, even in the case of higher number of input parameters (Fig. 5.11c). In the latter 6-parameter example, the total time of CWS optimization was 4.5 days.

5.6 Conclusions and perspectives

The presented sampling strategy can efficiently be used to generate output-space filling databases as surrogate models. The output samples in an OSF database (as we defined herein) cover well the output space,

whereas the distances between the neighbouring output samples are approximately equal all over the output space. Besides generally reducing the error of the nearest neighbour interpolation, such OSF databases provide valuable meta-information on the modeled problem, via the repartition of the samples. However, let us note that the precision of the yielded surrogate models using the NN interpolation definitely lags behind the precision of the functional kriging-based approach discussed in § 4. Consequently, we believe that the main benefit of the OSF method is the exploited meta-information.

An important contribution in the presented work is the introduction of the distance functions, providing the following benefits:

- One does not need functional kriging for the database generation, even if the output data are indeed functional.
- All distance functions can be assumed to behave similarly, i.e., the same random process can be used to predict all of them in the kriging framework. This yields a very simple realization.
- The inverse-mapped Voronoi cells and noise levels are defined by simple inequalities in terms of the distance functions. The inverse mappings appear to be very powerful tools for the quantitative characterization of the inverse problem.
- The distance matrix (with entries being the values of the distance functions at the sampled locations) can directly be used to perform MDS for the low-dimensional visualization of the output samples.

The research on the OSF sampling strategies and on the use of the databases is far from finished. Several perspectives have been occurred during the studies we carried out.

- If the central goal of the database generation is to obtain as precise as possible interpolations based on the sample-set, using NN interpolation, one might imagine a sampling strategy being directly driven by the step-by-step reduction of the interpolation error (5.29). The window functions are easy-to-compute and the distance functions can be approximated by kriging, similarly to the presented approach. Thus, by a slight modification of the sample-insertion criterion, the sampling algorithm would be able to generate databases optimized for NN interpolation.
- The sample-removal algorithm might be improved by relying on the ratio $d_{\max}/d_{\min}$ instead of $d_{\min}$. The sample which would decrease the most the ratio $d_{\max}/d_{\min}$ by its removal, would then be removed.
- One might consider to take the positivity-constraint of the distance functions into account in the kriging model ("positive-kriging", see also the perspectives in § 3.6 and, e.g., [Chilès and Delfiner, 1999]).
- We have applied the OSF sampling strategy for another physical problem: radar observations of forested areas. This is not presented in detail in the Dissertation, as only some preliminary results have been obtained so far, but a brief summary is given in § E.
- One might consider more sophisticated noise models in the inverse-mapping. So far we assumed a sort of uniform distribution for the noise. A simple extension would be to assume a non-uniform, but still "radial" distribution in the sense that the Probability Density Function (PDF) depends only on the distance from the true (noise-free) data. Let $\tilde{q}(t) = q_{\mathbf{p}_k}(t)$ the measured noisy data (being equal to the kth output sample in a database[13]) and $F(\delta)$ be a given cumulative distribution function defined as

$$F(\delta) = \mathrm{Prob}\left(\delta > \left\|q(t) - q_{\mathbf{p}_k}(t)\right\|\right), \tag{5.50}$$

  where $q(t)$ is the true (noise-free) output data. Let us perform the inverse mapping of n noise levels

$$0 < \delta_1 < \delta_2 < \cdots < \delta_n, \tag{5.51}$$

[13] As it was always assumed in the inverse mapping of noise levels, § 5.4.3.

yielding n noise patches Φ^{δ_i} $(i = 1, 2, \ldots, n)$. Let $w(\mathbf{p})$ a function over the input space $\mathbb{P}$ defined as

$$w(\mathbf{p}) = \frac{F(\delta_{i+1}) - F(\delta_i)}{|\Phi^{\delta_{i+1}} \setminus \Phi^{\delta_i}|}, \quad \text{if } \mathbf{p} \in \Phi^{\delta_{i+1}} \setminus \Phi^{\delta_i}, \tag{5.52}$$

where the symbol $|\cdot|$ denotes the cardinality of a set. The $w(\mathbf{p})$ function can then be considered as an estimator of a PDF over the input space, being related to the solution of the inverse problem.

In a more general sense, we believe that some links can be established between the OSF strategy and the modern research on uncertainty propagation. The latter still appears to be quite application-oriented; for lack of a general recent textbook, let us refer only the up-to-date theses [Blatman, 2009; Sudret, 2007].

- In the practice of kriging, stationary covariance models are much more widely applied than non-stationary ones. This is due to the several theoretical and computational challenges arising in the case of the latter. However, the assumption of stationarity is controvertible in many practical cases, when the modeled function appears to behave quite diversely in the different regions of its domain.

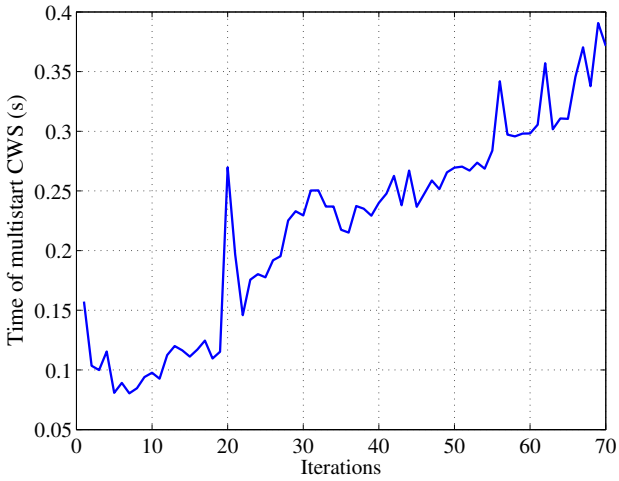
  Some attempts have been made to construct non-stationary covariance models (e.g., [Xiong et al., 2007]). A very preliminary idea might also be inspired by the OSF databases. As it was pointed out in this chapter, the "density" of the input samples is related to the "slope" of the forward operator. This density can be formalized via a density function $\zeta(\mathbf{p})$, being strictly positive and a higher $\zeta(\mathbf{p})$ indicates higher density, i.e, faster variation of $\mathcal{F}$. $\zeta(\mathbf{p})$ can be estimated, e.g., by using a sort of kernel density estimator [Wand and Jones, 1995]. An inhomogeneous distance measure $\mu(\cdot, \cdot)$ can then be defined in the input space as

$$\mu(\mathbf{p}_a, \mathbf{p}_b) = \|\mathbf{p}_b - \mathbf{p}_a\| \int_0^1 \zeta\left[\mathbf{p}_a + (\mathbf{p}_b - \mathbf{p}_a)u\right] \mathrm{d}u. \tag{5.53}$$

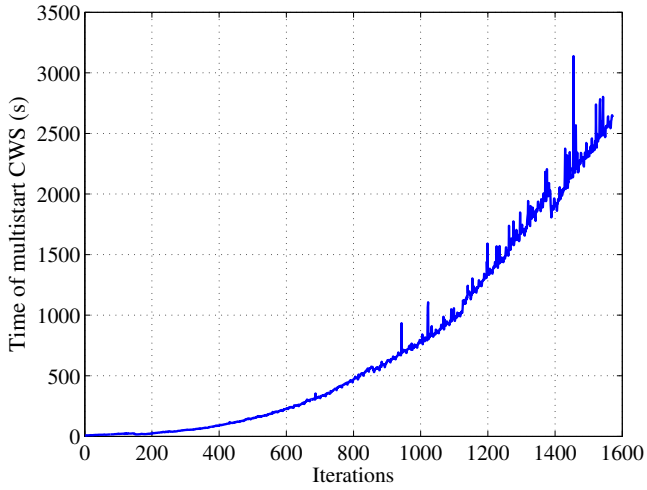
  This re-definition of the distance makes possible the construction of a non-stationary covariance model simply by defining the covariance as

$$\mathrm{cov}(\mathbf{p}_a, \mathbf{p}_b) \coloneqq k(\mu(\mathbf{p}_a, \mathbf{p}_b)), \tag{5.54}$$

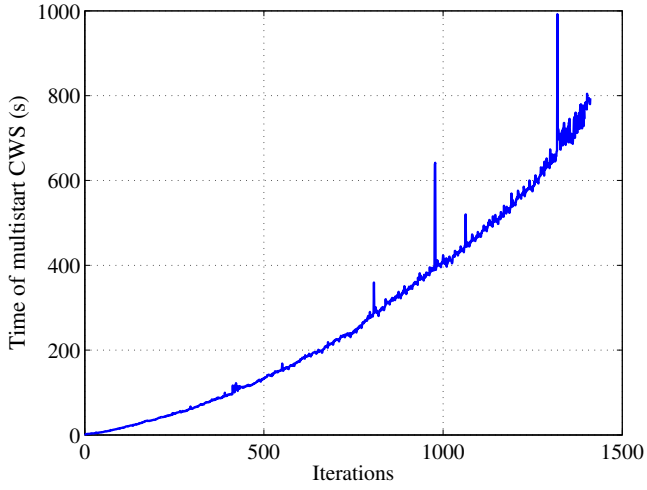
  where $k(\cdot)$ is an arbitrary stationary covariance model, e.g., the Matérn function (§ B). However, the inter-sample covariance matrix (B.11) is not necessarily positive definite any more, moreover, it might be singular, since $\mu(\cdot, \cdot)$ is only a semi-metric (the triangular inequality [Steen and J. Arthur Seebach, 1970] might be violated).



(a) 2-parameter example. (Scheme I.)



(b) 4-parameter example. (Scheme I.)



(c) 6-parameter example. (Scheme II.)

Figure 5.11: Time consumption of the CWS optimization in the different examples.

6

Thesis points: summary of the new scientific achievements

Thesis 1 and its short explanation

I have devised an inversion method for nondestructive evaluation by combining a numerical simulator of the electromagnetic phenomena and the "Efficient Global Optimization" (EGO) algorithm, which inversion scheme needs only a small number of forward simulations to solve an inverse problem even if the number of parameters to be retrieved is relatively high.

A classical inversion method is to compare the measured data to the output of an appropriate simulator and to reduce the discrepancy between the measured and simulated data. The solution of the (regularized) inverse problem is then those combinations of the input parameters of the simulator which yield the "smallest" discrepancy. One of the main bottlenecks of this optimization-based inversion approach is that the numerical simulators are usually computationally expensive. Thus, on one hand, the number of simulations needed to solve an inverse problem should be small. On the other hand, however, the yielded objective function (to be minimized) might be multimodal, claiming the use of a global optimization method.

I have applied the EGO algorithm for the first time in a classical optimization-based inversion scheme to solve eddy-current testing inverse problems. The EGO algorithm –being based on the kriging model of the objective function– has been designed for the optimization of expensive-to-evaluate objective functions as the number of observations needed to obtain the global minimum is usually small (compared to certain stochastic global optimization strategies).

To sum up the advantages of the approach, we can emphasize (i) the relatively small number of forward simulations needed to solve the inverse problem and (ii) the convergence of the EGO algorithm to the global minimum of the objective function. Further advantages are provided by the implementation based on a coordinatewise search algorithm on a discrete grid: (iii) the required precision of the solution of the inverse problem can be taken into account to some extent; (iv) the spatial discretization used in the integral equation-based numerical simulator can also be taken into account in the inversion scheme, thus, considerable computation time can be saved.

The detailed discussion of this thesis point is found in § 3. This research is reported in the following publications:

- Bilicz et al., 2009a (peer-reviewed journal paper): the characterization of a volumetric defect;
- Bilicz et al., 2009b (conference paper): characterization of a pair of surface cracks described by 6 parameters;
- Bilicz et al., 2009d (conference paper): a brief summary with illustration of surface crack inversion.

- Bilicz et al., 2011b (accepted conference paper): a brief summary in the frame of an overview of kriging-based surrogate modeling.

Thesis 2 and its short explanation

I have devised a surrogate modeling method for nondestructive evaluation which uses functional kriging to interpolate the forward operator based on a set of observations chosen adaptively in order to improve the precision of the surrogate model being built.

Given an expensive-to-run forward operator (e.g., realized by the numerical simulation of electromagnetic phenomena), a natural way of constructing a fast approximation (a surrogate model) is to evaluate the forward operator at certain well-chosen input values, then to fit an interpolator to the observed data. Obviously, the interpolator should provide results as close as possible to the true output of the forward model at unsampled input values.

I have applied the functional kriging technique to build interpolation-based surrogate models for electromagnetic nondestructive testing. Functional kriging –a recent extension of the original kriging theory– makes possible the prediction of whole functions as single entities, instead of pointwise predictions. Since the models involved in electromagnetic nondestructive evaluation often have a functional output, the proposed approach is convenient.

The precision provided by the surrogate model is strongly influenced by the choice of the observations on which the interpolation is based. I have devised an adaptive sampling strategy –using jackknife variance estimation– aiming at reducing the interpolation error of the functional kriging interpolator being built.

The detailed discussion of this thesis point is found in § 4. This research is reported in the following publications:

- Bilicz et al., 2010e (peer-reviewed journal paper): the use of functional kriging to predict the output data of surface crack ECT models based on a full-factorial design;
- Bilicz et al., 2010b (conference paper): the extension of the previous contribution with the adaptive sampling strategy.
- Bilicz et al., 2011b (accepted conference paper): a brief summary in the frame of an overview of kriging-based surrogate modeling.
- Bilicz et al., 2011a (accepted conference paper): an application of the functional kriging-based surrogate model to build a new surrogate for “inverse interpolation”.
- Vaskó et al., 2011 (accepted conference paper): functional kriging-based surrogate modeling applied to the direct problem of radar observations of forested areas.

Thesis 3 and its short explanation

I have devised an adaptive sampling method for the generation of output space-filling defect-databases for nondestructive evaluation, along with two techniques to exploit certain meta-information provided by such output space-filling databases for the quantitative characterization of the related inverse problem.

A database consisting of input parameter - output data pairs can be considered as a sort of discrete representation of the involved forward operator. Such databases can form the basis of cheap surrogate models, by fitting an interpolator to the stored samples. However, the choice of the samples (i.e., the sampling strategy) strongly influences the precision of the yielded surrogate model.

Classical Design-Of-Experiment provides numerous tools to generate sample sets satisfying certain criteria related to the repartition of the input samples (e.g., space-filling, uniformity). My main contribution in the frame of this thesis point is the extension of the input space-filling to the output space in a certain sense. I have devised a sampling strategy which aims at building a database whose output samples fill the output space evenly: (i) none of the output samples is "too close" to an other one; (ii) for all elements of the output space, an output sample exists "not too far". The sampling strategy consists in the alternating use of subsequent sample insertion and removal, controlled by the pairwise distances of the output samples. To reduce the computational cost of the procedure, the kriging interpolation is used to approximate those distances.

The benefit of such output space-filling databases is twofold: (i) generally, the interpolation error of the nearest neighbour interpolator is smaller compared to using input space-filling databases; (ii) the repartition of the input samples carries some meta-information on the modeled forward operator (e.g., indicates the regions where the forward operator is flat or varies rapidly).

Based on the distance functions (also involved in the sampling algorithm) and their kriging interpolation, I have devised two methods for the quantitative characterization of the underlying inverse problem. The first method aims at the inverse mapping of a given noise level from the output space to the input space. The inverse-mapped noise level characterizes the ill-posedness of the inverse problem by indicating how the uncertainty of the measured output influences the uncertainty of the reconstructed input parameters. The second method maps the Voronoi cells in the output space into the input space. To a certain extent, the inverse-mapped Voronoi cells can be considered as the representation of the "resolution" provided by the studied database when addressing the inverse problem.

An interesting study has also been carried out (briefly summarized in § E): the database generation strategy has been applied to the forward problem of radar observations of forested areas. The input parameters were the frequency and the incident angle of the incident wave, the output data consisted of certain quantities derived from the scattered EM field (back-scattering coefficient, attenuation, interferometric height).

The detailed discussion of this thesis point is found in § 5. This research is reported in the following publications:

- Bilicz, Lambert, and Gyimóthy, 2010b (peer-reviewed journal paper): basically contains all we presented in § 5 (except the MDS visualization);
- Bilicz et al., 2010d (peer-reviewed journal paper): a former and the present version of the OSF database generation, without the inverse mappings;
- Bilicz et al., 2010a (peer-reviewed paper in edited book): a former version of database generation (without the sample-removal step);
- Bilicz, Lambert, and Gyimóthy, 2010a (conference paper): focuses on the inverse mappings and the characterization of the inverse problems;
- Bilicz et al., 2010c (conference paper): a shorter version of the previous contribution;
- Thirion et al., 2010 (conference paper): application of the OSF sampling to forest characterization.
- Bilicz et al., 2011b (accepted conference paper): a brief summary in the frame of an overview of kriging-based surrogate modeling.

A

Mathematical model and numerical simulation of the studied ECT configuration

In this chapter, the principles of the ECT simulator applied in the numerical studies of this Dissertation are briefly described. Let us highlight that the development of this simulator is not part of the presented PhD work, but the whole program has been provided by József Pávó, Budapest University of Technology and Economics. A thorough discussion of the applied modeling and simulation tools can be found in his PhD dissertation [Pávó, 1994].

A.1 The mathematical model

The model of the electromagnetic interaction of the exciting coil with an infinitesimally thin crack is described in [Bowler, 1994]. This model assumes the cracks in the otherwise homogeneous conductor as mathematical surfaces that block the flow of the eddy currents across their surfaces, in other words the normal component of the electric field must be zero on the surfaces representing the cracks. In the same time, the magnetic field is assumed to be continuous on the two sides of these surfaces. Though the approach is not limited to planar cracks, we restrict ourself to this case, as our implementation can only handle cracks with planar surfaces.

The above boundary conditions can be satisfied by replacing the mathematical surface(s) of the crack(s) with current dipole distribution(s), having the same effect as the presence of the crack. In the case of one single crack, let us denote its surface by S and the current dipole distribution by $p(\mathbf{r})\hat{\mathbf{n}}$, respectively, with $\hat{\mathbf{n}}$ being the unit normal vector of S. Thus, the sum of the electromagnetic field generated by the exciting coil and the secondary sources in the homogeneous conductor will provide the field generated by the crack-probe interaction. Consequently, this field has to satisfy the above boundary conditions. The described considerations yield the following integral equation:

$$0 = E_n^i(\mathbf{r}) - j\omega\mu_0 \lim_{\mathbf{r}\to\mathbf{r}_\pm} \iint_S g(\mathbf{r}|\mathbf{r}')p(\mathbf{r}')\mathrm{d}\mathbf{r}', \quad \mathbf{r} \in S, \tag{A.1}$$

where $\mathbf{r}_\pm$ denotes the limiting values of the approaches towards the point $\mathbf{r}$ from the $\hat{\mathbf{n}}$ or $-\hat{\mathbf{n}}$ directions, ω is the angular frequency of the sinusoidal excitation of the probe coil and μ_0 is the permeability of the vacuum. $E_n^i = \hat{\mathbf{n}} \cdot \mathbf{E}^i$, where $\mathbf{E}^i$ is the so-called incident field, i.e., the electric field generated by the current of the exciting coil in the defect-free plate. $g(\mathbf{r}|\mathbf{r}') = \hat{\mathbf{n}} \cdot \mathbf{G}(\mathbf{r}|\mathbf{r}') \cdot \hat{\mathbf{n}}$, where $\mathbf{G}(\mathbf{r}|\mathbf{r}')$ is the Green's dyad satisfying

$$\nabla \times \nabla \times \mathbf{G}(\mathbf{r}|\mathbf{r}') - j\omega\mu_0\sigma_0\mathbf{G}(\mathbf{r}|\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')\mathbf{I}, \tag{A.2}$$

with σ_0 being the specific conductivity of the plate; $\delta(\mathbf{r}-\mathbf{r}')$ being the three dimensional Dirac delta function and $\mathbf{I}$ is the unit-dyad, respectively.

In the case of double cracks, the above considerations can be extended (see [Pávó, 2000]) by introducing the two crack surfaces S_1 and S_2, with the two corresponding unit normal vectors $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$, respectively. Instead of the previous single integral equation, here we have a system of two integral equations:

$$\begin{aligned} 0 &= E^i_{n1}(\mathbf{r}_1) - j\omega\mu_0 \iint_{S_2} g_{21}(\mathbf{r}_1|\mathbf{r}')p_2(\mathbf{r}')\mathrm{d}\mathbf{r}' - j\omega\mu_0 \lim_{\mathbf{r}\to\mathbf{r}_{1\pm}} \iint_{S_1} g_{11}(\mathbf{r}|\mathbf{r}')p_1(\mathbf{r}')\mathrm{d}\mathbf{r}', \quad \mathbf{r}_1 \in S_1, \\ 0 &= E^i_{n2}(\mathbf{r}_2) - j\omega\mu_0 \iint_{S_1} g_{12}(\mathbf{r}_2|\mathbf{r}')p_1(\mathbf{r}')\mathrm{d}\mathbf{r}' - j\omega\mu_0 \lim_{\mathbf{r}\to\mathbf{r}_{2\pm}} \iint_{S_2} g_{22}(\mathbf{r}|\mathbf{r}')p_2(\mathbf{r}')\mathrm{d}\mathbf{r}', \quad \mathbf{r}_2 \in S_2, \end{aligned} \tag{A.3}$$

where $E^i_{nk} = \hat{\mathbf{n}}_k \cdot \mathbf{E}^i$ $(k = 1, 2)$, with $\mathbf{E}^i$ being the incident field. $g_{kl}(\mathbf{r}|\mathbf{r}') = \hat{\mathbf{n}}_l \cdot \mathbf{G}(\mathbf{r}|\mathbf{r}') \cdot \hat{\mathbf{n}}_k$ $(k, l = 1, 2)$, where $\mathbf{G}(\mathbf{r}|\mathbf{r}')$ is the Green's dyad.

Once the current dipol density $p(\mathbf{r})$ (or $p_1(\mathbf{r})$ and $p_2(\mathbf{r})$) is determined via the above integral equation(s), the ΔZ impedance variation of the probe coil due to the presence of the crack(s) can be calculated based on the reciprocity theorem [Bowler, 1994; Harrington, 1961]:

$$\Delta Z = -\frac{1}{I^2} \iint_{S_k} E^i_n(\mathbf{r})p(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{A.4}$$

in the single-crack case, or

$$\Delta Z = -\frac{1}{I^2} \left(\iint_{S_1} E^i_{n1}(\mathbf{r})p_1(\mathbf{r})\mathrm{d}\mathbf{r} + \iint_{S_2} E^i_{n2}(\mathbf{r})p_2(\mathbf{r})\mathrm{d}\mathbf{r} \right), \tag{A.5}$$

in the case of two cracks, where I is the current of the exciting coil.

Since the probe coil is axisymmetric and the turns are parallel to the examined plate, the incident field can analytically be expressed by using the model published in [Dodd and Deeds, 1968].

A.2 The numerical simulation

The integral equations (A.1) and (A.3) are numerically solved by using the Method of Moments [Harrington, 1968]. The unknown current dipole density $p(\mathbf{r})$ (or $p_1(\mathbf{r})$ and $p_2(\mathbf{r})$) is approximated as a sum of linear basis functions (each of them being defined over a small rectangular surface element, specified by a regular grid over the whole surface of the crack). The integral equation is tested by weighting functions being the same as the basis functions (Galerkin method). The kernel of the integral equation (related to the Green's function) is evaluated by the formulae published in [Pávó and Miya, 1994]. The integral equation then boils down to a set of linear equations, with the unknowns being the coefficients of the basis functions in the expansion of the current dipole density.

If the sizes of the surface elements in the discretization scheme are fixed, one can compute the possible entries of the system matrix of the linear system of equations in advance. Furthermore, if the coil positions (relative to the crack(s)) are also pre-defined, similar pre-computation is possible for the elements of the excitation vector of the linear system. In so doing, considerable computation time can be saved, since whenever a new crack is to be simulated, one only recourses to the pre-computed data and assemble the equation system. The simulation time is then practically equal to the time of the solution of the yielded system of linear equations. We use this pre-defined discretization scheme, its effect on the forward operator are discussed in § 2.2.

A.3 Some impedance maps

The results of the described solution of the forward problem have been tested against measurements available for single artificial cracks (unfortunately, measured results on two parallel cracks are not available in the literature). These comparisons demonstrated the validity of the results [Pávó, 1994].

To give an idea to the reader how the coil impedance can change due to a defect in the examined specimen, the typical behaviour of the impedance signals is presented herein. The magnitude of the impedance variation is typically smaller than about 1 % of the magnitude of the total impedance. In the practice, this might raise difficulties in the measurement, however, we do not deal with such issues herein. Let us refer to the two canonical configurations defined in § 2.1, being used in the following examples.

In the case of one single crack, ΔZ is plotted in Fig. A.1. The crack is centred under the scanned rectangular surface, its depth is 40 % of the plate thickness and its length is set to three different values. The strongly nonlinear dependence of the coil impedance on the length of the crack is notable.

In Fig. A.2, double cracks with the same sizes are found in the plate. In function of the distance between them, the simulated impedance variation slightly differs. However, the signals corresponding to $h = 0.5$ mm and $h = 1$ mm are almost the same. This fact refers to the ill-posedness of the corresponding *inverse problem*.

In Fig. A.3, the whole surface scan is presented in the case of a single- and a double-crack configuration, respectively.

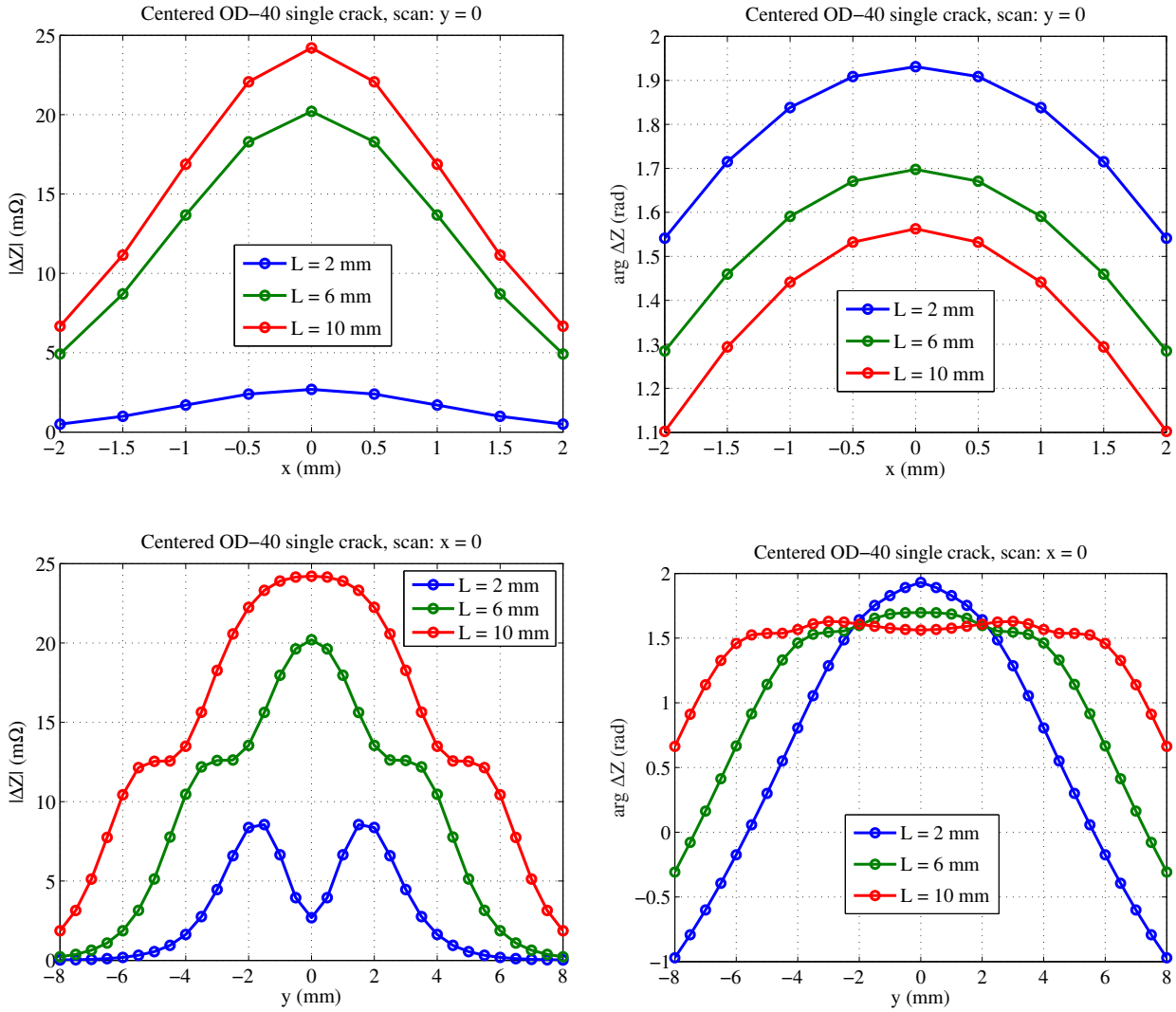


Figure A.1: Typical line scans in the case of one single OD crack with a depth of 40 %. The coil center moves along the x and y axes, respectively.

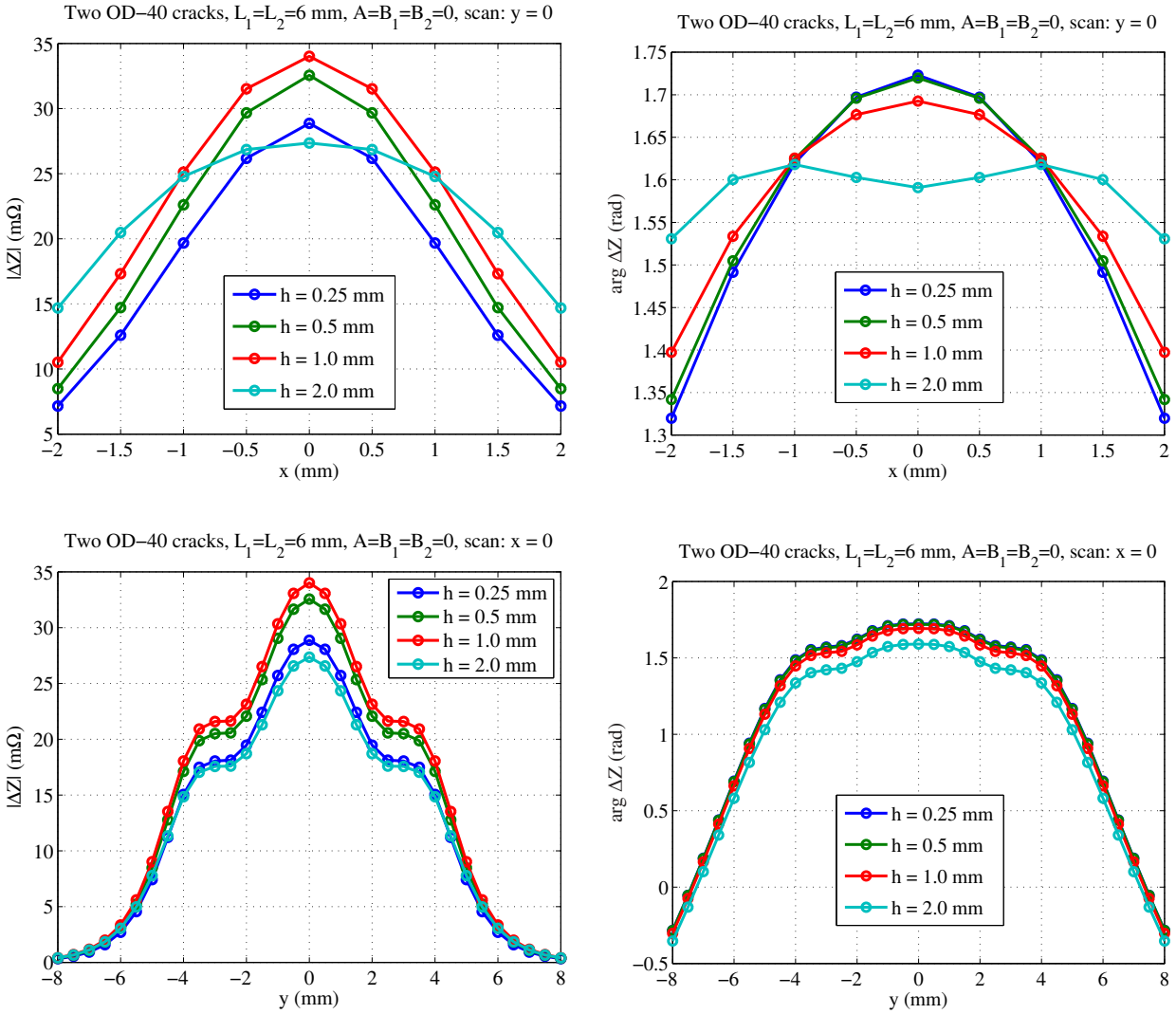
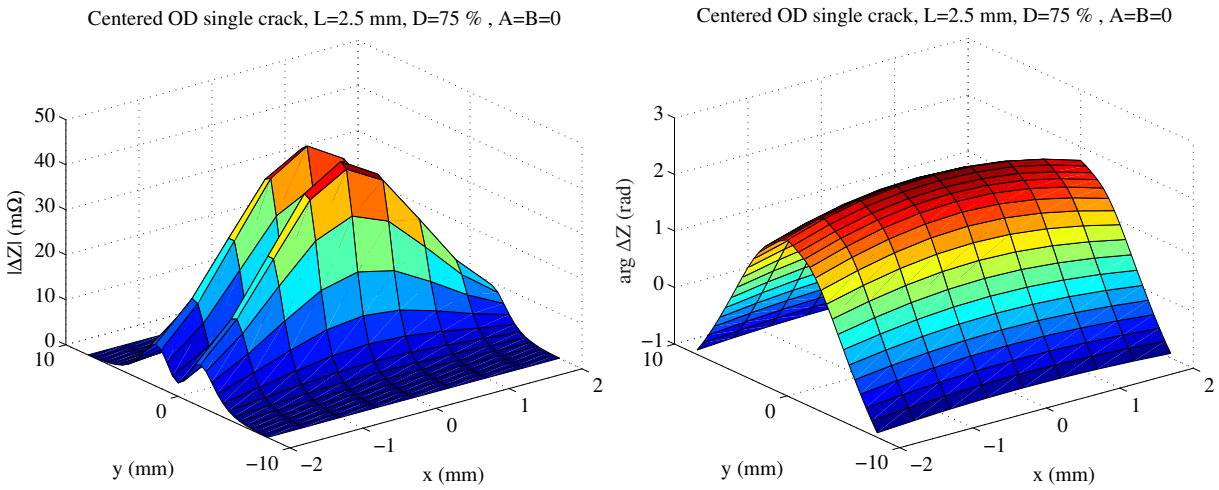
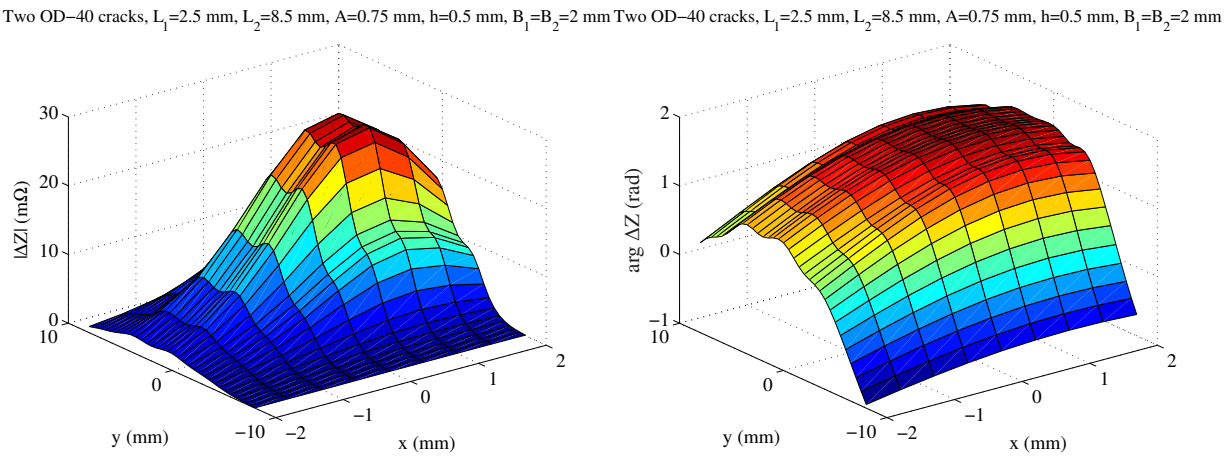


Figure A.2: Typical line scans in the case of double OD cracks with depths of 40 %. The coil center moves along the x and y axes, respectively.



(a) Centred single OD-type crack.



(b) Double OD-type cracks (both depths are equal to 40 % of the plate thickness).

Figure A.3: Typical surface scans.

B

Function approximation by kriging

In this chapter of the Appendix, a brief presentation on the kriging method is given. Let us emphasize that by now, kriging has become a quite diversified domain, herein we focus only on the formalism and techniques used in this Dissertation. The interested reader can find more detailed discussions in the textbooks, e.g., [Chilès and Delfiner, 1999; Cressie, 1993] (from the original domain, geostatistics), [Stein, 1999] (a rigorous mathematical approach), [Fang, Li, and Sudjianto, 2006; Santner, Williams, and Notz, 2003] (containing a short but thoughtful summary from an engineering viewpoint). Among the vast amount of journal papers, let us mention [Sacks et al., 1989] (first introducing kriging to model computer experiments) and [Kleijnen, 2009] as a quite recent overview.

B.1 Modeling by a Gaussian random process

Let $f(\mathbf{p})$ be a scalar function over the N-dimensional domain $\mathbb{P}$. The goal of kriging is to approximate $f(\mathbf{p})$ based on n observations

$$\mathbf{f} = [f(\mathbf{p}_1),\ f(\mathbf{p}_2),\ \ldots,\ f(\mathbf{p}_n)]^\mathsf{T}. \tag{B.1}$$

To this end, $f(\mathbf{p})$ is modeled by a Gaussian random process[1] $\xi(\mathbf{p})$, described by its mean

$$m(\mathbf{p}) = \mathbb{E}\left[\xi(\mathbf{p})\right] \tag{B.2}$$

and its covariance function

$$k(\mathbf{p}_i, \mathbf{p}_j) = \mathbb{E}\left[\Big(\xi(\mathbf{p}_i) - m(\mathbf{p}_i)\Big)\Big(\xi(\mathbf{p}_j) - m(\mathbf{p}_j)\Big)\right], \tag{B.3}$$

where $\mathbb{E}\left[\cdot\right]$ is the expected value operator. Let us assume that the mean function can be written as

$$m(\mathbf{p}) = \mathbf{u}(\mathbf{p})^\mathsf{T}\boldsymbol{\beta}, \tag{B.4}$$

where $\mathbf{u}(\mathbf{p})$ is a vector of l known functions:

$$\mathbf{u}(\mathbf{p})^\mathsf{T} = [u_1(\mathbf{p}),\ u_2(\mathbf{p}),\ \ldots,\ u_l(\mathbf{p})] \tag{B.5}$$

and $\boldsymbol{\beta}$ is a vector of unknown coefficients. The functions $u_i(\mathbf{p})$ must be linearly independent, usually they are low-order monomials in the components of $\mathbf{p}$.

Kriging provides a linear prediction of the modeling process $\xi(\mathbf{p})$, based on the observed variables, collected into a vector

$$\boldsymbol{\xi} = [\xi(\mathbf{p}_1),\ \xi(\mathbf{p}_2),\ \ldots,\ \xi(\mathbf{p}_n)]^\mathsf{T} \tag{B.6}$$

[1] Roughly speaking, a Gaussian random process is a "function" whose values are not scalars but random variables having a Gaussian distribution, moreover, all finite vectors consisting of these random variables have a multivariate Gaussian distribution.

in the form

$$\hat{\xi}(\mathbf{p}) = \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}}\boldsymbol{\xi}, \tag{B.7}$$

with the vector of n coefficients, $\boldsymbol{\lambda}(\mathbf{p})$. The requirements for this prediction are the followings:

- Let (B.7) be *unbiased*, i.e., let $\mathbb{E}\left[\hat{\xi}(\mathbf{p})\right] = \mathbb{E}\left[\xi(\mathbf{p})\right]$ hold. Based on the (B.4) assumption on the mean value, the unbiasedness entails that

$$\mathbf{u}(\mathbf{p})^{\mathsf{T}}\boldsymbol{\beta} = \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}} \underbrace{\begin{bmatrix} \mathbf{u}(\mathbf{p}_1)^{\mathsf{T}} \\ \mathbf{u}(\mathbf{p}_2)^{\mathsf{T}} \\ \vdots \\ \mathbf{u}(\mathbf{p}_n)^{\mathsf{T}} \end{bmatrix}}_{\mathbf{U}} \boldsymbol{\beta}, \quad \forall \boldsymbol{\beta}, \quad \text{i.e., } \mathbf{u}(\mathbf{p})^{\mathsf{T}} = \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}}\mathbf{U}. \tag{B.8}$$

  Let us note that matrix $\mathbf{U}$ is of full column rank l (provided that $l < n$ which is the usual case). By choosing the functions $u_i(\mathbf{p})$ to be monomials in the components of $\mathbf{p}$, this condition is fulfilled.

- Let (B.7) be the optimal prediction in the sense of minimal variance of prediction error, i.e., let the coefficients be the solution of the optimization problem

$$\boldsymbol{\lambda}(\mathbf{p}) = \min_{\boldsymbol{\chi}(\mathbf{p})\in\mathbb{R}^n} \mathbb{E}\left[\left(\xi(\mathbf{p}) - \boldsymbol{\chi}(\mathbf{p})^{\mathsf{T}}\boldsymbol{\xi}\right)^2\right]. \tag{B.9}$$

  Let us note that here we used that the mean of the prediction is zero, due to the unbiasedness condition (B.8).

The above optimization problem (B.9) under the constraint (B.8) can be solved via a Lagrangian formalization, yielding the following linear system of equations:

$$\begin{bmatrix} \mathbf{K} & \mathbf{U} \\ \mathbf{U}^{\mathsf{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}(\mathbf{p}) \\ \boldsymbol{\mu}(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} \mathbf{k}(\mathbf{p}) \\ \mathbf{u}(\mathbf{p}) \end{bmatrix}, \tag{B.10}$$

where the covariance matrix $\mathbf{K}$ and vector $\mathbf{k}(\mathbf{p})$ are defined as

$$\mathbf{K} = \left[k(\mathbf{p}_i, \mathbf{p}_j)\right]_{ij} \quad \text{and} \quad \mathbf{k}(\mathbf{p}) = \left[k(\mathbf{p}, \mathbf{p}_i)\right]_i, \quad \forall i, j = 1, 2, \ldots, n, \tag{B.11}$$

and vector $\boldsymbol{\mu}(\mathbf{p})$ consists of l Lagrange-multipliers. The prediction (B.7) is the Best Linear Unbiased Prediction (BLUP) provided that the coefficients $\boldsymbol{\lambda}(\mathbf{p})$ satisfy the equation system (B.10).

The variance of the error of the prediction (which is subjected to be minimized in (B.9)) can also be estimated in an analytical form, assuming that the coefficients $\boldsymbol{\lambda}(\mathbf{p})$ are chosen according to (B.10). This estimation –playing an important role in many kriging-based approaches– is denoted by $\hat{\sigma}^2(\mathbf{p})$ and expressed as

$$\hat{\sigma}^2(\mathbf{p}) \equiv \mathbb{E}\left[\left(\xi(\mathbf{p}) - \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}}\boldsymbol{\xi}\right)^2\right] = k(\mathbf{p}, \mathbf{p}) - \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}}\mathbf{k}(\mathbf{p}) - \boldsymbol{\mu}(\mathbf{p})^{\mathsf{T}}\mathbf{u}(\mathbf{p}). \tag{B.12}$$

Let us note that this expression does not depend on the actual observed values of the function $f(\mathbf{p})$. However, there is an indirect relation: the covariance function $k(\cdot, \cdot)$ is usually predicted based on the observations $\mathbf{f}$ (see § B.2).

The n observations $\mathbf{f}$ are considered as samples of the random variables $\boldsymbol{\xi}$. Thus, the prediction for the function $f(\mathbf{p})$ has a form similar to (B.7), replacing ξ by f:

$$\hat{f}(\mathbf{p}) = \boldsymbol{\lambda}(\mathbf{p})^{\mathsf{T}}\mathbf{f}. \tag{B.13}$$

Let us note that the choice of the functions $\mathbf{u}(\mathbf{p})$ is to reflect some prior knowledge on the modeled function. Since in the approaches presented in this Dissertation, prior information is not available (or cannot

appropriately be reflected by $\mathbf{u}(\mathbf{p})$) on the function $f(\mathbf{p})$, we simply assume a constant but unknown mean, i.e., $u(\mathbf{p}) = 1$ is chosen. Consequently, the matrix $\mathbf{U}$ becomes a vector of n ones:

$$\mathbf{U} = [1,\ 1,\ \ldots,\ 1]^{\mathsf{T}}. \tag{B.14}$$

The Lagrange multiplier $\boldsymbol{\mu}(\mathbf{p})$ also becomes a scalar, respectively.

B.2 Choice of the covariance

The covariance function is usually not known in the scalar or functional kriging approach, but one has to predict it from the observed data. Though usually one has no particular reason to assume a "uniform behaviour" of the modeled function $f(\mathbf{p})$ (or $q_{\mathbf{p}}(t)$) over the domain $\mathbb{P}$, the chosen covariance is homogeneous (also called stationary), i.e., invariant to translation in $\mathbb{P}$. The use of such stationary covariances is much easier than the use of non-stationary covariances; however, the latter might fit better to the modeled function. The stationary covariances depend only on the distance between the observation points:

$$k(\mathbf{p}_i, \mathbf{p}_j) \coloneqq k(h), \quad \text{where } h = \sqrt{\sum_{k=1}^{N} \left(\frac{p_{i,k} - p_{j,k}}{\rho_k} \right)^2}, \tag{B.15}$$

where $p_{i,k}$ and $p_{j,k}$ are the kth component of the vectors $\mathbf{p}_i$ and $\mathbf{p}_j$, respectively. The scaling factors ρ_k (also called characteristic correlation distances, or ranges of the covariance) are introduced to reflect the sensitivity of the modeled function to the different components of $\mathbf{p}$. The higher ρ_k, the less sensitive the modeled function to the kth component of $\mathbf{p}$.

In practice, the covariance is modeled by a parametrized function, whose parameters (called the "hyper-parameters") are tuned so that covariance model fits the observed data. Common choices are the rational, exponential or Gaussian functions as covariance models. In this Dissertation, the so-called Matérn function –being also a popular choice thanks to its flexibility– is used. The Matérn covariance has the form

$$k(h) = \frac{\sigma_0^2}{2^{\nu-1}\Gamma(\nu)} \left(2\sqrt{\nu}\, h\right)^{\nu} \mathcal{K}_{\nu}\left(2\sqrt{\nu}\, h\right), \tag{B.16}$$

where $\Gamma(\cdot)$ is the Gamma-function and $\mathcal{K}_{\nu}$ is the modified Bessel-function of the second kind of order ν. The covariance has two parameters: σ_0^2 is the variance of the process and ν controls the regularity. In Fig. B.1, the effect of the ν parameter is illustrated. Let us note that the smaller ν, the more "peaked" the covariance function at the origin. This regularity parameter makes the Matérn covariance flexible: it can be tuned to the given smoothness of the modeled function $f(\mathbf{p})$. Let us recall that kriging assumes that the modeled function is a sample path of a Gaussian random process $\xi(p)$. The effect of the parameter ν on the modeling process is illustrated in Fig. B.2. Let us also note that the use of the Matérn function (similarly to the other covariance models used in practice) ensures the positive definiteness of the covariance matrix (B.11) (provided that h is computed by using a "valid" distance function[2], e.g., (B.15)).

In Fig. B.3, the performance of the kriging prediction with different hyperparameters of the Matérn covariance is illustrated, highlighting the importance of the adequate estimation of the hyperparameters.

The proper choice of the hyperparameters of the covariance function is essential to obtain good kriging interpolations. This choice is usually traced back to an optimization problem. Let us note that the ranges ρ_k in the anisotropic distance function (B.15) are also to be chosen, thus, one has $2 + N$ hyperparameters in total to optimize, where N is the number of components of the vector $\mathbf{p}$. The number of independent

[2]Satisfying the axioms of non-negativity, symmetry and triangular inequality [Steen and J. Arthur Seebach, 1970].

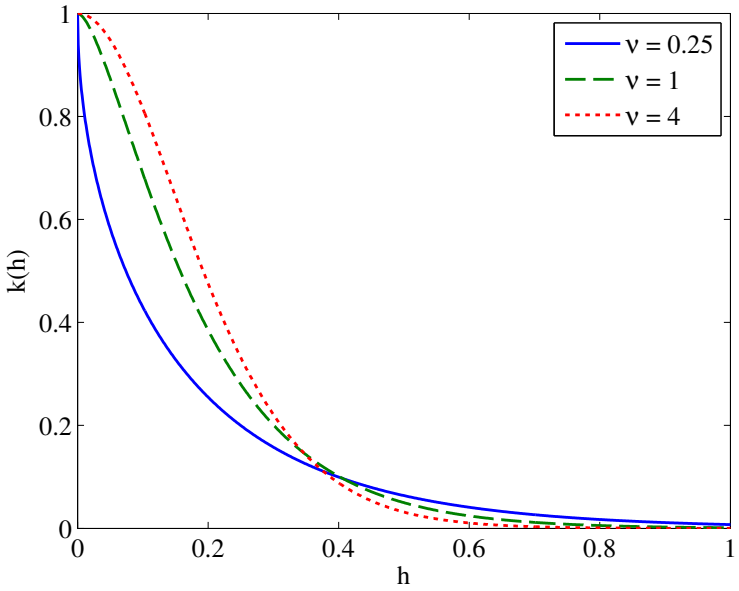


Figure B.1: The Matérn covariance function ($\sigma_0^2 = 1, \rho = 0.25$) with three different ν values. The behaviour at the origin is significantly influenced by ν. Let us note that the limit case $\nu \to \infty$ yields the Gaussian "bell-curve".

hyperparameters might be reduced when problems having certain symmetries are modeled. For instance, in our canonical examples with 6 and 8 parameters (§ 2.4), the ρ ranges of the two coordinates B_1, B_2, the two lengths L_1, L_2 and the two depths D_1, D_2, respectively, should be pairwise equal. Thus, in the 6-parameter example, only 5 hyperparameters are to be estimated.

In practice, three approaches mainly are used to estimate the hyperparameters: the maximum likelihood, the cross-validation and the variogram-fitting methods. The first two are used in this Dissertation (brief summaries are given below), the third one is classical in the original, geostatistical domain of kriging.

B.2.1 Parameter estimation by the Maximum Likelihood Method

The Maximum Likelihood (ML) method is a widely used approach not only for the covariance hyperparameter estimation, but more generally in statistics as well. The method assumes that the class of probability distribution (from which the observed data are considered to be drawn) is known, only its parameters are to be estimated. This is the case in the traditional kriging approach, as the observed vector (B.1) is assumed to be a sample of the random vector (B.6), having a multivariate Gaussian distribution.

The PDF of the multivariate Gaussian distribution is

$$\mathrm{pdf}(\mathbf{f}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{K}|}} \exp\left[-\frac{1}{2}(\mathbf{f}-\mathbf{m})^{\mathsf{T}}\mathbf{K}^{-1}(\mathbf{f}-\mathbf{m})\right], \tag{B.17}$$

where $\mathbf{f}$ is the vector of n observations of the random vector $\boldsymbol{\xi}$, $\mathbf{m}$ is the mean vector and $\mathbf{K}$ is the covariance matrix, so that

$$\mathbf{m} = \mathbb{E}[\boldsymbol{\xi}] \quad \text{and} \quad \mathbf{K} = \mathbb{E}\left[(\boldsymbol{\xi}-\mathbf{m})(\boldsymbol{\xi}-\mathbf{m})^{\mathsf{T}}\right], \tag{B.18}$$

and the symbol $|\cdot|$ denotes the determinant.

In general, the ML method estimates the distribution properties $\mathbf{m}$ and $\mathbf{K}$ by considering the probability density function (B.17) as a function of $\mathbf{m}$ and $\mathbf{K}$, with the observed $\mathbf{y}$ as parameter. Then, the estimated

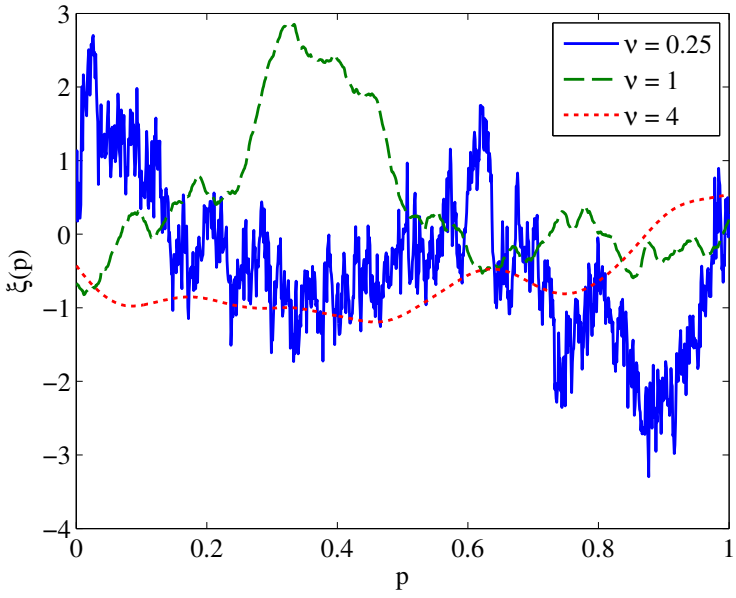


Figure B.2: Sample paths of zero-mean Gaussian processes (what the modeled function f is assumed to be) described by the Matérn covariance function ($\sigma_0^2 = 1$, $\rho = 0.25$) with three different ν values. The smaller the ν, the more irregular the process. Higher ν values are used when a smooth function is modeled.

mean vector and covariance matrix are found by maximizing the probability density function. In other words, ML results in the specific distribution from which the observed data are drawn the "most probably".

More specifically, let us take into account that the covariance is modeled by the Matérn function, thus the covariance matrix depends on the parameters

$$\boldsymbol{\Theta} = [\sigma_0^2, \nu, \rho_1, \ldots, \rho_N]. \tag{B.19}$$

The mean vector is assumed to have the form (B.4). Consequently, the likelihood function –the probability density as a function of Θ– when $\mathbf{f}$ is observed, can be written as

$$L(\boldsymbol{\Theta}|\mathbf{f}, \boldsymbol{\beta}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{K}(\boldsymbol{\Theta})|}} \exp\left[-\frac{1}{2}(\mathbf{f} - \mathbf{U}\boldsymbol{\beta})^{\mathsf{T}} \mathbf{K}(\boldsymbol{\Theta})^{-1} (\mathbf{f} - \mathbf{U}\boldsymbol{\beta})\right]. \tag{B.20}$$

The coefficient vector $\boldsymbol{\beta}$ is not known in the expression of the mean vector. A natural idea would be to extend the optimization for $\boldsymbol{\beta}$ as well, however, a more effective alternative is provided by the REML method [Stein, 1999]. The main idea of the REML method is to trace back the originally assumed multivariate Gaussian distribution to a zero-mean Gaussian distribution by a linear transformation. To this end, let us find the matrix $\mathbf{W}$ for which

$$\mathbf{WU} = \mathbf{0} \tag{B.21}$$

holds. $\mathbf{W}$ –a $(n - l) \times n$ matrix of full row rank $(n - l)$– can be found via, e.g., the QR-decomposition of $\mathbf{U}$:

$$\left[\begin{array}{c|c} \mathbf{Q}_1 & \mathbf{Q}_2 \end{array}\right] \left[\begin{array}{c} \mathbf{R} \\ \hline \mathbf{0} \end{array}\right] = \mathbf{U} \quad \Rightarrow \quad \mathbf{W} = \mathbf{Q}_2^{\mathsf{T}}, \tag{B.22}$$

where the matrix $[\mathbf{Q}_1|\mathbf{Q}_2]$ is orthogonal, $\mathbf{R}$ is upper triangular. A transformed observation vector $\mathbf{f}^\star$ is obtained from the original one by using the transformation matrix $\mathbf{W}$:

$$\mathbf{f}^\star = \mathbf{Wf}. \tag{B.23}$$

This vector $\mathbf{f}^\star$ is now assumed to be a sample of a zero-mean Gaussian distribution, characterized by a transformed covariance matrix (still depending on $\boldsymbol{\Theta}$):

$$\mathbf{K}(\boldsymbol{\Theta})^\star = \mathbf{W}\mathbf{K}(\boldsymbol{\Theta})\mathbf{W}^\mathsf{T}. \tag{B.24}$$

The transformed observation vector contains l fewer elements, but similarly, the number of unknown parameters is also l fewer since $\boldsymbol{\beta}$ does not occur any more in the likelihood function of the transformed distribution:

$$L^\star(\boldsymbol{\Theta}|\mathbf{W}\mathbf{f}) = \frac{1}{\sqrt{(2\pi)^{(n-l)}|\mathbf{W}\mathbf{K}(\boldsymbol{\Theta})\mathbf{W}^\mathsf{T}|}} \exp\left[-\frac{1}{2}(\mathbf{W}\mathbf{f})^\mathsf{T}(\mathbf{W}\mathbf{K}(\boldsymbol{\Theta})\mathbf{W}^\mathsf{T})^{-1}(\mathbf{W}\mathbf{f})\right]. \tag{B.25}$$

The REML estimation of the covariance hyperparameters is then the maximizer of the likelihood function:

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} L^\star(\boldsymbol{\Theta}|\mathbf{W}\mathbf{f}). \tag{B.26}$$

The optimization of the likelihood function might be a numerically challenging problem since $\mathbf{K}(\boldsymbol{\Theta})$ can appear to be ill-conditioned for certain values of $\boldsymbol{\Theta}$ (e.g., high ν and ρ_k parameters at the same time). However, one usually has prior knowledge on the hyperparameters (e.g., by rule of thumbs). Some regularization of the problem (B.26) can be carried out by using the so-called Maximum A Posteriori approach: one defines a prior distribution of $\boldsymbol{\Theta}$ according to the prior knowledge, then the *a posteriori* likelihood is optimized. The prior distribution is usually a multivariate Gaussian distribution[3] with no correlation between the hyperparameters:

$$\pi(\boldsymbol{\Theta}) = \prod_{j=1}^{N+2} \frac{1}{\sqrt{(2\pi)\sigma^2_{\text{prior},j}}} \exp\left[-\frac{(\Theta_j - \Theta_{\text{prior},j})^2}{2\sigma^2_{\text{prior},j}}\right], \tag{B.27}$$

where $\Theta_{\text{prior},j}$ and $\sigma_{\text{prior},j}$ are the prior mean and variance of the jth hyperparameter (of which the Matérn covariance has totally $N + 2$)[4]. The estimated hyperparameters are then obtained by maximizing the a posteriori likelihood function:

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \left[\pi(\boldsymbol{\Theta})L^\star(\boldsymbol{\Theta}|\mathbf{W}\mathbf{f})\right]. \tag{B.28}$$

In practice, the optimization problems (B.26) or (B.28) are obviously solved by using numerical methods, such as a gradient method or a gradient-free pattern-search approach. Let us note that for practical reasons, usually the logarithm of the objective functions in (B.26) and (B.28) is taken, then the so-called log-likelihood function is maximized. In our numerical studies, we use the gradient-free optimization routine `fminsearch` of MATLAB® for the covariance estimation.

B.2.2 Parameter estimation by Cross Validation (CV)

In the case of functional kriging (§ 4.2), the trace-covariance (4.17) cannot be involved in a likelihood function, thus, the ML (or REML) estimation cannot be used to fit a Matérn model to the trace-covariance. In this Dissertation, the CV method is applied instead. CV-like methods are also popular techniques in statistics (see, e.g., [Geisser, 1993]). The basic idea of such approaches is to predict the modeled function at observed points but not accounting to that certain observed value. By comparing the prediction and the true, observed value, conclusions can be drawn on the quality of the prediction. A particular version of CV, called Leave-One-Out Cross Validation (LOOCV) is used in our numerical studies.

[3] In practice, all computations are performed on the log-scaled hyperparameters (to avoid negative values), thus, the prior distribution on the hyperparameters is indeed multivariate log-normal.

[4] As a rule of thumb, the prior mean of the variance σ_0^2 of the process is equal to the variance of the observation vector $\mathbf{f}$, the prior mean of ν is set to $3N/2$, whereas the prior mean of the ranges ρ_k is chosen according to the range of the kth component of the vector $\mathbf{p}$.

Let $\mathbf{q}(t)$ be the vector of the n (functional) observations as defined in (4.2). Let $\mathbf{q}^{(-i)}(t)$ ($i = 1, 2, \ldots, n$) be the reduced vector of $n-1$ observations, derived from $\mathbf{q}(t)$ by leaving out the ith observation. The trace-covariance is modeled by a parametrized function (e.g., the Matérn function), thus, the trace-covariance matrix $\mathbf{K}_T(\boldsymbol{\Theta})$ and vector $\mathbf{k}_T(\mathbf{p}, \boldsymbol{\Theta})$ in (4.18) depends on the $\boldsymbol{\Theta}$ vector of hyperparameters of the model. Let $\boldsymbol{\lambda}^{(-i)}(\mathbf{p})$ be the vector of $n-1$ coefficients resulting from (4.18), assuming that the observation point $\mathbf{p}_i$ is not considered, i.e., the ith row and column of $\mathbf{K}_T(\boldsymbol{\Theta})$ and the ith row of $\mathbf{k}_T(\mathbf{p}, \boldsymbol{\Theta})$ are deleted. The LOOCV objective function is then written as

$$H(\boldsymbol{\Theta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\left\| q_{\mathbf{p}_i}(t) - \boldsymbol{\lambda}^{(-i)}(\mathbf{p}_i)^{\mathsf{T}} \mathbf{q}^{(-i)}(t) \right\|}{\left\| q_{\mathbf{p}_i}(t) \right\|}. \tag{B.29}$$

The CV estimation of the covariance parameters $\boldsymbol{\Theta}$ is the minimizer of the objective function:

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} H(\boldsymbol{\Theta}). \tag{B.30}$$

In certain cases, one might need to regularize the above optimization problem in order to avoid ill-conditioned covariance matrices due to non-reasonable hyperparameter values. In this simple CV scheme, there is no elegant way of incorporating a priori knowledge on the hyperparameters like the prior distribution in the maximum likelihood method. However, by heuristically introducing an additive regularization term, the optimization problem becomes handier:

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \left[H(\boldsymbol{\Theta}) + \frac{1}{N+2} \sum_{j=1}^{N+2} \frac{(\Theta_j - \Theta_{\text{prior},j})^2}{\Theta_{\text{prior},j}^2} \right], \tag{B.31}$$

where $\Theta_{\text{prior},j}$ is the a priori assumed value of the jth hyperparameter (of which the Matérn covariance has totally $N+2$). These priors are chosen similarly to the prior means in the ML estimation. In our numerical studies in § 4, the LOOCV objective function is minimized by using a gradient-free direct search method (the `fminsearch` routine of MATLAB®).

A pitfall of LOOCV is its high computational cost (n kriging models (each with $n-1$ samples) are needed) to evaluate the objective function in case of n samples. A possible reduction of the computation time is to partition the n samples into k disjunct groups, then, perform cross-validation leaving all samples of a group out sequentially (thus, k kriging models are needed). This technique is called k-fold cross-validation.

B.3 Derivation of the Expected Improvement expression

This section is related to the EGO algorithm, more precisely, the formula of the expected improvement (3.6). To derive that expression, we have to recourse to some basics of probability theory.

At a certain point $\mathbf{p}$, the unknown objective function f is modeled by a Gaussian random variable $\xi(\mathbf{p})$ in the kriging framework. One can predict the mean of $\xi(\mathbf{p})$ –this is $\hat{f}(\mathbf{p})$– and its variance, $\hat{\sigma}^2(\mathbf{p})$. The improvement –keeping in mind that the goal is the minimization of f and the current minimum among the observations is $f_{\min}$– can obviously be written as

$$I(\mathbf{p}) = f_{\min} - \xi(\mathbf{p}), \tag{B.32}$$

which is also a normally distributed random variable with a mean of $f_{\min} - \hat{f}(\mathbf{p})$ and with the same $\hat{\sigma}^2(\mathbf{p})$ estimated variance as $\xi(\mathbf{p})$. The probability density function of $I(\mathbf{p})$ is then written as

$$\text{pdf}(u) = \frac{1}{\hat{\sigma}(\mathbf{p})} \varphi\left(\frac{u - (f_{\min} - \hat{f}(\mathbf{p}))}{\hat{\sigma}(\mathbf{p})} \right) \tag{B.33}$$

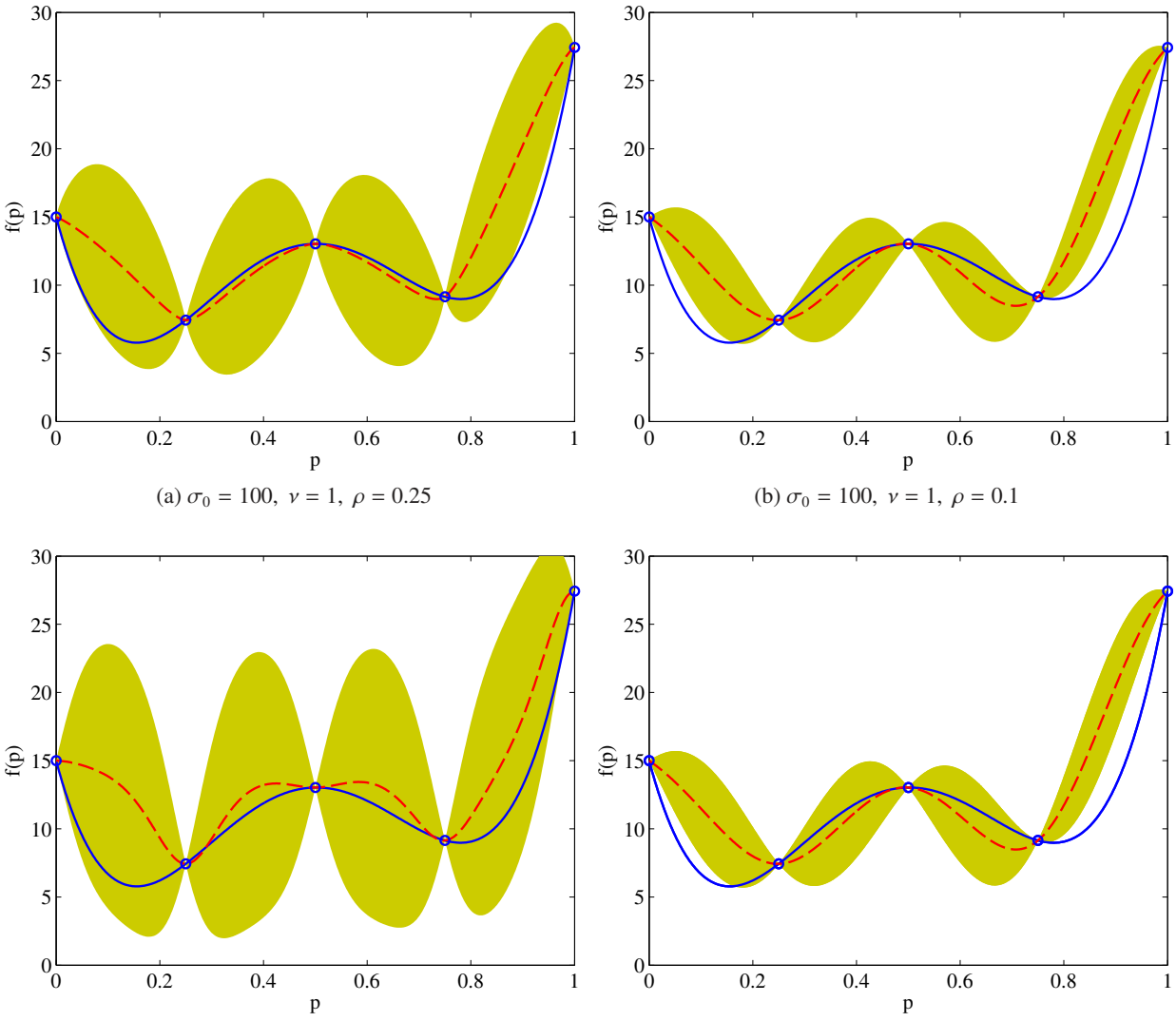
where $\varphi(\cdot)$ is the probability density function of the zero-mean, unit-variance normal distribution. The expected value of the improvement $I(\mathbf{p})$ can be analytically computed, however, the key point is that only positive values of $I(\mathbf{p})$ are considered as true improvements. Thus, when defining the expected value (as the first moment of the distribution), the negative semi-axis of the support has to be ignored:

$$\mathbb{E}\left[I(\mathbf{p})\right] = \int_0^\infty u \,\mathrm{pdf}(u)\mathrm{d}u = \int_0^\infty u \frac{1}{\hat{\sigma}(\mathbf{p})}\varphi\left(\frac{u - (f_{\min} - \hat{f}(\mathbf{p}))}{\hat{\sigma}(\mathbf{p})}\right)\mathrm{d}u. \tag{B.34}$$

The above integral can be expressed in a closed form, e.g., with the integration-by-parts technique. Due to the central role of this expression, a new notation $\eta(\mathbf{p})$ is introduced, and the final form is written as

$$\eta(\mathbf{p}) \equiv \mathbb{E}\left[I(\mathbf{p})\right] = \left(f_{\min} - \hat{f}(\mathbf{p})\right)\Phi\left(\frac{f_{\min} - \hat{f}(\mathbf{p})}{\hat{\sigma}(\mathbf{p})}\right) + \hat{\sigma}(\mathbf{p})\varphi\left(\frac{f_{\min} - \hat{f}(\mathbf{p})}{\hat{\sigma}(\mathbf{p})}\right), \tag{B.35}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function with zero-mean and unit-variance. This expression equivalently appears in (3.6).



(a) $\sigma_0 = 100,\ \nu = 1,\ \rho = 0.25$

(b) $\sigma_0 = 100,\ \nu = 1,\ \rho = 0.1$

(c) $\sigma_0 = 100,\ \nu = 4,\ \rho = 0.1$

(d) $\sigma_0 = 82.3,\ \nu = 2.74,\ \rho = 0.28$ (Obtained by maximum likelihood estimation (§ B.2.1)).

Figure B.3: Kriging predictions of the same one-variate toy function using the Matérn covariance. Solid line: function to be predicted (f); circles: samples ($f(p_i)$); dashed line: kriging prediction ($\hat{f}$). The estimated variance ($\hat{\sigma}^2$) is represented by the filled are, being bordered by the $\hat{f} \pm \sqrt{\hat{\sigma}^2}$ curves. The quality of the predictor strongly depends on the covariance model, thus, an appropriate choice of the hyperparameters is essential. ($f(p) = +579p^4 - 1110p^3 + 684.5p^2 - 141.071p + 15$)

C

Optimization by means of the Coordinatewise Search Algorithm

In this chapter, a presentation of the applied variant of the Coordinatewise Search (CWS) algorithm is given. This tool is used to optimize the auxiliary functions: the expected improvement in § 3.4 and the point-insertion criteria in §§ 4.6.2 and 5.2.3.

C.1 Discretization of continuous optimization problems

Let us first assume a general objective function $f(\mathbf{p})$, defined on the N-dimensional unit-hypercube:

$$\mathbb{U} = [0, 1]^N. \tag{C.1}$$

The goal is the to find the global maximizer of f over $\mathbb{U}$:

$$\mathbf{p}^\star_{\text{cont}} = \arg\max_{\mathbf{p}\in\mathbb{U}} f(\mathbf{p}). \tag{C.2}$$

The proposed CWS method operates on a rectangular grid, thus, let us define a discrete space $\mathbb{P}$ by means of a grid ($\mathbb{P}$ is the input space in the examples of this Dissertation, being indeed a finite set of vectors defined by a grid). Let us recall the definition (2.9):

$$\mathbb{P} = \left\{ \sum_{i=1}^{N} k_i \mathbf{e}_i \;\middle|\; \forall [k_1, k_2, \ldots, k_N] \in \mathbb{K} \right\}, \tag{C.3}$$

where $\mathbf{e}_i$ is the ith unit-step vector:

$$\mathbf{e}_i = [0,\, 0,\, \ldots,\, 0,\, \Delta p_i,\, 0,\, \ldots,\, 0], \quad \Delta p_i = \frac{1}{K_i}, \quad \forall i = \{1,\, 2,\, \ldots,\, N\}, \tag{C.4}$$

and $\mathbb{K}$ is an index-set:

$$\mathbb{K} = \{0,\, 1,\, \ldots,\, K_1\} \times \{0,\, 1,\, \ldots,\, K_2\} \times \cdots \times \{0,\, 1,\, \ldots,\, K_N\}. \tag{C.5}$$

The discretized optimization problem is formalized as

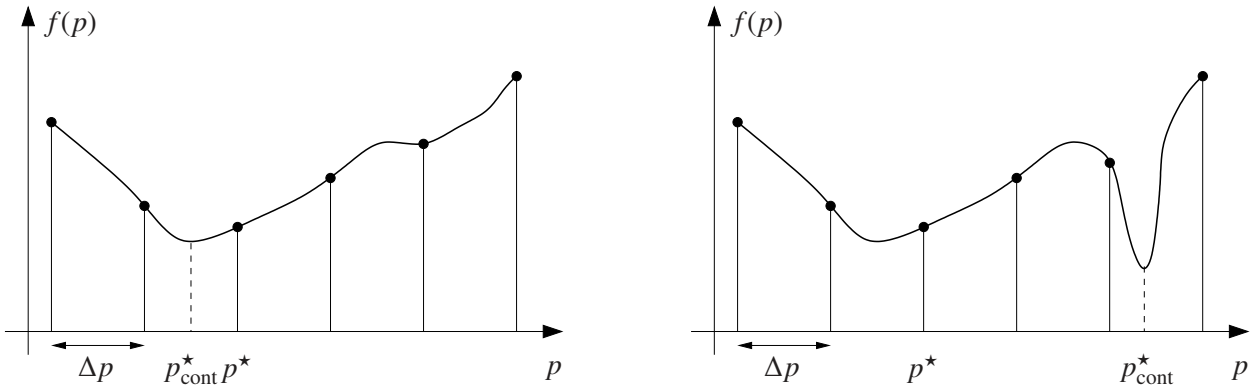
$$\mathbf{p}^\star = \arg\max_{\mathbf{p}\in\mathbb{P}} f(\mathbf{p}). \tag{C.6}$$

An important assumption has to be made here (see the related part of § 2.3 on the adequacy of the discretization). One must assume that the chosen discretization (i.e., the grid) is fine enough with respect to f, in the

sense that the global maximizer of f in $\mathbb{P}$ is "close to the true one" in $\mathbb{U}$, as it is illustrated by Fig. C.1a. Formally, we assume that

$$|p^{\star}_{\mathrm{cont},i} - p^{\star}_{i}| < \Delta p_i \quad \forall i = 1, 2, \ldots, N \tag{C.7}$$

holds, i.e., all component of the vectors $\mathbf{p}^{\star}_{\mathrm{cont}}$ and $\mathbf{p}^{\star}$ are closer to each other than the step size. However, one can easily imagine cases where this condition is violated, like the simple illustration in Fig. C.1b.



(a) Good performance. (b) Unfortunate case.

Figure C.1: Illustration of the discretized optimization problem in a simple one-parameter case with fictitious objective functions.

C.2 Local extrema and global extremum in the discretized optimization problem

The definition of the local extrema of the function $f(\mathbf{p})$ over the continuous domain $\mathbb{U}$ is straightforward. A point $\mathbf{p}_0 \in \mathbb{U}$ is called local maximizer (resp. minimizer) of f, if and only if a positive ε exists, such that

$$\|\mathbf{p} - \mathbf{p}_0\| < \varepsilon \Rightarrow f(\mathbf{p}_0) \geq f(\mathbf{p}) \quad (\text{resp. } f(\mathbf{p}_0) \leq f(\mathbf{p})). \tag{C.8}$$

The value $f(\mathbf{p}_0)$ is called local maximum (resp. minimum). The function f might have several local maxima and minima. The largest local maximum (resp. the smallest local minimum) is called the global maximum (resp. minimum) over $\mathbb{U}$.

However, these definitions cannot be applied in the case of the discrete optimization problem (C.6). Though one surely has an intuition about what local extremum means when the domain of f is defined by a discrete grid, we give a formal definition in the following. Let us define the neighbourhood of a point $\mathbf{p} \in \mathbb{P}$ as

$$\mathbb{A}_{\mathbf{p}} = \left\{ \mathbf{p} \pm \mathbf{e}_i \;\middle|\; \forall i = \{1, 2, \ldots, N\} \right\} \cap \mathbb{P}. \tag{C.9}$$

The number of neighbours is always between N (if $\mathbf{p}$ is in the "corner" of $\mathbb{P}$) and $2N$ (if $\mathbf{p}$ is an "inner" point of $\mathbb{P}$). A point $\mathbf{p}_0$ is called local maximizer (resp. minimizer) of f in the "grid-sense", if and only if

$$\mathbf{p} \in \mathbb{A}_{\mathbf{p}_0} \Rightarrow f(\mathbf{p}_0) \geq f(\mathbf{p}) \quad (\text{resp. } f(\mathbf{p}_0) \leq f(\mathbf{p})). \tag{C.10}$$

The value $f(\mathbf{p}_0)$ is called local maximum (resp. minimum). The function f might have several local maxima and minima in the grid-sense as well. The largest local maximum (resp. the smallest local minimum) is called the global maximum (resp. minimum) over $\mathbb{P}$ in the grid-sense.

C.3 Coordinatewise search on a discrete grid

The CWS algorithm operates on the discrete input space $\mathbb{P}$ defined by a rectangular grid. Among the several CWS variants, we use a heuristic combination of the discrete grid search [Powell, 1998] and the scheme proposed in [Csallner and Balogh, 2007] to solve (C.6). The optimization algorithm is designed to find a local maxima (or minima, herein we focus on the maximization), and can be considered as a sort of discretized version of the classical steepest descent optimization scheme.

Our CWS algorithm is the following:

1. *Initial guess:* Let $\mathbf{p}_1 \in \mathbb{P}$ be an arbitrarily chosen initial guess for $\mathbf{p}^\star$; $n = 1$.
2. *Jump to the "best" neighbour:* (Recursive definition.) Let us update our current guess $\mathbf{p}_n$ by choosing the vector corresponding to the highest value of the objective function:

$$\mathbf{p}_{n+1} = \arg\max_{\mathbf{p}\in\mathbb{A}_{\mathbf{p}_n}} f(\mathbf{p}); \tag{C.11}$$

3. *Stopping criterion:* If the update yields improvement, i.e., $f(\mathbf{p}_{n+1}) > f(\mathbf{p}_n)$ holds, then increase $n := n + 1$ and go to step 2. Otherwise stop the iteration and the solution is $\mathbf{p}^\star = \mathbf{p}_n$.

The algorithm always converges to a local maximizer. The global convergence of the method is obviously not guaranteed, however. To reduce the possibility of stalling in a local maximum, the search can sequentially be started from different initial points, as it is usual with such local search methods. A schematic illustration of the discrete grid and the related notations are shown in Fig. C.2.

Many other variants of the CWS-like methods also exist. The presented CWS method is in fact a pattern search, with the pattern defined by the neighbourhood (C.9). In contrary with other approaches (like the CWS proposed in [Csallner and Balogh, 2007]), the "size of the pattern", i.e., the step size is fixed in our implementation. Certain discrete grid methods (like the one discussed in [Powell, 1998]) do not seek thoroughly all neighbours but cyclically change the search directions and more than one step along the actual direction are allowed to be taken. Despite of its simplicity, the presented algorithm can effectively be applied for the solution of the auxiliary optimization problems arising in this Dissertation.

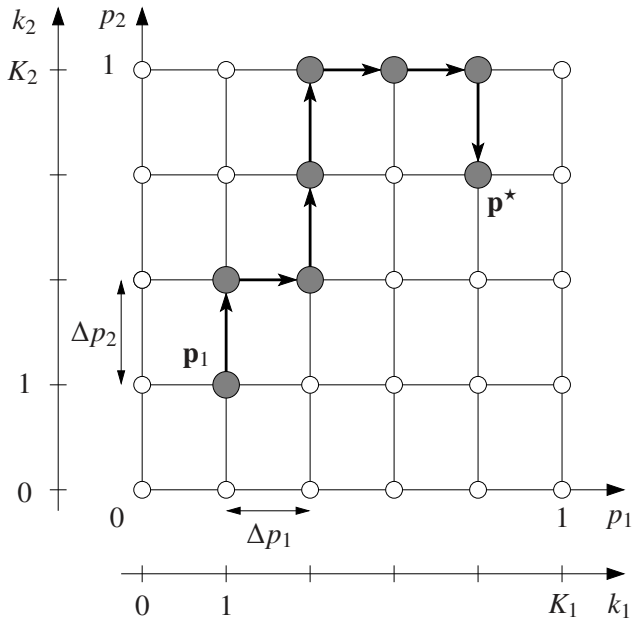


Figure C.2: Illustration of the CWS algorithm in the case of $N = 2$. The grid consists of 6×5 nodes, the solution is found in 7 iterations.

D

Some classical sampling and interpolation techniques

For the sake of completeness, certain classical sampling and interpolation techniques used in this Dissertation are formally defined in this chapter. As sampling method, we deal herein with the factorial and the Latin hypercube designs. We point out how these designs are modified in the case of discrete input domains (being involved in our numerical examples). Then, the piecewise linear interpolation based on a full-factorial design is discussed.

D.1 Factorial designs

Let us assume a setup with N input parameters collected into the input vector $\mathbf{u}$. Let the input space be continuous: $\mathbb{U} = [0, 1]^N$. A full-factorial design is then defined by the levels of each input parameter. Let the set of levels of the kth input parameter be

$$\mathbb{L}_k = \left\{u_k^{(1)}, u_k^{(2)}, \ldots, u_k^{(l_k)}\right\}, \quad 0 \leq u_k^{(1)} < u_k^{(2)} < \cdots < u_k^{(l_k)} \leq 1. \tag{D.1}$$

The full-factorial design $\mathbb{L}^{\mathrm{FF}}$ is the set of all input vectors given by the coordinates equal to the Cartesian product of the N level-sets:

$$\mathbb{FF} = \left\{\mathbf{u} = [u_1, u_2, \ldots, u_N] \;\middle|\; \forall (u_1, u_2, \ldots, u_N) \in \mathbb{L}_1 \times \mathbb{L}_2 \times \cdots \times \mathbb{L}_N\right\}. \tag{D.2}$$

The number of input vectors in such a full-factorial design is then equal to the product of the level numbers:

$$|\mathbb{FF}| = \prod_{k=1}^{N} l_k. \tag{D.3}$$

One often does not need all input vectors of a factorial design ($|\mathbb{L}^{\mathrm{FF}}|$ might be intractably huge). By deleting certain input vectors from $\mathbb{L}^{\mathrm{FF}}$ (according to a given rule), a fractional factorial design is obtained.

In practice, the levels of full-factorial designs are equispaced along each input parameter. Furthermore, it is also common to choose the first and last level as 0 and 1, respectively. In so doing, the levels of the kth input parameter are defined as

$$\mathbb{L}_k = \left\{\frac{i-1}{l_k - 1} \;\middle|\; \forall i = 1, 2, \ldots, l_k\right\}. \tag{D.4}$$

In this Dissertation, only full-factorial designs with such equispaced levels are involved in certain numerical examples.

The MATLAB® program has a routine for the generation of full-factorial designs. For instance, in the case of $N = 2$, a full-factorial design with 3 and 4 levels, respectively, can be generated by the following command:

```
>> FF = (fullfact([3 4])-1)*diag(1./([3 4]-1))

FF =

         0         0
    0.5000         0
    1.0000         0
         0    0.3333
    0.5000    0.3333
    1.0000    0.3333
         0    0.6667
    0.5000    0.6667
    1.0000    0.6667
         0    1.0000
    0.5000    1.0000
    1.0000    1.0000
```

The yielded design is plotted in Fig. D.1a.

D.2 Latin hypercube design

Latin hypercube sampling (LHS) (see, e.g. [McKay, Beckman, and Conover, 1979] or the textbooks [Fang, Li, and Sudjianto, 2006; Santner, Williams, and Notz, 2003]) are well known and widely used in various fields of DoE. The main goal of LHS is to represent all required levels of each input parameter in the design, but at a much smaller total number of samples than a full-factorial design can do.

Let us assume a setup with N input parameters ($2 \leq N$) collected to the input vector $\mathbf{u}$. Let the input space be continuous: $\mathbb{U} = [0, 1]^N$. Let us define a full-factorial design $\mathbb{FF}$ with the same n levels for each input parameters:

$$\mathbb{L}_k = \left\{ \frac{i-1}{n} + \frac{1}{2n} \;\middle|\; \forall i = 1, 2, \ldots, n \right\}, \; \forall k = 1, 2, \ldots, N. \tag{D.5}$$

Then a Latin hypercube design is obtained by choosing n points from the full-factorial design:

$$\mathbb{LH} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathbb{FF}\}, \tag{D.6}$$

so that the following criterion holds:

$$\Delta\mathbf{u}_{ij} = \mathbf{u}_i - \mathbf{u}_j, \quad \forall i, j = 1, 2, \ldots, n, \quad i \neq j$$
$$\Rightarrow \text{The vector } \Delta\mathbf{u}_{ij} \text{ has at least two non-zero elements.} \tag{D.7}$$

In words, one and only one point is chosen from each "grid-line" of the full-factorial design. Obviously, many repartitions of the points fulfil the criterion (D.7). The choice is then usually random. In many cases, additional criteria are used, for instance, maximizing the minimal distance between the points of the Latin hypercube design (maximin criterion).

The MATLAB® program has a routine for the generation of Latin hypercube designs. For instance, in the case of $N = 2$, a Latin hypercube design with 12 points can be generated by the following command:

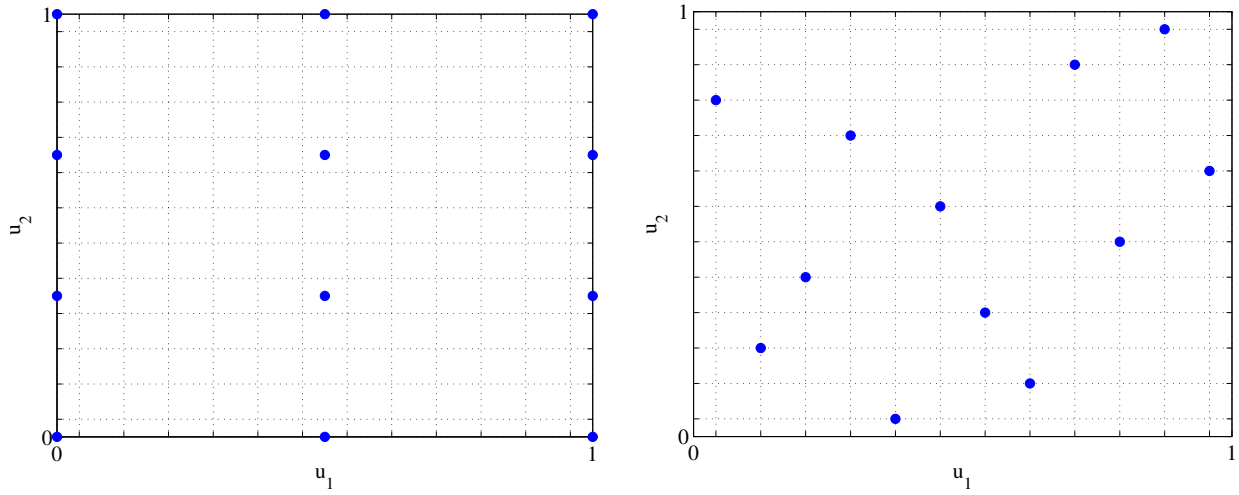
```
>> LH = lhsdesign(12, 2, 'smooth', 'off')

LH =

    0.7917    0.4583
    0.6250    0.1250
    0.8750    0.9583
    0.2917    0.7083
    0.2083    0.3750
    0.5417    0.2917
    0.1250    0.2083
    0.9583    0.6250
    0.4583    0.5417
    0.0417    0.7917
    0.3750    0.0417
    0.7083    0.8750
```

The option `'smooth', 'off'` ensures that the points are chosen as discussed above (i.e., derived from a full-factorial design). The routine `lhsdesign` randomly generates a Latin hypercube design, using a maximin criterion by default. The resulting design is shown in Fig. D.1b.

A considerable advantage of LHS is that for a given discretization of an input parameter (i.e., for a given n), the total number of points in the design is independent from the N number of input parameters, whereas an exponential dependence on N exists in the case of full-factorial design.



(a) Full-factorial design.

(b) Latin hypercube design.

Figure D.1: Two classical designs in case of $N = 2$, using the same $n = 12$ number of points.

D.3 Designs for discrete input spaces

Two classical sampling methods are presented in §§ D.1 and D.2. In both cases, a continuous input space $\mathbb{U} = [0, 1]^N$ was assumed. However, in our numerical examples, the input spaces –denoted by $\mathbb{P}$– are discrete, i.e., finite sets of the feasible input vectors (see § 2.4). Thus, the presented classical sampling methods need slight modifications to be applicable in the case of $\mathbb{P}$ as well.

We propose the following. First, the aimed sampling is carried out in the (fictive) continuous input space $\mathbb{U} = [0,1]^N$ as discussed in the previous sections. The corresponding design of n points (either a full-factorial or a Latin hypercube) is denoted by

$$\mathbb{S}_{\text{cont}} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}. \tag{D.8}$$

The design fitted to the given discrete input space $\mathbb{P}$ is then defined as

$$\mathbb{S}_{\text{disc}} = \left\{ \mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m \;\middle|\; \mathbf{p}_i \in \mathbb{P},\ \forall \mathbf{u}_k \in \mathbb{S}_{\text{cont}}\ \exists \mathbf{p}_l : \mathbf{p}_l = \arg\min_{\mathbf{p}\in\mathbb{P}} \|\mathbf{p} - \mathbf{u}_k\| \right\}. \tag{D.9}$$

In words, all points of the continuous design are represented by the nearest point of $\mathbb{P}$ in the discretized design. The discretized design might contain fewer samples than the continuous design, i.e., $m \leq n$. Furthermore, the equispaced levels of a continuous factorial design might become non-equispaced after the discretization. The Latin hypercube designs might lose the defining property (D.7), thus, the discretized LHS must be considered as an approximate LHS.

D.4 Piecewise multilinear interpolation

The piecewise multilinear interpolation is a popular tool to approximate a scalar function or even functional data. Let $q_{\mathbf{u}}(t)$ be a function-valued function of $\mathbf{u} \in \mathbb{U} = [0,1]^N$ with $t \in T$. Let us further assume that a full-factorial design $\mathbb{FF}$ with levels

$$\mathbb{L}_k = \left\{ u_k^{(1)} = 0, u_k^{(2)}, \ldots, u_k^{(l_k)} = 1 \right\},\ 0 < u_k^{(1)} < u_k^{(2)} < \cdots < u_k^{(l_k-1)} < 1,\ \forall k = 1, 2, \ldots, N \tag{D.10}$$

is given and the function to be interpolated is observed at all points of the full-factorial design. The idea of the piecewise linear interpolation method is to partition $\mathbb{U}$ into disjunct subdomains (whose union gives $\mathbb{U}$) being hyper-cuboids:

$$\mathbb{U}_{r_1 r_2 \ldots r_N} = \left[u_1^{(r_1)}, u_1^{(r_1+1)}\right] \times \left[u_2^{(r_2)}, u_2^{(r_2+1)}\right] \times \ldots \left[u_N^{(r_N)}, u_N^{(r_N+1)}\right], \tag{D.11}$$

where $r_i = 1, 2, \ldots, l_i - 1$ ($i = 1, 2, \ldots, N$). The approximation of $q_{\mathbf{u}}(t)$ in the subdomain with indices $r_1 r_2 \ldots r_N$ is given in a multilinear form (being linear in each input parameter):

$$\hat{q}_{\mathbf{u}}^{(r_1 r_2 \ldots r_N)}(t) = \sum_{s_1, s_2, \ldots, s_N = \{0,1\}} c_{s_1 s_2 \ldots s_N}(t) u_1^{s_1} u_2^{s_2} \ldots u_N^{s_N}. \tag{D.12}$$

The 2^N functional coefficients $c_{s_1 s_2 \ldots s_N}(t)$ are determined by using the 2^N observed functions at the corner points of the N-dimensional subdomain: the method interpolates the given data, i.e.,

$$\hat{q}_{\mathbf{u}}^{(r_1 r_2 \ldots r_N)}(t) = q_{\mathbf{u}}^{(r_1 r_2 \ldots r_N)}(t), \quad \forall \mathbf{u} \in \mathbb{FF}. \tag{D.13}$$

The extension of the method to the case of discrete input spaces (like in our examples, see § 2.4) is trivial.

E

Forest characterization by means of adaptive databases

A novel application of the OSF sampling strategy (§ 5) has recently been proposed in the framework of forest characterization problems based on radar observations. The study on this application is a joint work with *Laetitia Thirion-Lefèvre*, and currently a student (*András Vaskó*) is working on his master thesis in this domain under our supervision. In this chapter, we briefly summarize the basics and our preliminary results (also reported in [Thirion et al., 2010; Vaskó et al., 2011]). Some parts are literally quoted from those papers.

E.1 Introduction

In the frame of forest observation, e.g., in the purpose of biomass retrieval, simple scattering models are used to perform inversion. For instance, the Random Volume over Ground (RVoG) model [Cloude and Papathanassiou, 2003] is widely used to retrieve the mean height of the forest. However, the use of more complex scattering models (e.g., [Thirion, Colin, and Dahon, 2006]) for inversion appears to be difficult due to several bottlenecks, particularly, the high number of descriptive parameters and the computational cost of the numerical simulation. Since the simulators of the scattering phenomena of forests are similar black-box tools as the ones we considered in the frame of ECT, we attempt to make use of our adaptive sampling strategies also within the frame of forest observations. In particular, we have generated OSF databases. Our goal was twofold: besides illustrating that the OSF strategy works for other applications (different from ENDE), the structure of such OSF databases might carry some meta-information on the studied problem, which might be turned into valuable conclusions by the experts of the radar community.

E.2 Some numerical examples

E.2.1 The forest characterization setup

The studied forest is a maritime pine tree forest (called Nezer), and it is located in the south part of France. The ground truth data have been provided by the *Institut National de la Recherche Agronomique (INRA)* in the frame of the PYLA'2004 campaign, conducted by *Office National d'Études et de Recherches Aéronautiques (ONERA)*. Two different stands are available for study: stand 1, which is 52 years old and 26 m high and stand 2, being 26 years old and 18.5 m high. These two stands are accurately described in [Thirion, Colin, and Dahon, 2006]. For the radar configuration illustrated in Fig. E.1, we consider that the plane flies at an altitude of 3.5 km. For the interferometric study, the baseline components may vary, but the ambiguity height remains constant around 100 m.

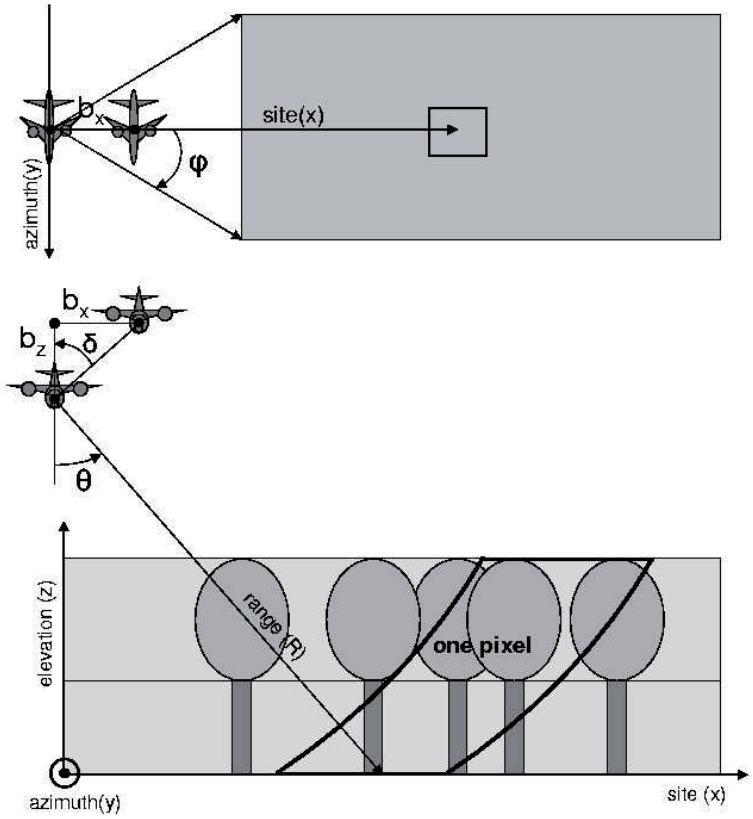


Figure E.1: Radar configuration: above view (top) and side view (bottom). Picture taken from [Thirion, Colin, and Dahon, 2006].

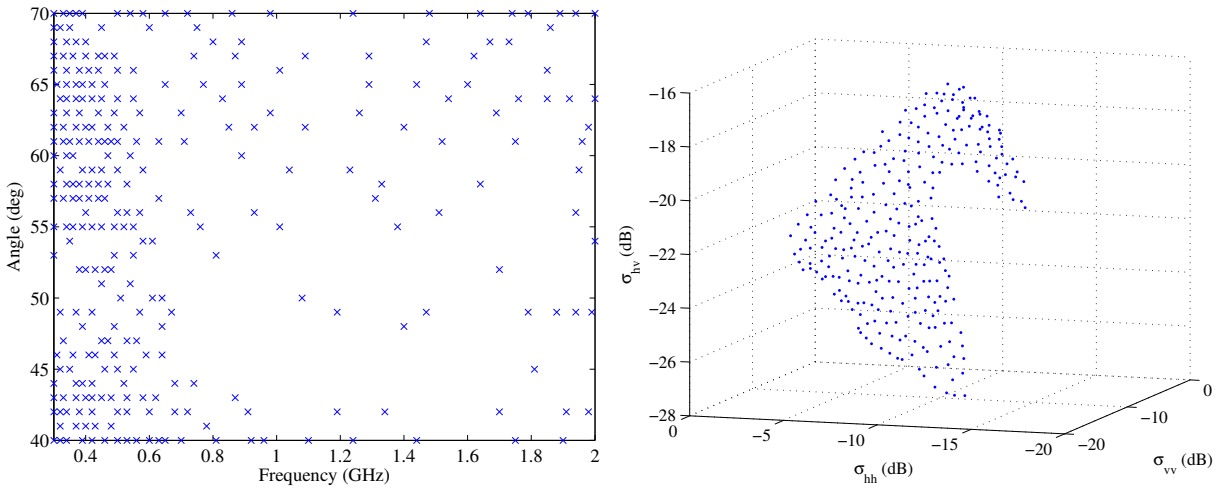
Table E.1: The variation domain and discretization of the input parameters (the input space).

	min	max	step
Frequency (GHz)	0.3	2.0	0.01
Incident angle (deg)	40	70	1

Our study focuses on three quantities: the polarimetric scattering coefficient σ_{qp} where p, q refer to the polarizations of the incident and scattered waves, which are either vertical or horizontal, the interferometric height h_{qp} and the total attenuation α_{pp} (α_{vv} or α_{hh}) of the forested area under study. A long term purpose is to retrieve the mean height of the forest under observation and to relate it to its mean attenuation. As a first step, herein we studied the behavior of σ_{qp}, h_{qp} and α_{pp} with the radar frequency f and the incidence angle θ_i.

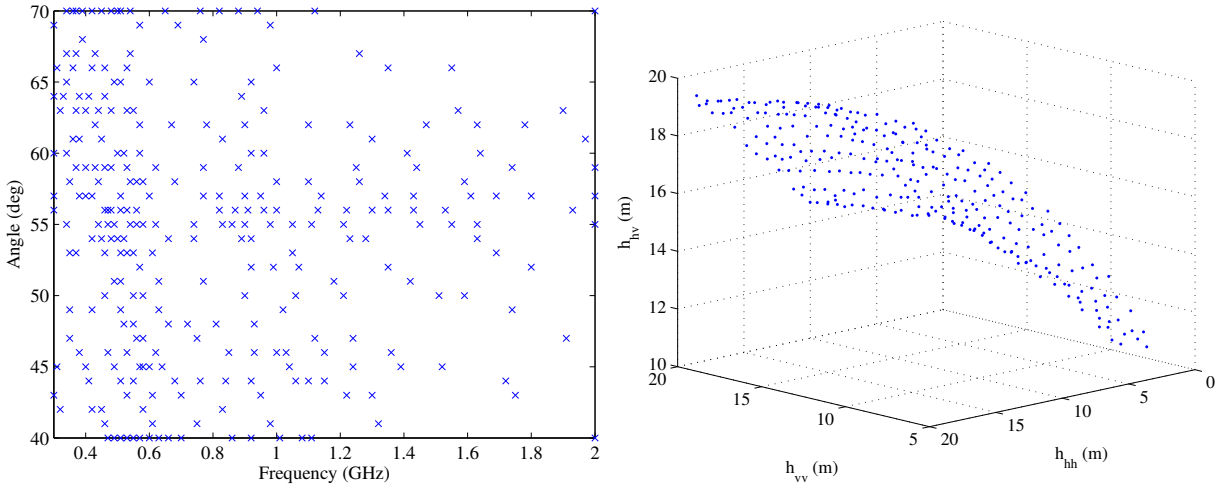
To do so, we have been defining two variation domains for these variables, leading to the input space $\mathbb{P}$, being characterized in Table E.1.

The output quantites (σ_{qp}, h_{qp} and α_{pp}) are treated separately, i.e., three different databases are generated. In all cases, the final number of samples in the database is 288. One output sample is a vector of $[\sigma_{vv}, \sigma_{hh}, \sigma_{hv}]$, $[h_{vv}, h_{hh}, h_{hv}]$ or $[\alpha_{vv}, \alpha_{hh}]$, respectively. The output space $\mathbb{Q}$ is embedded in 3-dimensional (3D) Euclidean space in the two first cases and it is 2D in the third case. The distance in $\mathbb{Q}$ is defined by the classical Euclidean norm in all cases.



(a) Input space. (b) Output space.

Figure E.2: Optimal database of the polarimetric scattering coefficients σ_{pq}.



(a) Input space. (b) Output space.

Figure E.3: Optimal database of the interferometric height h_{pq}.

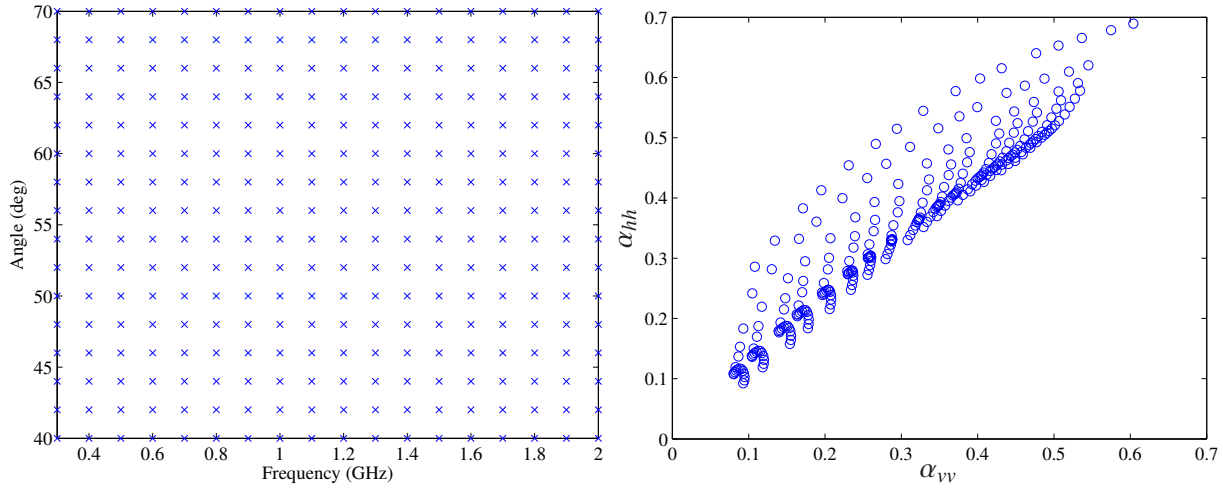
E.2.2 The obtained OSF databases and conclusions

The optimal database of the polarimetric scattering coefficient is presented in Fig. E.2. One can see that the output space (a 2D surface in a 3D space) is more or less evenly filled by samples, whereas in the input space, many more samples are concentrated in the region of lower frequencies. The reason of this is that σ is more sensitive to the frequency in this region than in the domain of higher frequencies.

In Fig. E.3, the database of the interferometric height h_{pq} is presented. Similar properties can be noticed as in the case of σ.

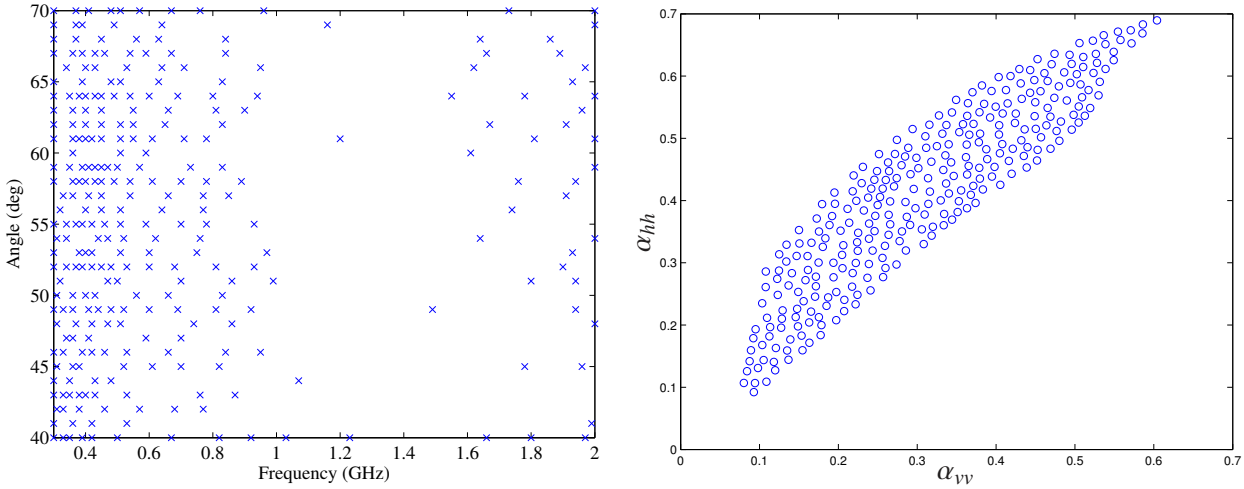
The total attenuation α_{pp} is easier to visualize as it has only two components in the output space. This is why we have generated a regular database (with a LHS design in the input space) as well. This database is shown in Fig. E.4. It can easily be seen that such a regular input sampling leads to quite distorted output sampling which is to be avoided. Our sampling strategy yields the optimal database shown in Fig. E.5.

As a conclusion of the studies, the repartition of the input samples in the case of an OSF sampling



(a) Input space. (b) Output space.

Figure E.4: Regular database of the total attenuation α_{pp}.



(a) Input space. (b) Output space.

Figure E.5: Optimal database of the total attenuation α_{pp}.

provides information about what kind of measurements is required to characterize the area. Indeed, we have observed that up to about 600 MHz, we need to consider different incident angles (more than 27 angles per 100 MHz for the interferometric heights and more than 40 angles per 100 MHz for the backscattering coefficients). This is not the case above 600 MHz where only some incident angles are required (10 times less for the backscattering coefficients, more than 3 times less for the interferometric heights). This result seems logical with regards to the structure of the forest. At high frequencies, mainly the crown of the forest interacts with the EM waves and if we can assume that the crown may be approximated as a random volume, consequently, few incidence angles are required to characterize this interaction. On the contrary, at low frequencies, the waves penetrate down to the trunks' level, and we need more incidence angles to describe the polarimetric response of these almost vertical elements. This result could also illustrate the relevancy of some measurements. Indeed, at high frequencies, we may expect that the measurements yield a quite good estimation of the forest behaviour. At low frequencies, the behaviour seems so versatile that we cannot expect to derive conclusions using only few measurements.

E.3 Perspectives

As future ambitions, the retrieval of the true mean height and the biomass of the forest can be mentioned. However, in such cases, the applied forward simulator might be stochastic. So far we have only preliminary knowledge and ideas on this domain; however, we believe that at some point, "stochastic kriging" should be involved in the generation and/or use of the appropriate surrogate model (see, e.g., [Ankenman, Nelson, and Staum, 2010]).

Another promising idea is to apply the surrogate modeling scheme using functional kriging (discussed in § 4) to construct precise interpolation-based surrogate models for radar observation setups. The functional data might be, e.g., the different back scattering coefficients in function of the radar frequency and incident angle, whereas the input parameters might be related to the forest-description, such as the age of the forest or the relative moisture of the ground, the trunks, etc.

Bibliography

Co-authored papers in international peer-reviewed journal

Bilicz, S., M. Lambert, and S. Gyimóthy (2010b), "Kriging-based generation of optimal databases as forward and inverse surrogate models", *Inverse Problems*, **26**(7), 074012 (cit. on pp. 67, 101).

Bilicz, S., E. Vazquez, M. Lambert, S. Gyimóthy, and J. Pávó (2009a), "Characterization of a 3D defect using the Expected Improvement algorithm", *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, **28**(4), 851–864 (cit. on pp. 19, 99).

Bilicz, S., M. Lambert, E. Vazquez, and S. Gyimóthy (2010d), "Combination of Maximin and Kriging Prediction Methods for Eddy-Current Testing Database Generation", *Journal of Physics: Conference Series*, **255**(1), 012003 (cit. on pp. 67, 101).

Bilicz, S., E. Vazquez, S. Gyimóthy, J. Pávó, and M. Lambert (2010e), "Kriging for eddy-current testing problems", *IEEE Transactions on Magnetics*, **46**(8), 3165–3168 (cit. on pp. 45, 100).

Co-authored paper in edited book

Bilicz, S., M. Lambert, E. Vazquez, and S. Gyimóthy (2010a), "A new database generation method combining maximin method and kriging prediction for eddy-current testing", in: *Electromagnetic Nondestructive Evaluation (XIII)*, ed. by J. Knopp, M. Blodgett, B. Wincheski, and N. Bowler, vol. 33, Studies in Applied Electromagnetics and Mechanics, Amsterdam: IOS Press, 199–206, ISBN: 978-1-60750-553-2 (cit. on pp. 67, 101).

Co-authored papers in proceedings of international conference

Bilicz, S., M. Lambert, and S. Gyimóthy (Aug. 2010a), "Inverse problem characterization using an adaptive database", in: *Electromagnetic Theory (EMTS), 2010 URSI International Symposium on*, Berlin, Germany, 509–512 (cit. on pp. 67, 101).

Bilicz, S., E. Vazquez, M. Lambert, S. Gyimóthy, and J. Pávó (Sept. 2008), "Characterization of a 3D defect using the Expected Improvement algorithm", in: *Proceedings of the 13th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, Graz, Austria, 157–162 (cit. on p. 19).

Bilicz, S., E. Vazquez, J. Pávó, M. Lambert, and S. Gyimóthy (July 2009b), "Eddy-current testing with the Expected Improvement optimization algorithm", in: *Proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, 1750–1755 (cit. on pp. 19, 99).

Bilicz, S., E. Vazquez, S. Gyimóthy, J. Pávó, and M. Lambert (Nov. 2009c), "Kriging for eddy-current testing problems", in: *17th Conference on the Computation of Electromagnetic Fields (COMPUMAG'09)*, Florianópolis, Brazil, 3–4 (cit. on p. 45).

Bilicz, S., E. Vazquez, M. Lambert, S. Gyimóthy, and J. Pávó (July 2009d), "The "Expected Improvement" global optimization algorithm for the solution of eddy-current testing inverse problems", in: *Proceedings of 14th International Workshop on Electromagnetic Nondestructive Evaluation (ENDE'09)*, Dayton, USA, 144–145 (cit. on pp. 19, 99).

Bilicz, S., S. Gyimóthy, M. Lambert, and J. Pávó (Sept. 2010b), "Adaptive kriging metamodels for expensive-to-run electromagnetic simulations", in: *Proceedings of the 14th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, CD-ROM, ISBN: 978-3-85125-133-3, IGTE, Graz, Austria, 214–219 (cit. on pp. 45, 100).

Bilicz, S., M. Lambert, S. Gyimóthy, J. Pávó, and G. Pichenot (June 2010c), "Characterization of eddy-current testing inverse problems using adaptive databases", in: *Proceedings of 15th International Workshop on Electromagnetic Nondestructive Evaluation (ENDE'10)*, Szczecin, Polland, 29–30 (cit. on pp. 67, 101).

Gyimóthy, S., I. Kiss, J. Pávó, and S. Bilicz (Nov. 2009), "Sensitivity based generation of optimized data set for ECT inversion", in: *17th Conference on the Computation of Electromagnetic Fields (COMPUMAG'09)*, Florianópolis, Brazil, 432–433 (cit. on p. 8).

Thirion, L., S. Bilicz, M. Lambert, and S. Gyimóthy (June 2010), "On the use of an optimal database for radar forest observation", in: *Proceedings of the 8th European Conference on Synthetic Aperture Radar*, Aachen, Germany, 873–876 (cit. on pp. 67, 101, 127).

Co-authored accepted papers

Bilicz, S., M. Lambert, S. Gyimóthy, and J. Pávó (July 2011a), "Kriging-based surrogate model for the solution of inverse problems in nondestructive testing", in: *18th Conference on the Computation of Electromagnetic Fields (COMPUMAG'11)*, Accepted. (cit. on pp. 45, 66, 100).

– (July 2011b), "Recent surrogate modeling approaches in electromagnetic nondestructive evaluation", in: *18th Conference on the Computation of Electromagnetic Fields (COMPUMAG'11)*, Accepted. (cit. on pp. 45, 100, 101).

Vaskó, A., L. Thirion-Lefevre, S. Bilicz, I. Champion, M. Lambert, and S. Gyimóthy (Sept. 2011), "Metamodel-based adaptive use of a coherent polarimetric backscattering simulator for the characterization of forested areas at low frequencies", in: *Proceedings of PIERS 2011 in Suzhou*, Accepted. (cit. on pp. 45, 100, 127).

Further references

Abascal, J. F. P. J., M. Lambert, D. Lesselier, and O. Dorn (2008), "3-D eddy-current imaging of metal tubes by gradient-based, controlled evolution of level sets", *IEEE Transactions on Magnetics*, **44**(12), 4721–4729 (cit. on p. 4).

Abubakar, A, W Hu, P. M. van den Berg, and T. M. Habashy (2008), "A finite-difference contrast source inversion method", *Inverse Problems*, **24**(6), 065004 (cit. on p. 4).

Abubakar, A. and P. M. van den Berg (2002), "The contrast source inversion method for location and shape reconstructions", *Inverse Problems*, **18**(2), 495 (cit. on p. 4).

Ammari, H., E. Iakovleva, D. Lesselier, and G. Perrusson (2007), "Music-type electromagnetic imaging of a collection of small three-dimensional inclusions", *SIAM Journal on Scientific Computing*, **29**, 674–709, 2 (cit. on p. 5).

Anger, G. (1990), *Inverse problems in differential equations*, New York and London: Plenum Press (cit. on p. 2).

Ankenman, B., B. L. Nelson, and J. Staum (2010), "Stochastic kriging for simulation metamodeling", *Operations Research*, **58**, 371–382, 2 (cit. on p. 131).

Badics, Z., J. Pávó, H. Komatsu, S. Kojima, Y. Matsumoto, and K. Aoki (1998a), "Fast flaw reconstruction from 3D eddy current data", *IEEE Transactions on Magnetics*, **34**(5), 2823–2828 (cit. on p. 4).

Badics, Z., H. Komatsu, Y. Matsumoto, and K. Aoki (1998b), "Inversion scheme based on optimization for 3-D eddy current flaw reconstruction problems", *Journal of Nondestructive Evaluation*, **17**(2), 67–78, 2 (cit. on p. 4).

Barton, R. (2009), "Simulation optimization using metamodels", in: *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 230 –238 (cit. on p. 8).

Bensetti, M., Y. Le Bihan, C. Marhand, and J. Pávó (2004), "Deposit characterization by eddy current nondestructive evaluation", *International Journal of Applied Electromagnetics and Mechanics*, **19**, 537–540 (cit. on p. 5).

Bettinger, R., P. Duchêne, L. Pronzato, and E. Thierry (2008), "Design of experiments for response diversity", *Journal of Physics: Conference Series*, **135**, 012017 (cit. on pp. 8, 68).

Biedermann, S. and H. Dette (2001), "Minimax optimal designs for nonparametric regression - a further optimality property of the uniform distribution", in: *mODa 6-advances in model-oriented design and analysis, Proceedings of the 6th international workshop on model-oriented design and analysis*, ed. by A. C. Atkinson, P. Hackl, and W. G. Müller, Springer-Verlag, 13–20 (cit. on p. 7).

Blatman, G. (2009), "Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis", PhD thesis, Université Blaise Pascal - Clermont II (cit. on p. 96).

Blitz, J. (1991), *Electrical and magnetic methods of nondestructive testing*, IOP Publishing (cit. on p. 1).

Bowler, J. R. (1994), "Eddy-current interaction with an ideal crack. I. The forward problem", *Journal of Applied Physics*, **75**(12), 8128–8137 (cit. on pp. 103, 104).

Bowler, J. R., S. J. Norton, and D. J. Harrison (1994), "Eddy-current interaction with an ideal crack. II. The inverse problem", *Journal of Applied Physics*, **75**(12), 8138 –8144 (cit. on p. 3).

Brus, D. J. and G. B. Heuvelink (2007), "Optimization of sample patterns for universal kriging of environmental variables", *Geoderma*, **138**(1-2), 86 –95 (cit. on pp. 7, 51).

Cacciola, M., S. Calcagno, F. C. Morabito, and M. Versaci (2007), "Swarm optimization for Imaging of corrosion by impedance measurements in eddy current test", *IEEE Transactions on Magnetics*, **43**(4), 1853–1856 (cit. on p. 4).

Canova, A, G Gruosso, and M Repetto (2003), "Magnetic design optimization and objective function approximation", *IEEE Transactions on Magnetics*, **39**(5), 2154–2162 (cit. on p. 9).

Chilès, J. and P. Delfiner (1999), *Geostatistics, Modeling Spatial Uncertainty*, Wiley (cit. on pp. 8, 42, 95, 109).

CIVA, "CIVA: State of the art simulation platform for NDE", URL: `http://www-civa.cea.fr` (cit. on p. 2).

Cloude, S. R. and K. P. Papathanassiou (2003), "Three-stage inversion process for polarimetric SAR interferometry", *Proc. Inst. Elect. Eng.Radar, Sonar Navig.* **150**, 125–134 (cit. on p. 127).

COMSOL, "COMSOL: Multiphysics Modeling and Simulation", URL: `http://www.comsol.com` (cit. on p. 2).

Couckuyt, I., F. Declercq, T. Dhaene, H. Rogier, and L. Knockaert (2010), "Surrogate-based infill optimization applied to electromagnetic problems", *International Journal of RF and Microwave Computer-Aided Engineering*, **20**(5), 492–501 (cit. on p. 9).

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, Wiley (cit. on p. 109).

Csallner, A. E. and J. Balogh (2007), "Optimization without derivatives: a simple direct search method", *Pollack Periodica*, **2**, 145–154 (cit. on p. 121).

CST, "CST: Complete Technology for 3D EM Simulation", URL: www.cst.com (cit. on p. 2).

Dalla'Rosa, A., A. Raizer, and L. Pichon (2008), "Optimal indoor transmitters location using TLM and Kriging methods", *IEEE Transactions on Magnetics*, **44**(6), 1354–1357 (cit. on p. 8).

Delicado, P., R. Giraldo, C. Comas, and J. Mateu (2009), "Statistics for spatial functional data: some recent contributions", *Environmetrics*, **21**(3-4), 224–239 (cit. on p. 46).

Dodd, C. V. and W. E. Deeds (1968), "Analytical solutions to eddy-current probe-coil problems", *Journal of Applied Physics*, **39**(6), 2829–2838 (cit. on p. 104).

Dorn, O. and D. Lesselier (2006), "Level set methods for inverse scattering", *Inverse Problems*, **22**(4), R67 (cit. on p. 4).

Dos Reis, D., M. Lambert, and D. Lesselier (2002), "Eddy-current evaluation of 3-D defects in a metal plate", *Inverse Problems*, **18**(6), 1857–1871 (cit. on p. 4).

Douvenot, R., M. Lambert, and D. Lesselier (June 2010a), "Metamodels as input of an optimization algorithm for solving an inverse eddy current testing problem", in: *Proceedings of 15th International Workshop on Electromagnetic Nondestructive Evaluation (ENDE'10)*, Szczecin, Polland, 25–26 (cit. on p. 6).

– (Aug. 2010b), "Particle optimization with metamodel for crack characterization", in: *Electromagnetic Theory (EMTS), 2010 URSI International Symposium on*, Berlin, Germany, 887–890 (cit. on p. 6).

Efron, B. and R. J. Tibshirani (1993), *An introduction to the bootstrap*, London: Chapman & Hall (cit. on p. 48).

Fang, K.-T., R. Li, and A. Sudjianto (2006), *Design and modeling for computer experiments*, London: Chapman & Hall/CRC (cit. on pp. 7, 67, 109, 124).

FLUX 3D, "FLUX 3D: Electromagnetic & thermal device design and Analysis Software", URL: http://www.cedrat.com (cit. on p. 2).

Franceschini, G., M. Lambert, and D. Lesselier (May 2009), "Adaptive database for eddy-current testing in metal tubes", in: *Proc. of the 8th International Symposium on Electric and Magnetic Fields (EMF 2009)*, 2 pp. (CD-ROM), Mondovì (cit. on pp. 6, 8).

Fukutomi, H., H. Huang, T. Takagi, and J. Tani (1998), "Identification of crack depths from eddy current testing signal", *IEEE Transactions on Magnetics*, **34**(5), 2893–2896 (cit. on p. 4).

Gao, Y. and X. Wang (2009), "Surrogate-based process optimization for reducing warpage in injection molding", *Journal of Materials Processing Technology*, **209**(3), 1302 –1309 (cit. on p. 9).

Geisser, S. (1993), *Predictive inference: an introduction*, Chapman & Hall/CRC (cit. on p. 114).

Giraldo, R., P. Delicado, and J. Mateu (2007), *Geostatistics for functional data: An ordinary kriging approach*, tech. rep., Universitat Politècnica de Catalunya (cit. on pp. 46, 48, 51).

Giraldo, R. H. (2009), "Geostatistical analysis of functional data", PhD thesis, Universitat Politècnica de Catalunya (cit. on p. 46).

Giunta, A. A., S. F. Wojtkiewicz Jr., and M. S. Eldred (Jan. 2003), "Overview of modern design of experiments methods for computational simulations", in: *Proc. of the 41th AIAA Aerospace Sciences Meeting And Exhibit*, Reno (cit. on p. 7).

Gyimóthy, S., I. Kiss, and J. Pávó (2009), "Adaptive sampling technique based on moving meshes for building data-equidistant inversion databases for NDT", *International Journal of Applied Electromagnetics and Mechanics*, **30**(3-4), 309–319 (cit. on pp. 5, 8).

Gyimóthy, S., Y. Le Bihan, and J. Pávó (2007), "Optimized database for training neural networks used in non-destructive testing", *International Journal of Applied Electromagnetics and Mechanics*, **25**(1-4), 717–721 (cit. on p. 5).

Gyimóthy, S. and J. Pávó (2005), "Qualification of the inverse problem of defect reconstruction using optimized mesh database", *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, **24**(2), 436–445 (cit. on p. 6).

Hadamard, J. (1952), *Lectures on Cauchy's problems in linear partial differential equations*, Mineola, New York: Dover Publication Inc. (cit. on pp. 2, 76).

Hammersley, J. M. and D. C. Handscomb (1975), *Monte carlo methods*, London: Methuen (cit. on p. 65).

Harrington, R. F. (1961), *Time-harmonic electromagnetic fields*, McGraw-Hill (cit. on p. 104).

– (1968), *Field computation by moment methods*, Macmillan (cit. on p. 104).

Henriksson, T., M. Lambert, and D. Lesselier (2010), "MUSIC-type algorithm for eddy-current nondestructive evaluation of small defects in metal plates", in: *15th International Workshop on Electromagnetic Non–Destructive Evaluation (ENDE'10)*, Szczecin, Polland, 151–152 (cit. on p. 5).

Hertog, D. d., J. P. C. Kleijnen, and A. Y. D. Siem (2006), "The correct Kriging variance estimated by bootstrapping", *Journal of the Operational Research Society*, **57**(4), 400–409 (cit. on p. 51).

Idier, J., ed. (2008), *Bayesian approach to inverse problems*, Wiley (cit. on p. 5).

Jin, R., W. Chen, and A. Sudjianto (2002), "On sequential sampling for global metamodeling in engineering design", in: *Proceedings of DETC'02 ASME 2002 Design Engineering Technical Conferences And Computers and Information in Engineering Conference*, Montreal, Canada (cit. on p. 8).

Jones, D. (2001), "A taxonomy of global optimization methods based on response surfaces", *Journal of Global Optimization*, **21**, 345–383 (cit. on p. 9).

Jones, D., M Schonlau, and W. Welch (1998), "Efficient global optimization of expensive black-box functions", *Journal of Global Optimization*, **13**(4), 455–492 (cit. on pp. 9, 20, 22, 26, 27).

Khuri, A. I. and J. A. Cornell (1996), *Response surfaces: designs and analyses*, CRC Press (cit. on p. 7).

Kleijnen, J. P. C. (2009), "Kriging metamodeling in simulation: A review", *European Journal of Operational Research*, **192**(3), 707–716 (cit. on pp. 8, 109).

Kleijnen, J. P. C. and v. W. C. M. Beers (2004), "Application-driven sequential designs for simulation experiments: Kriging metamodelling", *Journal of the Operational Research Society*, **55**, 876–883 (cit. on pp. 8, 23, 49, 51–53).

Knowles, J. (2006), "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems", *Evolutionary Computation, IEEE Transactions on*, **10**(1), 50 –66 (cit. on p. 9).

Kruskal, J. B. and M. Wish (1986), *Multidimensional Scaling*, SAGE (cit. on p. 76).

Lambert, M, D Lesselier, and B. J. Kooij (1998), "The retrieval of a buried cylindrical obstacle by a constrained modified gradient method in the H-polarization case and for Maxwellian materials", *Inverse Problems*, **14**(5), 1265 (cit. on p. 4).

Le Bihan, Y., J. Pávó, and C. Marchand (2008), "Characterization of small cracks in eddy current testing", *European Physical Journal - Applied Physics*, **43**(2), 231–237 (cit. on p. 5).

Leary, S., A Bhaskar, and A. Keane (2004), "A derivative based surrogate model for approximating and optimizing the output of an expensive computer simulation", *Journal Of Global Optimization*, **30**(1), 39–58 (cit. on p. 9).

Lebensztajn, L., C. A. R. Marretto, M. C. Costa, and J. L. Coulomb (2004), "Kriging: a useful tool for electromagnetic device optimization", *IEEE Transactions on Magnetics*, **40**(2), 1196–1199 (cit. on p. 8).

Mardia, K. V., J. T. Kent, C. R. Goodall, and J. A. Little (1996), "Kriging and splines with derivative information", *Biometrika*, **83**(1), 207–221 (cit. on p. 42).

Martin, J. D. and T. W. Simpson (2005), "Use of kriging models to approximate deterministic computer models", *AIAA JOURNAL*, **43**(4), 853–863 (cit. on p. 8).

McIntire, P., ed. (1986), *Nondestructive testing handbook*, 2nd ed., vol. 4, Electromagnetic Testing, Colombus, Ohio, USA: American Society for Nondestructive Testing (cit. on p. 1).

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code", *Technometrics*, **42**, 55–61 (cit. on pp. 7, 124).

Mockus, J., V. Tiesis, and A. Zilinskas (1978), "The application of Bayesian methods for seeking the extremum", in: *Towards Global Optimisation*, ed. by L. C. V. Dixon and G. P. Szego, vol. 2, Amsterdam: North-Holland, 117–129 (cit. on p. 9).

Monebhurrun, V., B. Duchêne, and D. Lesselier (1998), "Three-dimensional inversion of eddy current data for non-destructive evaluation of steam generator tubes", *Inverse Problems*, **14**(3), 707 (cit. on p. 4).

Morgans, R. C., A. C. Zander, C. H. Hansen, and D. J. Murphy (2008), "EGO shape optimization of horn-loaded loudspeakers", *Optimization and Engineering*, **9**, 361–374 (cit. on p. 9).

Morris, M. D., T. J. Mitchell, and D. Ylvisaker (1993), "Bayesian design and analysis of computer experiments: use of derivatives in surface prediction", *Technometrics*, **35**(3), 243–255 (cit. on p. 9).

Nair, S. M. and J. H. Rose (1990), "Reconstruction of three-dimensional conductivity variations from eddy current (electromagnetic induction) data", *Inverse Problems*, **6**(6), 1007–1030 (cit. on p. 3).

Norton, S. J. and J. R. Bowler (1993), "Theory of eddy current inversion", *Journal of Applied Physics*, **73**(2), 501–512 (cit. on p. 3).

Okabe, A., B. Boots, K. Sugihara, and S. N. Chiu (2000), *Spatial tessellations: concepts and applications of voronoi diagrams*, 2nd ed., Wiley (cit. on p. 77).

Park, W.-K. (2010), "On the imaging of thin dielectric inclusions buried within a half-space", *Inverse Problems*, **26**(7), 074008 (cit. on p. 5).

Park, W.-K. and D. Lesselier (2009), "MUSIC-type imaging of a thin penetrable inclusion from its multi-static response matrix", *Inverse Problems*, **25**(7), 075002 (cit. on p. 5).

Pávó, J. (2000), "Reconstruction of group of cracks in plate specimens using ECT impedance data", in: *Electromagnetic Nondestructive Evaluation (IV)*, ed. by S. S. Udpa, T. Takagi, J. Pávó, and R. Albanese, vol. 17, Studies in Applied Electromagnetics and Mechanics, Amsterdam: IOS Press, 204–211 (cit. on p. 104).

Pávó, J. and S. Gyimóthy (2007), "Adaptive inversion database for electromagnetic nondestructive evaluation", *NDT & E International*, **40**, 192–202 (cit. on pp. 5, 8).

Pávó, J and D Lesselier (2006), "Calculation of eddy current testing probe signal with global approximation", *IEEE Transactions on Magnetics*, **42**(4), 1419–1422 (cit. on p. 15).

Pávó, J. and K. Miya (1994), "Reconstruction of crack shape by optimization using eddy current field measurement", *IEEE Transactions on Magnetics*, **30**(5), 3407–3410 (cit. on pp. 4, 104).

Pávó, J. (1994), "Reconstruction of crack shape in a conducting plate using eddy current measurements", PhD thesis, University of Tokyo (cit. on pp. 4, 103, 105).

Picheny, V., D. Ginsbourger, O. Roustant, and R. Haftka (2010), "Adaptive designs of experiments for accurate approximation of target region", *Journal of Mechanical Design*, **132**(7), 071008 (cit. on pp. 8, 50).

Picheny, V. (2009), "Improving accuracy and compensating for uncertainty in surrogate modeling", PhD thesis, University of Florida (cit. on p. 8).

Powell, M. J. D. (1998), "Direct search algorithms for optimization calculations", *Acta Numerica*, **7**, 287–336 (cit. on p. 121).

Prémel, D. and A. Baussard (2002), "Eddy-current evaluation of three-dimensional flaws in flat conductive materials using a Bayesian approach", *Inverse Problems*, **18**(6), 1873 (cit. on p. 5).

Pronzato, L. (2008), "Optimal experimental design and some related control problems", *Automatica*, **44**(2), 303–325 (cit. on p. 7).

Queipo, N. V., R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. K. Tucker (2005), "Surrogate-based analysis and optimization", *Progress in Aerospace Sciences*, **41**(1), 1–28 (cit. on p. 8).

Quenouille, M. H. (1949), "Approximate tests of correlation in time series", *Journal of the Royal Statistical Society, Series B*, **11**, 18–44 (cit. on p. 48).

– (1956), "Notes on bias in estimation", *Biometrika*, **61**, 353–360 (cit. on p. 48).

Ramuhalli, P., L. Upda, and S. Upda (2001), "Neural network algorithm for electromagnetic NDE signal inversion", in: *Electromagnetic Nondetructive Evaluation (V)*, ed. by J. Pávó, G. Vértesy, T. Takagi, and S. S. Udpa, vol. 21, Amsterdam: IOS Press, 121–128 (cit. on p. 5).

Ryan, T. P. (2007), *Modern experimental design*, Wiley (cit. on pp. 7, 67).

Sabbagh, H. A., D. J. Radecki, S. Barkeshli, B. Shamee, J. A. Treece, and S. A. Jenkins (1990), "Inversion of eddy-current data and the reconstruction of 3-dimensional flaws", *IEEE Transactions on Magnetics*, **26**(2), 626–629 (cit. on p. 3).

Sacks, J., B. S. Schiller, and J. W. Welch (1989), "Designs for computer experiments", *Technometrics*, **31**, 41–47 (cit. on p. 7).

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989), "Design and analysis of computer experiments", *Statistical Science*, **4**, 409–435 (cit. on pp. 7, 109).

Santner, T. J., B. J. Williams, and W. I. Notz (2003), *The design and analysis of computer experiments*, Springer (cit. on pp. 7, 67, 109, 124).

Santosa, F. (1996), "A level-set approach for inverse problems involving obstacles", *ESAIM Control, Optimisation and Calculus of Variations*, **1**, 17–33 (cit. on p. 4).

Shao, J. and D. Tu (1995), *The jackknife and bootstrap*, Springer-Verlag (cit. on pp. 48, 65).

Simone, G. and F. C. Morabito (2000), "Plural defect separation using wavelet Hough transform", in: *Electromagnetic Nondestructive Evaluation (IV)*, ed. by S. S. Udpa, T. Takagi, J. Pávó, and R. Albanese, vol. 17, Studies in Electromagnetics and Mechanics, Amsterdam: IOS Press, 196–203 (cit. on p. 5).

Souriau, L, B Duchêne, D Lesselier, and R. E. Kleinman (1996), "Modified gradient approach to inverse scattering for binary objects in stratified media", *Inverse Problems*, **12**(4), 463 (cit. on p. 4).

Steen, L. and J. J. Arthur Seebach (1970), *Counterexamples in topology*, Holt, Rinehart and Wintson (cit. on pp. 96, 111).

Stein, M. (1999), *Interpolation of spatial data: some theory for kriging*, New York: Springer (cit. on pp. 8, 109, 113).

Sudret, B. (2007), "Uncertainty propagation and sensitivity analysis in mechanical models, contribution to structural reliability and stochastic spectral methods", Habilitation Thesis. Université Blaise Pascal - Clermont II (cit. on p. 96).

Sykulski, J. K. (2007), "New trends in optimization in electromagnetics", *Przeglad Elektrotechniczny*, **83**(6), 13–18 (cit. on p. 8).

Takagi, T., M. Uesaka, and K. Miya (1997), "Electromagnetic NDE research activities in JSAEM", in: *Electromagnetic Nondestructive Evaluation*, ed. by T. Takagi, J. R. Bowler, and Y. Yoshida, vol. 1, Studies in Applied Electromagnetics and Mechanics, IOS Press, 9–16 (cit. on p. 12).

Tamburrino, A. and G. Rubinacci (2002), "A new non-iterative inversion method for electrical resistance tomography", *Inverse Problems*, **18**(6), 1809 (cit. on p. 5).

Tamburrino, A., S. Ventre, and G. Rubinacci (2010), "Recent developments of a monotonicity imaging method for magnetic induction tomography in the small skin-depth regime", *Inverse Problems*, **26**(7), 074016 (cit. on p. 5).

Taniguchi, T., D. Kacprzak, S. Yamanda, M. Iwahara, and T. Miyagoshi (2001), "Wavelet-based image processing for inspection of printed circuit board using eddy-current testing technique", in: *Electromagnetic Nondestructive Evaluation (V)*, ed. by J. Pávó, G. Vértesy, T. Takagi, and S. S. Udpa, vol. 21, Studies in Electromagnetics and Mechanics, Amsterdam: IOS Press, 226–233 (cit. on p. 5).

Taniguchi, T., K. Nakamura, S. Yamada, and M. Iwahara (2002), "Image processing in eddy-current testing for extraction of orientations of defects", *International Journal of Applied Electromagnetics and Mechanics*, **14**, 503–506 (cit. on p. 5).

Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, SIAM (cit. on pp. 2, 14).

Thirion, L., E. Colin, and C. Dahon (2006), "Capabilities of a forest coherent scattering model applied to radiometry, interferometry and polarimetry at P- and L-bands", *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 849–862 (cit. on pp. 127, 128).

Trillon, A., A. Girard, J. Idier, Y. Goussard, F. Sirois, S. Dubost, and N. Paul (2010), "Eddy current tomography based on a finite difference forward model with additive regularization", *AIP Conference Proceedings*, **1211**(1), ed. by D. O. Thompson and D. E. Chimenti, 782–789 (cit. on p. 4).

Tukey, J. W. (1958), "Bias and confidence in not-quite large samples (abstract)", *The Annals of Mathematical Statistics*, **29**, 614 (cit. on p. 48).

van Beers, W. C. M. (2005), "Kriging metamodeling for simulation", PhD thesis, Tilburg University (cit. on p. 52).

van Beers, W. C. M. and J. P. C. Kleijnen (2008), "Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping", *European Journal of Operational Research*, **186**, 1099–1113 (cit. on p. 8).

van den Berg, P. M. and R. E. Kleinman (1997), "A contrast source inversion method", *Inverse Problems*, **13**(6), 1607 (cit. on p. 4).

Vazquez, E. and J. Bect (2007), "On the convergence of the expected improvement algorithm", *arXiv:0712.3744v1* (cit. on p. 22).

Vazquez, E. and E. Walter (Dec. 2005), "Estimating derivatives and integrals with Kriging", in: *44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05*, 8156–8161 (cit. on p. 42).

Vazquez, E. and J. Bect (2010), "Convergence properties of the expected improvement algorithm with fixed mean and covariance functions", *Journal of Statistical Planning and Inference*, **140**(11), 3088–3095 (cit. on p. 22).

Viana, F. A. C., C. Gogu, and R. T. Haftka (Aug. 2010), "Making the most out of surrogate models: tricks of the trade", in: *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2010*, Montreal, Canada, 28813 (cit. on p. 8).

Villemonteix, J., E. Vazquez, and E. Walter (2008), "An informational approach to the global optimization of expensive-to-evaluate functions", *Journal of Global Optimization*, **44**(4), 509–534 (cit. on p. 9).

Voronoi, M. G. (1908), "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire: recherches sur les parallélloèdres primitifs", *Journal für die Reine und Angewandte Mathematik*, **134**, 198–287 (cit. on p. 77).

Walter, E. and L. Pronzato (1997), *Identification of parametric models from experimental data*, Heidelberg: Springer (cit. on p. 52).

Wand, M. P. and M. C. Jones (1995), *Kernel smoothing*, Chapman & Hall (cit. on p. 96).

XFdtd, XFdtd: 3D EM Simulation, URL: `http://www.remcom.com/xf7` (cit. on p. 2).

Xiong, Y., W. Chen, D. Apley, and X. Ding (2007), "A non-stationary covariance-based Kriging method for metamodelling in engineering design", *International Journal For Numerical Methods In Engineering*, **71**(6), 733–756 (cit. on p. 96).

Zaoui, F., C. Marchand, and J. Pávó (2001), "Stochastic crack inversion by an integral approach", in: *Electromagnetic Nondestructive Evaluation (V)*, ed. by J. Pávó, G. Vértesy, T. Takagi, and S. S. Upda, vol. 21, Studies in Applied Electromagnetics and Mechanics, Amsterdam: IOS Press, 129–136 (cit. on pp. 4, 41).