Gaussian processes-based excursion set estimation for scalar or vector black box functions. Application to the calibration of a numerical wind turbine simulator.

Clément Duhamel

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et technologies de l'information, Informatique
Spécialité : Mathématiques Appliquées
Unité de recherche : Laboratoire Jean Kuntzmann

Estimation d'ensembles d'excursion par processus gaussiens pour des fonctions boîtes noires à valeurs scalaires ou vectorielles. Application à la calibration d'un simulateur numérique éolien

Gaussian processes-based excursion set estimation for scalar or vector black box functions. Application to the calibration of a numerical wind turbine simulator.

Présentée par :

Clément DUHAMEL

Direction de thèse :

Clémentine PRIEUR PROFESSEURE DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Directrice de thèse
Céline HELBERT MAITRESSE DE CONFERENCES HDR, Ecole Centrale de Lyon	Co-directrice de thèse
Miguel MUNOZ ZUNIGA INGENIEUR DE RECHERCHE, IFP Énergies Nouvelles	Co-encadrant de thèse
Delphine SINOQUET INGENIEURE DE RECHERCHE, IFP Énergies Nouvelles	Co-encadrante de thèse

Rapporteurs :

DAVID GINSBOURGER
PROFESSEUR, UNIVERSITÄT BERN
RODOLPHE LE RICHE
DIRECTEUR DE RECHERCHE, CNRS DELEGATION RHONE AUVERGNE

Thèse soutenue publiquement le **21 novembre 2024**, devant le jury composé de :

ADELINE LECLERCQ-SAMSON, PROFESSEURE DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Présidente
CLEMENTINE PRIEUR, PROFESSEURE DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Directrice de thèse
CELINE HELBERT, MAITRESSE DE CONFERENCES HDR, ECOLE CENTRALE LYON	Co-directrice de thèse
DAVID GINSBOURGER, PROFESSEUR, UNIVERSITÄT BERN	Rapporteur
RODOLPHE LE RICHE, DIRECTEUR DE RECHERCHE, CNRS DELEGATION RHONE AUVERGNE	Rapporteur
JOSSELIN GARNIER, PROFESSEUR, ECOLE POLYTECHNIQUE	Examinateur
JULIEN BECT, MAITRE DE CONFERENCES, UNIVERSITE PARIS-SACLAY	Examinateur

Invités :

MIGUEL MUNOZ ZUNIGA
INGENIEUR DE RECHERCHE, IFP ÉNERGIES NOUVELLES
DELPHINE SINOQUET
INGENIEURE DE RECHERCHE, IFP ÉNERGIES NOUVELLES

Remerciements

Je tiens à exprimer toute ma reconnaissance à toutes les personnes qui ont contribué à la réalisation de ce projet de thèse.

Je remercie sincèrement mes rapporteurs, David Ginsbourger et Rodolphe Le Riche, pour leur précieuse relecture et leurs remarques, qui ont permis d'améliorer la qualité de ce manuscrit. J'adresse également mes remerciements aux autres membres du jury, Adeline Leclercq Samson, Julien Bect et Josselin Garnier, pour les discussions à venir lors de la soutenance, lesquelles, j'en suis certain, seront particulièrement enrichissantes pour moi.

Je tiens à exprimer ma profonde gratitude à mes encadrants de thèse : ma directrice, Clémentine Prieur, ma co-directrice, Céline Helbert, mon promoteur, Miguel Munoz Zuniga, et ma co-promotrice, Delphine Sinoquet, pour leur soutien constant et leurs précieux conseils tout au long de ce projet de recherche. J'ai énormément apprécié collaborer avec chacun d'entre vous et je vous suis très reconnaissant de m'avoir donné la chance d'entreprendre ce projet, de m'avoir suivi régulièrement et de m'avoir soutenu dans les moments difficiles pour mener à bien cette recherche.

Je remercie mes collègues de l'Inria et du Laboratoire Jean Kuntzmann (LJK), ainsi que les assistantes d'équipes (Annie Simon, Luce Coelho et Elodie Toihein) pour leur soutien essentiel, et mes camarades du bureau 195 pour l'ambiance de travail conviviale et les bons moments passés ensemble. Merci également à Céline Acary-Robert pour son aide dans la prise en main des clusters de calcul de Gricad. J'exprime également ma reconnaissance envers mes collègues de l'IFP Énergies Nouvelles, en particulier Nicolas Guy, Morgane Menz et Jean-Lou Pfister, pour leur aide précieuse sur l'application au simulateur éolien. Je remercie aussi les différents membres des groupes de recherche CIROQUO et RT-UQ (anciennement Mascot Num), avec qui j'ai pu échanger de manière enrichissante.

Un grand merci à mes anciens professeurs, Carole Saya, Christian Mejia-Taulé et Frédérique Durand, pour m'avoir transmis leur savoir et leur passion pour les mathématiques. Je suis profondément reconnaissant envers mes parents, ma sœur, mon frère et le reste de ma famille pour leur soutien indéfectible, ainsi qu'à mes amis, en particulier Amélie, Anthony, Clément, Paul, Sébastien, Thibault, et tous ceux rencontrés pendant la préparation à l'agrégation à Grenoble, pour leur amitié et leurs encouragements.

Enfin, je remercie les institutions qui ont rendu cette thèse possible, notamment l'Inria, l'IFP Énergies Nouvelles, le LJK, Gricad et toutes les autres organisations qui ont contribué à mon parcours académique.

À toutes celles et ceux mentionnés ici, votre soutien et votre collaboration ont été essentiels à la réussite de ce travail, et je vous en suis infiniment reconnaissant.

Mots Clés : Modélisation par processus gaussiens, Estimation d'ensembles d'excursion, Plans d'expériences séquentiels

Résumé

De nombreuses questions industrielles sont liées aux problèmes d'estimation d'ensembles d'excursion, sous la forme de l'estimation d'un ensemble de valeurs d'entrée admissibles de modèle, correspondant par exemple à des schémas de conception optimale admissibles. L'ensemble d'excursion que nous cherchons à estimer est défini comme l'ensemble des valeurs du modèle dans l'espace des entrées (espace de design) satisfaisant une ou des contraintes données sur les sorties du modèle. Dans ce qui suit, nous considérons le cadre d'un simulateur numérique de type boîte noire, pouvant être évalué en tout point de l'espace de design, mais telle que chacune de ces évaluations soit coûteuse en temps de calcul.

Une manière efficace d'estimer un ensemble d'excursion consiste à modéliser la fonction boîte noire comme la réalisation d'un processus gaussien. La construction séquentielle d'un plan d'expériences, c'est-a-dire un ensemble de points de l'espace de design ainsi que les évaluations associées par la fonction boîte noire, permet l'apprentissage progressif du modèle de substitution. Les points ajoutés au plan d'expériences séquentiel sont choisis en fonction de l'optimisation d'un critère d'acquisition, dépendant du modèle de substitution à l'étape courante. De nombreux types de critères d'acquisition peuvent être envisagés suivant l'objectif fixé sur la fonction boîte noire (connaissance globale, optimisation, estimation d'un ensemble d'excursion, etc.). Il existe également une classe de critères d'acquisition appelée "Stepwise Uncertainty Reduction" (SUR) ayant pour but d'anticiper l'impact de l'ajout de la prochaine évaluation au plan d'expériences sur une mesure d'incertitude bien choisie.

L'objectif de cette thèse est d'étudier les critères d'acquisition adaptés à l'estimation d'ensembles d'excursion, pour des fonctions boîtes noires à sorties scalaires ou vectorielles, et d'appliquer ces contributions à une application de calibration d'un simulateur numérique pour la conception d'éoliennes. La première partie se concentre, dans le cas de sorties scalaires, sur l'amélioration d'un critère d'acquisition adapté à l'estimation d'ensembles d'excursion en une version SUR. Le but de cette démarche est de combiner la robustesse du critère choisi due à son caractère exploratoire avec les bonnes performances en termes d'exploitation des stratégies de type SUR. La seconde partie se concentre sur le cadre de travail de modèles de type boîte noire avec sortie vectorielle, et pour lesquels toutes les composantes de sortie sont évaluées simultanément (données isotopiques). L'objectif de cette seconde partie est de développer plusieurs critères adaptés à l'estimation simultanée de chacun des ensembles d'excursion pour chaque composante de sortie du modèle. Parmi les critères proposés, on distingue des critères inspirés du cadre scalaire utilisant un modèle de substitution sur chaque composante de sortie du modèle et un autre critère utilisant un modèle de substitution de type Multi-Output Gaussian Process (MOGP) ayant pour objectif de prendre en compte la corrélation entre les différentes composantes de sortie de la fonction boîte noire. Les différentes stratégies proposées sont appliquées à la calibration d'un simulateur numérique pour la conception d'éoliennes. Le but de cette application est de trouver l'ensemble des paramètres d'entrée du simulateur (coefficients de raideur de certains matériaux), tels que les modes vibratoires calculés par le simulateur ne soient pas trop éloignés pour une norme adaptée et par rapport à des seuils fixés, des modes observés issus des données expérimentales.

Keywords: Gaussian process regression, Excursion set estimation, Sequential design of experiments

Abstract

Many industrial issues are related to excursion set estimation problems, in the form of estimating a set of feasible model input values, corresponding for example to feasible optimal design schemes. The excursion set we seek to estimate is defined as the set of model values in the input space (design space) satisfying a given constraint(s) on the model outputs. In the following, we consider the framework of a black box numerical simulator, which can be evaluated at any point in the design space, but for a high computational time.

An efficient way of estimating an excursion set is to model the black box function as the realization of a Gaussian process. The sequential construction of a design of experiments, i.e. a set of points in the design space and the associated evaluations by the black box function, enables the progressive learning of the surrogate model. The points added to the sequential design of experiments are chosen according to the optimization of an acquisition criterion, that depends on the surrogate model at the current stage. Many types of acquisition criteria can be considered, depending on the objective set for the black box function (global knowledge, optimization, estimation of an excursion set, etc.). There is also a class of acquisition criteria called "Stepwise Uncertainty Reduction" (SUR), whose aim is to anticipate the impact of adding the next evaluation to the experimental design on a well-chosen uncertainty measure.

The aim of this thesis is to study acquisition criteria suitable for estimating excursion sets, for black box functions with scalar or vector outputs, and to apply these contributions to the calibration of a numerical simulator for wind turbine design. The first part focuses, in the case of scalar outputs, on the improvement of an acquisition criterion adapted to the estimation of excursion sets into a SUR version. The aim of this approach is to combine the robustness of the chosen criterion due to its exploration property with good performance in terms of the exploitation of SUR-type strategies. The second part focuses on the framework of black box models with vector output, and for which all output components are evaluated simultaneously (isotopic data). The aim of this second part is to develop several criteria adapted to the simultaneous estimation of each of the excursion sets for each output component of the model. Among the proposed criteria, we distinguish between criteria inspired by the scalar framework using a surrogate model on each model output component, and another criterion using a Multi-Output Gaussian Process (MOGP) type surrogate model whose aim is to take into account the correlation between the different output components of the black box function. The different proposed strategies are applied to the calibration of a numerical simulator for wind turbine design. The aim of this application is to find a set of input parameters of the simulator (stiffness coefficients for certain materials), such that the vibration modes calculated by the simulator are not too far apart, for an adapted norm and in relation to fixed thresholds, from the observed modes derived from experimental data.

Contents

Introduction **1**

1 Background to Gaussian process regression **9**

1.1 Introduction . . . 9

1.2 Principle of GPR . . . 11

1.3 Covariance structure . . . 12

1.3.1 Covariance structure type . . . 12

1.3.2 Parameter estimation . . . 14

1.4 Examples . . . 15

1.5 Initial design of experiments . . . 15

1.6 Sequential DoE and enrichment criteria . . . 17

1.6.1 Global enrichment criteria . . . 19

1.6.2 Goal-oriented enrichment criteria for optimization . . . 19

1.7 Extension to vector-valued models . . . 20

1.7.1 Principle of multi-output Gaussian processes . . . 20

1.7.2 Covariance structure . . . 22

1.7.3 Autokrigeability . . . 24

2 Background and tools for estimating excursion sets **28**

2.1 Excursion set estimation and adapted enrichment criteria . . . 28

2.1.1 Excursion set estimation framework . . . 28

2.1.2 Deviation number criterion . . . 31

2.1.3 tmse, tMMSE and tIMSE criteria . . . 31

2.1.4 Bichon and Ranjan criteria . . . 32

2.2 Stepwise Uncertainty Reduction (SUR) strategies for excursion set estimation 35

2.2.1 Introduction to SUR strategies . . . 35

2.2.2 SUR Excursion measure variance and Integrated Bernouilli variance strategies . . . 37

2.2.3 Basics on Vorob'ev Theory and corresponding SUR strategy . . . 38

2.2.4 Extension of SUR Excursion measure variance and Integrated Bernoulli variance strategies for vector-valued models . . . 40

3 A SUR version of the Bichon criterion **43**

3.1 Introduction . . . 44

3.2 The framework for estimating excursion sets . . . 45

3.2.1 Some reminders on Gaussian process regression . . . 45

3.2.2 Towards more exploration: the Bichon criterion . . . 47

3.3 SUR Bichon criterion . . . 49
3.3.1 Reminders on SUR strategies . . . 49
3.3.2 Formulation of the SUR Bichon criterion . . . 50
3.4 Numerical experiments . . . 51
3.4.1 Implementation choices . . . 51
3.4.2 Performance tests on Branin-rescaled 2D function . . . 52
3.4.3 Performance tests on Hartmann-rescaled 6D function . . . 55
3.5 Conclusion . . . 57
3.6 Acknowledgements . . . 58
3.7 Appendices . . . 58
A Basics on Vorob'ev Theory and corresponding SUR strategy . . . 58
B Kriging update formulas . . . 60
C Proof of the explicit formula for SUR Bichon . . . 60
D Robustness of estimators with respect to the GP stationarity assumption. . . . 62
3.8 Additions to the article . . . 63
A Influence of the κ parameter . . . 63
B Use of Importance Sampling for evaluating the SUR Bichon criterion . . . 63

4 Excursion set estimation on vector valued models 66

4.1 The framework for estimating excursion sets . . . 67
4.1.1 Reminders on Gaussian process regression and scalar Bichon criterion . . . 67
4.1.2 Excursion sets in the case of several outputs . . . 68
4.2 Two natural extensions of Bichon criterion . . . 70
4.2.1 Alternating Scalar Bichon criterion . . . 70
4.2.2 Pareto Scalar Bichon criterion . . . 70
4.3 Vector output extension to Bichon criterion . . . 72
4.3.1 Reminders on multi-output Gaussian process regression . . . 72
4.3.2 The proposed Vector Bichon criterion . . . 73
4.4 Numerical experiments . . . 74
4.4.1 Implementation choices . . . 75
4.4.2 Performance tests based on the 2D Branin function . . . 76
4.4.3 Performance tests based on the 4D Hartmann function . . . 84
4.5 Conclusion . . . 88
4.6 Appendices . . . 90
A Proof of the integral formulation for Vector Bichon criterion . . . 90
B Estimation of $H : (\alpha_1, \alpha_2, \rho) \mapsto \int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t$ . . . 92

5 Multi-output excursion set estimation applied to the pre-calibration of a wind turbine numerical model 98

5.1 Introduction . . . 99
5.1.1 Excited deformation modes in wave theory . . . 99
5.1.2 The case of a wind turbine structure . . . 99
5.2 The wind turbine numerical model . . . 101
5.2.1 Presentation of the DTU 10MW Reference Wind Turbine . . . 101
5.2.2 Operational Modal Analysis (OMA) . . . 103
5.2.3 Model overview . . . 105

5.3 Mode matching . . . 107
5.3.1 Matching example and mode matching algorithm . . . 107
5.3.2 Checking the mode matching . . . 109
5.4 Pre-calibration as a multiple excursion set estimation problem . . . 110
5.5 Pre-calibration using only the two main modes (Formulation 1) . . . 112
5.6 Pre-calibration using frequencies and modes (Formulation 2) . . . 115
5.7 Conclusion . . . 118
5.8 Appendix . . . 120
A Checking mode matching with respect to vibration frequencies . . . 120

Conclusion/perspectives **125**

List of Figures

1.1 Illustration of GP realizations conditioned by observations. 10
1.2 On the left, comparison of different types of covariance kernel for fixed parameters $\theta = \sigma = 1$, as a function of the distance between the two points under consideration. On the right, comparison of examples of associated trajectories on the interval $[0, 1]$, with a constant zero trend. 14
1.3 Comparison of different types of GPR, based on 5 observation points uniformly distributed over $[0, 1]$, for a black box function simulated from a centered Gaussian process of kernel covariance Matérn 5/2 ($\theta = 0.2$ and $\sigma = 1$). 16
1.4 Comparison of an LHS plan and an LHS plan optimized by the Maximin criterion, for 10 points on $[0, 1]^2$. 18
1.5 Diagram of the sequential DoE construction, coupled with GPR. 18

2.1 Representation of Feasibility Function (bottom) for a given example of a GP sample path conditioned on 5 evaluations of the g function, a threshold T set to 0 and $\epsilon(x) := \sigma_n(\mathbf{x})$ (top). The height of the gray area represents, for each value of x, the value of FF(x). 33

3.1 Diagram of the sequential DoE construction, coupled with GPR. 47
3.2 Representation of Feasibility Function (bottom) for a given example of a GP sample path conditioned on 5 evaluations of the g function, a threshold T set to 0 and $\epsilon(x) := \sigma_n(\mathbf{x})$ (top). 48
3.3 Representation of the Branin-rescaled function on $\mathbb{X}$. 52
3.4 Line plots (with logarithmic scale) of the approximation error $\mathrm{Err}(\hat{\Gamma}_i(\chi_{j,n}))$ for the different criteria during 20 iterations, for Branin-rescaled function inversion ($d = 2$) with $T = 10$, for 100 different initial DoEs of size 10 of type LHS Maximin, for $\kappa = 1$ and with n.points= 10^4. Left column: Bichon with naive estimator (a) and Vorob'ev estimator (d). Middle column: SUR Bichon with naive estimator (b) and Vorob'ev estimator (e). Right column: SUR Vorob'ev with naive estimator (c) and Vorob'ev estimator (f). 53
3.5 Representation of two $\hat{\Gamma}$ estimators for each of the two criteria after 20 iterations, in comparison with the true excursion set (top right), for Branin-rescaled function inversion ($d = 2$) with $T = 10$, for a particular initial DoE of size 10 where SUR Bichon outperforms SUR Vorob'ev. Left column: SUR Bichon with naive estimator (a) and Vorob'ev estimator (d). Middle column: SUR Vorob'ev with naive estimator (b) and Vorob'ev estimator (e). Right column: true excursion set and Branin-rescaled contour lines (c). 54

3.6 Line plots (with logarithmic scale) of the approximation error $\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_i(\chi_{j,n})\Delta\Gamma^\star)/\mathbb{P}_{\mathbb{X}}(\Gamma^\star)$ (with a Sobol' sequence of size 10^4) for the different criteria during 600 iterations, for Hartmann-rescaled function inversion ($d = 6$) with $T = -1.6$, for 50 different initial DoEs of size 30 of type LHS Maximin, for $\kappa = 1$ and with n.points= 10^4. Left column: SUR Bichon with naive estimator (a) and Vorob'ev estimator (c). Right column: SUR Vorob'ev with naive estimator (b) and Vorob'ev estimator (d). 56
3.7 Pairwise projection plot of $\hat{\Gamma}_1\Delta\Gamma^\star$ for a 5.10^3-Sobol' sequence, for the two criteria after 600 iterations, for Hartmann-rescaled function inversion ($d = 6$) with $T = -1.6$, for a particular initial DoE of size 30 where SUR Bichon outperforms on SUR Vorob'ev, for $\kappa = 1$ and for n.points= 10^4. Left column: SUR Bichon. Right column: SUR Vorob'ev. 57
3.8 At left and center, representation of the two $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ estimators after 100 iterations (first line) and 200 iterations (second line) for SUR Vorob'ev, and with the contour lines of p_n (first column) and m_n (second column), in comparison with the true excursion set (top right), for Loggruy function inversion ($d = 2$) with $T = 10\,log_{10}(3.6)$, for a particular initial DoE of size 10. 64

4.1 Left and middle: level sets for each component of the 2D function, and the partial excursion set $\Gamma_1^\star$ for threshold $T_1 = 10$ (left), $\Gamma_2^\star$ for threshold $T_2 = 10$ (middle). On the right, comparison of these partial excursion sets with the global excursion set $\mathbf{\Gamma^\star}$. 69
4.2 Example of a Pareto front. 71
4.3 Representation of the model based on the 2D Branin function with the first component on the left and the second one on the right and the contour of excursion sets in blue (with $\boldsymbol{T} = (10, 10)$). 77
4.4 Averaged partial relative errors and of their sum with respect to the number of iterations when estimating excursion sets of the enhanced 2D Branin test function with $\boldsymbol{T} = (10, 10)$. The three enrichment stategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Averages are evaluated over 40 LHS Maximin intial DoEs randomly chosen. 77
4.5 Functional boxplots of partial relative errors and of their sum, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin test function with $\boldsymbol{T} = (10, 10)$. 79
4.6 Standard "Data profiles" of partial relative errors with threshold set equal to 20%, 10% and 5%, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10)$. 80
4.7 Enrichment calculation time (assuming that the evaluation time of $\boldsymbol{g}$ is negligible) for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10)$. 81
4.8 "Data profiles" as a function of total computing time of partial relative errors for evaluation times of $\boldsymbol{g}$ of 3h, 10min and 1min, with $C = 10\%$, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10)$. 82

4.9 Averaged partial relative errors and of their sum with respect to the number of iterations when estimating excursion sets of the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10\,000)$. The three enrichment stategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Averages are evaluated over 40 LHS Maximin intial DoEs randomly chosen. 83
4.10 Functional boxplots of partial relative errors for the first component, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10\,000)$. 83
4.11 Averaged partial relative errors and of their sum with respect to the number of iterations when estimating excursion sets of the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$. The three enrichment stategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 20 and with 200 enrichment iterations. Averages are evaluated over 40 LHS Maximin intial DoEs randomly chosen. 85
4.12 Functional boxplots of partial relative errors and of their sum, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$. 86
4.13 Standard "Data profiles" of partial relative errors with $C = 20\%$, 10% and 5%, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$. 87
4.14 Enrichment calculation time (assuming that the evaluation time of $\boldsymbol{g}$ is negligible) for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$. 88
4.15 "Data profiles" as a function of total computing time of partial relative errors for evaluation times of $\boldsymbol{g}$ of 3h, 10min and 1min, with $C = 10\%$, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the enhanced 4D Hartmann test function with $\boldsymbol{T} = (-1.6, -1)$. 89
4.16 Representation of the level sets of H (with $\kappa = 1$) on a grid of size 50×50 for $\alpha := (\alpha_1, \alpha_2) \in [-4, 4]$, and for ρ values of -1, -0.5, 0, 0.5 and 1. 94
4.17 Comparison between a regular grid and an irregular grid resulting from a cubic transformation, on $[-1, 1]^2$, with 15 points per dimension. 95

5.1 Vibrational modes of a circular membrane (CC BY-SA-ND 4.0; [Russell, 2018]). 99
5.2 Tacoma Narrows bridge collapse, Nov. 1940. 99
5.3 Illustration of offshore wind turbines. 100
5.4 Schematic diagram of the wind turbine simulator. 100
5.5 Plots of the DTU 10 MW RWT geometry from [Bak et al., 2013]. 101
5.6 2D mesh around the FFA-W3-336 airfoil fitted with a Gurney flap from [Bak et al., 2013]. 102
5.7 Representation of the blade's composite stratification, divided into 11 circumferential and 100 radial regions from [Bak et al., 2013]. 103
5.8 OMA summary diagram. 104
5.9 Schematic diagram of the black box wind model. 106

5.10 On the left, comparison of simulated (red) and reference (black) frequencies for $\Theta = (1.05, 1.1, 1.1)$. On the right, representation of the table of dissimilarity measure values between simulated and reference modes. 108
5.11 Boxplots (in logarithmic scale) of the minimum matching ratio for Θ values from a Maximin LHS design of size 100 on $[0.8, 1.2] \times [0.6, 1.4]^2$. 109
5.12 Schematic diagram of Formulation 1 of the pre-calibration problem using only the two main modes. 111
5.13 Representation of the logarithm of dissimilarity measures $\text{Meas}_1(\Theta)$ (left) and $\text{Meas}_2(\Theta)$ (right), on a grid of size 30 by 30 for the first two components of Θ on $[0.8, 1.2] \times [0.6, 1.4]$. 111
5.14 Schematic diagram of Formulation 2 of the pre-calibration problem using modes and frequencies. 112
5.15 Representation of the logarithm of the sum of dissimilarity measures $\text{Meas}_i^{\lambda}(\Theta)$ (left) and $\text{Meas}_i(\Theta)$ (right), on a grid of size 30 by 30 for the first two components of Θ on $[0.8, 1.2] \times [0.6, 1.4]$. 113
5.16 Representation of the enrichment of 3 initial LHS Maximin DoEs of size 5 after 30 iterations for the different enrichment criteria, for the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$. Boundaries of partial excursion sets are overlaid, calculated from a 30×30 grid. 114
5.17 Line plots and means of partial relative errors and of their sum with the number of iterations when estimating excursion sets of the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$. The three enrichment strategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Means are evaluated over 10 LHS Maximin intial DoE randomly chosen. 115
5.18 Standard "Data profiles" (DPs) of partial relative errors for the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$. DPs are represented for the different criteria, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations. 116
5.19 Representation of $\theta_i^{(1)}$ and $\theta_i^{(2)}$ hyperparameters of SOGP models associated with Alternating and Pareto Scalar criteria, and of θ_i and θ_{kOut} hyperparameters of MOGP model associated with Vector criterion, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations, for the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$. 117
5.20 Representation of the enrichment of 3 initial LHS Maximin DoEs of size 5 after 30 iterations for the different enrichment criteria, for the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$. Boundaries of partial excursion sets are overlaid, calculated from a 30×30 grid. 119
5.21 Line plots and means of partial relative errors and of their sum with the number of iterations when estimating excursion sets of the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$. The three enrichment strategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Means are evaluated over 10 LHS Maximin intial DoE randomly chosen. 120
5.22 Standard "Data profiles" (DPs) of partial relative errors for the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$. DPs are represented for the different criteria, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations. 121

5.23 Representation of $\theta_i^{(1)}$ and $\theta_i^{(2)}$ hyperparameters of SOGP models associated with Alternating and Pareto Scalar criteria, and of θ_i and θ_{kOut} hyperparameters of MOGP model associated with Vector criterion, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations, for the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$. 122
5.24 Boxplot (with logarithmic scale) of the distribution of the values of Equation (5.12) for 10 000 random permutations on $\{1, \dots 13\}$ and comparison with the same measure for the identity permutation and for the permutation resulting from mode matching algorithm (Algorithm 1). 123

List of Tables

1.1 Summary of some Matérn-type covariance kernels. 13
1.2 Summary of hyperparameters for the different GPR models used. 15
1.3 Summary of possible structures for $\mathbf{B}_q$ matrices (UT and LT stand for Upper Triangular and Lower Triangular). 24

3.1 Summary of empirical distributions of the $\mathrm{Err}(\hat{\Gamma}_i(\chi_{j,n}))$ for the different criteria after 10 and 20 iterations, for Branin-rescaled function inversion ($d = 2$) with $T = 10$, for 100 different initial DoEs of size 10 of type LHS Maximin. 54
3.2 Summary of empirical distributions of the approximation error $\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}\Delta\Gamma^\star)/\mathbb{P}_{\mathbb{X}}(\Gamma^\star)$ (with a Sobol' sequence of size 10^4) for the different criteria after 99, 300 and 600 iterations, for Hartmann-rescaled function inversion ($d = 6$) with $T = -1.6$, for 50 different initial DoEs of size 30 of type LHS Maximin, for $\kappa = 1$ and with n.points= 10^4. 57
3.3 Characteristic values of the empirical distributions of the comparison measure $\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}\Delta\Gamma^\star)$ (with a grid 200×200) for the different criteria after 20 iterations, in the case of the inversion of the Branin-rescaled function ($d = 2$) with $T = 10$, for 100 different initial plans of size 10 of type LHS Maximin, for $\kappa \in \{0.1, 0.5, 1, 2, 10\}$ and with $n.points = 10\,000$. 64

4.1 Absolute errors of different interpolation methods chosen with 200 interpolation points per dimension and for 10^5 points chosen randomly on $[-5, 5]^2 \times [-1, 1]$. 96
4.2 Absolute errors of different interpolation methods chosen with 200 interpolation points per dimension and for 10^5 points chosen randomly on $[-50, 50]^2 \times [-1, 1]$. 96
4.3 Comparison of interpolation function evaluation computation times for 10^4 randomly selected points, between the different interpolation methods selected. . . 96
4.4 Values and absolute errors of the Stalker method (with cubic grid) for points, mainly chosen outside or on the boundary of the interpolation domain, in the context of interpolation of H on $[-50, 50]^2 \times [-1, 1]$ with 200 interpolation points per dimension. 97

5.1 Summary table of $|\lambda_i^\star|$ for $i \in \{1, \ldots, 13\}$, in percentages relative to $\sum |\lambda_i^\star|$. . . 107
5.2 Summary table of the distribution of the values of Equation (5.12) for 10 000 random permutations on $\{1, \ldots 13\}$ and comparison with the same measure for the identity permutation and for the permutation resulting from mode matching algorithm (Algorithm 1). 123

List of Acronyms and Symbols

Acronyms

a.s.	almost surely (probability measure theory)
AUV	Autonomous Underwater Vehicles
BEM	Blade Element Momentum (theory)
BHM	Bayesian History Matching
BLUP	Best Linear Unbiaised Predictor
CFD	Computational Fluid Dynamics (theory)
CLT	Classical Lamination Theory
DoE	Design of Experiments
e.g.,	for example
EFF	Expected Feasibility Function
EGO	Efficient Global Optimization
EI	Expected Improvement
FF	Feasibility Function
GP	Gaussian Process
GPR	Gaussian Process Regression
ICM	Intrinsic Coregionalization Model
IMSE	Integrated Mean Squared Error
LCM	Linear Model of Coregionalization
LHS	Latin Hypercube Sampling
LOO	Leave-One-Out
MLE	Maximum Likelihood Estimator
MMSE	Maximum Mean Squared Error
MOGP	Multi-Output Gaussian Process

mse	mean squared error
MTGP	Multi-Task Gaussian Process
OMA	Operational Modal Analysis
PI	Probability Improvement
RWT	Reference Wind Turbine
SLFM	Semi-parametric Latent Factor Model
SOGP	Single-Output Gaussian Process
SSI	Stochastic Subspace Identification
SUR	Stepwise Uncertainty Reduction
tIMSE	targeted Integrated Mean Squared Error
tMMSE	targeted Maximum Mean Squared Error
tmse	targeted mean squared error
U	Deviation number criterion
VEFF	Vector Expected Feasibility Function

Symbols

$\#A$	Cardinal of set A
$\{n, \dots, m\}$	Set of natural numbers between n and m.
ln	Natural logarithm
$\mathbb{E}\left[X \mid \mathcal{E}\right]$	Conditional expectation of random variable X given event $\mathcal{E}$
$\mathcal{N}(\mu, \sigma)$	Normal distribution of mean μ and standard deviation σ
$\mathcal{S}_p^+(\mathbb{R})$	Real symmetric positive semi-definite matrices of size p
$\det(M)$	Determinant of matrix M
$\mathrm{Var}\left[X \mid \mathcal{E}\right]$	Conditional variance of random variable X given event $\mathcal{E}$
$A \,\Delta\, B$	Symmetric difference between A and B sets
$(.)^+$	Positive part function ($= \max(., 0)$)
$[a, b]$	Real interval between a and b
$\mathbb{E}[X]$	Expectation of random variable X
$\mathbb{N}$	Natural Numbers
$\mathbb{R}$	Real numbers
$\mathrm{Var}[X]$	Variance of random variable X
ϕ	Cumulative distribution function of standard normal distribution
φ	Probability density function of standard normal distribution

Introduction

Contexte

De nombreuses problématiques industrielles actuelles sont liées à l'estimation d'ensembles d'excursion, notamment dans le cadre de problèmes de conception optimale complexes, où l'on recherche des solutions réalisables. La quantité d'intérêt du problème est généralement une sortie d'un simulateur numérique coûteux en temps de calcul, modélisée par une fonction boîte noire notée g, représentant un ou plusieurs phénomènes physiques complexes. L'objectif est de trouver les valeurs de paramètres d'entrée de g telles que la quantité d'intérêt respecte une certaine contrainte, par exemple reste en dessous d'un seuil fixé. Parmi les nombreux exemples d'applications, on peut citer le cas du paramétrage du contrôle d'un système de dépollution d'un véhicule ([El Amri et al., 2020]), récemment étudié dans le cadre d'une thèse IFP Énergies Nouvelles.

D'un point de vue mathématique, un problème d'estimation d'ensemble d'excursion consiste à estimer, à partir d'un nombre limité d'évaluations potentiellement coûteuses de g, l'ensemble défini par

$$\Gamma^\star := \Big\{\mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) \leq T\Big\}, \tag{1}$$

où $\mathbb{X}$ désigne l'espace des entrées appelé espace de design et T un seuil prédéfini. Pour résoudre ce problème, il est possible de définir de façon séquentielle un ensemble de points d'évaluations dans $\mathbb{X}$ et d'évaluations correspondantes par g, que l'on appelle plan d'expériences (noté DoE pour *Design of Experiments*). L'objectif d'une telle approche est d'adapter le DoE au fur et à mesure de l'enrichissement, afin de limiter le nombre d'évaluations coûteuses de g (voir par exemple [Ginsbourger, 2017]). Pour cela, on sélectionne séquentiellement les nouveaux points à ajouter au DoE en optimisant un critère d'acquisition qui exploite l'information fournie par un modèle de substitution du modèle g. Un modèle de substitution est un modèle moins coûteux en temps de calcul, et défini à partir d'un nombre limité d'évaluations du vrai modèle boîte noire g.

Parmi les différents modèles de substitution possibles, nous nous intéressons à la régression par processus gaussiens (notée GPR pour *Gaussian Process Regression*), qui consiste à considérer le modèle g comme la réalisation d'un processus gaussien. L'avantage d'un tel modèle de substitution, qui le rend particulièrement populaire, est le fait qu'il fournisse à la fois une prédiction, mais également une estimation de l'erreur associée. De plus, un processus gaussien est entièrement déterminé par ses deux premiers moments (moyenne et covariance), et ces moments peuvent être mis à jour conditionnellement à des observations de façon analytique, ce qui est particulièrement intéressant pour la procédure d'enrichissement séquentielle.

Les critères d'acquisition adaptés à l'estimation d'un ensemble d'excursion incluent des critères standards comme le nombre de déviation noté "U" et le critère Bichon qui ont pour objectif d'estimer la frontière de l'ensemble $\Gamma^\star$. Les stratégies de réduction progressive de l'incertitude (notées SUR pour *Stepwise Uncertainty Reduction*) sont généralement plus efficaces, puisqu'elles anticipent l'impact de l'ajout de nouveaux points au DoE sur la réduction

de l'incertitude de prédiction. Parmi elles, on peut par exemple citer la stratégie SUR Vorob'ev inspirée de la théorie des ensembles aléatoires.

Motivation : application au simulateur d'éoliennes

L'application qui motive notre travail est proposée par IFP Énergies Nouvelles, et concerne la calibration d'un simulateur numérique pour la conception d'éoliennes. Le simulateur numérique en question étudie les fréquences de vibration et les modes de déformation de la structure mécanique d'une éolienne à l'arrêt et en réponse à des charges de vent. L'objectif de l'étude de la signature vibratoire (fréquences et modes) d'une éolienne est d'identifier les défauts structurels (défauts de masse ou de raideur, usure prématurée, désalignement de composants, etc.). L'éolienne étudiée est la DTU 10-MW Reference Wind Turbine (DTU 10MW RWT).

Le simulateur numérique peut être considéré comme un modèle boîte noire vectoriel, prenant en entrée des paramètres du système correspondant aux coefficients de raideur de certains matériaux, et renvoyant en sortie les fréquences de vibration et les modes de déformation de la structure mécanique. Le but est de trouver l'ensemble des valeurs des paramètres d'entrée Θ du simulateur, telles que les fréquences de vibration $\lambda_i(\Theta)$ et les modes de déformation $\mathrm{Mod}_i(\Theta)$ calculés par le simulateur ne soient pas trop éloignés, selon une certaine mesure de similarité et pour des seuils fixés, des fréquences de référence $\lambda_i^\star$ et des modes $\mathrm{Mod}_i^\star$ calculés à partir de données expérimentales. Dans le cahier des charges pour l'application IFPEN, il est demandé de rechercher simultanément chacun des ensembles d'excursions partiels pour plusieurs composantes de sortie. Nous allons aborder ce problème de pré-calibration selon deux formulations : l'une prenant en compte les deux modes principaux ensemble, et l'autre traitant séparément les modes et les fréquences, avec une mesure de similarité distincte pour chaque approche.

Les données du simulateur sont supposées être isotopiques, ce qui signifie que les différentes composantes de sortie du simulateur sont évaluées simultanément pour un point d'évaluation donné. Cette hypothèse signifie qu'un point d'évaluation commun pour toutes les composantes de sortie du simulateur doit être choisi pour l'enrichissement du plan d'expérience, afin d'utiliser correctement toutes les informations fournies par les simulations coûteuses du simulateur.

Problématiques et défis de la thèse

Les objectifs principaux de cette thèse sont d'étudier les critères d'acquisition adaptés à l'estimation des ensembles d'excursion, pour les fonctions boîtes noires à sortie scalaire ou vectorielle, et d'appliquer ces contributions à l'application présentée ci-dessus concernant la calibration d'un simulateur numérique pour la conception d'éoliennes.

Tout d'abord, dans le cadre des fonctions boîtes noires à sortie scalaire, nous avons remarqué que le critère SUR Vorob'ev n'est pas toujours robuste, notamment en raison de son faible caractère exploratoire. En effet, lors de la recherche d'un ensemble d'excursion composé de plusieurs composantes connexes et pour un nombre raisonnable de simulations, la stratégie d'enrichissement basée sur le critère SUR Vorob'ev manque parfois certaines composantes connexes. Nous proposons donc de résoudre ce problème en développant une version SUR du critère Bichon. L'idée sous-jacente du critère SUR Bichon est de proposer une stratégie SUR facile à mettre en œuvre qui combine la robustesse du critère Bichon (en raison de sa nature exploratoire) avec les performances reconnues des stratégies SUR (en termes d'exploitation). Il est important de rappeler que chaque critère d'acquisition définit un certain équilibre entre

l'exploration de l'espace de design et son exploitation dans un but spécifique, dans notre cas, l'estimation d'un ensemble d'excursion.

La seconde partie, motivée par l'application à la conception d'éoliennes, se concentre sur le cadre des modèles boîte noire avec sortie vectorielle, pour lesquels toutes les composantes de sortie sont évaluées simultanément (données isotopiques). Dans ce cadre, le modèle boîte noire vectoriel est noté $g := (g_1, \ldots, g_p)^\top$, et pour un vecteur de seuils $T := (T_1, \ldots, T_p)^\top$, les ensembles d'excursion partiels sont définis par

$$\Gamma_i^\star := \{\mathbf{x} \in \mathbb{X},\, g_i(\mathbf{x}) \leq T_i\}, \tag{2}$$

et l'ensemble d'excursion global $\boldsymbol{\Gamma^\star}$ est défini comme l'intersection de tous les ensembles d'excursion partiels :

$$\boldsymbol{\Gamma^\star} := \{\mathbf{x} \in \mathbb{X},\, \boldsymbol{g}(\mathbf{x}) \leq \boldsymbol{T}\} = \bigcap_{i=1}^{p} \Gamma_i^\star. \tag{3}$$

L'étude de [Fossum et al., 2021] se concentre sur l'estimation de l'ensemble d'excursion global $\Gamma^\star$ en utilisant des extensions des critères SUR de la variance de la mesure d'excursion et de la variance de Bernoulli intégrée. Ces critères utilisent un modèle de substitution de processus Gaussien à sorties multiples et une généralisation de la probabilité de couverture. L'estimation de chacun des ensembles d'excursion partiels $\Gamma_i^\star$ est différente de l'estimation de l'ensemble d'excursion global, car elle nécessite non plus de découvrir seulement les frontières de l'ensemble d'excursion global, mais toutes les frontières de chacun des ensembles d'excursion partiels. Pour ce faire, trois nouvelles stratégies d'enrichissement basées sur le critère Bichon sont introduites. Les deux premières utilisent des critères scalaires basés sur des modèles de substitution indépendants, tandis que la troisième utilise un processus gaussien à sortie vectorielle. Les différentes stratégies étudiées sont ensuite mises en œuvre dans le cadre de l'application susmentionnée, à savoir la calibration d'un simulateur numérique pour la conception d'éoliennes.

Organisation du manuscrit

Ce manuscrit est structuré en cinq chapitres. Les deux premiers chapitres présentent des outils classiques issus de la littérature scientifique, fournissant ainsi les bases nécessaires à la compréhension de la suite du manuscrit. Les trois chapitres suivants présentent les contributions spécifiques de cette recherche. L'organisation du manuscrit est la suivante :

- Le Chapitre 1 fournit un aperçu de la régression par processus gaussiens (GPR), et de son utilisation dans le cadre de DoEs séquentiels pour la modélisation des fonctions boîtes noires couteuses. La régression par processus gaussiens (GPR) est un exemple de modèle de substitution, construit à partir d'un échantillon entrées/sorties du modèle boîte noire. Ce modèle de substitution offre la possibilité de définir un critère d'enrichissement peu coûteux sur l'ensemble des entrées du modèle, permettant ainsi l'ajout efficace de points pertinents au plan d'expériences.

- Le Chapitre 2 introduit le problème d'estimation d'un ensemble d'excursion et explore divers critères standards d'enrichissement de DoEs dédiés à ce cadre. La classe des critères *Stepwise Uncertainty Reduction* (SUR) est également abordée, ces critères anticipant l'impact de l'ajout de points au plan d'expériences séquentiel en minimisant l'espérance d'une incertitude résiduelle conditionnelle. Plusieurs exemples de stratégies SUR sont présentés, y compris une extension adaptée au cadre des modèles vectoriels.

- Le Chapitre 3 propose un nouveau critère SUR pour l'enrichissement des DoEs basé sur le critère Bichon dans le contexte de l'estimation d'un ensemble d'excursion via GPR. Ce nouveau critère est comparé aux critères usuels comme le critère SUR Vorob'ev ou le critère Bichon. Des simulations numériques sur des fonctions tests classiques mettent en évidence le bon comportement exploratoire et la robustesse du nouveau critère. Les résultats de ce chapitre ont été publiés dans la revue *Statistics and Computing* ([Duhamel et al., 2023]).

- Le Chapitre 4 étudie l'estimation des ensembles d'excursion dans le cadre d'une fonction boîte noire vectorielle. Il introduit trois critères d'enrichissement du plan d'expériences inspirés par le critère Bichon, adaptés à l'estimation simultanée de chacun des ensembles d'excursion partiels pour chaque composante de sortie du modèle. Ces critères sont comparés sur des exemples analytiques avec plusieurs composantes de sortie.

- Le Chapitre 5 applique la méthodologie proposée au Chapitre 4 à la pré-calibration d'un simulateur numérique d'éolienne. Le simulateur numérique reproduit les fréquences de vibration et les modes de déformation de l'éolienne en réponse aux charges de vent. L'objectif de cette application est de trouver l'ensemble des valeurs des paramètres d'entrée du simulateur (coefficients de raideur de certains matériaux) telles que les sorties simulées soient suffisamment proches des fréquences et modes de référence, pour une mesure de dissimilarité et des seuils spécifiés.

Parmi les contributions scientifiques apportées durant cette thèse, on compte un article publié dans la revue *Statistics and Computing*, des participations (présentations orales) a des congrès scientifiques internationaux (SAMO 2022, SIAM UQ 2022, Journées de la Statistiques 2023 de la SFdS et SIAM UQ 2024), ainsi que de nombreuses participations à des évenements nationaux (Journées MASCOT NUM et CIROQUO) (1 présentation orale et 5 poster sessions). De plus, un projet d'article sur les Chapitres 4 et 5 est en cours, avec une soumission prévue avant la fin de l'année 2024.

Introduction

Context

Many industrial problems are linked to the estimation of excursion sets, particularly in the context of complex optimal design problems, where feasible solutions are sought. The quantity of interest in the problem is usually an output of a computationally time-consuming numerical simulator, modeled by a black-box function denoted g, representing one or more complex physical phenomena. The objective is to find input parameter values of g such that the quantity of interest respects a certain constraint, e.g., remains below a fixed threshold. Among the many examples of applications, we can cite the case of parameterizing the control of a vehicle pollution control system ([El Amri et al., 2020]), recently studied as part of an IFP Énergies Nouvelles thesis.

From a mathematical point of view, an excursion set estimation problem consists in estimating, from a limited number of potentially costly evaluations of g, the set defined by

$$\Gamma^\star := \Big\{\mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) \leq T\Big\}, \tag{4}$$

where $\mathbb{X}$ denotes the space of inputs called the design space and T a predefined threshold. To solve this problem, it is possible to sequentially define a set of evaluation points in $\mathbb{X}$ and corresponding evaluations by g, which we call Design of Experiments (DoE). The aim of such an approach is to adapt the DoE as enrichment proceeds, in order to limit the number of costly evaluations of g (see for example [Ginsbourger, 2017]). This is achieved by sequentially selecting new points to be added to the DoE by optimizing an acquisition criterion that exploits the information provided by a surrogate model of the g model. A surrogate model is one that is less costly in terms of computation time, and is defined on the basis of a limited number of evaluations of the true g black box model.

Among the various possible surrogate models, we are interested in Gaussian Process Regression (GPR), which consists in considering the g model as the realization of a Gaussian process. The advantage of such a surrogate model, which makes it particularly popular, is that it provides both a prediction and an estimate of the associated error. Furthermore, a Gaussian process is entirely determined by its first two moments (mean and covariance), and these moments can be updated analytically conditional on observations, which is particularly interesting for the sequential enrichment procedure.

Acquisition criteria suitable for estimating an excursion set include standard criteria such as Deviation number denoted "U" and Bichon criterion, which aim to estimate the boundary of the $\Gamma^\star$ set. Stepwise Uncertainty Reduction strategies (SUR) are generally more effective, since they anticipate the impact of adding new points to the DoE on reducing prediction uncertainty. SUR Vorob'ev strategy is one example, inspired by random set theory.

Motivation: application to wind turbine simulator

The application that motivates our work is proposed by IFP Énergies Nouvelles, and concerns the calibration of a numerical simulator for wind turbine design. The numerical simulator in question studies the vibration frequencies and deformation modes of the mechanical structure of a wind turbine at standstill and in response to wind loads. The aim of studying the vibration signature (frequencies and modes) of a wind turbine is to identify structural faults (mass or stiffness defects, premature wear, component misalignment, etc.). The wind turbine studied is the DTU 10-MW Reference Wind Turbine (DTU 10MW RWT).

The numerical simulator can be considered a vector-valued black box model, taking as input system parameters corresponding to the stiffness coefficients of certain materials, and returning as output the vibration frequencies and deformation modes of the mechanical structure. The aim is to find the set of values for the input parameters Θ for the simulator, such that the simulated vibration frequencies $\lambda_i(\Theta)$ and deformation modes $\mathrm{Mod}_i(\Theta)$ closely match, within specified thresholds, the reference frequencies $\lambda_i^\star$ and modes $\mathrm{Mod}_i^\star$ obtained from experimental data. In the specifications for the IFPEN application, we are asked to simultaneously search each of the partial excursion sets for several output components. We will approach this pre-calibration problem using two formulations: one taking into account the two main modes together, and the other treating modes and frequencies separately, with a distinct dissimilarity measure for each approach.

Simulator data are assumed to be isotopic, which means that the different simulator output components are evaluated simultaneously for a given evaluation point. This assumption means that a common evaluation point for all simulator output components must be chosen for DoE enrichment, in order to make proper use of all the information provided by expensive simulations of the simulator.

Issues and challenges of the thesis

The main objectives of this thesis are to study acquisition criteria suitable for estimating excursion sets, for black box functions with scalar or vector output, and to apply these contributions to the application presented above concerning the calibration of a numerical simulator for wind turbine design.

First of all, in the context of black box functions with scalar output, we noted that SUR Vorob'ev criterion is not always robust, notably due to its weak exploratory character. Indeed, when searching for an excursion set composed of several connected components and for a reasonable number of simulations, the enrichment strategy based on SUR Vorob'ev criterion sometimes misses certain connected components. We therefore propose to solve this problem by developing a SUR version of Bichon criterion. The idea behind SUR Bichon criterion is to propose an easy-to-implement SUR strategy that combines the robustness of Bichon criterion (due to its exploratory nature) with the recognized performance of SUR strategies (in terms of exploitation). It's important to remember that each acquisition criterion defines a certain balance between exploring the design space and exploiting it for a specific purpose, in our case the estimation of an excursion set.

The second part, motivated by application to wind turbine design, focuses on the framework of black box models with vector output, for which all output components are evaluated simultaneously (isotopic data). In this framework, the vector black box model is denoted $g := (g_1, \ldots, g_p)^\top$, and for a vector of thresholds $T := (T_1, \ldots, T_p)^\top$, the partial excursion sets are defined by

$$\Gamma_i^\star := \{\mathbf{x} \in \mathbb{X},\, g_i(\mathbf{x}) \leq T_i\}, \tag{5}$$

and the global excursion set $\boldsymbol{\Gamma}^\star$ is defined as the intersection of all partial excursion sets :

$$\boldsymbol{\Gamma}^\star := \{\mathbf{x} \in \mathbb{X},\, \boldsymbol{g}(\mathbf{x}) \leq \boldsymbol{T}\} = \bigcap_{i=1}^{p} \Gamma_i^\star. \tag{6}$$

The study of [Fossum et al., 2021] focuses on estimating the global excursion set $\Gamma^\star$ using extensions of the SUR Excursion measure variance and Integrated Bernoulli variance criteria. These criteria use a multi-output Gaussian process surrogate model and a generalization of coverage probability. Estimating each of the partial excursion sets $\Gamma_i^\star$ is different from estimating the global excursion set, as it requires discovering both the global and partial boundaries. To achieve this, three new enrichment strategies based on Bichon criterion are introduced. The first two use scalar criteria based on independent surrogate models, while the third uses a multi-output Gaussian process. The various strategies studied are then implemented in the context of the aforementioned application, which concerns the calibration of a numerical simulator for wind turbine design.

Manuscript organization

This manuscript is structured in five chapters. The first two chapters present classic tools from the scientific literature, providing the foundations for understanding the rest of the manuscript. The following three chapters present the specific contributions of this research. The manuscript is organized as follows:

- Chapter 1 provides an overview of Gaussian process regression (GPR), and its use in sequential DoEs for modeling expensive black box functions. Gaussian Process Regression (GPR) is an example of a surrogate model, built from a sample input/output black box model. This surrogate model offers the possibility of defining an inexpensive enrichment criterion on all model inputs, enabling the efficient addition of relevant points to the DoE.

- Chapter 2 introduces the problem of estimating an excursion set and explores various standard DoEs enrichment criteria dedicated to this framework. The class of *Stepwise Uncertainty Reduction* (SUR) criteria is also discussed, which anticipate the impact of adding points to the sequential experimental design by minimizing the expectation of a conditional residual uncertainty. Several examples of SUR strategies are presented, including an extension adapted to the framework of vector models.

- Chapter 3 proposes a new SUR criterion for DoE enrichment based on Bichon criterion in the context of excursion set estimation via GPR. This new criterion is compared to usual criteria such as SUR Vorob'ev criterion or Bichon criterion. Numerical simulations on classical test functions highlight the good exploratory behavior and robustness of the new criterion. The results of this chapter have been published in the journal *Statistics and Computing* ([Duhamel et al., 2023]).

- Chapter 4 studies the estimation of excursion sets within the framework of a vector black box function. It introduces three DoE enrichment criteria inspired by Bichon criterion, adapted for the simultaneous estimation of each partial excursion set for each model output component. These criteria are compared on analytical examples with several output components.

- Chapter 5 applies the methodology proposed in Chapter 4 to the pre-calibration of a numerical wind turbine simulator. The numerical simulator reproduces the vibration

frequencies and deformation modes of the wind turbine in response to wind loads. The aim of this application is to find the set of values for the input parameters of the simulator (stiffness coefficients of certain materials) such that the simulated outputs are sufficiently close to the reference frequencies and modes, for a measure of dissimilarity and specified thresholds.

Among the scientific contributions made during this thesis are an article published in the journal *Statistics and Computing*, participations (oral presentations) at international scientific congresses (SAMO 2022, SIAM UQ 2022, Journées de la Statistique 2023 of the SFdS and SIAM UQ 2024), as well as numerous participations in national events (Journées MASCOT NUM and CIROQUO) (1 oral presentation and 5 poster sessions). In addition, a draft article on Chapters 4 and 5 is underway, with submission expected before the end of the year 2024.

Chapter 1

Background to Gaussian process regression

Outlines

Surrogate models are approximations of the output of the simulator built from a sample of simulations, which can replace the original expensive simulator with a less time-consuming version. Among surrogate models, Gaussian Process Regression (GPR) is very popular: the simulator (black box function) is considered as a realization of a Gaussian process (GP). The use of such a GPR model, coupled with a strategy of sequential addition of points evaluated by the black box function, called sequential design of experiments (DoE), is commonly used to obtain information on the black box function in a time-efficient way. This chapter seeks to provide an overview of GPR and its use in the context of sequential DoEs for modeling black box functions.

Contents

1.1 Introduction **9**

1.2 Principle of GPR **11**

1.3 Covariance structure **12**

1.3.1 Covariance structure type 12

1.3.2 Parameter estimation 14

1.4 Examples **15**

1.5 Initial design of experiments **15**

1.6 Sequential DoE and enrichment criteria **17**

1.6.1 Global enrichment criteria 19

1.6.2 Goal-oriented enrichment criteria for optimization 19

1.7 Extension to vector-valued models **20**

1.7.1 Principle of multi-output Gaussian processes 20

1.7.2 Covariance structure 22

1.7.3 Autokrigeability 24

1.1 Introduction

In the 1950s, Daniel Krige developed an empirical approach ([Krige, 1951]) to solve problems of bias when estimating the grade of a block of ore from a limited number of samples around the block, notably in South African gold mines. Subsequently, Georges Matheron brought

a theoretical formalization ([Matheron, 1963]) to this method, introducing a key tool called the variogram to analyze the degree of spatial dependency of observations. This led to the development of the variogram-based estimation method known as Kriging.

Geostatistics, which emerged from these advances, applies Kriging theory to data located in a geographical space. This approach integrates probabilistic tools to study spatially correlated phenomena. From the 1970s to the present day, geostatistics has spread to other fields such as meteorology, hydrology and oceanography. The method has also been applied to time-consuming industrial calculation codes, thus generalizing the geostatistical approach to functions with input parameter vectors of dimension $d > 3$ and with data not necessarily geographical.

Kriging, as introduced by Georges Matheron, is an interpolation technique formulated as a linear combination of observations. This method is based on determining the optimal weights to assign to neighboring observations in order to predict the value at an unobserved point. From this point of view, Kriging can be interpreted as the Best Linear Unbiased Predictor (BLUP). This means it provides the best possible unbiased linear prediction for unobserved values, taking into account the spatial dependency structure of the data.

A Bayesian approach examining the Kriging problem in function space is also possible. This perspective uses Gaussian Processes (GPs) to describe the distribution of possible functions that could represent the data. The prior Gaussian distribution is conditioned to the simulation results at design points. It describes the distribution of these functions taking into account the observed data (see Figure 1.1). Thanks to the Gaussian property, the distribution is totally characterized by its mean and covariance. This is known as Gaussian process regression (GPR) or Bayesian approach of Kriging.

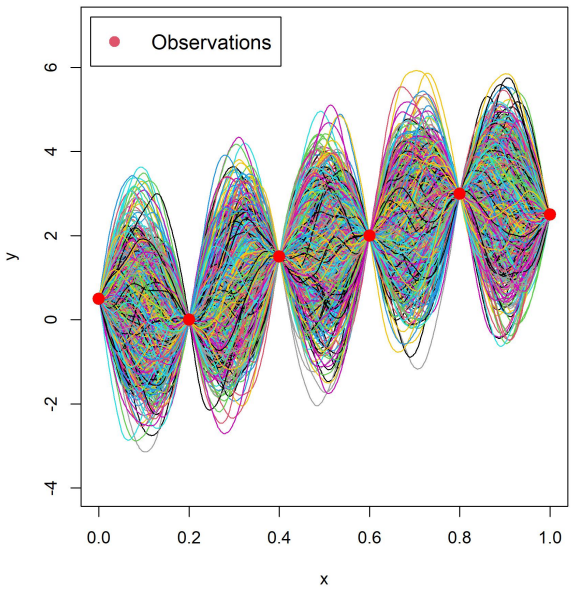


Figure 1.1: Illustration of GP realizations conditioned by observations.

In summary, although these two points of view, BLUP and GPR, use different approaches, they lead to the same predictor. These two perspectives offer complementary angles for understanding how Kriging works and how can spatial dependence of data be taken into

account to perform accurate interpolations. In this thesis we focus mainly on GPR, and refer to [Matheron, 1963], [Cressie, 1990] and [Baillargeon, 2005] for more details on the BLUP point of view.

Many other interpolation methods have also been developed and are frequently used. Some are deterministic, such as barycentric methods [Arnaud and Emery, 2000] and spline methods [Wahba, 1990], and some are stochastic, such as multiple linear regression [Jobson, 1991]. However, GPR distinguishes itself from these other interpolation methods by taking into account the spatial dependency structure of the data. GPR also offers the advantage of providing, due to its stochastic nature, an estimate of the error associated with the model, which, thanks to the Bayesian viewpoint of Kriging, provides a credible interval for the prediction.

1.2 Principle of GPR

In this section, we assume that the reader is familiar with the notions of Gaussian processes (GPs) and stationary processes, and refer to the books by [Breiman, 1992] and [Lindgren, 2012] for further details if necessary.

Let $\mathbb{X}$ be a compact set of $\mathbb{R}^d$ ($d \in \mathbb{N}^\star$) and $g : \mathbb{X} \to \mathbb{R}$ a black box function, whose analytical expression is unknown but which can be evaluated at any point of $\mathbb{X}$ at a heavy computational cost. In the context of GPR, g is considered as a realization of a GP ξ defined on a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., $g(\mathbf{x}) = \xi(\mathbf{x}, \omega)$ for a given ω in Ω. More precisely, the process is written as the sum of a deterministic part and a stochastic part:

$$\xi(\mathbf{x}) := m(\mathbf{x}) + \mathbf{Z}(\mathbf{x}), \quad \forall\, \mathbf{x} \in \mathbb{X}$$

with m the trend of ξ (deterministic part) and $\mathbf{Z}$ a stationary GP (stochastic part), of zero mean, known covariance kernel $k : \mathbb{X}^2 \to \mathbb{R}$ and in particular variance function $\sigma^2(\mathbf{x}) := k(\mathbf{x}, \mathbf{x})$ for any $\mathbf{x}$ in $\mathbb{X}$. We simply recall that a stationary GP is a GP which requires any joint distribution to be translation-invariant. To limit the complexity of estimating the trend function m, the choice of this latter is often parametrized as a linear combination of known basis functions $(f_i)_{i=1}^l$ with coefficients $\boldsymbol{\beta} := \{\beta_i\}_{i=1}^l$ to be estimated. The choice of the covariance kernel associated with the GP $\mathbf{Z}$ (see Section 1.3) is crucial since it determines the predictor regularity and the dependency structure of the data.

Let us denote $g(\chi_n) := (g(\mathbf{x}^{(1)}), ..., g(\mathbf{x}^{(n)}))^\top$ the evaluations of g on an initial design of experiments (DoE) $\chi_n := (\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)})$ belonging to $\mathbb{X}^n$. The random vector $\xi(\chi_n)$ then corresponds to the finite-dimensional distribution of the process $(\xi(\mathbf{x}), \mathbf{x} \in \mathbb{X})$ on χ_n and we define $\mathcal{E}_n$ as the event $\xi(\chi_n) = g(\chi_n)$. $K := (k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}))_{1 \leq i,j \leq n}$ is the covariance matrix of $\xi(\chi_n)$ and $\mathbf{k}(\mathbf{x})$ the cross-covariance matrix between $\xi(\chi_n)$ and $\xi(\mathbf{x})$ defined by $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x}, \mathbf{x}^{(1)}), ..., k(\mathbf{x}, \mathbf{x}^{(n)}))^\top$ for any $\mathbf{x}$ in $\mathbb{X}$. We denote $\mathbf{f}(\mathbf{x}) := (f_1(\mathbf{x}), ..., f_l(\mathbf{x}))^\top$ the evaluation vector of $\mathbf{f}$ on $\mathbf{x}$ defining the trend and $\mathbb{F} \in \mathbb{R}^{n \times l}$ the matrix with $\mathbf{f}(\mathbf{x}^{(i)})^\top$ as i^{th} row. When $\boldsymbol{\beta}$ is known, the process ξ conditioned on the event $\mathcal{E}_n$ is still Gaussian ([O'Hagan, 1978]) with mean, variance and covariance respectively denoted m_n, σ_n^2, and k_n given by

$$m_n(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + \mathbf{k}(\mathbf{x})^\top K^{-1}(g(\chi_n) - \mathbb{F}\boldsymbol{\beta}), \tag{1.1}$$

$$\sigma_n^2(\mathbf{x}) = \sigma^2(\mathbf{x}) - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}), \tag{1.2}$$

$$k_n(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}'). \tag{1.3}$$

We notice that the best linear unbiased predictor (BLUP) (with respect to mean quadratic error) is given by (1.1) with variance (1.2) and covariance (1.3). Also, the error $\sigma_n^2(\mathbf{x})$ of the model at a point $\mathbf{x}$ does not depend on the evaluations of g on the DoE.

When $\boldsymbol{\beta}$ is unknown and estimated by the maximum likelihood estimator (MLE)

$$\widehat{\boldsymbol{\beta}} := \left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\mathbb{F}^\top K^{-1} g(\chi_n),$$

formulas (1.1), (1.2) and (1.3) become

$$m_n(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \widehat{\boldsymbol{\beta}} + \mathbf{k}(\mathbf{x})^\top K^{-1}\left(g(\chi_n) - \mathbb{F}\widehat{\boldsymbol{\beta}}\right), \tag{1.4}$$

$$\sigma_n^2(\mathbf{x}) = \sigma^2(\mathbf{x}) - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}) + \left(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\right)\left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\left(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\right)^\top, \tag{1.5}$$

$$k_n(\mathbf{x},\mathbf{x}') = k(\mathbf{x},\mathbf{x}') - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}') + \left(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\right)\left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\left(\mathbf{f}(\mathbf{x}')^\top - \mathbf{k}(\mathbf{x}')^\top K^{-1}\mathbb{F}\right)^\top. \tag{1.6}$$

Their interpretation as conditional expectation, covariance and variance is still possible in a Bayesian framework with a non informative prior distribution on the parameter $\boldsymbol{\beta}$ ([Helbert et al., 2009]).

The main advantage of this method is that predictive mean and variance functions are analytically tractable, and this is due to the Gaussian nature of the processes used.

1.3 Covariance structure

The choice of covariance structure, given by the kernel k, plays a key role in Gaussian process regression. Recalling that for all $(\mathbf{x},\mathbf{x}') \in \mathbb{X}^2$,

$$k(\mathbf{x},\mathbf{x}') := \mathrm{Cov}(\xi(\mathbf{x}),\xi(\mathbf{x}')),$$

k defines the reciprocal influence of design space points on each other, with respect to Gaussian process modeling. In practice, we often choose a class of parameter-dependent covariance structures. Its parameters are then estimated, based on the available model observations. The two main methods for estimating covariance parameters are maximum likelihood estimation and cross-validation. In practice, the type of covariance structure used is chosen on the basis of expert knowledge of the expected regularity for the black box function under consideration.

1.3.1 Covariance structure type

Most of the time, the study of multidimensional kernels on $\mathbb{X} \subset \mathbb{R}^d$, really boils down to the study of one-dimensional kernels by tensor product. For $\mathbf{x} := (x_1,\ldots,x_d)^\top$ and $\mathbf{x} := (x_1',\ldots,x_d')^\top$ belonging to $\mathbb{X}$,

$$k(\mathbf{x},\mathbf{x}') := \prod_{i=1}^{d} k_i(\mathbf{x}_i,\mathbf{x}_i'),$$

with k_i one-dimensional kernels.

It is important to note that any positive definite symmetrical function can be used to define a covariance kernel, but we will present some of the main types of covariance kernels commonly used in the literature. Recall that the stationary nature of the GPs used implies that the covariance kernel $k(\mathbf{x},\mathbf{x}')$, evaluated at two points, depends only on $\mathbf{x}-\mathbf{x}'$. In this section, we consider isotropic kernels, a special case of stationary kernels for which the covariance kernel $k(\mathbf{x},\mathbf{x}')$ is a function of the distance $h_{\mathbf{x},\mathbf{x}'} := \|\mathbf{x}-\mathbf{x}'\|$ between the two points $\mathbf{x}$ and $\mathbf{x}'$, where $\|.\|$ denotes the Euclidean norm on $\mathbb{X}$.

The most basic type of kernel is the Gaussian (or squared exponential) kernel (see e.g. [Rasmussen et al., 2006]) given by

$$k(\mathbf{x}, \mathbf{x}') := \sigma^2 \exp\left(-\frac{h_{\mathbf{x},\mathbf{x}'}^2}{2\theta^2}\right) \tag{1.7}$$

This covariance kernel corresponds to infinitely differentiable process trajectories. [Stein, 1999] argues that imposing such strict regularity constraints is often unrealistic for modeling physical processes and suggests the use of the Matérn kernel class, which is also particularly widely used in the literature.

The Matérn ν type covariance kernel, of parameters $(\sigma,\ \theta) \in (\mathbb{R}_+^*)^2$ is defined, for all $(\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2$, by:

$$k_\nu(\mathbf{x}, \mathbf{x}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{h_{\mathbf{x},\mathbf{x}'}}{\theta}\right) K_\nu \left(\sqrt{2\nu}\frac{h_{\mathbf{x},\mathbf{x}'}}{\theta}\right), \tag{1.8}$$

where $\Gamma(.)$ denotes the Euler Gamma function, $K_\nu(.)$ the modified Bessel function of the second kind ([Abramowitz and Stegun, 1965]).

In the case where ν is a positive half-integer (i.e., of the form $p + 1/2$ with $p \in \mathbb{N}$), there exists a simplified expression from [Abramowitz and Stegun, 1965] of Eq. (1.8) as the product of an exponential and a polynomial of order p. For p a natural number, the simplified expression is given by

$$k_{\nu=p+\frac{1}{2}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\sqrt{2\nu}h_{\mathbf{x},\mathbf{x}'}}{\theta}\right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^{p} \frac{(p+i)!}{i!(p-i)!} \left(\frac{\sqrt{8\nu}h_{\mathbf{x},\mathbf{x}'}}{\theta}\right)^{p-i}. \tag{1.9}$$

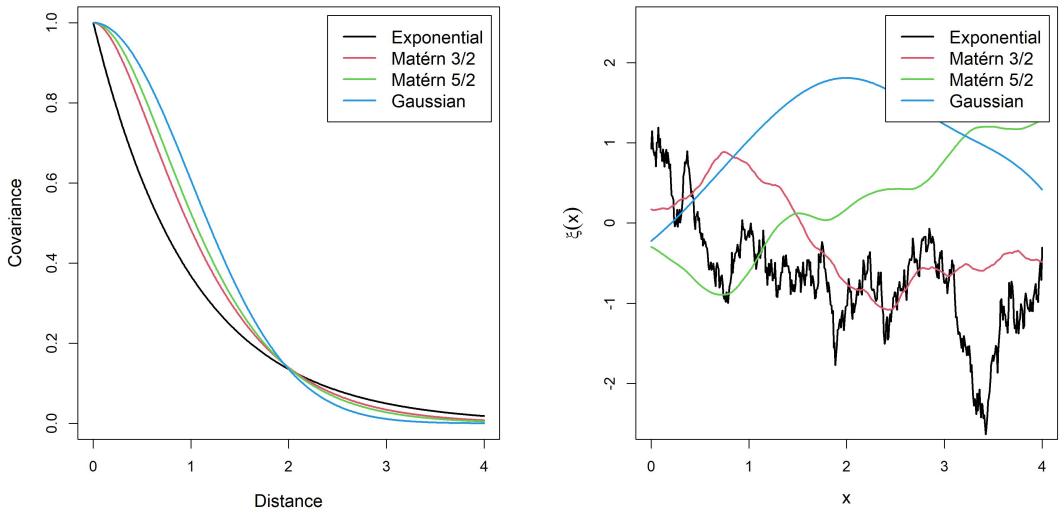
In the case of $\nu = 1/2$, the kernel is said to be exponential, and when ν is made to tend towards $+\infty$, the kernel tends towards the Gaussian kernel. Table 1.1 summarizes the expressions of the Matérn ν type covariance kernels and the regularity of the associated trajectories (see [Paciorek, 2003]), for different classical values of ν.

Name	ν values	Expression of $k_\nu(\mathbf{x}, \mathbf{x}')$	Trajectory regularity
Exponential	$\frac{1}{2}$	$\sigma^2 \exp\left(-\frac{h_{\mathbf{x},\mathbf{x}'}}{\theta}\right)$	$\mathcal{C}^0$
Matérn 3/2	$\frac{3}{2}$	$\sigma^2 \left(1 + \frac{\sqrt{3}h_{\mathbf{x},\mathbf{x}'}}{\theta}\right) \exp\left(-\frac{\sqrt{3}h_{\mathbf{x},\mathbf{x}'}}{\theta}\right)$	$\mathcal{C}^1$
Matérn 5/2	$\frac{5}{2}$	$\sigma^2 \left(1 + \frac{\sqrt{5}h_{\mathbf{x},\mathbf{x}'}}{\theta} + \frac{5h_{\mathbf{x},\mathbf{x}'}^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}h_{\mathbf{x},\mathbf{x}'}}{\theta}\right)$	$\mathcal{C}^2$
Gaussian	$+\infty$	$\sigma^2 \exp\left(-\frac{h_{\mathbf{x},\mathbf{x}'}^2}{2\theta^2}\right)$	$\mathcal{C}^\infty$

Table 1.1: Summary of some Matérn-type covariance kernels.

The different types of covariance kernel presented above are illustrated in Figure 1.2. The covariance between two points is plotted (Figure 1.2a) as a function of the distance between them, with $\theta = \sigma = 1$. Figure 1.2b shows examples of trajectories associated with these different types of covariance structure.

Other classes of covariance kernels include $\gamma-$exponential functions, rational quadratic functions and piecewise polynomial functions (see [Rasmussen et al., 2006] for more details).



(a) Covariance kernel

(b) Example of trajectories

Figure 1.2: On the left, comparison of different types of covariance kernel for fixed parameters $\theta = \sigma = 1$, as a function of the distance between the two points under consideration. On the right, comparison of examples of associated trajectories on the interval $[0, 1]$, with a constant zero trend.

1.3.2 Parameter estimation

The choice of covariance kernel parameters, also known as hyperparameters, is also important to ensure that the GPR model used is suitable for the studied black box function. In practice, these hyperparameters are estimated on the basis of evaluations already carried out of the black box function on the latent DoE. Two general estimation methods can be distinguished.

The first method by maximum likelihood consists in maximizing a likelihood function, aiming to find hyperparameters where, within the assumed statistical model, the data observed on the latent DoE are the most probable (see for example [Ginsbourger et al., 2009], [Gorbach et al., 2017] and [Gauchy, 2022]). Hyperparameter values that maximize the likelihood function are called the maximum likelihood estimates (MLE). The maximum likelihood method is a general approach that is not exclusively limited to hyperparameter estimation in the context of Gaussian process regression (see [Stigler, 2007], [Rossi, 2018] for more general explanations of this method).

The second method by cross-validation (see for example [Gorbach et al., 2017]) involves segmenting the entire DoE into several distinct sets, then using some to train the surrogate model and others to evaluate its performance, by minimizing an estimated generalization error of the model. In the case of K-fold cross-validation, for a regression dataset consisting of inputs $\boldsymbol{X}$ and corresponding outputs $\boldsymbol{y}$, the generalization error is

$$-\frac{1}{K}\sum_{k=1}^{K}\log p_{\boldsymbol{\theta}}\left(\boldsymbol{y}_k \mid \boldsymbol{X}, \boldsymbol{y}_{-k}\right) \tag{1.10}$$

with $\log p_{\boldsymbol{\theta}}\left(\boldsymbol{y}_k \mid \boldsymbol{X}, \boldsymbol{y}_{-k}\right)$ the log-likelihood of the outputs $\boldsymbol{y}_k$ of the k^{th} block knowing the outputs $\boldsymbol{y}_{-k}$ of the other blocks and all latent DoE evaluation points $\boldsymbol{X}$. When the blocks have

size 1, we deal with Leave-One-Out (LOO) cross validation. As with the maximum likelihood method, cross-validation methods are more general and not only applicable to the case of hyperparameter optimization for Gaussian process regression (see for example [Allen, 1974] and [Stone, 1974]).

1.4 Examples

The principle of GPR and the notion of covariance structure are illustrated below on examples performed in the **R** language with the help of the *Dicekriging* package.

To illustrate the notion of GPR, we simulate a trajectory of a centered Gaussian process of covariance kernel of type Matérn 5/2 with a correlation length θ equal to 0.2 and a scaling parameter σ equal to 1 (Figure 1.3a). This trajectory is then used as a black box function, with 5 evaluation points chosen uniformly over $[0, 1]$, to apply and compare the different GPR methods (known parameters, MLE with constant trend and MLE with linear trend) on a simple example.

The results for the different types of GPR (Figures 1.3b, 1.3c and 1.3d) show that the prediction error associated with the model taking into account knowledge of the constant trend is better than that of the other two models where the (constant or linear) trend must be estimated. In practice, the black box function to be estimated is not a simulated trajectory of a GP and the trend is therefore unknown, which justifies the interest of models with trend estimation. In the case of the model MLE with constant trend, the trend is relatively well estimated compared with the case of the model MLE with linear trend (Table 1.2). In the latter case, the poor estimation of θ, linked to the small number of observations compared to the complexity of hyperparameter optimization, results in a very poor prediction error (Figure 1.3c). In practice, the trend of the black box function to be estimated is not necessarily constant, and the model MLE with linear trend may sometimes still be more suitable than the model MLE with constant trend.

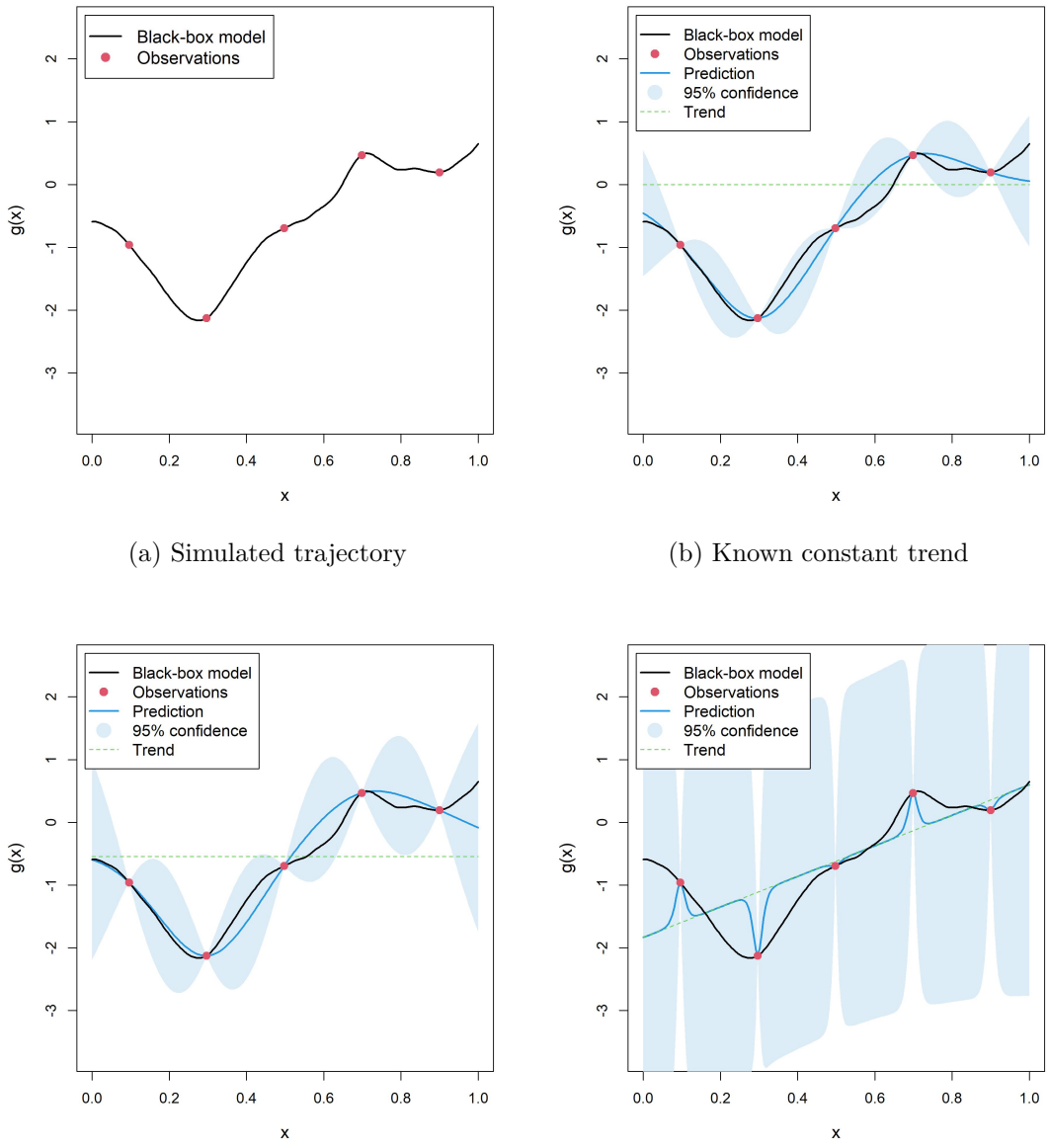
GPR type	m or $\widehat{m}$	θ or $\widehat{\theta}$	σ or $\widehat{\sigma}$
Known parameters	0	0.2	1
MLE with constant trend	-0.55	0.19	0.85
MLE with linear trend	$-1.84 + 2.44x$	0.0134	0.37

Table 1.2: Summary of hyperparameters for the different GPR models used.

This example illustrates the concept of GPR. In the presented case, the type of covariance kernel used corresponds well, by definition, to that of the trajectory. However, in practice, the choice of kernel type is a difficult step that can lead to poor predictions (see Section 1.3).

1.5 Initial design of experiments

In the context of an expensive industrial model (e.g. a computational simulator), approximating a black box function g by the use of GPR requires an initial Design of experiments (DoE) which is then sequentially enriched (see e.g. [Ginsbourger, 2017]). The aim of such an approach is to adapt the DoE as enrichment proceeds, in order to limit the number of costly g evaluations. The first phase of the sequential enrichment strategy is to define a small initial DoE (generally $5d$ to $10d$ points). A first GP is ajusted to this initial DoE (see formulas of



(a) Simulated trajectory

(b) Known constant trend

(c) Unknown constant trend

(d) Unknown linear trend

Figure 1.3: Comparison of different types of GPR, based on 5 observation points uniformly distributed over $[0, 1]$, for a black box function simulated from a centered Gaussian process of kernel covariance Matérn 5/2 ($\theta = 0.2$ and $\sigma = 1$).

Section 1.2). New points are then sequentially selected one by one by optimizing an acquisition criterion which exploits information given by the GP (see next section). The choice of initial DoE has an impact on further enrichment. Most initial DoE involve geometrical considerations, in the sense that the aim is to spread the DoE points as evenly as possible in the design space. In some cases, sequential constructions are used, but without taking into account the surrogate model provided by GPR. Some of the main types of initial DoE are presented below.

Minimax and Maximin distance designs

The Minimax distance design proposed by [Johnson et al., 1990] minimizes the maximum distance between all points in $\mathbb{X}$ and their closest point among the DoE. It minimizes the quantity :

$$\Phi_{\mathrm{mM}}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) := \max_{\mathbf{x} \in \mathbb{X}} \min_{i} \|\mathbf{x} - \mathbf{x}^{(i)}\|. \tag{1.11}$$

In other words, it comes down to finding a configuration $(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$ that minimizes the radius of n balls of the same radius that completely cover the space $\mathbb{X}$. This method is relatively time-consuming, since it requires numerous evaluations of the norm present in (1.11).

Another method for distributing points, which is less computationally intensive, involves choosing a design that maximizes the minimum distance between any two DoE points. This approach, known as the Maximin distance design ([Johnson et al., 1990]), aims to keep the DoE points as far apart as possible by maximizing the quantity

$$\Phi_{\mathrm{Mm}}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) := \min_{i \neq j} \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|, \tag{1.12}$$

where $\|.\|$ denotes, for example, the Euclidean norm on $\mathbb{X}$. This is equivalent to finding a configuration $(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$ that maximizes the radius of n non-intersecting balls of the same radius with respective centers $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$.

Latin hypercube sampling designs

In statistical sampling, a Latin square involves a square grid with one sample per row and column. The Latin hypercube sampling (LHS) introduced by [McKay et al., 1979] extends this concept to an arbitrary number of dimensions, ensuring that each sample is the only one in its axis-aligned hyperplane. The range of each component is divided into n equally probable intervals and n sampling points are strategically placed to meet the Latin hypercube criteria. The main advantage of this method is the good distribution of DoE point projections on each dimension. It should also be noted that this sampling method does not require more samples as the number of dimensions increases.

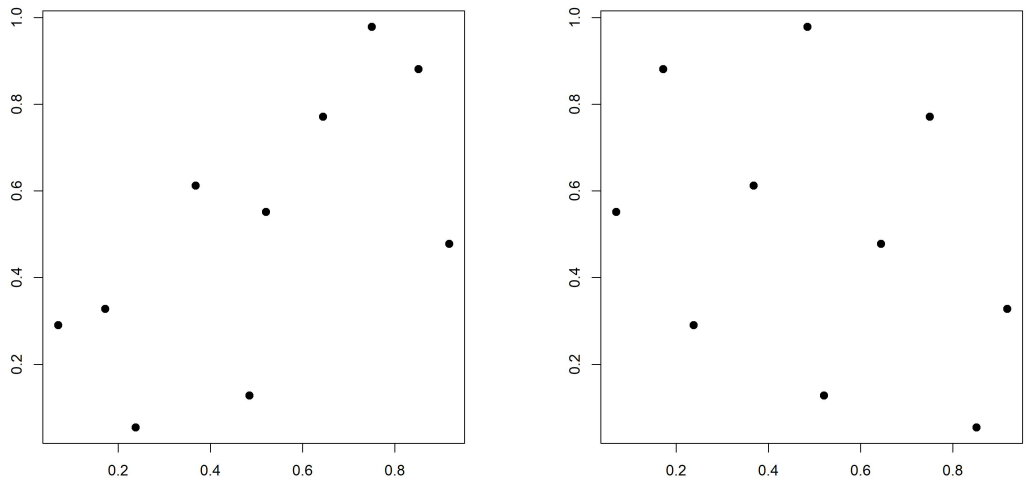
Numerous LHS designs can thus be proposed, including some with poor space-filling properties, such as the one with all points distributed along a straight diagonal line of the design hypercube. Various algorithms can be used to optimize the space-filing character of LHS designs (see [Damblin et al., 2013] and [Dupuy et al., 2015]), including Minimax and Maximin criteria applied to a preliminary set of different LHS designs. Figure 1.4 shows a comparison between a classic LHS plan and an LHS plan optimized by the Maximin criterion. This comparison highlights the space-filling nature of the optimized LHS plan compared with the classic LHS plan, which nevertheless has a good distribution of DoE point projections on each dimension.

Other designs

Many other types of design can be cited, such as low-discrepancy sequences ([Sobol', 1967], [Halton, 1960], etc.). See [Pronzato and Müller, 2012] for a full review of possible designs that can be used as initial DoE. Note that the low-discrepancy criterion can also be used to optimize LHS, in place of the Maximin and Minimax criteria.

1.6 Sequential DoE and enrichment criteria

As a reminder, the initial phase of the sequential enrichment strategy involves defining a small initial DoE, fitting a GP to this DoE. Then, the steps for updating the GP and optimizing the



(a) LHS design

(b) LHS Maximin optimized design

Figure 1.4: Comparison of an LHS plan and an LHS plan optimized by the Maximin criterion, for 10 points on $[0,1]^2$.

criterion follow one another until the stopping criterion is verified, and then a final GP update is performed. Figure 1.5 provides a synthetic scheme of this sequential strategy. The choice of enrichment criterion is crucial and depends on the model's objective (global knowledge, optimization or excursion set estimation). This is the subject of this section, and of Chapter 2 for the case of excursion set estimation. The stopping criterion is also an important element of the sequential enrichment strategy via GPR. It is usually a fixed simulation budget, or defined as a threshold on the remaining uncertainty, calculated according to the objective.

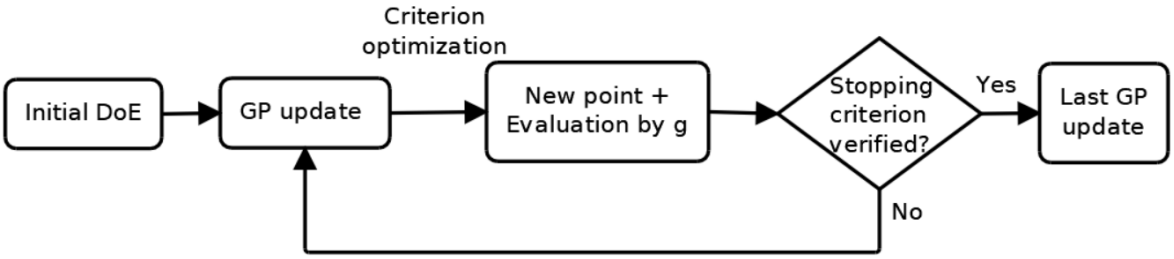


Figure 1.5: Diagram of the sequential DoE construction, coupled with GPR.

We distinguish enrichment criteria according to the associated objective: global knowledge, optimization, estimation of excursion sets. Criteria linked to global knowledge of the function and those associated with an optimization objective for the black box function are presented below. Criteria related to the goal of estimating excursion sets, which are pivotal in this thesis, are deferred to the next chapter with a detailed introduction to the context of excursion set estimation.

1.6.1 Global enrichment criteria

Among the enrichment criteria used to build a sequential design of experiments, there is one type of criterion, called *global criteria*, that provide global knowledge of the black box model.

mse criterion

The "mean squared error" (mse) criterion (see for example [Jin et al., 2002]) is the most natural global enrichment criterion. It consists in maximizing the Kriging variance, conditional on the first n observations. In other words, we are looking for the point in the design space for which the uncertainty associated with the Kriging model is as high as possible. The criterion is given by the following formula:

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\arg\max}\, \sigma_n^2(\mathbf{x}). \tag{1.13}$$

Note that the Kriging variance σ_n^2 does not depend on the values taken by the model on the design of experiments (see Section 1.2). Therefore, this criterion can also be used to construct initial DoE (see [Abtini, 2018]). The same applies to the next two criteria.

MMSE criterion

The "Maximum Mean Squared Error" (MMSE) criterion from [Sacks and Schiller, 1988] is a one-step ahead version of the mse criterion. We look for the point in the design space which, when added to the DoE, would give the best possible improvement in the overall Kriging error (in the sense of the L^∞ norm on the design space). The MMSE is formulated as follows:

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\arg\min} \Big\{ \max_{\mathbf{y}\in\mathbb{X}} \sigma_{n,\mathbf{x}}^2(\mathbf{y}) \Big\}. \tag{1.14}$$

where $\sigma_{n,\mathbf{x}}^2$ denotes the Kriging variance conditional on the first n evaluations and the addition of $\mathbf{x}$ (see update formulas of Section 1.2 with $\mathbf{x}^{(n+1)} = \mathbf{x}$). The disadvantage of this criterion is that it often has many local minima and is difficult to optimize.

IMSE criterion

The "Integrated Mean Squared Error" (IMSE) criterion introduced by [Sacks et al., 1989] is an alternative to the MMSE criterion, replacing the maximum of the Kriging variance used to quantify the error obtained, by the integral of the Kriging variance. By default, we use the renormalized restriction of the Lebesgue measure on $\mathbb{X}$ (denoted by $\mathbb{P}_\mathbb{X}$) for integration, but it is possible to choose other measures on $\mathbb{X}$. The resulting criterion is given by :

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\arg\min} \left\{ \int_\mathbb{X} \sigma_{n,\mathbf{x}}^2(\mathbf{y})\, \mathrm{d}\mathbb{P}_\mathbb{X}(\mathbf{y}) \right\}. \tag{1.15}$$

Unlike the MMSE criterion, the IMSE criterion is more regular thus simpler to optimize.

1.6.2 Goal-oriented enrichment criteria for optimization

Among the criteria used to build sequential experimental designs, we historically find those associated with the objective of optimizing the black box model.

PI criterion

The "Probability Improvement" (PI) criterion, whose original idea came from [Kushner, 1964], measures the probability that the posterior GP modeling the black box model is smaller than a real constant c, for example strictly less than the current minimum $g_{\min}$ in the optimization context. The criterion is written:

$$\mathbf{x}^{(n+1)} \in \arg\max_{\mathbf{x}\in\mathbb{X}} \mathrm{PI}(\mathbf{x}) \quad \text{with} \quad \mathrm{PI}(\mathbf{x}) := \mathbb{P}\big[\xi(x) \leq c \,\big|\, \mathcal{E}_n\big]. \tag{1.16}$$

The function PI can be rewritten more simply as:

$$\mathrm{PI}(\mathbf{x}) = \phi\bigg(\frac{c - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\bigg),$$

with ϕ the cumulative distribution function (cdf) of the standard normal distribution.

The choice of c is crucial: if it is too small, the search will be too global, while if it is too large, the search will be too local. An example of a possible choice for c is $c := g_{\min} - 0.25|g_{\min}|$, with $g_{\min} := (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$ (see [Jones, 2001] for details).

EI criterion

The "Expected Improvement" (EI) criterion introduced in [Jones et al., 1998] is an alternative to the PI criterion, which does not require parameter calibration. It quantifies the potential improvement provided by the surrogate model over the current minimum $g_{\min}$. The EI criterion is formulated as :

$$\mathbf{x}^{(n+1)} \in \arg\max_{\mathbf{x}\in\mathbb{X}} \mathrm{EI}(\mathbf{x}) \quad \text{with} \quad \mathrm{EI}(\mathbf{x}) := \mathbb{E}\big[(g_{\min} - \xi(\mathbf{x}))^+ \,\big|\, \mathcal{E}_n\big], \tag{1.17}$$

where $(.)^+ : y \mapsto \max(0, y)$ denotes the positive value function. One can express the expected improvement in explicit form:

$$\mathrm{EI}(\mathbf{x}) = (g_{\min} - m_n(\mathbf{x}))\,\phi\bigg(\frac{g_{\min} - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\bigg) + \sigma_n(\mathbf{x})\,\varphi\bigg(\frac{g_{\min} - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\bigg),$$

with φ the probability distribution function (pdf) of the standard normal distribution. Note that for a fixed $\mathbf{x}$ in the design space,

$$\frac{\partial \mathrm{EI}}{\partial m_n} = -\phi\left(\frac{g_{\min} - m_n}{\sigma_n}\right) < 0 \quad \text{and} \quad \frac{\partial \mathrm{EI}}{\partial \sigma_n} = \varphi\left(\frac{g_{\min} - m_n}{\sigma_n}\right) > 0,$$

which means that the smaller m_n is, the larger EI is, and the larger σ_n is, the larger EI is.

The use of the EI criterion thus makes it possible to obtain a compromise between local and global optimization, which is in fact equivalent to an exploration-exploitation compromise of the model in an optimization context. The Bayesian optimization algorithm associated with the construction of a sequential experimental design using the EI criterion is known as the "Efficient Global Optimization" (EGO) algorithm (see [Jones et al., 1998] for more details).

1.7 Extension to vector-valued models

1.7.1 Principle of multi-output Gaussian processes

Vector-valued Gaussian process regression, also known as multi-output Gaussian process (MOGP) regression is a generalization of GPR to build a surrogate of a deterministic vector-valued function $\boldsymbol{g} : \mathbb{X} \subset \mathbb{R}^d \mapsto \mathbb{R}^p$. MOGP is based on the fundamental belief that the output

components are correlated in some manner. Therefore, a crucial aspect of MOGP is to take advantage of correlations between output components to mutualize the information acquired on each output and obtain more accurate predictions than if the output components were modeled independently.

In the following, we are interested in a framework where prediction importance is the same for all model components (symmetric MOGP) and where the evaluation points are the same for each component of model output (isotopic data). This often occurs when the p output responses at a point $\mathbf{x}$ can be obtained through a single simulation (see for example [Liu et al., 2018] for more details on MOGP).

Let us assume that $\boldsymbol{g} := (g_1, \ldots, g_p)^\top$ is the realization of a vector-valued Gaussian process $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_p)^\top$. This process is characterized by its mean $M := (M_1, \ldots, M_p) : \mathbb{R}^d \to \mathbb{R}^p$ and covariance $K := (K_{i,j})_{1 \leq i,j \leq p} : \mathbb{R}^d \times \mathbb{R}^d \to \mathcal{S}_p^+(\mathbb{R})$ functions (a priori, both assumed known here) defined for $(i, j) \in \{1, \ldots, p\}^2$ by

$$M_i(\mathbf{x}) := \mathbb{E}(\xi_i(\mathbf{x})) \quad \text{and} \quad K_{i,j}(\mathbf{x}, \mathbf{x}') := \mathrm{Cov}(\xi_i(\mathbf{x}), \xi_j(\mathbf{x}')), \tag{1.18}$$

with $\mathcal{S}^+(\mathbb{R})$ denoting real symmetric positive semi-definite matrices of size p. We note $\Sigma(\mathbf{x}) := K(\mathbf{x}, \mathbf{x})$ the covariance matrix of $\boldsymbol{\xi}(\mathbf{x})$. Let $\chi_n := (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})^\top$ denote the DoE at step n. We define

$$\boldsymbol{\xi}(\chi_n) := \left(\xi_1(\mathbf{x}^{(1)}), \ldots, \xi_1(\mathbf{x}^{(n)}), \ldots, \xi_p(\mathbf{x}^{(1)}), \ldots, \xi_p(\mathbf{x}^{(n)})\right)^\top \in \mathbb{R}^{pn \times 1}, \tag{1.19}$$

$$\boldsymbol{g}(\chi_n) := \left(g_1(\mathbf{x}^{(1)}), \ldots, g_1(\mathbf{x}^{(n)}), \ldots, g_p(\mathbf{x}^{(1)}), \ldots, g_p(\mathbf{x}^{(n)})\right)^\top \in \mathbb{R}^{pn \times 1}, \tag{1.20}$$

and $\boldsymbol{\mathcal{E}}_n$ the event $\boldsymbol{\xi}(\chi_n) = \boldsymbol{g}(\chi_n)$. The process $\boldsymbol{\xi}(\chi_n)$ is characterized by

$$\boldsymbol{M}(\chi_n) := \left(M_1(\mathbf{x}^{(1)}), \ldots, M_1(\mathbf{x}^{(n)}), \ldots, M_p(\mathbf{x}^{(1)}), \ldots, M_p(\mathbf{x}^{(n)})\right)^\top \in \mathbb{R}^{pn \times 1}. \tag{1.21}$$

and

$$\mathbf{K}_{\chi_n, \chi_n} := \begin{bmatrix} K_{11}(\chi_n, \chi_n) & \cdots & K_{1p}(\chi_n, \chi_n) \\ \vdots & \ddots & \vdots \\ K_{p1}(\chi_n, \chi_n) & \cdots & K_{pp}(\chi_n, \chi_n) \end{bmatrix} \in \mathbb{R}^{pn \times pn}, \tag{1.22}$$

with $K_{ii'}(\chi_n, \chi_n) = \mathbb{E}[(\xi_i(\chi_n) - \mathbb{E}[\xi_i(\chi_n)])(\xi_{i'}(\chi_n) - \mathbb{E}[\xi_{i'}(\chi_n)])^\top] \in \mathbb{R}^{n \times n}$ the cross-covariance matrix between $\xi_i(\chi_n)$ and $\xi_{i'}(\chi_n)$. In the same way, we note $\mathbf{K}_{\chi_n}(\mathbf{x}) \in \mathbb{R}^{pn \times p}$ the cross-covariance matrix between $\boldsymbol{\xi}(\chi_n)$ and $\boldsymbol{\xi}(\mathbf{x})$ defined from blocks

$$K_{ii'}(\mathbf{x}, \chi_n) := (K_{ii'}(\mathbf{x}, \mathbf{x}^{(1)}), \ldots, K_{ii'}(\mathbf{x}, \mathbf{x}^{(n)}))^\top \in \mathbb{R}^{1 \times n}. \tag{1.23}$$

Surrogate update formulas, which are extensions of Formulas (1.1), (1.2) and (1.3) to the vector case, are then given (see for example [Liu et al., 2018]) by:

$$M_n(\mathbf{x}) = M(\mathbf{x}) + \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n, \chi_n}^{-1} \big(\boldsymbol{g}(\chi_n) - \boldsymbol{M}(\chi_n)\big), \tag{1.24}$$

$$\Sigma_n(\mathbf{x}) = \Sigma(\mathbf{x}) - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n, \chi_n}^{-1} \mathbf{K}_{\chi_n}(\mathbf{x}), \tag{1.25}$$

$$K_n(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}, \mathbf{x}') - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n, \chi_n}^{-1} \mathbf{K}_{\chi_n}(\mathbf{x}'). \tag{1.26}$$

where M_n, Σ_n and K_n denote the mean, variance and covariance functions, respectively, of the process $\boldsymbol{\xi}_n := \boldsymbol{\xi} \,|\, \boldsymbol{\mathcal{E}}_n$.

In the case where the a priori trend is no longer known, but assumed to be a linear combination of basis functions with coefficients to be estimated, update formulas exist ([Helterbrand and Cressie, 1994]) and are extension of Formulas (1.4), (1.5), and (1.6).

More precisely, for an integer $1 \leq i \leq p$, assume that the i^{th} component of M is a linear combination of l_i functions, denoted as vector $F_i(\mathbf{x}) \in \mathbb{R}^{l_i \times 1}$ for any $\mathbf{x} \in \mathbb{X}$. We note $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^{(\Sigma_{i=1}^p l_i) \times 1}$ with $\beta_i \in \mathbb{R}^{1 \times l_i}$ the coefficients for the i^{th} component and $\mathbf{F}(\mathbf{x})$ the block diagonal matrix defined by $\mathbf{F}(\mathbf{x}) := \text{diag}(F_1(\mathbf{x})^\top, \ldots, F_p(\mathbf{x})^\top) \in \mathbb{R}^{p \times (\Sigma_{i=1}^p l_i)}$. We then have:

$$M(\mathbf{x}) = \mathbf{F}(\mathbf{x})\, \boldsymbol{\beta}. \tag{1.27}$$

Noting

$$\mathbf{F}(\chi_n) := \text{diag}(\mathbf{F}_1(\chi_n), \ldots, \mathbf{F}_p(\chi_n)) \in \mathbb{R}^{pn \times (\Sigma_{i=1}^p l_i)},$$

with $\mathbf{F}_i(\chi_n) := \left(F_i(\mathbf{x}^{(1)}), \ldots, F_i(\mathbf{x}^{(n)})\right)^\top \in \mathbb{R}^{n \times l_i}$, these update formulas are given by:

$$M_n(\mathbf{x}) = \mathbf{F}(\mathbf{x})\, \widehat{\boldsymbol{\beta}} + \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} (\boldsymbol{g}(\chi_n) - \mathbf{F}(\chi_n)\, \widehat{\boldsymbol{\beta}}), \tag{1.28}$$

$$\begin{aligned}\Sigma_n(\mathbf{x}) = \Sigma(\mathbf{x}) - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{K}_{\chi_n}(\mathbf{x}) + (\mathbf{F}(\mathbf{x}) - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n)) \\ (\mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n))^{-1} (\mathbf{F}(\mathbf{x}) - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n))^\top,\end{aligned} \tag{1.29}$$

$$\begin{aligned}K_n(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}, \mathbf{x}') - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{K}_{\chi_n}(\mathbf{x}') + (\mathbf{F}(\mathbf{x}) - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n)) \\ (\mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n))^{-1} (\mathbf{F}_{\chi_n}(\mathbf{x}') - \mathbf{K}_{\chi_n}(\mathbf{x}')^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n))^\top,\end{aligned} \tag{1.30}$$

with $\widehat{\boldsymbol{\beta}} := (\mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n))^{-1} \mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \boldsymbol{G}(\chi_n)$ the maximum likelihood estimator.

The sequential procedure in the vector case is similar to the scalar one (Section 1.5, Figure 1.5): an initial DoE is generated and an initial MOGP is fitted, then they are followed by a succession of MOGP updates and optimizations from a well-chosen enrichment criterion. An example of a vector criterion adapted to the estimation of excursion sets will be presented in the next chapter, Section 2.2.4.

1.7.2 Covariance structure

The choice of the covariance structure is a complex matter, often necessitating the formulation of simplifying assumptions for practical implementation. The main kernels to take correlation among the components of $\boldsymbol{g}$ into account are outlined below.

Separable kernels

One simple way to model this correlation is to use separable kernels, which can be formulated as a product of a kernel function in the input space and a kernel function that encodes the interactions between the output components (see for example [Liu et al., 2018] and [Alvarez et al., 2012]). Typically,

$$\forall\, 1 \leq i, j \leq p,\ K(\mathbf{x}, \mathbf{x}')_{i,j} = k(\mathbf{x}, \mathbf{x}')\, k_p(i, j), \tag{1.31}$$

where k and k_p are respectively scalar kernels on $\mathbb{X}^2$ and $\{1, \ldots, p\}^2$. Equivalently, we can consider the matrix expression:

$$\left(K(\mathbf{x}, \mathbf{x}')_{i,j}\right)_{1 \leq i,j \leq p} := k(\mathbf{x}, \mathbf{x}')\, \mathbf{B}, \tag{1.32}$$

with $\mathbf{B}$ a $p \times p$ symmetric positive definite matrix. We then recognize the intrinsic coregionalization model (ICM) (see [Goovaerts, 1997]).

A natural generalization of this model, known as the linear model of coregionalization (LMC) ([Goulard and Voltz, 1992]) is given by

$$\left(K(\mathbf{x},\mathbf{x}')_{i,j}\right)_{1\leq i,j\leq p} := \sum_{q=1}^{Q} k_q(\mathbf{x},\mathbf{x}')\,\mathbf{B}_q, \tag{1.33}$$

with k_q scalar kernels on $\mathbb{X}^2$ and $\mathbf{B}_q$ $p \times p$ symetric and positive definite matrices. This covariance structure comes from a model where outputs $g_1, \ldots, g_p$ are assumed to be linear combinations of Q independent latent Gaussian processes. ICM is therefore a special case of LMC with only one latent process.

In ICM or LMC, the choice of scalar kernels k_q can be made among the classical scalar kernels (see Section 1.3.1). As explained in [Alvarez et al., 2012] and in [Liu et al., 2018], different possible structures exist for $\mathbf{B}_q$ in order to reduce the number of parameters to be estimated. Some of the main possible structures are presented below:

- "Independant"
$$\mathbf{B}_q = I, \tag{1.34}$$

- "Semi-parametric latent factor model" (SLFM) from [Teh et al., 2005]
$$\mathbf{B}_q = a^q (a^q)^\top \quad \text{with} \quad a^q := \{a_i^q\}_{1\leq i\leq p}, \tag{1.35}$$

- "Spherical parametrization" from [Osborne, 2007]
$$\mathbf{B}_q = \mathrm{diag}(\mathbf{e}) S^{\mathrm{T}} S \mathrm{diag}(\mathbf{e}) \tag{1.36}$$
with $\mathbf{e} \in \mathbb{R}^p$ scale factors of each output and S upper triangular matrix whose i^{th} column is associated to a spherical representation of points in $\mathbb{R}^p$ (see e.g., [Pelamatti et al., 2024]),

- "Free-form" from [Bonilla et al., 2007] which requires a Cholesky factorization of $\mathbf{B}_q$,
$$\mathbf{B}_q = LL^{\mathrm{T}}, \tag{1.37}$$

- "Low rank free form" from [Bonilla et al., 2007]
$$\mathbf{B}_q = \tilde{L}\tilde{L}^{\mathrm{T}} \tag{1.38}$$
with $\tilde{L} \in \mathbb{R}^{p\times\tilde{p}}$ and $1 \leq \tilde{p} \leq p$. This type of parameterization with $Q = 1$ (ICM) is called Multi-task Gaussian Process (MTGP).

Once the $\mathbf{B}_q$ structure has been chosen, hyperparameters can be estimated by maximum likelihood (see e.g. [Bonilla et al., 2007]). Table 1.3 summarizes the number of hyperparameters to be estimated for each type of $\mathbf{B}_q$.

Other kernels

Convolution processes extend the possibilities for modeling correlations between output components, compared with separable models such as LMC. The idea of convolution processes ([Ver Hoef and Barry, 1998]) is to model dependencies between output components by convolving a basic process with an output-dependent smoothing kernel. The convoluted process remains Gaussian, which considerably facilitates its analytical treatment. What is more,

Name	$\mathbf{B}_q$	Features	Number of hyperparameters
Independant	I	I identity matrix	0
Semi-parametric latent factor model (SLFM)	$a^q(a^q)^\top$	$a^q \in \mathbb{R}^p$	p
Spherical parametrization	$\text{diag}(\mathbf{e})S^{\mathrm{T}}S\text{diag}(\mathbf{e})$	$\mathbf{e} \in \mathbb{R}^p$ and $S \in \mathbb{R}^{p\times p}$ particular UT	$\frac{p(p-1)}{2}$
Free-form	LL^{T}	$L \in \mathbb{R}^{p\times p}$ LT	$\frac{p(p+1)}{2}$
Low rank free-form	$\tilde{L}\tilde{L}^{\mathrm{T}}$	$\tilde{L} \in \mathbb{R}^{p\times \tilde{p}}$ LT	$\frac{\tilde{p}(\tilde{p}+1)}{2}$

Table 1.3: Summary of possible structures for $\mathbf{B}_q$ matrices (UT and LT stand for Upper Triangular and Lower Triangular).

unlike LMCs which share the same hyperparameters for all latent processes, convolution processes allow the use of individual hyperparameters for each output component. This enables more flexible modeling of the relationships between different output components, particularly useful in situations where outputs have complex dependencies.

References [Fricker et al., 2013], and [Alvarez and Lawrence, 2011] provide further details on convolution processes, while [Liu et al., 2018] offers an overview of the various options available for modeling the correlation in a MOGP.

1.7.3 Autokrigeability

The ICM model is more restrictive than the LMC models, since it assumes that each base covariance $k_q(\mathbf{x}, \mathbf{x}')$ contributes equally to the construction of autocovariances and cross-covariances for the outputs. In the case of ICM, for isotopic data with noise-free observations, the elementary properties of the Kronecker product (see [Horn and Johnson, 2012]) make it possible to simplify calculations of the inverse of $\mathbf{K}_{\chi_n,\chi_n}$ in the surrogate model update formulas, compared with the complete inversion of the $\mathbf{K}_{\chi_n,\chi_n}$ matrix required in the case of LMC (see [Alvarez et al., 2012]). These properties significantly reduce computational complexity, making the model more efficient for large covariance matrices.

It can also be shown that, in this case, predictions for the i^{th} simulator output depend only on the $g_i(\chi_n)$ observations of this same component (see for example [Wackernagel, 2003], [Helterbrand and Cressie, 1994] and [Bonilla et al., 2007]). In other words, prediction with the ICM model is equivalent to independent prediction for each output. This property is known as autokrigeability.

In practice, the same shared covariance hyperparameters transcribe a certain dependence between outputs (see [Teh et al., 2005] and [Liu et al., 2018]), and allows a better performance of the ICM model compared to independent models ([Liu et al., 2018]) when a small amount of data is available. The advantage of this model over two independent models is that it reduces the number of parameters to be estimated by taking into account the correlation between model output components, making it more robust when less data is available. In the case of heterotopic data (different surrogate model evaluation points for each output component), the autokrigeability property no longer applies, allowing the ICM model to demonstrate its effectiveness by learning the correlations between outputs and using them for predictions.

In the following, we present calculations to obtain the autokrigeability property, using the

elementary properties of the Kronecker product (see for example [Horn and Johnson, 2012]), in the case of an ICM model with isotopic data and no noise. We distinguish the case of a known constant trend from the more general case of an unknown linear trend. These calculations are inspired by [Bonilla et al., 2007] and [Helterbrand and Cressie, 1994] for the cases of known constant trend and unknown linear trend respectively. We use the various notations introduced in Section 1.7.1.

Calculations for a known constant trend

It is assumed that the trend M and the covariance K a priori are both known. It is important to remember that in the case of the ICM model:

$$\left(K(\mathbf{x},\mathbf{x}')_{i,j}\right)_{1\leq i,j\leq p} := k(\mathbf{x},\mathbf{x}')\,\mathbf{B}, \tag{1.39}$$

with $\mathbf{B}\in\mathbb{R}^{p\times p}$ symmetric positive definite matrix, called coregionalization and k continuous kernel with values in $\mathbb{R}$ defined on $\mathbb{X}$. Just remember that $\mathbf{K}_{\chi_n,\chi_n}\in\mathbb{R}^{pn\times pn}$ denotes the block covariance matrix on χ_n and $\mathbf{K}_{\chi_n}(\mathbf{x})\in\mathbb{R}^{pn\times p}$ the covariance matrix between $\mathbf{x}$ and χ_n. We then have

$$\mathbf{K}_{\chi_n,\chi_n} = \mathbf{B}\otimes k(\chi_n,\chi_n) \quad \text{and} \quad \mathbf{K}_{\chi_n}(\mathbf{x}) = \mathbf{B}\otimes k(\chi_n,\mathbf{x}) \tag{1.40}$$

with $k(\chi_n,\chi_n) := (k(\mathbf{x}^{(k)},\mathbf{x}^{(l)})_{1\leq k,l\leq n}\in\mathbb{R}^{n\times n}$ and $k(\chi_n,\mathbf{x}) := \left(k(\mathbf{x}^{(1)},\mathbf{x}),\ldots,k(\mathbf{x}^{(n)},\mathbf{x})\right)^\top \in \mathbb{R}^{n\times 1}$. According to Equation (1.24) we have :

$$\begin{aligned}
M_n(\mathbf{x}) &= M(\mathbf{x}) + \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1}\big(\boldsymbol{g}(\chi_n) - \boldsymbol{M}(\chi_n)\big)\\
&= M(\mathbf{x}) + \left(\mathbf{B}\otimes k(\chi_n,\mathbf{x})\right)^\top \left(\mathbf{B}\otimes k(\chi_n,\chi_n)\right)^{-1}\big(\boldsymbol{g}(\chi_n) - \boldsymbol{M}(\chi_n)\big)\\
&= M(\mathbf{x}) + \left(\mathbf{B}^\top\otimes k(\chi_n,\mathbf{x})^\top\right)\left(\mathbf{B}^{-1}\otimes k(\chi_n,\chi_n)^{-1}\right)\big(\boldsymbol{g}(\chi_n) - \boldsymbol{M}(\chi_n)\big)\\
&= M(\mathbf{x}) + \left(\mathbf{B}^\top\mathbf{B}^{-1}\right)\otimes\left(k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}\right)\big(\boldsymbol{g}(\chi_n) - \boldsymbol{M}(\chi_n)\big)\\
&= M(\mathbf{x}) + I_p\otimes\left(k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}\right)\big(\boldsymbol{g}(\chi_n) - \boldsymbol{M}(\chi_n)\big)
\end{aligned} \tag{1.41}$$

where $I_p\in\mathbb{R}^{p\times p}$ is the identity matrix and $\boldsymbol{g}(\chi_n)\in\mathbb{R}^{pn\times 1}$ and $\boldsymbol{M}(\chi_n)\in\mathbb{R}^{pn\times 1}$ respectively represent the column vectors of the $\boldsymbol{g}$ model and M trend evaluations on the χ_n DoE by concatenating the p blocks of size n corresponding to the evaluations of each $\boldsymbol{g}$ component on χ_n. Using index notation to denote the components of each vector quantity, we obtain for all $i\in\{1,\ldots,p\}$

$$M_{n,i}(\mathbf{x}) = M_i(\mathbf{x}) + k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}\big(g_i(\chi_n) - M_i(\chi_n)\big). \tag{1.42}$$

The prediction at a fixed $\mathbf{x}$ point and for a given i component depends only on the i^{th} component's evaluations of $\boldsymbol{g}$ on the DoE.

However, when updating the multivariate covariance matrix, the $\mathbf{B}$ term does not disappear. Indeed, according to Equation (1.25) and reusing the calculation performed in (1.41), we obtain :

$$\begin{aligned}
K_n(\mathbf{x},\mathbf{x}') &= K(\mathbf{x},\mathbf{x}') - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1}\mathbf{K}_{\chi_n}(\mathbf{x}')\\
&= K(\mathbf{x},\mathbf{x}') - I_p\otimes\left(k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}\right)\mathbf{K}_{\chi_n}(\mathbf{x}')\\
&= K(\mathbf{x},\mathbf{x}') - I_p\otimes\left(k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}\right)\left(\mathbf{B}\otimes k(\chi_n,\mathbf{x}')\right)\\
&= K(\mathbf{x},\mathbf{x}') - \left(I_p\mathbf{B}\right)\otimes\left(k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}k(\chi_n,\mathbf{x}')\right)\\
&= K(\mathbf{x},\mathbf{x}') - \left(k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}k(\chi_n,\mathbf{x}')\right)\mathbf{B}
\end{aligned} \tag{1.43}$$

because $\left(k(\chi_n,\mathbf{x})^\top k(\chi_n,\chi_n)^{-1}k(\chi_n,\mathbf{x})\right)\in\mathbb{R}$. Unlike the update for the mean, the covariance update therefore depends on the coregionalization matrix.

Calculations for an unknown linear trend

We place ourselves in the case where the a priori trend is no longer assumed to be known, but a linear combination of base functions with coefficients to be estimated (Equation (1.27)). We retain the case of an ICM (Equation (1.39)) with isotopic (and noise-free) data.

The formulas for updating the surrogate model are given in Equations (1.28), (1.29) and (1.30). Combining Equation (1.28) with the calculation performed in Equation (1.41), we see that M_n depends on $\boldsymbol{B}$ only through the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ defined by

$$\widehat{\boldsymbol{\beta}} := \left(\mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n)\right)^{-1} \mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \boldsymbol{g}(\chi_n).$$

All we need to do is study the dependence of $\widehat{\boldsymbol{\beta}}$ on $\mathbf{B}$.

Let us take a simplified case where $F_1 = \ldots = F_p = F^\star$ with $F^\star : \mathbb{X} \to \mathbb{R}^{l^\star \times 1}$, and $l_1 = \ldots = l_p = l^\star$ for $l^\star \in \mathbb{N}^\star$. In this case, we have :

$$\mathbf{F}(\chi_n) = I_p \otimes F^\star(\chi_n) \tag{1.44}$$

where $F^\star(\chi_n) := (F^\star(\mathbf{x}^{(1)}), \ldots, F^\star(\mathbf{x}^{(n)}))^\top \in \mathbb{R}^{n \times l^\star}$. Reinjecting (1.44) and (1.40) into the expression of $\widehat{\boldsymbol{\beta}}$, we obtain:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{F}(\chi_n)\right)^{-1} \mathbf{F}(\chi_n)^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \boldsymbol{g}(\chi_n) \\
&= \Big(\big(I_p \otimes F^\star(\chi_n)\big)^\top \big(\mathbf{B} \otimes k(\chi_n,\chi_n)\big)^{-1} \big(I_p \otimes F^\star(\chi_n)\big)\Big)^{-1} \big(I_p \otimes F^\star(\chi_n)\big)^\top \big(\mathbf{B} \otimes k(\chi_n,\chi_n)\big)^{-1} \boldsymbol{g}(\chi_n) \\
&= \Big(\big(I_p \otimes F^\star(\chi_n)^\top\big) \big(\mathbf{B}^{-1} \otimes k(\chi_n,\chi_n)^{-1}\big) \big(I_p \otimes F^\star(\chi_n)\big)\Big)^{-1} \big(I_p \otimes F^\star(\chi_n)^\top\big) \big(\mathbf{B}^{-1} \otimes k(\chi_n,\chi_n)^{-1}\big) \boldsymbol{g}(\chi_n) \\
&= \Big(\mathbf{B}^{-1} \otimes F^\star(\chi_n)^\top k(\chi_n,\chi_n)^{-1} F^\star(\chi_n)\Big)^{-1} \big(\mathbf{B}^{-1} \otimes F^\star(\chi_n)^\top k(\chi_n,\chi_n)^{-1}\big) \boldsymbol{g}(\chi_n) \\
&= \Big(\mathbf{B} \otimes \big(F^\star(\chi_n)^\top k(\chi_n,\chi_n)^{-1} F^\star(\chi_n)\big)^{-1}\Big) \big(\mathbf{B}^{-1} \otimes F^\star(\chi_n)^\top k(\chi_n,\chi_n)^{-1}\big) \boldsymbol{g}(\chi_n) \\
&= I_p \otimes \Big(\big(F^\star(\chi_n)^\top k(\chi_n,\chi_n)^{-1} F^\star(\chi_n)\big)^{-1} F^\star(\chi_n)^\top k(\chi_n,\chi_n)^{-1}\Big) \boldsymbol{g}(\chi_n),
\end{aligned} \tag{1.45}$$

hence the expected result that $\widehat{\beta}_i$ does not depend on $g_j(\chi_n)$ if $i \neq j$. The result demonstrated above generalizes ([Helterbrand and Cressie, 1994]) to the case where F_i are not necessarily equal. However, in the case where we impose restrictions between the β_i, for example $\beta_1 = \beta_2$, the result no longer holds (see [Helterbrand and Cressie, 1994]).

Finally, the autokriging property is not guaranteed in the case of heterotopic data, i.e., when the DoE evaluation points are no longer necessarily common to all simulator output components.

Summary:

The aim of Chapter 1 is to introduce Gaussian process regression (GPR) as a probabilistic model for deterministic numerical simulators, which are costly to evaluate. GPR involves the choice of both a trend and covariance structure, depending on parameters to be fitted from input/output samples. We illustrate GPR on simple analytical examples. Then, sequential design of experiments (DoE) based on GPR is introduced, with a specific focus on enrichment criteria. This chapter presents criteria suitable for black box global knowledge and black box optimization. Finally, the extension of Gaussian process regression (GPR) to vector-valued functions $\boldsymbol{g} : \mathbb{X} \subset \mathbb{R}^d \mapsto \mathbb{R}^p$ is presented, it is known as Multi-output Gaussian process (MOGP) regression. Different joint correlation kernels are presented that enhance predictive accuracy compared to separate output modeling.

The following chapter focuses on the study of enrichment criteria for the purpose of excursion set estimation, starting with an introduction to the framework of excursion set estimation.

Chapter 2

Background and tools for estimating excursion sets

Outlines

The first part of this chapter introduces excursion set estimation and explores various standard DoE enrichment criteria dedicated to this framework. The second part concerns the class of Stepwise Uncertainty Reduction (SUR) criteria. These criteria anticipate the impact of adding points to the sequential DoE, minimizing the expectation of a conditional residual uncertainty. Several examples of SUR strategies are presented, with an extension adapted to vector-valued model framework.

Contents

2.1 Excursion set estimation and adapted enrichment criteria **28**
2.1.1 Excursion set estimation framework 28
2.1.2 Deviation number criterion 31
2.1.3 tmse, tMMSE and tIMSE criteria 31
2.1.4 Bichon and Ranjan criteria 32
2.2 Stepwise Uncertainty Reduction (SUR) strategies for excursion set estimation **35**
2.2.1 Introduction to SUR strategies 35
2.2.2 SUR Excursion measure variance and Integrated Bernouilli variance strategies 37
2.2.3 Basics on Vorob'ev Theory and corresponding SUR strategy . . . 38
2.2.4 Extension of SUR Excursion measure variance and Integrated Bernoulli variance strategies for vector-valued models 40

2.1 Excursion set estimation and adapted enrichment criteria

2.1.1 Excursion set estimation framework

Nowadays, many industrial challenges arise from the difficulty of identifying feasible solutions, particularly when solving complex problems related to industrial product design. In this manuscript, the set of feasible solutions is called excursion set. Excursion set estimation consists in determining input parameter values that guarantee that a quantity of interest remains within specified limits, e.g., below a threshold. In practice, this quantity of interest often comes from a computationally expensive numerical model, typically a black box function representing a complex physical phenomenon. This is also known as an inversion

problem ([Chevalier, 2013]) and this problematic arises in various fields of application. For example, in the context of vehicle pollution control, this has been explored in the study of [El Amri et al., 2020]. Another example is the application to Autonomous Underwater Vehicles (AUV) sampling with salinity and temperature excursion sets from [Fossum et al., 2021]. The application covered by this thesis concerns the design of wind turbines. We refer to Chapter 5 for more details on this application.

Mathematical formulation

Recall that $\mathbb{X}$ denotes a compact set in $\mathbb{R}^d$ and $g : \mathbb{X} \to \mathbb{R}$ is a black box function, requiring a significant amount of time for computation. The objective of an excursion set estimation problem is to determine the set defined by:

$$\Gamma^\star := \Big\{\mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) \leq T\Big\} \tag{2.1}$$

with T a prescribed threshold. Given the high computational cost of evaluating the black box function g, estimating the set $\Gamma^\star$ must be done by limiting the number of evaluations of g.

Under assumptions of continuity of the black box function g, the problem of estimating $\Gamma^\star$ is in fact equivalent ([Bect et al., 2012]) to that of estimating

$$\partial\Gamma^\star := \Big\{\mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) = T\Big\}. \tag{2.2}$$

Indeed, estimation of both $\Gamma^\star$ and $\partial\Gamma^\star$ requires the selection of sampling points to refine our knowledge of function g in a neighborhood of $\partial\Gamma^\star$. Besides, still under the assumption of continuity of g, and if T belongs to the image of g, level set $\partial\Gamma^\star$ is a curve, a surface, or more generally a hypersurface of $\mathbb{X}$ (of dimension $d-1$), depending on d the dimension of $\mathbb{X}$. Thus, the difficulty of $\Gamma^\star$ estimation strongly increases with the dimension d of design space. It may also be noted that, in practice, the direction of the inequality in Equation (2.1) has limited significance, since an estimate of $\Gamma^\star$ allows an estimate of

$$\mathbb{X}\backslash\Gamma^\star := \Big\{\mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) > T\Big\}, \tag{2.3}$$

and vice versa.

Differences with estimating a probability of failure

Let us now look at a slightly different problem. Consider a probability measure on $\mathbb{X}$, for example Lebesgue measure on $\mathbb{X}$, denoted by $\mathbb{P}_\mathbb{X}$. We are now interested in estimating a failure probability of a system (see for example [Vazquez and Bect, 2009] and [Bect et al., 2012]). This translates into the estimates not of $\Gamma^\star$ (resp. $\mathbb{X}\backslash\Gamma^\star$) but of

$$\mathbb{P}_\mathbb{X}(\Gamma^\star) := \mathbb{P}_\mathbb{X}\Big(\mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) \leq T\Big), \tag{2.4}$$

(resp. $\mathbb{P}_\mathbb{X}(\mathbb{X}\backslash\Gamma^\star)$). The problem of estimating a failure probability is broader than that of estimating an excursion set, since it requires the selection of sampling points to refine knowledge of the volume $\mathbb{P}_\mathbb{X}(\Gamma^\star)$, which does not necessarily imply good knowledge of the level set $\partial\Gamma^\star$. On the other hand, it is clear that good knowledge of level set $\partial\Gamma^\star$ implies good knowledge of the volume $\mathbb{P}_\mathbb{X}(\Gamma^\star)$ but in the case where we are interested in volume, it might not be necessary to characterize $\partial\Gamma^\star$ with precision.

Ways of estimating an excursion set

Recall (see Chapter 1) that the Gaussian process used as surrogate model is denoted ξ, that the event given by the n first observations of the model in a sequential DoE is denoted $\mathcal{E}_n$, and that m_n and σ_n denote respectively the Kriging mean and standard deviation conditional on these observations. At the end of sequential DoE enrichment, and following a final update of the surrogate model based on observations, m_n provides an estimate of the black box function g. Thus, it is common and natural to use, as an estimator of $\Gamma^\star$, the set

$$\hat{\Gamma}_1 := \Big\{\mathbf{x} \in \mathbb{X},\, m_n(\mathbf{x}) \leq T\Big\}, \tag{2.5}$$

which we call naive estimator of $\Gamma^\star$. From Vorob'ev random set theory ([Molchanov, 2005]), it is also possible to use as estimator of $\Gamma^\star$ the Vorob'ev expectation (see Section 2.2.3) of random set

$$\Gamma \,\big|\, \mathcal{E}_n := \Big\{\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\Big\}\,\Big|\, \mathcal{E}_n. \tag{2.6}$$

In the following, we denote Vorob'ev estimator by $\hat{\Gamma}_2$, while naive estimator is denoted either by $\hat{\Gamma}_1$, or more concisely, by $\hat{\Gamma}$ if Vorob'ev estimator is not employed.

Let p_n be the coverage probability defined by

$$p_n(\mathbf{x}) := \mathbb{P}(\mathbf{x} \in \Gamma \,|\, \mathcal{E}_n) = \mathbb{P}(\xi(\mathbf{x}) \leq T \,|\, \mathcal{E}_n) = \phi\Big(\frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big), \tag{2.7}$$

where ϕ denotes the cumulative distribution function of standard normal distribution, we obtain that

$$\begin{aligned}
\hat{\Gamma}_1 &= \Big\{\mathbf{x} \in \mathbb{X},\, m_n(\mathbf{x}) \leq T\Big\} \\
&\overset{\text{a.s.}}{=} \Big\{\mathbf{x} \in \mathbb{X},\, \frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})} \geq 0 \;\text{ and }\; \sigma_n(\mathbf{x}) \neq 0\Big\} \\
&= \Big\{\mathbf{x} \in \mathbb{X},\, \phi\Big(\frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) \geq \phi(0) \;\text{ and }\; \sigma_n(\mathbf{x}) \neq 0\Big\} \quad \text{as } \phi \text{ increases} \\
&\overset{\text{a.s.}}{=} \Big\{\mathbf{x} \in \mathbb{X},\, p_n(\mathbf{x}) \geq \frac{1}{2}\Big\}.
\end{aligned} \tag{2.8}$$

In other words, the estimator $\hat{\Gamma}_1$ is almost surely equal to the set of $\mathbf{x}$ for which the coverage probability is greater than or equal to $1/2$.

Goal-oriented enrichment criteria

We recall that goal-oriented enrichment criteria are criteria adapted to a particular objective, such as optimization (see Section 1.6.2). These enrichment criteria are said to be adaptive, since they take into account model observations. This is commonly referred to as *active learning* (see for example [Bryan et al., 2005] and [Echard et al., 2011]). The aim of the following sections is to present the main enrichment criteria oriented to the estimation of excursion sets or failure probabilities (volume of an excursion set). By default, the proposed criteria are considered to be suitable for estimating excursion sets, and we will indicate specifically whether any of the criteria mentioned are rather suited to estimate a failure probability.

2.1.2 Deviation number criterion

The deviation number or "U" criterion was introduced in [Echard et al., 2011] as part of an active learning reliability method. It is written as :

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\arg\min}\, \mathrm{U}(\mathbf{x}) \quad \text{with} \quad \mathrm{U}(\mathbf{x}) := \frac{|m_n(\mathbf{x}) - T|}{\sigma_n(\mathbf{x})}. \tag{2.9}$$

The choice of this "U" criterion is quite natural in a context of excursion set estimation, since it simultaneously gives importance to points $\mathbf{x}$ of the design space, such that $m_n(\mathbf{x})$ is close to T and such that $\sigma_n(\mathbf{x})$ is sufficiently large. This brings us back to the trade-off between exploration and exploitation, in a context where exploitation corresponds to the estimation of an excursion set.

Using strict monotonicity and symmetry about the point $(0, 1/2)$ of the function ϕ enables us to show that

$$\begin{aligned}
\underset{\mathbf{x}\in\mathbb{X}}{\arg\min}\, \mathrm{U}(\mathbf{x}) &:= \underset{\mathbf{x}\in\mathbb{X}}{\arg\min} \left\{ \frac{|m_n(\mathbf{x}) - T|}{\sigma_n(\mathbf{x})} \right\} \\
&= \underset{\mathbf{x}\in\mathbb{X}}{\arg\min} \left\{ \left| \frac{m_n(\mathbf{x}) - T}{\sigma_n(\mathbf{x})} - 0 \right| \right\} \\
&= \underset{\mathbf{x}\in\mathbb{X}}{\arg\min} \left\{ \left| \phi\left(\frac{m_n(\mathbf{x}) - T}{\sigma_n(\mathbf{x})} \right) - \phi(0) \right| \right\} \\
&= \underset{\mathbf{x}\in\mathbb{X}}{\arg\min} \left\{ \left| p_n(\mathbf{x}) - \frac{1}{2} \right| \right\}.
\end{aligned} \tag{2.10}$$

So, minimizing the deviation number means finding a point $\mathbf{x}$ such that the probability of coverage is as close as possible to $1/2$.

According to [Echard et al., 2011], the "U" enrichment criterion gives more importance to points located in a neighborhood close of the border, thus favorising exploitation to exploration, contrarly to Bichon and Ranjan criteria (see Section 2.1.4). Other enrichment criteria based on coverage probability and very similar to the "U" criterion can be cited, such as entropy or probability of incorrect classification ([Bryan et al., 2005]). However, similarly to the "U" criterion, they have poor exploration properties.

2.1.3 tmse, tMMSE and tIMSE criteria

Three criteria called tmse, tIMSE and tMMSE, presented in [Picheny et al., 2010], generalize the mse, MMSE and IMSE criteria (see Section 1.6.1), to the estimation of excursion sets. They incorporate weights corresponding to the probability that the surrogate model belongs to the interval $[T - \varepsilon, T + \varepsilon]$, where ε is a "relatively" small positive real number. The added prefix "t" stands for "targeted", indicating that the focus is on the area of interest relative to the excursion set.

tmse criterion

Recall that the mse criterion improves the DoE by adding a point where the Kriging variance is maximum. In the tmse criterion, the improvement is weighted according to the probability that the surrogate model, at this point, belongs to the interval $[T-\varepsilon, T+\varepsilon]$, given the previous

observations. The expression of the tmse is given for a positive real number ε by

$$\begin{aligned}\mathbf{x}^{(n+1)} \in \arg\max_{\mathbf{x}\in\mathbb{X}} \operatorname{tmse}(\mathbf{x}) \quad \text{with} \quad \operatorname{tmse}(\mathbf{x}) &:= \mathbb{E}\big[\sigma_n^2(\mathbf{x})\mathbf{1}_{[T-\varepsilon,T+\varepsilon]}(\xi(\mathbf{x}))\,\big|\,\mathcal{E}_n\big] \\ &= \sigma_n^2(\mathbf{x})\,\mathbb{P}\big(\xi(\mathbf{x})\in[T-\varepsilon,T+\varepsilon]\,\big|\,\mathcal{E}_n\big).\end{aligned} \tag{2.11}$$

tMMSE criterion

For the tMMSE criterion, the approach is similar to that of the MMSE criterion, aiming to minimize the maximum Kriging variance after updating the DoE with the point under consideration. However, in addition to this, we incorporate the weighting of the tmse criterion, linked to the estimation of excursion sets. The criterion is formulated as follows:

$$\begin{aligned}\mathbf{x}^{(n+1)} \in \arg\min_{\mathbf{x}\in\mathbb{X}} \left\{ \max_{\mathbf{y}\in\mathbb{X}} \operatorname{tmse}_{\mathbf{x}}(\mathbf{y}) \right\} \quad \text{with} \quad \operatorname{tmse}_{\mathbf{x}}(\mathbf{y}) &:= \mathbb{E}\big[\sigma_{n,\mathbf{x}}^2(\mathbf{y})\mathbf{1}_{[T-\varepsilon,T+\varepsilon]}(\xi(\mathbf{y}))\,\big|\,\mathcal{E}_n\big] \\ &= \sigma_{n,\mathbf{x}}^2(\mathbf{y})\,\mathbb{P}\big(\xi(\mathbf{y})\in[T-\varepsilon,T+\varepsilon]\,\big|\,\mathcal{E}_n\big),\end{aligned} \tag{2.12}$$

where $\sigma_{n,\mathbf{x}}^2$ is defined as the Kriging variance conditioned on $\mathcal{E}_n$ enlarged with point $\mathbf{x}$.

tIMSE criterion

The tIMSE criterion is obtained by replacing the maximum in the tMMSE criterion with an integral to measure the total error on the design space instead of the maximum error. The formulation of the tIMSE criterion is as follows:

$$\mathbf{x}_{n+1} \in \arg\min_{\mathbf{x}\in\mathbb{X}} \left\{ \int_{\mathbb{X}} \operatorname{tmse}_{\mathbf{x}}(\mathbf{y})\,\mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{y}) \right\}, \tag{2.13}$$

where $\mathbb{P}_{\mathbb{X}}$ denotes Lesbesgue measure on $\mathbb{X}$.

Although Kriging variance remains independent of model values (see Section 1.2), these criteria now depend on these values with the introduction of weights. Within these criteria, we also observe the balance between exploration and exploitation. This is the distinction between global and local search of the excursion set. The three criteria presented above depend on the fixed parameter $\varepsilon > 0$ and that the choice of this parameter will have an impact on the compromise. In practice, according to [Picheny et al., 2010], the value of the parameter ε has moderate impact on the criteria and their use in a sequential enrichment strategy, except for very large or very small values.

2.1.4 Bichon and Ranjan criteria

Formulation and origin

Bichon and Ranjan criteria, introduced in [Bichon et al., 2008] and [Ranjan et al., 2008] are goal-oriented criteria for the DoE enrichment. These criteria are an adaptation of EI from [Jones et al., 1998], introduced in the context of global optimization, to the excursion set estimation framework.

As a reminder, the original idea of EI is to select a point $\mathbf{x}$ which improves the current mimimum $g_{\min}$, i.e., where $\xi(\mathbf{x})$ is below $g_{\min}$, while taking into account the uncertainty of the surrogate model. The idea behind both Bichon and Ranjan criteria is to adapt this strategy to the excursion set estimation framework by considering both the variability of the surrogate model and the potential improvement in the knowledge of the excursion set boundary. This is the exploration-exploitation compromise.

To introduce Bichon criterion ([Bichon et al., 2008]), it is necessary to define, for a fixed $\mathbf{x}$ belonging to $\mathbb{X}$, a random variable $\mathrm{FF}(\mathbf{x})$ (Feasibility Function) defined by

$$\begin{aligned}\mathrm{FF}(\mathbf{x}) :&= \epsilon(\mathbf{x}) - \min\left\{|T - \xi(\mathbf{x})|, \epsilon(\mathbf{x})\right\} \\ &= \left(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\right)^+,\end{aligned} \tag{2.14}$$

with $(\,.\,)^+ := \max(\,.\,, 0)$ and $\epsilon(\mathbf{x})$ a fixed function. This function EFF represents the distance between the surrogate model (i.e. the response under GP assumptions) and the bounds of the interval $[T - \varepsilon(\mathbf{x}), T + \varepsilon(\mathbf{x})]$ only if the surrogate model belongs to this interval and is 0 otherwise. In practice, the interval width $\epsilon(\mathbf{x})$ is chosen proportional to the posterior standard deviation $\sigma_n(\mathbf{x})$, leading in particular to a null value of the criterion for the points already present in the DoE. An example is given in Figure 2.1. The feasibility function is drawn for one sample path of $\xi|\,\mathcal{E}_n$. Its maximization aims to select points close to the boundary of the excursion set naive estimate or points associated to high values of $\epsilon(\mathbf{x})$. Then, the average of

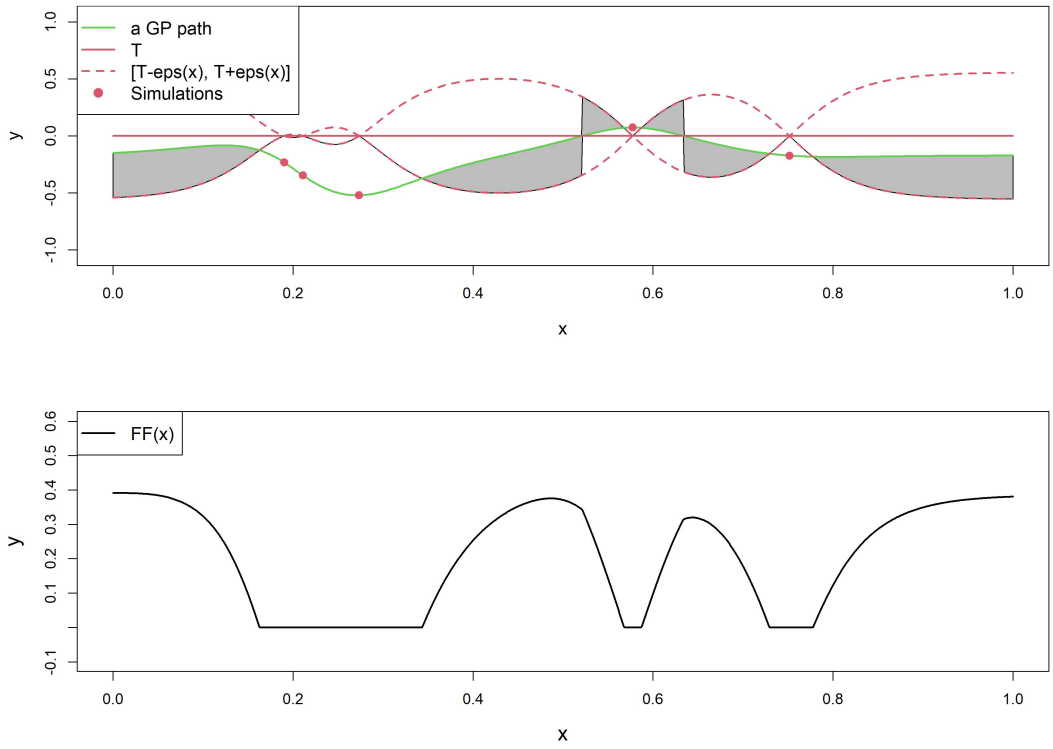


Figure 2.1: Representation of Feasibility Function (bottom) for a given example of a GP sample path conditioned on 5 evaluations of the g function, a threshold T set to 0 and $\epsilon(x) := \sigma_n(\mathbf{x})$ (top). The height of the gray area represents, for each value of x, the value of FF(x).

FF over all sample paths gives the Expected Feasibility Function (EFF)

$$\mathrm{EFF}(\mathbf{x}) := \mathbb{E}\Big[\left(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\right)^+ \Big|\, \mathcal{E}_n\Big]. \tag{2.15}$$

The new selected point according to Bichon criterion is

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}}\ \mathrm{EFF}(\mathbf{x}). \tag{2.16}$$

Ranjan criterion is an alternative to Bichon criterion given by

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}}\ \mathrm{EFF}_2(\mathbf{x}) \quad \text{with} \quad \mathrm{EFF}_2(\mathbf{x}) := \mathbb{E}\Big[\big(\epsilon(\mathbf{x})^2 - |T-\xi(\mathbf{x})|^2\big)^+ \,\Big|\, \mathcal{E}_n\Big]. \tag{2.17}$$

The representation as mean distance of the surrogate model to the bounds of the interval $[T-\varepsilon(\mathbf{x}), T+\varepsilon(\mathbf{x})]$ (Figure 2.1) is less obvious due to the introduction of squares. However, this criterion is also widely used in the context of estimating excursion sets.

Statistical interpretation

To interpret (2.15) and (2.17), it is possible to make a heuristic analogy with the theory of statistical tests ([Dagnelie, 1992]). Let $\mathbf{x}$ be fixed, suppose that $\xi(\mathbf{x})|\,\mathcal{E}_n \sim \mathcal{N}(m_n(\mathbf{x}), \sigma_n(\mathbf{x}))$ with $m_n(\mathbf{x})$ unknown and $\sigma_n(\mathbf{x}) > 0$ known and let us define the following statistical test

$$\mathrm{H}_0 : m_n(\mathbf{x}) = T \quad \text{against} \quad \mathrm{H}_1 : m_n(\mathbf{x}) \neq T. \tag{2.18}$$

We choose $\upsilon_{\mathbf{x}}^{\delta} := \left(\frac{\xi(\mathbf{x})-T}{\sigma_n(\mathbf{x})}\right)^{\delta} \Big|\, \mathcal{E}_n$ as the conditional test statistic which follows standard normal distribution under H_0 for $\delta = 1$ (Bichon criterion) and χ^2 distribution with 1 degree of freedom (denoted χ_1^2) for $\delta = 2$ (Ranjan criterion). Consequently, we refute Hypothesis H_0 at order α if

$$|\upsilon_{\mathbf{x}}^{\delta}| > \kappa_{\alpha}^{\delta}, \tag{2.19}$$

with $\kappa_{\alpha}^{1} := q_{1-\frac{\alpha}{2}}$ the quantile of order $1-\frac{\alpha}{2}$ of the standard normal distribution and $\kappa_{\alpha}^{2} := z_{1-\alpha}$ the quantile of order $1-\alpha$ of the χ_1^2 distribution.

However, our goal is not to refute the hypothesis that $m_n(\mathbf{x}) = T$ but rather to select the $\mathbf{x}$ for which, on average conditioned on the event "H_0 is plausible" (i.e., $\kappa_{\alpha}^{\delta} - |\upsilon_{\mathbf{x}}^{\delta}| > 0$), the quantity $\kappa_{\alpha}^{\delta} - \upsilon_{x}^{\delta}$ is the largest possible, weighted by the probability that H_0 is plausible. Multiplying $\mathbb{E}\big[(\kappa_{\alpha}^{\delta} - |\upsilon_{\mathbf{x}}^{\delta}|)^+ \big|\, \mathcal{E}_n\big]$ by $\sigma_n(\mathbf{x})^{\delta}$ leads to both Bichon and Ranjan criteria with $\epsilon(\mathbf{x}) := \kappa_{\alpha}^{\delta}\sigma_n(\mathbf{x})^{\delta}$, and has the effect to increase the exploration ability of the criterion.

Explicit formulas

Finally, explicit formulations of Bichon and Ranjan criteria (see [Bect et al., 2012] for a proof) can be calculated based on the posterior Kriging mean and variance, the threshold T and the width $\varepsilon(\mathbf{x})$.

$$\begin{aligned}
\mathrm{EFF}(\mathbf{x}) = (m_n(\mathbf{x}) - T)&\left[2\,\phi\Big(\frac{T-m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \phi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right] \\
&- \sigma_n(\mathbf{x})\left[2\,\varphi\Big(\frac{T-m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \varphi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \varphi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right] \\
&+ \epsilon(\mathbf{x})\left[\phi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right],
\end{aligned} \tag{2.20}$$

and

$$\begin{aligned}\mathrm{EFF}_2(\mathbf{x}) = 2\sigma_n(\mathbf{x})(m_n(\mathbf{x}) - T)\left[\varphi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \varphi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right] \\ + \sigma_n(\mathbf{x})(T^+ - m_n(\mathbf{x}))\varphi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \sigma_n(\mathbf{x})(T^- - m_n(\mathbf{x}))\varphi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) \\ + \big(\epsilon(\mathbf{x})^2 - \sigma_n(\mathbf{x})^2 - (T - m_n(\mathbf{x}))^2\big)\left[\phi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right],\end{aligned} \tag{2.21}$$

with $T^{\pm} := T \pm \epsilon(\mathbf{x})$, φ and ϕ represent respectively the probability density and cumulative distribution functions of the standard normal distribution. In practice, the enrichment of the DoE is done by maximizing the criterion given by Equation (2.20) for Bichon criterion and the one given by Equation (2.21) for Ranjan criterion.

Link with tmse and tIMSE criteria

Bichon criterion, defined by equation (2.15), can be rewritten as:

$$\mathrm{EFF}(\mathbf{x}) = \mathbb{E}\Big[\big(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\big)\mathbf{1}_{[T-\varepsilon,T+\varepsilon]}(\xi(\mathbf{x}))\,\Big|\,\mathcal{E}_n\Big]. \tag{2.22}$$

Thus, Bichon criterion is actually a variation of the tmse criterion (see Section 2.1.3), by replacing $\sigma_n(\mathbf{x})^2$ by $\big(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\big)$. In Figure 2.1, if instead of taking the grey vertical average distance on the graph, we simply consider the value of $\sigma_n(\mathbf{x})^2$, then we obtain an interpretation of the tmse criterion. Since the tIMSE criterion is an integral criterion based on the tmse criterion, it is reasonable to ask whether an integral version of Bichon criterion would be as natural to formulate. In reality, the dependence of the quantity $\big(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\big)$ on $\xi(\mathbf{x})$ in Bichon criterion makes things relatively more complex than for the tIMSE criterion (see Chapter 2 for an introduction to a SUR version of Bichon criterion).

2.2 Stepwise Uncertainty Reduction (SUR) strategies for excursion set estimation

In this section, we present Stepwise Uncertainty Reduction (SUR) strategies, which anticipate the successive impacts of selecting the next sampling points. A general introduction to this type of strategy is first given, followed by some specific examples of SUR strategies adapted to the estimation of excursion sets.

2.2.1 Introduction to SUR strategies

Origin and formulation

k-step lookahead strategies are optimal Bayesian enrichment strategies (for a finite horizon of k new points). The aim of these strategies is to anticipate the successive impacts of choosing the next sampling points, in order to optimally select the points that lead to minimization of a Bayesian risk, chosen by assuming the successive addition of the k points considered to the DoE (see for example [Bect et al., 2012]). In practice, the use of such criteria is not possible for large values of k, since it generally requires the calculation of $k + 1$ nested conditional expectations. Stepwise Uncertainty Reduction (SUR) strategies correspond to the simplified case where $k = 1$.

More precisely, let us introduce a residual uncertainty $\mathcal{H}_n$, computed with the GP model conditioned on $\mathcal{E}_n$. From this residual uncertainty, the conditional residual uncertainty $\mathcal{H}_{n+1}(\mathbf{x})$ is defined as the updated uncertainty when adding $\mathbf{x}$ to DoE χ_n. This conditional uncertainty is then a measurable function of the random variable $\xi(\mathbf{x})|\,\mathcal{E}_n$ and corresponds to the Bayesian risk of k-step lookahead strategies when $k = 1$. The associated SUR strategy is then defined by

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmin}}\, \mathcal{J}_n(\mathbf{x}) \quad \text{with} \quad \mathcal{J}_n(\mathbf{x}) := \mathbb{E}\big[\mathscr{H}_{n+1}(\mathbf{x})\big]. \tag{2.23}$$

Note that the expectation in $\mathcal{J}_n(\mathbf{x})$ is relative to the distribution of $\xi(\mathbf{x})|\,\mathcal{E}_n$. Eq. (2.23) means that evaluating the surrogate model at $\mathbf{x}^{(n+1)}$ decreases at most the expected residual uncertainty.

SUR strategies are numerically more complex to implement, but are generally more efficient than other goal oriented strategies ([Bect et al., 2012] and [Chevalier, 2013]), for the same number of evaluations. Theoretical convergence results for these strategies under certain assumptions have been provided in [Bect et al., 2019]. A reduction in the numerical complexity of SUR strategies is frequently used through the use of Kriging update formulas introduced in [Chevalier, 2013].

Kriging update formulas

In the context of SUR strategies, the quantity $\mathcal{J}_n(\mathbf{x})$ in Equation (2.23) for a fixed $\mathbf{x}$ is usually simplified thanks to Kriging update formulas and more particularly using Kriging standard deviation. Indeed, contrary to the trend, the Kriging standard deviation does not depend on surrogate model observations. For instance the recurrence formula, used in [Chevalier, 2013] is efficient for calculating Kriging model in the context of universal Kriging and when Kriging parameters β and θ do not need to be re-estimated. These Kriging update formulas are given for all $\mathbf{y}, \mathbf{y}'$ in $\mathbb{X}^2$ by

$$m_{n+1}(\mathbf{y}) = m_n(\mathbf{y}) + k_n(\mathbf{y}, \mathbf{x}^{(n+1)})\, k_n(\mathbf{x}^{(n+1)}, \mathbf{x}^{(n+1)})^{-1} \left(g(\mathbf{x}^{(n+1)}) - m_n(\mathbf{x}^{(n+1)})\right), \tag{2.24}$$

$$\sigma^2_{n+1}(\mathbf{y}) = \sigma^2_n(\mathbf{y}) - k^2_n(\mathbf{y}, \mathbf{x}^{(n+1)})\, \sigma^2_n(\mathbf{x}^{(n+1)})^{-1}, \tag{2.25}$$

$$k_{n+1}(\mathbf{y}, \mathbf{y}') = k_n(\mathbf{y}, \mathbf{y}') - k_n(\mathbf{y}, \mathbf{x}^{(n+1)}) k_n(\mathbf{x}^{(n+1)}, \mathbf{x}^{(n+1)})^{-1} k_n(\mathbf{y}', \mathbf{x}^{(n+1)}), \tag{2.26}$$

with $\mathbf{x}^{(n+1)}$ the $(n+1)^{\text{th}}$ observation point and can be demonstrated from Equations (1.1), (1.2) and (1.3) using Schur's complement formula ([Horn and Johnson, 2012]). These kriging update formulas can be interpreted as kriging formulas on the DoE χ_n with a single observation $(\mathbf{x}^{(n+1)}, g(\mathbf{x}^{(n+1)}))$.

As for SUR strategies, it is possible to generalize these formulas to the case of simultaneous addition of q points ([Chevalier, 2013]). The advantage of these formulas is that the expressions of m_n, σ_n, and k_n are reused to reduce computational time. It is particularly useful in SUR strategies where numerous evaluations of Kriging formulas are necessary for multiple assessments of the sampling criterion in the context of its minimization (Equation (2.23)). Finally, it can be shown that these Kriging formulas still coincide with Gaussian process conditional formulas in the context of universal Kriging (see Appendix A of [Chevalier, 2013] for a proof).

In the following, several classic examples of SUR strategies are presented in the context of estimating excursion sets or estimating a failure probability. SUR criteria presented below are defined from their conditional residual uncertainty $\mathcal{H}_{n+1}(\mathbf{x})$ (or their residual uncertainty $\mathcal{H}_n$) and the SUR formulation of Equation (2.23).

2.2.2 SUR Excursion measure variance and Integrated Bernouilli variance strategies

SUR Excursion measure variance

Recall that $\mathbb{P}_{\mathbb{X}}$ denotes a probability measure on $\mathbb{X}$, for example Lebesgue measure on $\mathbb{X}$, and let $\Gamma^\star$ and Γ denote respectively the excursion set we are looking at (Equation (2.1)) and the associated random set induced by ξ (Equation (2.6)). Let $\gamma^\star$ be the volume of $\Gamma^\star$ and γ the random variable modeling the volume of Γ:

$$\gamma^\star := \mathbb{P}_{\mathbb{X}}(\Gamma^\star) = \mathbb{P}_{\mathbb{X}}\Big(\mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) \leq T\Big), \tag{2.27}$$

and

$$\gamma := \mathbb{P}_{\mathbb{X}}(\Gamma) = \mathbb{P}_{\mathbb{X}}\Big(\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\Big) = \int_{\mathbb{X}} \mathbf{1}_{\{\xi(\mathbf{z}) \leq T\}}\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \tag{2.28}$$

Recall that p_n denotes the coverage probability of Γ conditioned on $\mathcal{E}_n$, defined by

$$p_n(\mathbf{x}) := \mathbb{P}(\mathbf{x} \in \Gamma \,|\, \mathcal{E}_n) = \mathbb{P}(\xi(\mathbf{x}) \leq T \,|\, \mathcal{E}_n) = \phi\Big(\frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big). \tag{2.29}$$

With these notations, it is easy to show using Fubini Tonelli Theorem that

$$\mathbb{E}[\gamma |\, \mathcal{E}_n] = \int_{\mathbb{X}} p_n(\mathbf{z})\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \tag{2.30}$$

which gives an estimator of the volume $\gamma^\star$ of the set $\Gamma^\star$.

The SUR Excursion measure variance criterion ([Bect et al., 2012]) is a natural SUR strategy when estimating a failure probability, i.e., when estimating the volume $\gamma^\star$. This SUR strategy is defined by considering the conditional variance of the excursion set volume as the residual uncertainty. More precisely, we have

$$\mathcal{H}_n := \mathrm{Var}[\gamma |\, \mathcal{E}_n] \quad \text{and} \quad \mathcal{H}_{n+1}(\mathbf{x}) := \mathrm{Var}[\gamma |\, \xi(\mathbf{x}),\, \mathcal{E}_n] \tag{2.31}$$

This criterion was initially considered impractical, as it might require Monte Carlo approximations, involving conditional simulations of Gaussian random fields. An explicit computation of this volume variance was proposed by [Chevalier et al., 2014a], making the criterion usable in practice.

SUR Integrated Bernouilli variance

The SUR Integrated Bernoulli variance criterion was initially introduced in [Bect et al., 2012] as an alternative to the initially impractical Excursion Measure Variance criterion before becoming a standard criterion. The idea was to provide an upper bound for the SUR Excursion measure variance criterion.

From Equations (2.30) and (2.31) and using Cauchy-Schwarz Inequality then Fubini Tonelli Theorem, we obtain

$$\begin{aligned}
\mathrm{Var}[\gamma |\, \mathcal{E}_n] &= \mathbb{E}\Big[\big(\gamma - \mathbb{E}[\gamma |\, \mathcal{E}_n]\big)^2 \Big|\, \mathcal{E}_n\Big] \\
&= \mathbb{E}\bigg[\Big(\int_{\mathbb{X}} \mathbf{1}_{\{\xi(\mathbf{z}) \leq T\}} - p_n(\mathbf{z})\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z})\Big)^2 \bigg|\, \mathcal{E}_n\bigg] \\
&\leq \mathbb{E}\bigg[\int_{\mathbb{X}} \big(\mathbf{1}_{\{\xi(\mathbf{z}) \leq T\}} - p_n(\mathbf{z})\big)^2\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}) \bigg|\, \mathcal{E}_n\bigg] \\
&= \int_{\mathbb{X}} \mathrm{Var}[\mathbf{1}_{\{\xi(\mathbf{z}) \leq T\}} |\, \mathcal{E}_n]\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}) \\
&= \int_{\mathbb{X}} p_n(\mathbf{z})(1 - p_n(\mathbf{z}))\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}).
\end{aligned} \tag{2.32}$$

Thus, the SUR Integrated Bernouilli variance criterion is the SUR strategy defined by posing

$$\mathcal{H}_n := \int_{\mathbb{X}} p_n(\mathbf{z})(1 - p_n(\mathbf{z}))\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}) \quad \text{and} \quad \mathcal{H}_{n+1} := \int_{\mathbb{X}} p_{n+1}(\mathbf{z})(1 - p_{n+1}(\mathbf{z}))\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \quad (2.33)$$

with

$$p_{n+1}(\mathbf{z}) := \mathbb{P}(\mathbf{z} \in \Gamma \,|\, \xi(\mathbf{x}),\, \mathcal{E}_n). \quad (2.34)$$

Inequality (2.32) shows that, as the uncertainty associated with the Integrated Bernoulli variance tends towards 0, the uncertainty associated with the Excursion measure variance also tends towards 0. Additionally, the function $p \mapsto p(1-p)$ is zero at $p = 0$ and $p = 1$, and reaches a maximum at $p = 1/2$. Therefore, the uncertainty associated with the Integrated Bernoulli variance is high when large regions of $\mathbb{X}$ are uncertain ($p_n(\mathbf{z}) \simeq 1/2$) and low when large regions of $\mathbb{X}$ are well determined ($p_n(\mathbf{z}) \simeq 0$ or 1).

An explicit formulation of the SUR Integrated Bernoulli variance criterion has been provided by [Chevalier et al., 2014a], enabling its practical application without the need for Monte Carlo methods on $\xi(\mathbf{x})$ to estimate m_{n+1} and subsequently derive p_{n+1}.

In the case of the SUR Integrated Bernoulli variance criterion, when the associated residual uncertainty $\mathcal{H}_n$ (Equation (2.33)) tends towards 0, this means that for each point $\mathbf{z}$ of $\mathbb{X}$, $p_n(\mathbf{z}) = 0$ or 1, indicating that all points are perfectly classified. The SUR Integrated Bernoulli variance criterion therefore seems better suited to estimating excursion sets than the SUR Excursion measure variance criterion, which is more appropriate for estimating volume $\gamma^\star$. However, the latter criterion is still derived from a criterion suitable for estimating a failure probability. The next section introduces Vorob'ev theory of random sets from [Molchanov, 2005], including in particular an extension of the notion of variance for a random set, called Vorob'ev deviation. This section then presents a SUR criterion from [Chevalier, 2013] that uses Vorob'ev deviation on random set Γ.

2.2.3 Basics on Vorob'ev Theory and corresponding SUR strategy

Vorob'ev expectation

In this part, the notion of expectation for random closed sets in the sense of Vorob'ev is drawn from [Molchanov, 2005] and has been further adopted in [Azzimonti, 2016] in the case of Bayesian set estimation based on random field priors. The framework is a compact set $\mathbb{X} \subset \mathbb{R}^d$ and a random closed set Γ of $\mathbb{X}$. We note $\mathcal{C}$ the set of all compacts of $\mathbb{X}$. It is recalled that $\Gamma : \Omega \to \mathcal{C}$ is a random closed set if it is a measurable function for the Borel σ-algebra on $\mathcal{C}$ with respect to the Fell topology, on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The Fell topology is generated by sets of the form $\{C \in \mathcal{C},\, C \cap U\}$ for any open set $U \subset \mathbb{X}$ and $\{C \in \mathcal{C},\, C \cap F\}$ for any closed set $F \subset \mathbb{X}$. This is equivalent to saying that

$$\forall\, C \in \mathcal{C},\, \{w \in \Omega, \Gamma(w) \cap C \neq \varnothing\} \in \mathcal{F}. \quad (2.35)$$

Let us define the parametric family $\{Q_\alpha\}_{\alpha \in [0,1]}$ by:

$$Q_\alpha := \{\mathrm{x} \in \mathbb{X} : p(\mathrm{x}) := \mathbb{P}(\mathrm{x} \in \Gamma) \geq \alpha\},\, \forall\, \alpha \in [0,1]. \quad (2.36)$$

The elements of $\{Q_\alpha\}_{\alpha \in [0,1]}$ are called the Vorob'ev quantiles of the random closed set Γ and the function p is called the coverage function of Γ.

The Vorob'ev approach from [Molchanov, 2005] is only one among several approaches to generalize the notion of expectation to random closed sets. From the parametric family of Vorob'ev quantiles (Eq. (2.36)), the expectation of Γ in the sense of Vorob'ev is then defined

as the Vorob'ev quantile of measure equal (or the closest higher one) to the expectation of the measure of Γ, as specified in the definition below.

Definition 1. Vorob'ev expectation of a random closed set Γ is the set $Q_{\alpha^\star}$, where Vorob'ev threshold $\alpha^\star$ is defined by

$$\forall\, \alpha > \alpha^\star,\ \mathbb{P}_{\mathbb{X}}(Q_\alpha) < \mathbb{E}[\mathbb{P}_{\mathbb{X}}(\Gamma)] \leq \mathbb{P}_{\mathbb{X}}(Q_{\alpha^\star}), \tag{2.37}$$

where $\mathbb{P}_{\mathbb{X}}$ denotes Lebesgue measure on $\mathbb{X}$.

Remark 1.
• Based on Equation (2.36), the function $\alpha \mapsto \mathbb{P}_{\mathbb{X}}(Q_\alpha)$ is decreasing on $[0,1]$.
• The uniqueness of $\alpha^\star$ in the definition is easily checked. The existence of such $\alpha^\star$ in the definition of Vorob'ev expectation is based on the fact that $\alpha \mapsto \mathbb{P}_{\mathbb{X}}(Q_\alpha)$ is decreasing and left continuous as the coverage function p has upper semi-continuity property (see [Molchanov, 2005] page 23).
• The continuity of the function $\alpha \mapsto \mathbb{P}_{\mathbb{X}}(Q_\alpha)$ ensures equality $\mathbb{P}_{\mathbb{X}}(Q_{\alpha^\star}) = \mathbb{E}[\mathbb{P}_{\mathbb{X}}(\Gamma)]$ in the definition of Vorob'ev expectation.

In the particular case where Γ is given by $\{\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\} \,|\, \mathcal{E}_n$ with ξ a stochastic process indexed by $\mathbb{X}$ with continuous trajectories conditioned on the event $\mathcal{E}_n$ corresponding to n evaluations of ξ and T a fixed threshold, Γ is a random closed set ([Molchanov, 2005, page 3]). A sufficient condition to obtain a stochastic process with continuous trajectories is to consider a separable Gaussian process with continuous mean and covariance kernel of type Matérn 3/2 or 5/2 ([Paciorek, 2003, pages 35 and 44]). Moreover, in this case, the function $\alpha \mapsto \mathbb{P}_{\mathbb{X}}(Q_\alpha)$ is continuous and so the equality $\mathbb{P}_{\mathbb{X}}(Q_{\alpha^\star}) = \mathbb{E}[\mathbb{P}_{\mathbb{X}}(\Gamma)|\, \mathcal{E}_n]$ is verified. It is also important to recall that naive estimator $\hat{\Gamma}_1$ is almost surely equal to the median of Vorob'ev (quantile of order 1/2) in the case of the restriction of the Lebesgue measure for $\mathbb{P}_{\mathbb{X}}$. Indeed, by noting ϕ the cumulative distribution function of the standard normal distribution,

$$\begin{aligned}
\hat{\Gamma}_1 &= \{\mathbf{x} \in \mathbb{X},\, m_n(\mathbf{x}) \leq T\} \\
&\overset{\text{a.s.}}{=} \left\{\mathbf{x} \in \mathbb{X},\, \frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})} \geq 0 \ \text{ and } \ \sigma_n(\mathbf{x}) \neq 0\right\} \\
&= \left\{\mathbf{x} \in \mathbb{X},\, \phi\Big(\frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) \geq \phi(0) \ \text{ and } \ \sigma_n(\mathbf{x}) \neq 0\right\} \quad \text{as } \phi \text{ increases} \\
&\overset{\text{a.s.}}{=} \left\{\mathbf{x} \in \mathbb{X},\, p_n(\mathbf{x}) \geq \frac{1}{2}\right\} \\
&= Q_{\frac{1}{2}},
\end{aligned} \tag{2.38}$$

where p_n is the coverage function $p_n(\mathbf{x}) := \mathbb{P}(\xi(\mathbf{x}) \leq T|\, \mathcal{E}_n)$.

Repeating the previous calculation and replacing 1/2 by Vorob'ev threshold $\alpha^\star$, we obtain:

$$Q_{\alpha^\star} \overset{\text{a.s.}}{=} \left\{\mathrm{x} \in \mathbb{X},\, m_n(\mathrm{x}) \leqslant T - \phi^{-1}(\alpha^\star)\, \sigma_n(\mathrm{x})\right\} \tag{2.39}$$

Vorob'ev deviation

The concept of Vorob'ev deviation is used to define residual uncertainty $\mathcal{H}_n(\mathbf{x})$ in a SUR strategy. Let us start by introducing the notion of distance between two random closed sets.

For all random closed sets Γ_1, Γ_2 defined on $\mathbb{X}$, the average distance $\mathrm{d}_{\mathbb{P}_\mathbb{X}}$ with respect to a measure $\mathbb{P}_\mathbb{X}$ between Γ_1 and Γ_2 is defined by:

$$d_{\mathbb{P}_\mathbb{X}}(\Gamma_1, \Gamma_2) := \mathbb{E}\left[\mathbb{P}_\mathbb{X}(\Gamma_1 \Delta \Gamma_2)\right] \tag{2.40}$$

where Δ is the random symmetric difference: $\Gamma_1 \Delta \Gamma_2 := (\Gamma_1 \backslash \Gamma_2) \cup (\Gamma_2 \backslash \Gamma_1)$. In addition, the function $\mathrm{d}_{\mathbb{P}_\mathbb{X}}$ checks the properties of a distance.

Proposition 1. Noting $Q_{\alpha^\star}$ the Vorob'ev expectation of the random closed set Γ and assuming that $\alpha^\star \geq \frac{1}{2}$, it results that, for any measurable set M included in $\mathbb{X}$ such that $\mathbb{P}_\mathbb{X}(\mathrm{M}) = \mathbb{E}[\mathbb{P}_\mathbb{X}(\Gamma)]$,

$$d_{\mathbb{P}_\mathbb{X}}(\Gamma, Q_{\alpha^\star}) \leq d_{\mathbb{P}_\mathbb{X}}(\Gamma, \mathrm{M}). \tag{2.41}$$

This proposition from [Molchanov, 2005] justifies the choice of Vorob'ev quantile family to define Vorob'ev expectation and also allows to define Vorob'ev deviation. Moreover, when $\mathbb{E}[\mathbb{P}_\mathbb{X}(\Gamma)] = \mathbb{P}_\mathbb{X}(Q_{\alpha^\star})$, which arises for example in the case of $\Gamma := \{\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\}$ with ξ a stochastic process indexed by $\mathbb{X}$ with continuous trajectories, the condition $\alpha^\star \geq \frac{1}{2}$ is no longer necessary (see [El Amri, 2019, page 28] for the proof).

Definition 2. Vorob'ev deviation of random set Γ is defined as the quantity $\mathrm{d}_{\mathbb{P}_\mathbb{X}}(\Gamma, Q_{\alpha^\star})$. Vorob'ev deviation quantifies the variability of random closed set Γ relative to its Vorob'ev expectation.

SUR Vorob'ev criterion

Once the basic elements of Vorob'ev theory are introduced, the associated SUR strategy is simply defined from the definition of SUR strategies via Equation (3.16) by taking:

$$\mathcal{H}_n := \mathbb{E}\left[\mathbb{P}_\mathbb{X}(\Gamma \Delta Q_{n,\alpha_n^\star}) \middle|\, \mathcal{E}_n\right] \quad \text{and} \quad \mathcal{H}_{n+1}(\mathbf{x}) := \mathbb{E}\left[\mathbb{P}_\mathbb{X}(\Gamma \Delta Q_{n+1,\alpha_{n+1}^\star}) \middle|\, \xi(\mathbf{x}),\, \mathcal{E}_n\right] \tag{2.42}$$

where $Q_{n,\alpha_n^\star}$ denotes Vorob'ev expectation conditioned on $\mathcal{E}_n$ and $Q_{n+1,\alpha_{n+1}^\star}$ Vorob'ev expectation conditioned on $\mathcal{E}_n$ and the addition of the point $(\mathbf{x}, \xi(\mathbf{x}))$ to the DoE. The idea behind (2.42) is to take as residual uncertainty, the variation with respect to Vorob'ev expectation of random closed set Γ, with $\Gamma := \{\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\}$. With the assumption that $\alpha_{n+1}^\star = \alpha_n^\star$ and by re-injecting the quantity $\mathcal{H}_{n+1}(\mathbf{x})$ of (2.42) in the $\mathcal{J}_n$ criterion (2.23), it is possible to find a simplified formulation involving only an integral of a simple quantity [Chevalier, 2013]. This quantity is dependent on the cumulative distribution functions of the standard normal distribution and the bivariate centered normal distribution with given covariance matrix. Such a formulation then allows less time consuming computations and therefore is implemented in the package *KrigInv* ([Chevalier et al., 2014b]).

2.2.4 Extension of SUR Excursion measure variance and Integrated Bernoulli variance strategies for vector-valued models

We present in this section excursion set estimation methodologies that have been developed in the context of a vector-valued black box model. Extensions of the SUR criteria of Excursion measure variance and Integrated Bernoulli variance have been proposed in [Fossum et al., 2021] in the context of vector-valued black box functions and multi-output Gaussian process surrogate models (see Chapter 1, Section 1.7 for the introduction of MOGPs models).

We begin by recalling that in this framework, it is possible to assume that the vector-valued black box model $\boldsymbol{g} := (g_1, \ldots, g_p)^\top$ is the realization of a multi-output Gaussian process

$\boldsymbol{\xi} := (\xi_1, \ldots, \xi_p)^\top$. In this case, $M := (M_1, \ldots, M_p)^\top$ and K denote the trend vector and covariance of $\boldsymbol{\xi}$ respectively, and $\Sigma(\mathbf{x}) := K(\mathbf{x}, \mathbf{x})$ denotes the variance at a fixed point $\mathbf{x}$. Also, the DoE is defined by $\chi_n := (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$ and $\boldsymbol{\mathcal{E}_n}$ is the event $\boldsymbol{\xi}(\chi_n) = \boldsymbol{g}(\chi_n)$.

Let $\boldsymbol{T} := (T_1, \ldots, T_p)^\top$ be a fixed threshold vector. For i integer between 1 and p, the partial excursion sets are defined by

$$\Gamma_i^\star := \{\mathbf{x} \in \mathbb{X},\, g_i(\mathbf{x}) \leq T_i\}, \tag{2.43}$$

and the global excursion set $\boldsymbol{\Gamma^\star}$ is defined as the intersection of all partial excursion sets:

$$\boldsymbol{\Gamma^\star} := \{\mathbf{x} \in \mathbb{X},\, \boldsymbol{g}(\mathbf{x}) \leq \boldsymbol{T}\} = \bigcap_{i=1}^{p} \Gamma_i^\star. \tag{2.44}$$

In a similar way to the scalar framework (Section 2.2.2), we can also define random set

$$\boldsymbol{\Gamma} := \{\mathbf{x} \in \mathbb{X},\, \boldsymbol{\xi}(\mathbf{x}) \leq \boldsymbol{T}\}, \tag{2.45}$$

and random variable $\boldsymbol{\gamma}$ as the volume of $\boldsymbol{\Gamma}$:

$$\boldsymbol{\gamma} := \mathbb{P}_{\mathbb{X}}(\boldsymbol{\Gamma}) = \mathbb{P}_{\mathbb{X}}\Big(\mathbf{x} \in \mathbb{X},\, \boldsymbol{\xi}(\mathbf{x}) \leq \boldsymbol{T}\Big) = \int_{\mathbb{X}} \mathbf{1}_{\{\boldsymbol{\xi}(\mathbf{z}) \leq \boldsymbol{T}\}}\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \tag{2.46}$$

where $\mathbb{P}_{\mathbb{X}}$ denotes Lebesgue measure on $\mathbb{X}$. The coverage probability of $\boldsymbol{\Gamma}$ conditioned on $\boldsymbol{\mathcal{E}_n}$ is defined by

$$\boldsymbol{p}_n(\mathbf{x}) := \mathbb{P}(\mathbf{x} \in \boldsymbol{\Gamma} \,|\, \boldsymbol{\mathcal{E}_n}) = \mathbb{P}(\boldsymbol{\xi}(\mathbf{x}) \leq \boldsymbol{T} \,|\, \boldsymbol{\mathcal{E}_n}) = \phi_p(\boldsymbol{T}, M_n(\mathbf{x}), \Sigma_n(\mathbf{x})), \tag{2.47}$$

where $\phi_p(\,.\,, M_n(\mathbf{x}), \Sigma_n(\mathbf{x}))$ is the distribution function of a multivariate normal distribution with mean $M_n(\mathbf{x})$ and variance $\Sigma_n(\mathbf{x})$. Using Fubini-Tonelli Theorem, it can be readily shown that

$$\mathbb{E}[\boldsymbol{\gamma} |\, \mathcal{E}_n] = \int_{\mathbb{X}} \boldsymbol{p}_n(\mathbf{z})\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \tag{2.48}$$

which gives an estimator of the volume $\boldsymbol{\gamma^\star} := \mathbb{P}_{\mathbb{X}}(\boldsymbol{\Gamma^\star})$.

The SUR strategy of Excursion measure variance is defined analogously to the scalar case with Equation (2.23) and the residual uncertainty

$$\mathcal{H}_n := \mathrm{Var}[\boldsymbol{\gamma} |\, \boldsymbol{\mathcal{E}_n}] \quad \text{and} \quad \mathcal{H}_{n+1}(\mathbf{x}) := \mathrm{Var}[\boldsymbol{\gamma} |\, \boldsymbol{\xi}(\mathbf{x}),\, \boldsymbol{\mathcal{E}_n}]. \tag{2.49}$$

The SUR strategy of the Integrated Bernoulli variance is also defined in a similar way to the scalar case with Equation (2.23) and the residual uncertainty

$$\mathcal{H}_n := \int_{\mathbb{X}} \boldsymbol{p_n}(\mathbf{z})(1 - \boldsymbol{p_n}(\mathbf{z}))\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}) \quad \text{and} \quad \mathcal{H}_{n+1} := \int_{\mathbb{X}} \boldsymbol{p_{n+1}}(\mathbf{z})(1 - \boldsymbol{p_{n+1}}(\mathbf{z}))\, \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \tag{2.50}$$

with

$$\boldsymbol{p_{n+1}}(\mathbf{z}) := \mathbb{P}(\mathbf{z} \in \boldsymbol{\Gamma} \,|\, \boldsymbol{\xi}(\mathbf{x}),\, \boldsymbol{\mathcal{E}_n}). \tag{2.51}$$

Explicit formulations of these two criteria have also been proposed in [Fossum et al., 2021]. Similarly to the scalar case, for the SUR Excursion measure variance strategy, the explicit formulation avoids the use of conditional Gaussian random field simulations, while for the SUR Integrated Bernoulli variance strategy, the explicit formulation avoids the use of $\boldsymbol{\xi}(\mathbf{x})$ simulations to estimate M_{n+1} and deduce $\boldsymbol{p_{n+1}}$.

Summary:

The aim of this chapter is to introduce the background of excursion set estimation and explore various existing enrichment criteria tailored to this context within the framework of constructing sequential design of experiments using Gaussian processes. Among the presented criteria are the Deviation Number criterion, the range of the three criteria tmse, tMMSE, and tIMSE, as well as Bichon and Ranjan criteria. Also, the Stepwise Uncertainty Reduction (SUR) class of criteria stands out from the others. These criteria consist in anticipating the impact of adding the next point to the sequential experimental design, in order to select the points that lead to minimization of the expectation of a conditionnal residual uncertainty. Several examples of SUR strategies are presented, with an extension adapted to vector-valued model framework.

The aim of the next chapter is to develop a SUR version of Bichon criterion. The underlying idea is to propose an easy-to-implement SUR strategy that combines the robustness of Bichon criterion (due to its exploratory nature) with the recognized performance of SUR strategies (in terms of exploitation). Chapter 3 thus presents the construction of the criterion, its simplification into an explicit formulation, and robustness and performance tests on 2D and 6D analytical examples.

Chapter 3

A SUR version of the Bichon criterion

Outlines

In the context of excursion set estimation via GPR, we propose in this chapter a new SUR criterion for DoE enrichment, based on the Bichon criterion ([Bichon et al., 2008]). The aim is to propose a more efficient version of the Bichon criterion, and to compare it with classical criteria such as the Vorob'ev SUR criterion ([Chevalier, 2013]). We will see that numerical simulations on classical test functions highlight the good exploratory behavior of the new criterion on the different zones of the design space, as well as a certain robustness with regard to the stationarity assumption made with the use of GPR. The results of this chapter have given rise to a published article referenced as [Duhamel et al., 2023]. As this chapter is a complete article, some sections present similarities with what was presented in chapters 1 and 2.

Contents

3.1 Introduction **44**

3.2 The framework for estimating excursion sets **45**

3.2.1 Some reminders on Gaussian process regression 45

3.2.2 Towards more exploration: the Bichon criterion 47

3.3 SUR Bichon criterion **49**

3.3.1 Reminders on SUR strategies 49

3.3.2 Formulation of the SUR Bichon criterion 50

3.4 Numerical experiments **51**

3.4.1 Implementation choices 51

3.4.2 Performance tests on Branin-rescaled 2D function 52

3.4.3 Performance tests on Hartmann-rescaled 6D function 55

3.5 Conclusion **57**

3.6 Acknowledgements **58**

3.7 Appendices **58**

A Basics on Vorob'ev Theory and corresponding SUR strategy 58

B Kriging update formulas 60

C Proof of the explicit formula for SUR Bichon 60

D Robustness of estimators with respect to the GP stationarity assumption. 62

3.8 Additions to the article **63**

A Influence of the κ parameter 63
B Use of Importance Sampling for evaluating the SUR Bichon criterion 63

Title: A SUR version of the Bichon criterion for excursion set estimation

Journal: Statistics and Computing

Link: https://link.springer.com/article/10.1007/s11222-023-10208-4

Authors: Clément Duhamel[§,†], Céline Helbert[¶], Miguel Munoz Zuniga[†], Clémentine Prieur[§], Delphine Sinoquet[†]

§ Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP[⋆], LJK, 38000 Grenoble, France
[⋆] Institute of Engineering Univ. Grenoble Alpes

¶ Univ. Lyon, Ecole Centrale de Lyon, CNRS UMR 5208, Institut Camille Jordan, 69130 Ecully, France

† IFP Énergies Nouvelles, 92852 Rueil-Malmaison, France

3.1 Introduction

Nowadays, many industrial issues are related to a problem of excursion set estimation for instance, to find feasible solutions of complex optimal design problems. This problem consists in finding, the set of input parameter values such that a quantity of interest defined from its outputs respects a constraint, for example remains below a threshold. In general, the quantity of interest is an output of a numerical model, computationally expensive, which is often a black box function, representing the complex physical phenomenon. The problem is also known as an inversion problem ([Chevalier, 2013]). For example, the application to a vehicle pollution control system, allowing compliance with pollutant emission norms was studied in [El Amri et al., 2020].

An effective way to solve this kind of problems is to replace the costly black box function of interest by a surrogate model based on Gaussian processes. The advantage of a Gaussian process is that it is entirely determined by its two first moments: mean and covariance functions. Also, the formulas for updating mean and covariance functions conditionally on observations are easily tractable. The set of evaluation points and the corresponding evaluations of the black box function is called Design of Experiments (DoE) and the choice of new evaluation points is made sequentially by the optimization of an acquisition criterion that depends on the Gaussian process (see for example [Bect et al., 2012] and [Moustapha et al., 2021]). Acquisition criteria are useful to select the runs which provide the best information considering a given objective: improvement of the predictive quality of the whole response surface, optimizing a quantity of interest, quantification of a failure probability, estimating an excursion set (inversion), etc. Acquisition criteria suitable for inversion include: the deviation number denoted U ([Echard et al., 2011]), the Bichon criterion also known as Expected Feasibility Function ([Bichon et al., 2008]), and the Ranjan criterion ([Ranjan et al., 2008]). The two last criteria are adaptations of the classical optimization-oriented Expected Improvement criterion ([Jones et al., 1998]) for excursion set estimation. The U criterion is the ratio of the absolute deviation of the prediction mean from the threshold defining the excursion set, to the value of the prediction standard deviation. All these criteria are based on an exploration targeted to a better knowledge of the boundary of the excursion set.

In addition, there is a more elaborate and in general more efficient class of criteria that anticipate the impact of adding new points to the DoE: the Stepwise Uncertainty Reduction (SUR) strategies ([Bect et al., 2012]). For example, SUR strategies based on the volume of

the excursion set can be cited as particularly suitable for the inversion framework. It is shown in [Bect et al., 2012] that those SUR criteria provide better performances compared to other criteria.

[Chevalier, 2013] introduced a SUR strategy based on Vorob'ev random set theory ([Molchanov, 2005]), that goes beyond taking into account the volume of excursion sets like other SUR strategies. However, we have noticed that SUR Vorob'ev criterion (SUR Vorob'ev) is not robust enough in the sense that it lacks exploration such that with a reasonable number of simulations it sometimes misses some of connected components of the set. Moreover, SUR Vorob'ev requires some approximations about the Vorob'ev threshold. Therefore we propose to tackle these issues by a SUR version of the Bichon criterion (SUR Bichon), which is easier to set up and more robust than SUR Vorob'ev. It should be noted that a SUR version of the U criterion could have been envisaged but this would require simulation of the observations for the estimation of the associated criterion and therefore a higher computational cost.

This article is divided into three main sections. In section 3.2, the framework of excursion set estimation is recalled. Details on the construction of the sequential DoE based on Gaussian process regression are given as well as details on the Bichon acquisition criterion. Section 3.3 is dedicated to the new SUR Bichon criterion, with some reminders on SUR strategies beforehand, and a simplified and easy-to-implement formulation of this new criterion. Numerical aspects are discussed in section 3.4 with tests of SUR Bichon performances, compared to those of SUR Vorob'ev and Bichon for several analytical examples. Appendices present technical proofs, theoretical results on kriging and bases of Vorob'ev theory.

3.2 The framework for estimating excursion sets

3.2.1 Some reminders on Gaussian process regression

Let $\mathbb{X}$ be a compact set of $\mathbb{R}^d$ ($d \in \mathbb{N}^\star$) and $g : \mathbb{X} \to \mathbb{R}$ a black box function, whose analytical expression is unknown but which can be evaluated at any point of $\mathbb{X}$ at a heavy computational cost. The objective of an excursion set estimation problem is to estimate the domain defined by

$$\Gamma^\star := \Big\{ \mathbf{x} \in \mathbb{X},\, g(\mathbf{x}) \leq T \Big\} \tag{3.1}$$

with T a fixed threshold, while limiting the number of costly evaluations of g.

Surrogate models, also known as meta-models, are approximations of the output of the simulator built from a sample of simulations and that are not expensive to evaluate. Therefore they can replace the original expensive simulator in a time-saving manner. Among surrogate models, Gaussian Process Regression (GPR) is very popular: g is considered as a realization of a Gaussian process (GP) ξ defined on a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. $g(\mathbf{x}) = \xi(\mathbf{x}, \omega)$ for a given ω in Ω. This type of surrogate models gives, in addition to a prediction, an associated prediction error estimate.

More precisely, the process is written as the sum of a deterministic part and a stochastic part:

$$\xi(\mathbf{x}) := m(\mathbf{x}) + \mathbf{Z}(\mathbf{x}), \quad \forall\, \mathbf{x} \in \mathbb{X} \tag{3.2}$$

with m the trend of ξ (deterministic part) and $\mathbf{Z}$ a stationary GP, of zero mean, known covariance kernel $k : \mathbb{X}^2 \to \mathbb{R}$ and in particular variance function $\sigma^2(\mathbf{x}) := k(\mathbf{x}, \mathbf{x})$ for any $\mathbf{x}$ in $\mathbb{X}$ (stochastic part). To limit the complexity of estimating the trend function m, the choice of this latter is often parametrized as a linear combination of known basis functions $(f_i)_{i=1}^l$ with coefficients $\boldsymbol{\beta} := \{\beta_i\}_{i=1}^l$ to be estimated. The choice of the covariance kernel associated with the GP $\mathbf{Z}$ is crucial since it determines the predictor regularity. Different implementation choices are detailed in Section 3.4.1.

Let us denote $g(\chi_n) := (g(\mathbf{x}^{(1)}), ..., g(\mathbf{x}^{(n)}))^\top$ the evaluations of g on an initial design of experiments $\chi_n := (\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)})$ belonging to $\mathbb{X}^n$. The random vector $\xi(\chi_n)$ then corresponds to the finite-dimensional distribution of the process $(\xi(\mathbf{x}), \mathbf{x} \in \mathbb{X})$ on χ_n and we define $\mathcal{E}_n$ as the event $\xi(\chi_n) = g(\chi_n)$. $K := (k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}))_{1 \leq i,j \leq n}$ is the covariance matrix on χ_n and $\mathbf{k}(\mathbf{x})$ the covariance vector between $\mathbf{x}$ and χ_n defined by $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x}, \mathbf{x}^{(1)}), ..., k(\mathbf{x}, \mathbf{x}^{(n)}))^\top$ for any $\mathbf{x}$ in $\mathbb{X}$. We denote $\mathbf{f}(\mathbf{x}) := (f_1(\mathbf{x}), ..., f_l(\mathbf{x}))^\top$ the evaluation vector of $\mathbf{f}$ on $\mathbf{x}$ defining the trend and $\mathbb{F} \in \mathbb{R}^{n \times l}$ the matrix with $\mathbf{f}(\mathbf{x}^{(i)})^\top$ as i^{th} row. When $\boldsymbol{\beta}$ is known, the process ξ conditioned on the event $\mathcal{E}_n$ is still Gaussian ([O'Hagan, 1978]) with mean, variance and covariance respectively denoted m_n, σ_n^2, and k_n given by

$$m_n(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + \mathbf{k}(\mathbf{x})^\top K^{-1}(g(\chi_n) - \mathbb{F}\boldsymbol{\beta}), \tag{3.3}$$

$$\sigma_n^2(\mathbf{x}) = \sigma^2(\mathbf{x}) - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}), \tag{3.4}$$

$$k_n(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}'). \tag{3.5}$$

We notice that the best linear unbiased predictor (BLUP) (with respect to mean quadratic error) is given by (3.3) with variance (3.4) and covariance (3.5).

When $\boldsymbol{\beta}$ is unknown and estimated by the maximum likelihood estimator (MLE)

$$\widehat{\boldsymbol{\beta}} := \left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\mathbb{F}^\top K^{-1} g(\chi_n), \tag{3.6}$$

formulas (3.3), (3.4) and (3.5) become

$$m_n(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \widehat{\boldsymbol{\beta}} + \mathbf{k}(\mathbf{x})^\top K^{-1}(g(\chi_n) - \mathbb{F}\widehat{\boldsymbol{\beta}}), \tag{3.7}$$

$$\begin{aligned}\sigma_n^2(\mathbf{x}) = \sigma^2(\mathbf{x}) - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}) + \\ \left(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\right)\left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\left(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\right)^\top,\end{aligned} \tag{3.8}$$

$$\begin{aligned}k_n(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}') + \\ \left(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\right)\left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\left(\mathbf{f}(\mathbf{x}')^\top - \mathbf{k}(\mathbf{x}')^\top K^{-1}\mathbb{F}\right)^\top.\end{aligned} \tag{3.9}$$

Their interpretation as conditional expectation, covariance and variance is still possible in a Bayesian context with a non informative prior distribution on the parameter $\boldsymbol{\beta}$ ([Helbert et al., 2009]).

In order to save costly evaluations of function g, a sequential strategy of enrichment of the DoE is classically used (see e.g. [Ginsbourger, 2017]). Figure 3.1 provides a generic scheme of a sequential strategy. The stopping criteria can be a budget of simulations or a threshold on the remaining uncertainty on the estimation of the excursion set.

Among enrichment criteria, one can distinguish criteria that lead to an overall improvement of the model from goal-oriented criteria, which are adapted to particular frameworks such as optimization or inversion. The classical Mean Squared Error (MSE) criterion ([Jin et al., 2002]), aims to select the point which has the highest prediction variance, as well as its integral versions IMSE and MMSE ([Picheny et al., 2010]) standing for Maximum Mean Squared Error. Among goal oriented criteria, classical Expected Improvement (EI) from [Jones et al., 1998], allows for global optimization. Several inversion-adapted criteria can be cited: deviation number ([Echard et al., 2011]), ratio of the distance of the prediction mean to the threshold to the kriging standard deviation; Bichon criterion also known as Expected Feasibility Function (EFF) ([Bichon et al., 2008]), and Ranjan criterion ([Ranjan et al., 2008]).

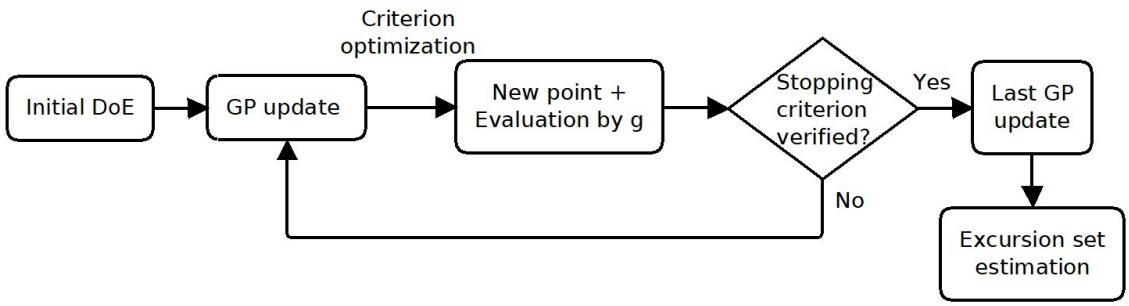


Figure 3.1: Diagram of the sequential DoE construction, coupled with GPR.

3.2.2 Towards more exploration: the Bichon criterion

The Bichon criterion (Bichon), originally presented in [Bichon et al., 2008] is a goal-oriented criterion for the DoE enrichment. This criterion is an adaptation of EI from [Jones et al., 1998], introduced in the context of global optimization, to the inversion framework.

As a reminder, the original idea of EI is to select a point $\mathbf{x}$ that allows an improvement of $g_{\min} - \xi(\mathbf{x})$ with $g_{\min}$ the current minimum observed on the DoE, while taking into account the uncertainty of the surrogate model. The idea behind Bichon is to adapt this strategy to the excursion set estimation framework by considering both the variability of the surrogate model and the potential improvement in the knowledge of the excursion set boundary. This is the exploration-exploitation compromise.

To introduce Bichon, it is necessary to define, for a fixed $\mathbf{x}$ belonging to $\mathbb{X}$, a random variable $\mathrm{FF}(\mathbf{x})$ (Feasibility Function) defined by

$$\begin{aligned} \mathrm{FF}(\mathbf{x}) : &= \epsilon(\mathbf{x}) - \min\left\{|T - \xi(\mathbf{x})|, \epsilon(\mathbf{x})\right\} \\ &= \left(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\right)^+, \end{aligned} \tag{3.10}$$

with $(\,.\,)^+ := \max(\,.\,, 0)$. This function represents the distance of the surrogate model to the bounds of the interval $[T - \varepsilon(\mathbf{x}), T + \varepsilon(\mathbf{x})]$ only if the surrogate model belongs to this interval and is 0 otherwise. In practice, the interval width $\epsilon(\mathbf{x})$ is chosen proportional to the kriging standard deviation $\sigma_n(\mathbf{x})$, leading in particular to a null value of the criterion for the points already present in the DoE. An example is given in Figure 3.2. The feasibility function is drawn for one sample path of $\xi|\,\mathcal{E}_n$. Its maximization aims to select points close to the boundary of the excursion set estimate or points associated to high values of $\epsilon(\mathbf{x})$. Then, the average of FF over all sample paths, gives the Expected Feasibility Function (EFF)

$$\mathrm{EFF}(\mathbf{x}) := \mathbb{E}\Big[\left(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\right)^+ \Big|\, \mathcal{E}_n\Big]. \tag{3.11}$$

The new selected points according Bichon are

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}}\ \mathrm{EFF}(\mathbf{x}). \tag{3.12}$$

To interpret (3.11), it is possible to make a heuristic analogy with the theory of statistical tests ([Dagnelie, 1992]). Let $\mathbf{x}$ be fixed, suppose that $\xi(\mathbf{x})|\,\mathcal{E}_n \sim \mathcal{N}(m_n(\mathbf{x}), \sigma_n(\mathbf{x}))$ with $m_n(\mathbf{x})$ unknown and $\sigma_n(\mathbf{x}) > 0$ known and let us define the following statistical test

$$\mathrm{H}_0 : m_n(\mathbf{x}) = T \quad \text{against} \quad \mathrm{H}_1 : m_n(\mathbf{x}) \neq T \tag{3.13}$$

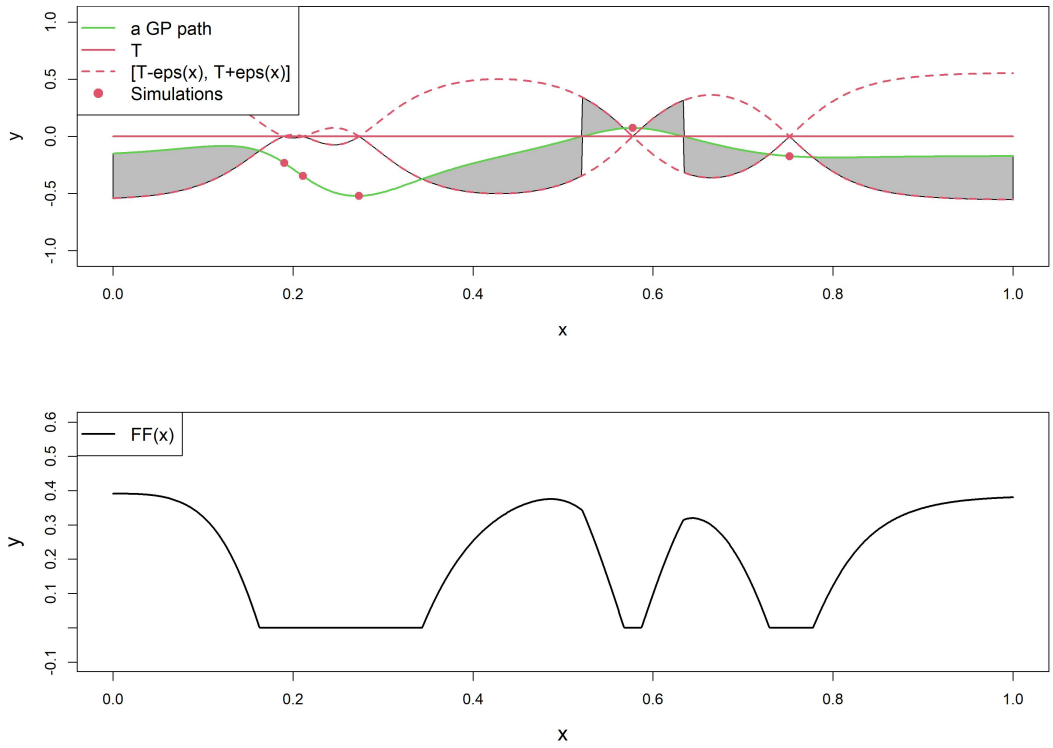


Figure 3.2: Representation of Feasibility Function (bottom) for a given example of a GP sample path conditioned on 5 evaluations of the g function, a threshold T set to 0 and $\epsilon(x) := \sigma_n(\mathbf{x})$ (top).

We choose $\upsilon_{\mathbf{x}} := \frac{\xi(\mathbf{x})-T}{\sigma_n(\mathbf{x})}\Big| \mathcal{E}_n$ as the test statistic which follows standard normal distribution under H_0. Consequently, if we want to refute Hypothesis H_0 at order α it is necessary that

$$|\upsilon_{\mathbf{x}}| > \kappa \tag{3.14}$$

with $\kappa := q_{1-\frac{\alpha}{2}}$ the quantile of order $1-\frac{\alpha}{2}$ of the standard normal distribution.

However, what we wish to do is not to refute the hypothesis that $m_n(\mathbf{x}) = T$ but rather to select, among the $\mathbf{x}$ for which the hypothesis H_0 is plausible i.e. $\kappa - |\upsilon_{\mathbf{x}}| > 0$, the $\mathbf{x}$ for which the quantity $\kappa - |\upsilon_{\mathbf{x}}|$ is the largest in average. Multiplying $(\kappa - |\upsilon_{\mathbf{x}}|)^+$ by $\sigma_n(\mathbf{x})$ leads to Bichon criterion with $\epsilon(\mathbf{x}) := \kappa\sigma_n(\mathbf{x})$, and has the effect to increase the exploration ability of the criterion.

Finally, an explicit formulation of Bichon (see [Bect et al., 2012] for a proof) can be calculated based on the posterior kriging trend and variance, the threshold T and the width $\varepsilon(\mathbf{x})$

$$\begin{aligned}\mathrm{EFF}(\mathbf{x}) = (m_n(\mathbf{x}) - T)\left[2\,\phi\Big(\frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \phi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right] \\ - \sigma_n(\mathbf{x})\left[2\,\varphi\Big(\frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \varphi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \varphi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right] \\ + \epsilon(\mathbf{x})\left[\phi\Big(\frac{T^+ - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\Big)\right],\end{aligned} \tag{3.15}$$

with $T^{\pm} := T \pm \epsilon(\mathbf{x})$, φ and ϕ represent respectively the probability density and cumulative distribution functions of the standard normal distribution. In practice, the enrichment of the DoE is done by maximizing the criterion given by Equation (3.15). The main objective of this work is to propose a SUR version of this goal-oriented criterion, in order to obtain a SUR method that is simpler to set up and more robust than SUR Vorob'ev.

3.3 SUR Bichon criterion

This section focuses on an adaptation of the Bichon criterion to a SUR strategy. The new SUR criterion we introduce can be implemented in the framework of GPR without any approximation unlike SUR Vorob'ev (see Appendix A and [Chevalier, 2013]). Moreover, SUR Bichon performs better than Bichon criterion and corrects the lack of robustness observed when applying SUR Vorob'ev, at least on test cases studied in Section 3.4.

3.3.1 Reminders on SUR strategies

SUR strategies aim at maximizing the mean uncertainty reduction induced by new evaluations. Let us introduce a residual uncertainty $\mathcal{H}_n$, computed with the GP model conditioned on $\mathcal{E}_n$. From this residual uncertainty, the conditional residual uncertainty $\mathcal{H}_{n+1}(\mathbf{x})$ is defined as the updated uncertainty when adding $\mathbf{x}$ to DoE χ_n. This conditional uncertainty is then a measurable function of the random variable $\xi(\mathbf{x})|\,\mathcal{E}_n$. The associated SUR strategy is then defined by

$$\mathbf{x}^{(n+1)} \in \operatorname*{argmin}_{\mathbf{x}\in\mathbb{X}} \mathcal{J}_n(\mathbf{x}) \quad \text{with} \quad \mathcal{J}_n(\mathbf{x}) := \mathbb{E}\big[\mathscr{H}_{n+1}(\mathbf{x})\big]. \tag{3.16}$$

Note that the expectation in $\mathcal{J}_n(\mathbf{x})$ is relative to the distribution of $\xi(\mathbf{x})|\,\mathcal{E}_n$. Eq. (3.16) means that evaluating the surrogate model at $\mathbf{x}^{(n+1)}$ will decrease at most, the expected residual uncertainty.

More details on SUR strategies and their origin from k-step lookahead strategies can be found in [Bect et al., 2012]. Among classical SUR strategies based on GPR, we can quote for example different criteria using the excursion set volume, presented in [Bect et al., 2012]. A more complex criterion requiring notions about the random set theory of Vorob'ev ([Molchanov, 2005]) introduced in [Chevalier, 2013], can also be cited. We refer to Appendix A for more details on Vorob'ev theory and the associated SUR Vorob'ev strategy.

SUR strategies are numerically more complex to implement, but are generally more efficient than other goal oriented strategies ([Bect et al., 2012] and [Chevalier, 2013]), for the same number of evaluations. A reduction in the numerical complexity of SUR strategies is frequently used through the use of kriging update formulas introduced in [Chevalier, 2013] (see Appendix B).

3.3.2 Formulation of the SUR Bichon criterion

Let $\mathbb{P}_{\mathbb{X}}$ a probability measure on $\mathbb{X}$. Following the formalism of SUR strategies given in section 3.3.1, we first define a residual uncertainty $\mathcal{H}_n$ by

$$\begin{aligned}\mathcal{H}_n &:= \int_{\mathbb{X}} \mathrm{EFF}(\mathbf{z})\,\mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}) \\ &= \int_{\mathbb{X}} \mathbb{E}\Big[\big(\kappa\sigma_n(\mathbf{z}) - |T-\xi(\mathbf{z})|\big)^+ \Big|\, \mathcal{E}_n\Big]\,\mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \quad \text{with} \quad \kappa > 0.\end{aligned} \tag{3.17}$$

The corresponding conditional residual uncertainty is then defined by

$$\mathcal{H}_{n+1}(\mathbf{x}) := \int_{\mathbb{X}} \mathbb{E}\Big[\big(\kappa\sigma_{n+1}(\mathbf{z}) - |T-\xi(\mathbf{z})|\big)^+ \Big|\, \xi(\mathbf{x}),\, \mathcal{E}_n\Big]\,\mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \quad \text{with} \quad \kappa > 0, \tag{3.18}$$

with $\sigma_{n+1}(\mathbf{z})$ the kriging standard deviation computed from Equation (3.34) in Appendix B and $\mathbf{x}$ being the $n+1^{\text{th}}$ observation point. In practice, the chosen probability measure $\mathbb{P}_{\mathbb{X}}$ is the Lebesgue measure restricted and normalized on $\mathbb{X}$ and the coefficient κ is usually set to 1 like in Bichon ([Bect et al., 2012]). The residual uncertainty $\mathcal{H}_n$ represents the average (with respect to $\mathbb{P}_{\mathbb{X}}$) of mean distances (in positive values) of $\xi(\mathbf{z})$ to the bounds of the interval $[T-\kappa\sigma_n(\mathbf{z}), T+\kappa\sigma_n(\mathbf{z})]$, conditioned on $\mathcal{E}_n$. An overall reduction in σ_n leads to a decrease in the uncertainty $\mathcal{H}_n$. The same applies to the addition of a new point close to the boundary defined by the threshold T.

The problem is to find

$$\mathbf{x}^{(n+1)} \in \operatorname*{argmin}_{\mathbf{x}\in\mathbb{X}} \mathcal{J}_n(\mathbf{x}) \quad \text{with} \quad \mathcal{J}_n(\mathbf{x}) := \mathbb{E}\bigg[\int_{\mathbb{X}} \mathbb{E}\Big[\big(\kappa\sigma_{n+1}(\mathbf{z}) - |T-\xi(\mathbf{z})|\big)^+ \Big|\, \xi(\mathbf{x}),\, \mathcal{E}_n\Big]\,\mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z})\bigg], \tag{3.19}$$

with $\kappa > 0$. The first expectation is relative to $\xi(\mathbf{x})|\,\mathcal{E}_n$ and the second one is relative to $\xi(\mathbf{z})$ knowing $\xi(\mathbf{x})$, $\mathcal{E}_n$.

Lemma 1.

$$\mathcal{J}_n(\mathbf{x}) = \int_{\mathbb{X}} \mathbb{E}\Big[\big(\kappa\sigma_{n+1}(\mathbf{z}) - |T-\xi(\mathbf{z})|\big)^+ \,\Big|\, \mathcal{E}_n\Big]\,\mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}). \tag{3.20}$$

Proof. The integrand of the chosen residual uncertainty in (3.18) is a positive quantity, by positivity of the expectation. So, by re-injecting the expression of uncertainty (3.18) into (3.16), then applying Fubini–Tonnelli theorem (thanks to σ-finite measures and positive integrand), we obtain

$$\mathcal{J}_n(\mathbf{x}) = \int_{\mathbb{X}} \mathbb{E}\bigg[\mathbb{E}\Big[\big(\kappa\sigma_{n+1}(\mathbf{z}) - |T-\xi(\mathbf{z})|\big)^+ \Big|\, \xi(\mathbf{x}),\, \mathcal{E}_n\Big]\bigg]\,\mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbf{z}), \tag{3.21}$$

Then, in (3.21) the two expectations are reduced in one to obtain (3.20). □

Finally, Proposition 2 below provides an explicit formula for the integrand of (3.20).

Proposition 2. For all $\mathbf{x}$, $\mathbf{z}$ belonging to $\mathbb{X}^2$, we have

$$\begin{aligned}&\mathbb{E}\Big[\big(\kappa\sigma_{n+1}(\mathbf{z}) - |T-\xi(\mathbf{z})|\big)^+ \,\Big|\, \mathcal{E}_n\Big] \\ &\quad= (m_n(\mathbf{z})-T)\bigg[2\,\phi\Big(\frac{T-m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\bigg] \\ &\qquad- \sigma_n(\mathbf{z})\bigg[2\,\varphi\Big(\frac{T-m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \varphi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \varphi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\bigg] \\ &\qquad+ \epsilon_{\mathbf{x}}(\mathbf{z})\bigg[\phi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\bigg]\end{aligned} \tag{3.22}$$

where $\epsilon_{\mathbf{x}}(\mathbf{z}) := \kappa\sigma_{n+1}(\mathbf{z})$, $T^{\pm} := T \pm \epsilon_{\mathbf{x}}(\mathbf{z})$, φ and ϕ denote the probability density and cumulative distribution functions of the standard normal distribution, respectively.

The dependency in $\mathbf{x}$ in Equation (3.22) is only given via $\epsilon_{\mathbf{x}}(\mathbf{z}) = \kappa\sigma_{n+1}(\mathbf{z})$, therefore only via $\sigma_{n+1}(\mathbf{z})$, which is independent of the model evaluation on $\mathbf{x}$ according to the kriging formulas. In practice, kriging update formulas (see Equation (3.34) Appendix B) will be used to get a fast evaluation of $\epsilon_{\mathbf{x}}(\mathbf{z})$. The proof of Proposition 2 is postponed to Appendix C.

3.4 Numerical experiments

The performances of SUR Bichon are illustrated on two analytical examples, and compared to SUR Vorob'ev and standard Bichon performances. The chosen test functions are the rescaled Branin function in dimension 2 and rescaled Hartmann function in dimension 6 ([Picheny et al., 2013]). The choice of the threshold T for each of these functions is discussed later.

Several DoE enrichment strategies can be considered: an enrichment by Bichon $\chi_{0,n}$, one using SUR Bichon $\chi_{1,n}$ and one using SUR Vorob'ev $\chi_{2,n}$, all three after n iterations. The criteria performance is evaluated through two different estimators: naive estimator $\hat{\Gamma}_1 := m_n^{-1}(]-\infty, T])$ and Vorob'ev estimator noted $\hat{\Gamma}_2$, which corresponds to the Vorob'ev expectation (Appendix A). The performances of the different criteria are then compared after n iterations with the approximation error $\mathrm{Err}(\hat{\Gamma}_i(\chi_{j,n})) := \mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_i(\chi_{j,n})\Delta\Gamma^\star)/\mathbb{P}_{\mathbb{X}}(\Gamma^\star)$, for $(i,j) \in \{1,2\}^2$. This error measures the relative volume of the symmetric difference between estimator $\hat{\Gamma}_i(\chi_{j,n})$ and true excursion set $\Gamma^\star$ defined in (3.1).

3.4.1 Implementation choices

As mentioned earlier, the choices of trend m and covariance kernel k are fundamental. In the following, the trend is chosen as a single constant term m, see [Roustant et al., 2012] for more details. A classical kernel product of type Matérn 5/2 is chosen:

$$k(\mathbf{x},\mathbf{x}') := \mathrm{Cov}(\mathbf{Z}(\mathbf{x}),\mathbf{Z}(\mathbf{x}')) = \sigma_{\mathrm{c}}^2 \prod_{i=1}^{d} R_{\text{Matérn }5/2}(h_i,\theta_i), \quad \forall\,(\mathbf{x},\mathbf{x}') \in \mathbb{X}^2, \tag{3.23}$$

with a vector of parameters θ belonging to $\mathbb{R}_+^{\star d}$, estimated by maximizing the likelihood at each iteration, $h_i = |\mathbf{x}_i - \mathbf{x}'_i|$, σ_{c} a fixed parameter and

$$R_{\text{Matérn }5/2}(h_i,\theta_i) := \left(1 + \frac{\sqrt{5}|h_i|}{\theta_i} + \frac{5h_i^2}{3\theta_i^2}\right)\exp\left(-\frac{\sqrt{5}|h_i|}{\theta_i}\right). \tag{3.24}$$

This choice leads to trajectories of class C^2 ([Paciorek, 2003]).

The implementation of SUR Bichon from formulas (3.20) and (3.22), is greatly inspired by the implementation of various SUR criteria in the package *KrigInv* ([Chevalier et al., 2014b]). In addition, the chosen measure $\mathbb{P}_{\mathbb{X}}$ is the renormalized Lebesgue measure restricted to $\mathbb{X}$, which is possible because $\mathbb{X}$ is compact. In (3.20), the integration is performed using a Sobol' sequence with *n.points* integration points (package *randtoolbox* [Dutang and Savicky, 2013]). The criterion is optimized with the genetic algorithm Genoud (with $pop.size = 50d$) (package *rgenoud*, [Mebane Jr and Sekhon, 2011]). Unless explicitly stated, κ in (3.22) is set to 1, initial DoEs are obtained by LHS optimized by maximizing minimal distances between the points (Latin Hypercube Sampling, [Dupuy et al., 2015]) with size to be specified in the following, and *n.points* defined above is set to 10^4.

The volume of $\Gamma^\star$ and $\hat{\Gamma}_i(\chi_{j,n})\Delta\Gamma^\star$ are approached using a quasi-Monte Carlo methods ([Lemieux, 2009]) with a Sobol' sequence of size 10^4.

3.4.2 Performance tests on Branin-rescaled 2D function

The Branin-rescaled function, defined in [Picheny et al., 2013] on $\mathbb{X} := [0,1]^2$, is represented in Figure 3.3. The $\Gamma^\star$ excursion set is defined by the upper bound $T = 10$ on the function values, which leads to 3 disconnected areas and the volume of $\Gamma^\star$ represents 15.74% of the total volume of $\mathbb{X}$. The tests are performed on 100 different initial DoEs of size 10. 20 iterations (1 simulation per iteration) are run for SUR Bichon, SUR Vorob'ev and Bichon.

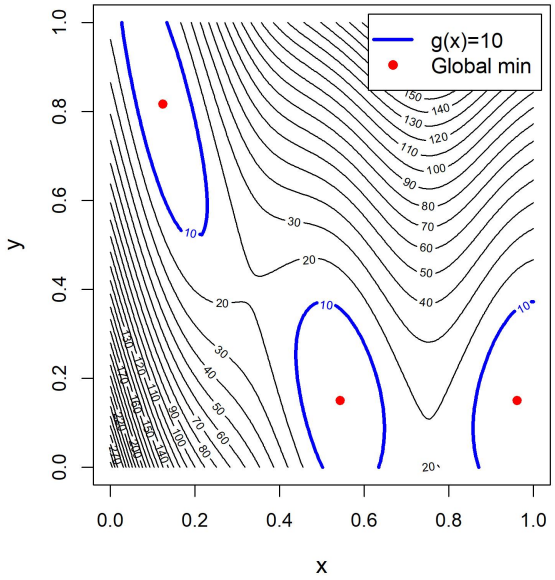
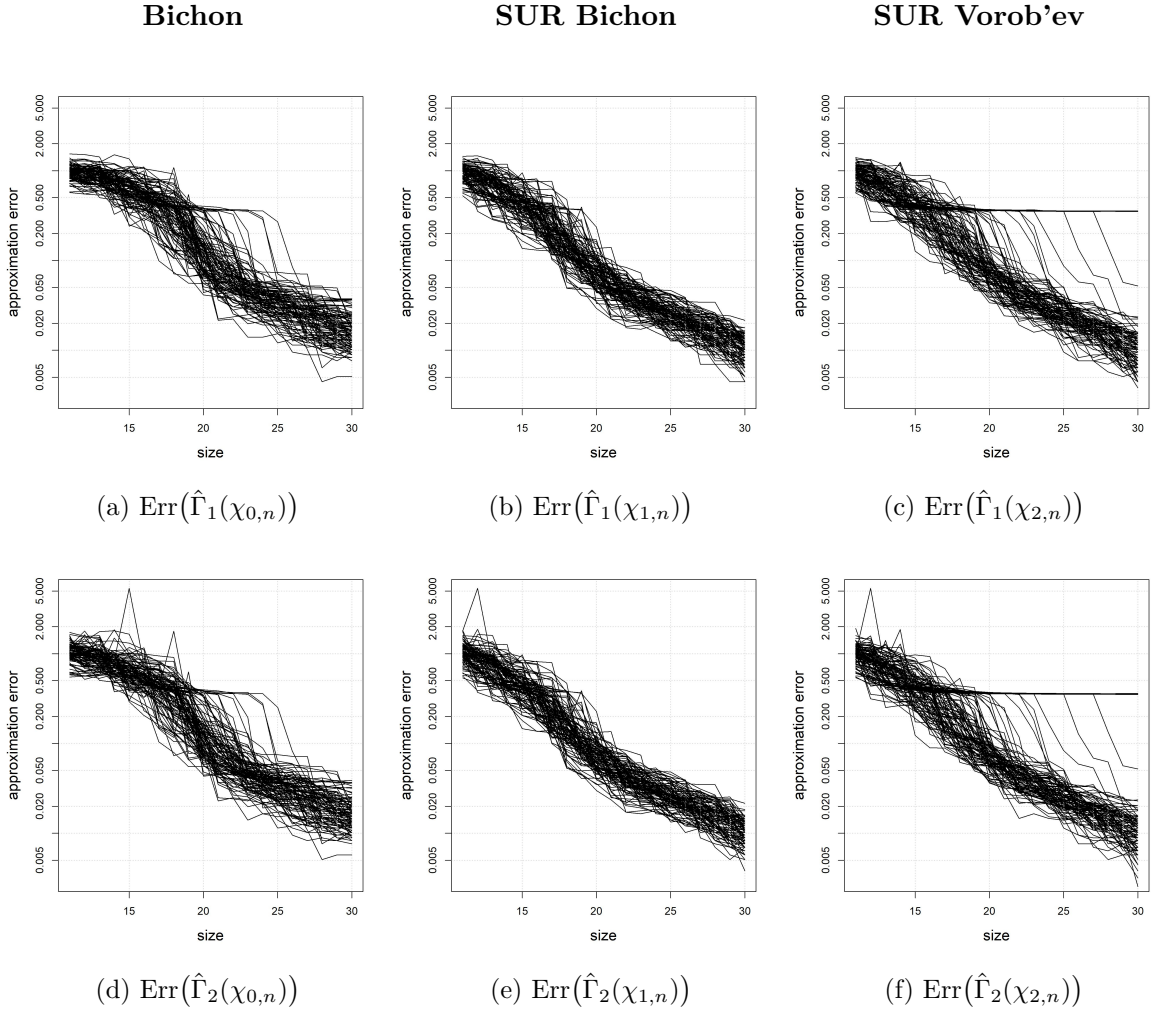


Figure 3.3: Representation of the Branin-rescaled function on $\mathbb{X}$.

Black solid lines (line plot) of Figure 3.4 represent the approximation error for each of the 100 initial DoEs of size 10. This makes it possible to visualize the enrichment performances throughout the iterations. Bichon seems to perform less well than the other two SUR criteria: the median is higher. SUR Bichon appears more robust than SUR Vorob'ev throughout the enrichment, no matter which estimator is chosen. Indeed, in the case of SUR Vorob'ev, several extreme cases present stagnation of the approximation error. These rare stagnations are due to the late discovery of one of the three components of the excursion set as illustrated below. The logarithmic scale of the graph seems to show a stagnation of the curve for these few cases, but it is in fact a faster progression towards a threshold value, due to the simplification of the problem to the two remaining components. Also, the approximation error with $\hat{\Gamma}_1$ seems more robust than with $\hat{\Gamma}_2$ (see Appendix D for more details), and this can be explained by the fact that estimators $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are respectively based on an extension of median and mean concepts to sets (Appendix A).

We focus on one run of the enrichment with SUR Vorob'ev associated with one of the outliers of Figure 3.4. The associated $\hat{\Gamma}_1$ (resp. $\hat{\Gamma}_2$) estimators are represented as green full (resp. dotted) line on Figure 3.5, after 20 iterations. This figure shows that SUR Vorob'ev misses one of the three areas of the exact excursion set $\Gamma^\star$ presented in Figure 3.3. The use of SUR Bichon allows a better exploration of the design space, which here allows to detect



(a) $\mathrm{Err}\big(\hat{\Gamma}_1(\chi_{0,n})\big)$ (b) $\mathrm{Err}\big(\hat{\Gamma}_1(\chi_{1,n})\big)$ (c) $\mathrm{Err}\big(\hat{\Gamma}_1(\chi_{2,n})\big)$

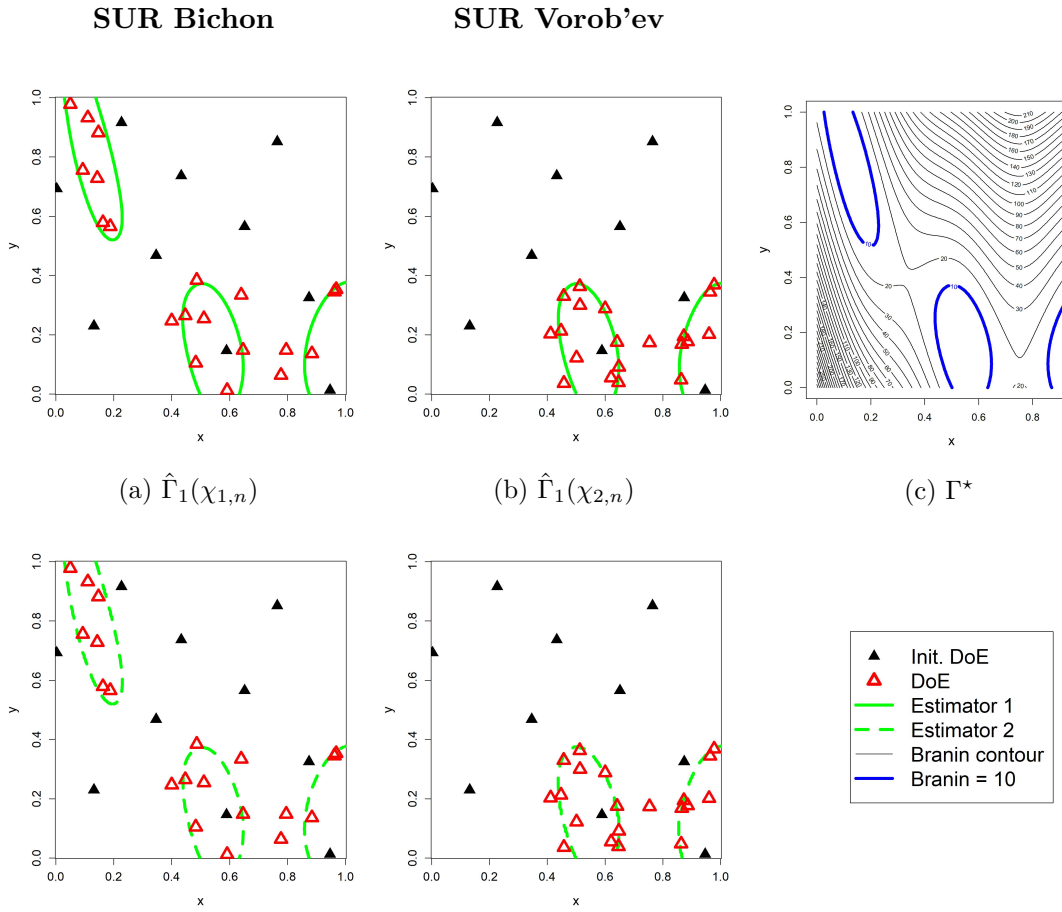
(d) $\mathrm{Err}\big(\hat{\Gamma}_2(\chi_{0,n})\big)$ (e) $\mathrm{Err}\big(\hat{\Gamma}_2(\chi_{1,n})\big)$ (f) $\mathrm{Err}\big(\hat{\Gamma}_2(\chi_{2,n})\big)$

Figure 3.4: Line plots (with logarithmic scale) of the approximation error $\mathrm{Err}\big(\hat{\Gamma}_i(\chi_{j,n})\big)$ for the different criteria during 20 iterations, for Branin-rescaled function inversion ($d = 2$) with $T = 10$, for 100 different initial DoEs of size 10 of type LHS Maximin, for $\kappa = 1$ and with n.points= 10^4. Left column: Bichon with naive estimator (a) and Vorob'ev estimator (d). Middle column: SUR Bichon with naive estimator (b) and Vorob'ev estimator (e). Right column: SUR Vorob'ev with naive estimator (c) and Vorob'ev estimator (f).

the three areas of $\hat{\Gamma}^\star$. We notice that there is very few differences between the two types of estimators and given the robustness of naive estimator $\hat{\Gamma}_1$ compared to $\hat{\Gamma}_2$ (Appendix D), naive estimator is kept for the remaining tests in dimension 2.

The characteristic statistical values of the empirical distribution for the error approximation with naïve $\hat{\Gamma}_1$ estimator are given in Table 3.1 for SUR Bichon, SUR Vorob'ev and Bichon. This table confirms the poor performance of Bichon in relation to the two SUR criteria. In the following, Bichon is set aside to focus on the comparison of the two SUR Bichon and SUR Vorob'ev. It can be also seen that only for the quantile 5%, the results are slightly better for SUR Vorob'ev. All others results show SUR Bichon is more efficient than SUR Vorob'ev: the outliers of SUR Vorob'ev deteriorate the characteristic values especially mean or standard deviation.

To summarize, the study of the performances of SUR Bichon on Branin-rescaled function, showed that for $T = 10$ the sought excursion set with three connected components is better



(a) $\hat{\Gamma}_1(\chi_{1,n})$ (b) $\hat{\Gamma}_1(\chi_{2,n})$ (c) $\Gamma^\star$

(d) $\hat{\Gamma}_2(\chi_{1,n})$ (e) $\hat{\Gamma}_2(\chi_{2,n})$

Figure 3.5: Representation of two $\hat{\Gamma}$ estimators for each of the two criteria after 20 iterations, in comparison with the true excursion set (top right), for Branin-rescaled function inversion ($d = 2$) with $T = 10$, for a particular initial DoE of size 10 where SUR Bichon outperforms SUR Vorob'ev. Left column: SUR Bichon with naive estimator (a) and Vorob'ev estimator (d). Middle column: SUR Vorob'ev with naive estimator (b) and Vorob'ev estimator (e). Right column: true excursion set and Branin-rescaled contour lines (c).

Crit. (×100)	**ite** $= 10$			**ite** $= 20$		
	SUR B.	SUR V.	B.	SUR B.	SUR V.	B.
Mean	7.82	12.35	15.07	1.09	3.18	1.71
Median	7.24	7.78	11.94	1.08	1.08	1.59
Quantile 5%	4.36	4.24	5.53	0.57	0.51	0.89
Quantile 95%	12.94	36.42	37.87	1.59	35.13	3.25
Standard Deviation	3.21	10.79	9.97	0.34	8.17	0.73
Interquartile Range	3.80	5.38	9.24	0.52	0.64	0.89

Table 3.1: Summary of empirical distributions of the $\mathrm{Err}(\hat{\Gamma}_i(\chi_{j,n}))$ for the different criteria after 10 and 20 iterations, for Branin-rescaled function inversion ($d = 2$) with $T = 10$, for 100 different initial DoEs of size 10 of type LHS Maximin.

detected when using SUR Bichon. Indeed, the latter, unlike SUR Vorob'ev, makes it possible to avoid extreme cases for which one of the three connected components is completely missed.

3.4.3 Performance tests on Hartmann-rescaled 6D function

In this section the inversion results for the Hartmann-rescaled function ([Picheny et al., 2013]) on $\mathbb{X} := [0,1]^6$ are presented. The excursion set $\Gamma^\star$ is defined by the upper bound $T = -1.6$, with a volume that represents 15.45% of the total volume of $\mathbb{X}$. Results of clustering methods (not presented here) suggest that $\Gamma^\star$ is composed of only one large connected component. The tests are performed on 50 different initial DoEs of size 30 and 600 iterations of DoE enrichment are run for both SUR Bichon and SUR Vorob'ev.

Line plots on Figure 3.3 show the approximation error along the iterations of the enrichment. Firstly, it can be observed that the improvement of the approximation error during the enrichment is slower than in dimension 2, which is consistent with the increasing difficulty of the problem in higher dimension. Moreover, we observe that SUR Bichon performs better than SUR Vorob'ev whatever the chosen estimator, with respect to the robustness to outliers but also on average, especially from the 300^{th} iteration. In addition, it is clear from Figure 3.3 (a) and (c) that naive estimator $\hat{\Gamma}_1$ gives a more robust approximation error than $\hat{\Gamma}_2$, with SUR Bichon enrichment. We thus decide to present the numerical results in the following of the section for $\hat{\Gamma}_1$ only.

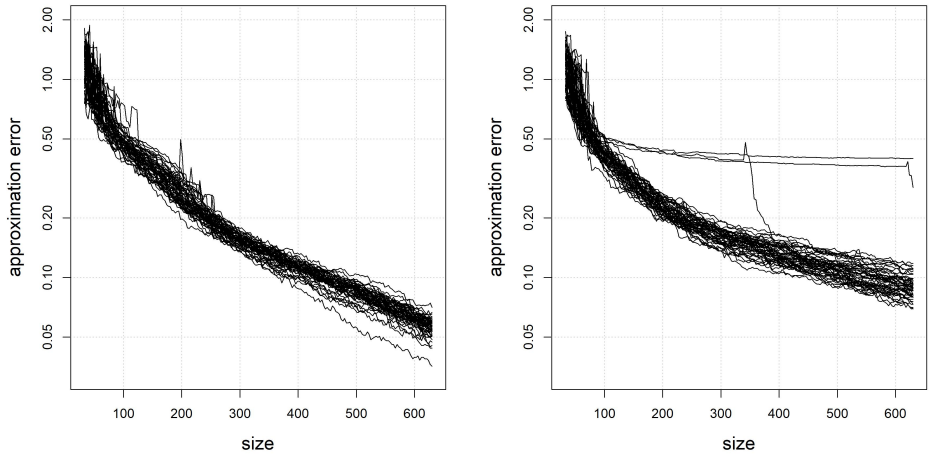
We focus on one of the outliers of Figure 3.3 (b) for which the enrichment strategy is based on SUR Vorob'ev. For an extreme case, we represent in pairwise projection the points of a Sobol' sequence of size 5.10^3 on $\mathbb{X}$ belonging to $\hat{\Gamma}_1 \Delta \Gamma^\star$, with $\hat{\Gamma}_1$ the estimator obtained after 600 iterations (see Figure 3.7). There are only 55 misclassified points observed for SUR Bichon, against 247 in the case of SUR Vorob'ev. Moreover, among the 247 points for the case SUR Vorob'ev, 191 correspond to $\hat{\Gamma} \backslash \Gamma^\star$ (unfeasible points that are predicted feasible), whereas the remaining misclassified points correspond to feasible points that were predicted unfeasible (as for 2D example). This allows to further illustrate the robustness of SUR Bichon compared to SUR Vorob'ev. A comparable study was carried out on the other extreme runs as well as on the non-extreme ones. The results showed that the configuration for the other extreme run is comparable to that in Figure 7 while the configurations for the non-extreme runs are all more or less similar and relatively balanced in terms of number of misclassified points between SUR Bichon and SUR Vorob'ev criteria.

The characteristic statistical values of the empirical distribution for the approximation error are given in Figure 3.2 for both SUR Bichon and SUR Vorob'ev. It can be confirmed that except at the beginning where enrichment is not yet sufficient, SUR Bichon performs better for all indicators than SUR Vorob'ev, and not only in terms of robustness (see e.g. the quantile of order 5% or the median).

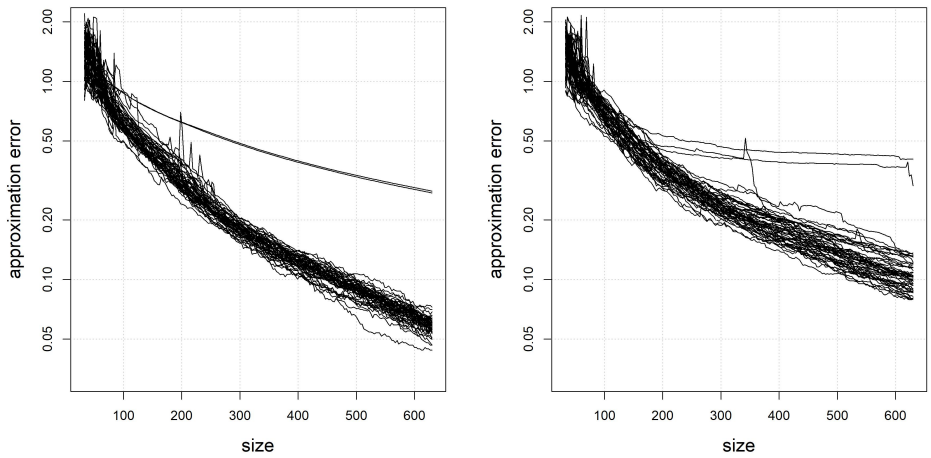
In summary, the tests on Hartmann-rescaled function in dimension 6 with $T = -1.6$, further highlight the robustness of SUR Bichon compared to SUR Vorob'ev, in dimension higher than 2. In addition to the robustness, it was also observed that in this 6 dimensional case, beyond about 300 iterations, SUR Bichon performs better than SUR Vorob'ev, even without considering the outliers.

The robustness of SUR Bichon can be explained by its exploratory capability forced by the kriging standard deviation in factor of the SUR Bichon formulation (see Section 3.2.2). Besides, the Vorob'ev expectation is strongly dependent on the stationarity assumption of the underlying Gaussian process (Appendix D), which has an influence on the enrichment, since the enrichment with SUR Vorob'ev is based on the Vorob'ev deviation (Appendix A), and then on the Vorob'ev expectation. This could explain the lack of robustness of SUR Vorob'ev. Indeed, in practice the stationarity hypothesis is never rigorously checked (unless the model

SUR Bichon **SUR Vorob'ev**



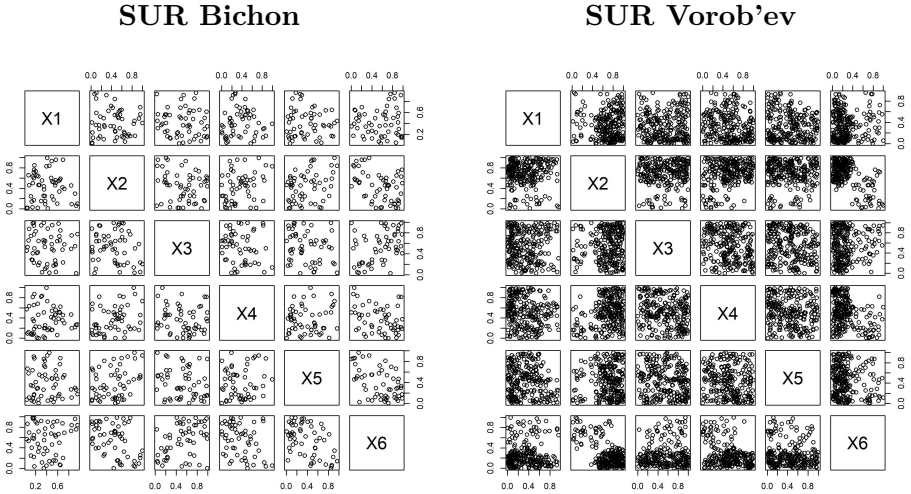
(a) $\mathrm{Err}\big(\hat{\Gamma}_1(\chi_{1,n})\big)$ (b) $\mathrm{Err}\big(\hat{\Gamma}_1(\chi_{2,n})\big)$



(c) $\mathrm{Err}\big(\hat{\Gamma}_2(\chi_{1,n})\big)$ (d) $\mathrm{Err}\big(\hat{\Gamma}_2(\chi_{2,n})\big)$

Figure 3.6: Line plots (with logarithmic scale) of the approximation error $\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_i(\chi_{j,n})\Delta\Gamma^\star)/\mathbb{P}_{\mathbb{X}}(\Gamma^\star)$ (with a Sobol' sequence of size 10^4) for the different criteria during 600 iterations, for Hartmann-rescaled function inversion ($d = 6$) with $T = -1.6$, for 50 different initial DoEs of size 30 of type LHS Maximin, for $\kappa = 1$ and with n.points= 10^4. Left column: SUR Bichon with naive estimator (a) and Vorob'ev estimator (c). Right column: SUR Vorob'ev with naive estimator (b) and Vorob'ev estimator (d).

is defined as a given realization of a stationary Gaussian process). Eventually, the calculation of the Vorob'ev criterion is sensitive to the determination of the Vorob'ev threshold $\alpha^\star$ (see Appendix A). However, we have verified that the determination of $\alpha^\star$ was not the problem for the robustness of SUR Vorob'ev by checking the numerical simplicity of the minimum search for the function $\alpha \mapsto \big|\mathbb{E}\big[\mathbb{P}_{\mathbb{X}}(\Gamma)|\mathcal{E}_n\big] - \mathbb{P}_{\mathbb{X}}(Q_\alpha)\big|$.



(a) $\hat{\Gamma}_1(\chi_{1,n})\Delta\hat{\Gamma}^\star$ (b) $\hat{\Gamma}_1(\chi_{2,n})\Delta\hat{\Gamma}^\star$

Figure 3.7: Pairwise projection plot of $\hat{\Gamma}_1\Delta\Gamma^\star$ for a 5.10^3-Sobol' sequence, for the two criteria after 600 iterations, for Hartmann-rescaled function inversion ($d = 6$) with $T = -1.6$, for a particular initial DoE of size 30 where SUR Bichon outperforms on SUR Vorob'ev, for $\kappa = 1$ and for n.points$= 10^4$. Left column: SUR Bichon. Right column: SUR Vorob'ev.

Crit. / (×100)	**ite** $= 99$		**ite** $= 300$		**ite** $= 600$	
	SUR B.	SUR V.	SUR B.	SUR V.	SUR B.	SUR V.
Mean	40.42	35.05	13.92	16.55	5.61	9.94
Median	40.32	34.72	13.92	15.28	5.66	8.83
Quantile 5%	34.14	28.84	12.09	12.74	4.62	7.17
Quantile 95%	46.67	44.40	15.38	29.42	6.45	11.63
Standard Deviation	3.98	4.48	0.99	6.14	0.61	5.26
Interquartile Range	5.26	4.56	1.13	1.81	0.68	1.75

Table 3.2: Summary of empirical distributions of the approximation error $\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}\Delta\Gamma^\star)/\mathbb{P}_{\mathbb{X}}(\Gamma^\star)$ (with a Sobol' sequence of size 10^4) for the different criteria after 99, 300 and 600 iterations, for Hartmann-rescaled function inversion ($d = 6$) with $T = -1.6$, for 50 different initial DoEs of size 30 of type LHS Maximin, for $\kappa = 1$ and with n.points$= 10^4$.

3.5 Conclusion

In the framework of solving inversion problems using Gaussian Process Regression, we have proposed a new SUR criterion based on the Bichon criterion for DoE enrichment. Numerical simulations have demonstrated its good exploratory behavior, as far as its robustness from different points of view. Indeed, our new criterion is robust with the stationarity assumption of the underlying Gaussian process. Moreover, it is robust to the geometry of the set to be retrieved, in particular in terms of number of connected components.

3.6 Acknowledgements

The authors are very grateful to the reviewers for their relevant and helpful comments. The authors would like to express their gratitude to Morgane Menz from IFP Énergies Nouvelles for her valuable help on numerical aspects and in particular for the study of the robustness of estimators with respect to the GP stationarity assumption. All the computations presented in this paper were performed using the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities. We are also grateful of the useful feedback from Julien Bect and the CIROQUO consortium partners. CIROQUO is an Applied Mathematics consortium, gathering partners in technological and academia in the development of advanced methods for Computer Experiments (https://doi.org/10.5281/zenodo.6581216). This work was carried out as part of the "HPC/AI/HPDA Convergence for the Energy Transition" Inria/IFPEN joint research laboratory (High-performance computation / Artificial Intelligence / High-Performance Data Analysis).

3.7 Appendices

A Basics on Vorob'ev Theory and corresponding SUR strategy

Vorob'ev expectation

In this part, the notion of expectation for random closed sets in the sense of Vorob'ev is defined from [Molchanov, 2005]. The framework is a compact set $\mathbb{X} \subset \mathbb{R}^d$ and a random closed set Γ of $\mathbb{X}$. It is recalled that $\Gamma : \Omega \to \mathcal{C}$ is a random closed set if it is a measurable function on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in the set of all compacts of $\mathbb{X}$ in the sense that:

$$\forall\, C \in \mathcal{C},\ \{w \in \Omega, \Gamma(w) \cap C \neq \varnothing\} \in \mathcal{F}. \tag{3.25}$$

Let us define the parametric family $\{Q_\alpha\}_{\alpha\in[0,1]}$ of Vorob'ev quantiles is defined by:

$$Q_\alpha := \{\mathrm{x} \in \mathbb{X} : p(\mathrm{x}) := \mathbb{P}(\mathrm{x} \in \Gamma) \geq \alpha\},\ \forall\, \alpha \in [0,1]. \tag{3.26}$$

The elements of $\{Q_\alpha\}_{\alpha\in[0,1]}$ are called the Vorob'ev quantiles of the random closed set Γ and the function p is called the coverage function of Γ.

To define the expectation of the random closed set Γ, [Molchanov, 2005] comes back to the expectation of a real random variable: the measure of the Γ set $\mathbb{P}_\mathbb{X}(\Gamma)$. From the parametric family of Vorob'ev quantiles (Eq. (3.26)), the expectation of Γ in the sense of Vorob'ev is then defined as the Vorob'ev quantile of measure equal (or the closest one higher) to the expectation of the measure of Γ. More precisely, the Vorob'ev expectation of a random closed set Γ is the set $Q_{\alpha^\star}$, where $\alpha^\star$ is defined as the Vorob'ev threshold by

$$\forall\, \alpha > \alpha^\star,\ \mathbb{P}_\mathbb{X}(Q_\alpha) < \mathbb{E}[\mathbb{P}_\mathbb{X}(\Gamma)] \leq \mathbb{P}_\mathbb{X}(Q_{\alpha^\star}), \tag{3.27}$$

where $\mathbb{P}_\mathbb{X}$ denotes the Lebesgue measure on $\mathbb{X}$. $\alpha^\star$ is called the Vorob'ev threshold.

Remark 2.
• Based on Equation (3.26), the function $\alpha \mapsto \mathbb{P}_\mathbb{X}(Q_\alpha)$ is decreasing on $[0,1]$.
• The uniqueness of $\alpha^\star$ in the definition is easily checked. The existence of such $\alpha^\star$ in the definition of Vorob'ev expectation is based on the decreasing and continuity to the left of the function $\alpha \mapsto \mathbb{P}_\mathbb{X}(Q_\alpha)$ which is itself guaranteed by the superior semi-continuity of the coverage function p (see [Molchanov, 2005] page 23).
• The continuity of the function $\alpha \mapsto \mathbb{P}_\mathbb{X}(Q_\alpha)$ ensures equality $\mathbb{P}_\mathbb{X}(Q_{\alpha^\star}) = \mathbb{E}[\mathbb{P}_\mathbb{X}(\Gamma)]$ in the definition of Vorob'ev expectation.

In the particular case where Γ is given by $\Gamma := \{\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\}$ with ξ a stochastic process indexed by $\mathbb{X}$ with continuous trajectories conditioned on the event $\mathcal{E}_n$ corresponding to n evaluations of ξ and T a fixed threshold, Γ is a random closed set ([Molchanov, 2005, page 3]). A sufficient condition to obtain a stochastic process with continuous trajectories is to consider a separable Gaussian process with continuous mean and covariance kernel of type Matérn 3/2 or 5/2 ([Paciorek, 2003, pages 35 and 44]). Moreover, in this case, the function $\alpha \mapsto \mathbb{P}_{\mathbb{X}}(Q_\alpha)$ is continuous and so the equality $\mathbb{P}_{\mathbb{X}}(Q_{\alpha^\star}) = \mathbb{E}[\mathbb{P}_{\mathbb{X}}(\Gamma)|\, \mathcal{E}_n]$ is verified. It is also important to notice that naive estimator $\hat{\Gamma}_1$ is almost surely equal to the median of Vorob'ev (quantile of order 1/2). Indeed, by noting ϕ the distribution function of the standard normal distribution,

$$\begin{aligned}
\hat{\Gamma}_1 &= \{\mathbf{x} \in \mathbb{X},\, m_n(\mathbf{x}) \leq T\} \\
&\overset{\text{a.s.}}{=} \left\{\mathbf{x} \in \mathbb{X},\, \frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})} \geq 0 \ \text{ and } \ \sigma_n(\mathbf{x}) \neq 0\right\} \\
&= \left\{\mathbf{x} \in \mathbb{X},\, \phi\left(\frac{T - m_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\right) \geq \phi(0) \ \text{ and } \ \sigma_n(\mathbf{x}) \neq 0\right\} \quad \text{as } \phi \text{ increases} \\
&\overset{\text{a.s.}}{=} \left\{\mathbf{x} \in \mathbb{X},\, p_n(\mathbf{x}) \geq \frac{1}{2}\right\} \\
&= Q_{\frac{1}{2}},
\end{aligned} \tag{3.28}$$

where p_n is the coverage function $p_n(\mathbf{x}) := \mathbb{P}(\xi(\mathbf{x}) \leq T|\, \mathcal{E}_n)$.

Repeating the previous calculation and replacing 1/2 by the Vorob'ev threshold $\alpha^\star$, we obtain :

$$Q_{\alpha^\star} \overset{\text{a.s.}}{=} \left\{\mathrm{x} \in \mathbb{X},\, m_n(\mathrm{x}) \leqslant T - \phi^{-1}(\alpha^\star)\, \sigma_n(\mathrm{x})\right\} \tag{3.29}$$

Vorob'ev deviation

The introduction of the concept of Vorob'ev deviation is used to define residual uncertainty $\mathcal{H}_n(\mathbf{x})$ in a SUR strategy. Let us start by introducing the notion of distance between two random closed sets.

The average distance $\mathrm{d}_{\mathbb{P}_{\mathbb{X}}}$ with respect to a measure $\mathbb{P}_{\mathbb{X}}$ on all pairs of random closed sets included in $\mathbb{X}$ is defined by: for all random closed sets Γ_1, Γ_2 defined on $\mathbb{X}$,

$$d_{\mathbb{P}_{\mathbb{X}}}(\Gamma_1, \Gamma_2) := \mathbb{E}\left[\mathbb{P}_{\mathbb{X}}(\Gamma_1 \Delta \Gamma_2)\right] \tag{3.30}$$

where Δ is the random symmetric difference: $\forall\, \omega \in \Omega$, $\Gamma_1 \Delta \Gamma_2(\omega) := (\Gamma_1 \backslash \Gamma_2)(\omega) \cup (\Gamma_2 \backslash \Gamma_1)(\omega)$. In addition, the function $\mathrm{d}_{\mathbb{P}_{\mathbb{X}}}$ checks the properties of a distance.

The following proposition [Molchanov, 2005] justifies the choice of the Vorob'ev quantile family to define the Vorob'ev expectation and also allows to define the Vorob'ev deviation. Moreover, when $\mathbb{E}[\mathbb{P}_{\mathbb{X}}(\Gamma)] = \mathbb{P}_{\mathbb{X}}(Q_{\alpha^\star})$ (especially in the case of $\Gamma := \{\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\}$ with ξ a stochastic process indexed by $\mathbb{X}$ with continuous trajectories), the condition $\alpha^\star \geq \frac{1}{2}$ is no longer necessary (see [El Amri, 2019, page 28] for the proof)

Proposition 3. Noting $Q_{\alpha^\star}$ the Vorob'ev expectation of the random closed set Γ and assuming that $\alpha^\star \geq \frac{1}{2}$, it results that: for any measurable set M included in $\mathbb{X}$ such that $\mathbb{P}_{\mathbb{X}}(\mathrm{M}) = \mathbb{E}[\mathbb{P}_{\mathbb{X}}(\Gamma)]$,

$$d_{\mathbb{P}_{\mathbb{X}}}(\Gamma, Q_{\alpha^\star}) \leq d_{\mathbb{P}_{\mathbb{X}}}(\Gamma, \mathrm{M}) \tag{3.31}$$

The Vorob'ev deviation of the random set Γ is defined as the quantity $\mathrm{d}_{\mathbb{P}_{\mathbb{X}}}(\Gamma, Q_{\alpha^\star})$. The Vorob'ev deviation quantifies the variability of the random closed set Γ relative to its Vorob'ev expectation.

SUR Vorob'ev criterion

Once the basic elements of Vorob'ev theory are introduced, the associated SUR strategy is simply defined from the definition of SUR strategies via Equation (3.16) by taking:

$$\mathcal{H}_n := \mathbb{E}\big[\mathbb{P}_{\mathbb{X}}(\Gamma \Delta Q_{n,\alpha_n^\star}) \big|\, \mathcal{E}_n\big] \quad \text{and} \quad \mathcal{H}_{n+1}(\mathbf{x}) := \mathbb{E}\big[\mathbb{P}_{\mathbb{X}}(\Gamma \Delta Q_{n+1,\alpha_{n+1}^\star}) \big|\, \xi(\mathbf{x}),\, \mathcal{E}_n\big] \tag{3.32}$$

where $Q_{n,\alpha_n^\star}$ denotes the Vorob'ev expectation conditioned on $\mathcal{E}_n$ and $Q_{n+1,\alpha_{n+1}^\star}$ the Vorob'ev expectation conditioned on $\mathcal{E}_n$ and the addition of the point $(\mathbf{x}, \xi(\mathbf{x}))$ to the DoE. The idea behind (3.32) is to take as residual uncertainty, the variation with respect to the Vorob'ev expectation of the random closed set Γ, with $\Gamma := \{\mathbf{x} \in \mathbb{X},\, \xi(\mathbf{x}) \leq T\}$. With the assumption that $\alpha_{n+1}^\star = \alpha_n^\star$ and by re-injecting the quantity $\mathcal{H}_{n+1}(\mathbf{x})$ of (3.32) in the $\mathcal{J}_n$ criterion (3.16), it is possible to find a simplified formulation involving only an integral of a simple quantity [Chevalier, 2013]. This quantity is dependent on the cumulative distribution functions of the standard normal distribution and the bivariate centered normal distribution with given covariance matrix. Such a formulation then allows less time consuming computations and therefore is implemented in the package *KrigInv* ([Chevalier et al., 2014b]).

B Kriging update formulas

In the context of SUR strategies, the quantity $\mathcal{J}_n(\mathbf{x})$ in Equation (3.16) for a fixed $\mathbf{x}$ is usually simplified thanks to formulas of kriging and conditionally on the point $(\mathbf{x}, \xi(\mathbf{x}))$ added to the DoE, and more particularly using the kriging standard deviation. Indeed, contrary to the trend, the kriging standard deviation does not depend on surrogate model observations. For instance the recurrent formula, used in [Chevalier, 2013] is efficient for calculating kriging model in the context of universal kriging and when the kriging parameters β and θ do not need to be re-estimated. These kriging update formulas are given for all $\mathbf{y}, \mathbf{y}'$ in $\mathbb{X}^2$ by

$$m_{n+1}(\mathbf{y}) = m_n(\mathbf{y}) + k_n\big(\mathbf{y}, \mathbf{x}^{(n+1)}\big)\, k_n\big(\mathbf{x}^{(n+1)}, \mathbf{x}^{(n+1)}\big)^{-1} \Big(g\big(\mathbf{x}^{(n+1)}\big) - m_n\big(\mathbf{x}^{(n+1)}\big)\Big), \tag{3.33}$$

$$\sigma_{n+1}^2(\mathbf{y}) = \sigma_n^2(\mathbf{y}) - k_n^2\big(\mathbf{y}, \mathbf{x}^{(n+1)}\big)\, \sigma_n^2\big(\mathbf{x}^{(n+1)}\big)^{-1}, \tag{3.34}$$

$$k_{n+1}(\mathbf{y}, \mathbf{y}') = k_n(\mathbf{y}, \mathbf{y}') - k_n\big(\mathbf{y}, \mathbf{x}^{(n+1)}\big) k_n\big(\mathbf{x}^{(n+1)}, \mathbf{x}^{(n+1)}\big)^{-1} k_n\big(\mathbf{y}', \mathbf{x}^{(n+1)}\big), \tag{3.35}$$

with $\mathbf{x}^{(n+1)}$ the $n+1^{\,\text{th}}$ observation point.

As for SUR strategies, it is possible to generalize these formulas in the case of simultaneous additions of q points ([Chevalier, 2013]). The advantage of these formulas is that the expressions of m_n, σ_n, and k_n are reused to reduce computational time. It is particularly useful in SUR strategies where many evaluations of the kriging formulas may be required for the numerous evaluations of the sampling criterion (in the context of its minimization (3.16)). Finally, it can be shown that these kriging formulas still coincide with the Gaussian process conditional formulas in the context of universal kriging (see Appendix A of [Chevalier, 2013] for a proof).

C Proof of the explicit formula for SUR Bichon

The interest of this appendix is to propose a demonstration of Proposition 2, allowing to give an explicit expression of SUR Bichon. Let us start by stating and proving an intermediate lemma.

Lemma 2. Let N be a standard Gaussian random variable and $(a, b) \in \mathbb{R}^2$ such that $a < b$, then:

$$\mathbb{E}\big[N\mathbf{1}_{[a,b]}(N)\big] = -\varphi(b) + \varphi(a) \tag{3.36}$$

where φ is the probability density function of the standard normal distribution.

Proof.

$$\mathbb{E}\big[N\mathbf{1}_{[a,b]}(N)\big] = \frac{1}{\sqrt{2\pi}}\int_a^b te^{-\frac{t^2}{2}}\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}}\Big[-e^{-\frac{t^2}{2}}\Big]_a^b = -\varphi(b) + \varphi(a) \tag{3.37}$$

□

Proposition 1. *For all* $\mathbf{x}$, $\mathbf{z}$ *belonging to* $\mathbb{X}^2$, *we have:*

$$\begin{aligned}
&\mathbb{E}\Big[\big(\kappa\sigma_{n+1}(\mathbf{z}) - |T - \xi(\mathbf{z})|\big)^+ \,\Big|\, \mathcal{E}_n\Big] \\
&\quad= (m_n(\mathbf{z}) - T)\left[2\,\phi\Big(\frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right] \\
&\qquad- \sigma_n(\mathbf{z})\left[2\,\varphi\Big(\frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \varphi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \varphi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right] \\
&\qquad+ \epsilon_{\mathbf{x}}(\mathbf{z})\left[\phi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right]
\end{aligned} \tag{3.38}$$

where $\epsilon_{\mathbf{x}}(\mathbf{y}) := \kappa\sigma_{n+1}(\mathbf{z})$, $T^{\pm} := T \pm \epsilon_{\mathbf{x}}(\mathbf{z})$, φ *and* ϕ *denote the probability density and cumulative distribution functions of the standard normal distribution, respectively.*

Proof. For the proof only and for the sake of lightening the notations, we note $\mathbb{E}_n$ the conditional expectation $\mathbb{E}\left[\,.\,|\,\mathcal{E}_n\right]$. In addition, the expression to be calculated is separated into three terms that are calculated separately. Specifically:

$$\begin{aligned}
&\mathbb{E}_n\big[\big(\kappa\sigma_{n+1}(\mathbf{z}) - |T - \xi(\mathbf{z})|\big)^+\big] = \mathbb{E}_n\Big[\big(\epsilon_{\mathbf{x}}(\mathbf{z}) - |T - \xi(\mathbf{z})|\big)\mathbf{1}_{[T^-,T^+]}\big(\xi(\mathbf{z})\big)\Big] \\
&\quad= \mathbb{E}_n\Big[\epsilon_{\mathbf{x}}(\mathbf{z})\,\mathbf{1}_{[T^-,T^+]}\big(\xi(\mathbf{z})\big)\Big] - \mathbb{E}_n\Big[\big(T - m_n(\mathbf{z}) + m_n(\mathbf{z}) - \xi(\mathbf{z})\big)\mathbf{1}_{[T^-,T[}\big(\xi(\mathbf{z})\big)\Big] \\
&\qquad+ \mathbb{E}_n\Big[\big(T - m_n(\mathbf{z}) + m_n(\mathbf{z}) - \xi(\mathbf{z})\big)\mathbf{1}_{[T,T^+]}\big(\xi(\mathbf{z})\big)\Big] \\
&\quad= \underbrace{\mathbb{E}_n\Big[\epsilon_{\mathbf{x}}(\mathbf{z})\,\mathbf{1}_{[T^-,T^+]}\big(\xi(\mathbf{z})\big)\Big]}_{①} + \underbrace{\big(T - m_n(\mathbf{z})\big)\Big[\mathbb{E}_n\Big[\mathbf{1}_{[T,T^+]}\big(\xi(\mathbf{z})\big)\Big] - \mathbb{E}_n\Big[\mathbf{1}_{[T^-,T]}\big(\xi(\mathbf{z})\big)\Big]\Big]}_{②} \\
&\qquad\underbrace{-\Big[\mathbb{E}_n\Big[\big(m_n(\mathbf{z}) - \xi(\mathbf{z})\big)\mathbf{1}_{[T^-,T]}\big(\xi(\mathbf{z})\big)\Big] - \mathbb{E}_n\Big[\big(m_n(\mathbf{z}) - \xi(\mathbf{z})\big)\mathbf{1}_{[T,T^+]}\big(\xi(\mathbf{z})\big)\Big]\Big]}_{③}
\end{aligned} \tag{3.39}$$

The calculation of the three terms separately is as follows:

$$\begin{aligned}
①: &= \mathbb{E}_n\Big[\epsilon_{\mathbf{x}}(\mathbf{z})\,\mathbf{1}_{[T^-,T^+]}\big(\xi(\mathbf{z})\big)\Big] \\
&= \epsilon_{\mathbf{x}}(\mathbf{z})\,\mathbb{E}_n\left[\mathbf{1}_{\left[\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}, \frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\right]}\Big(\frac{\xi(\mathbf{z}) - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right] \\
&= \epsilon_{\mathbf{x}}(\mathbf{z})\left[\phi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right]
\end{aligned} \tag{3.40}$$

$$
\begin{aligned}
② :&= (T - m_n(\mathbf{z}))\Big[\mathbb{E}_n\Big[\mathbf{1}_{[T,T^+]}(\xi(\mathbf{z}))\Big] - \mathbb{E}_n\Big[\mathbf{1}_{[T^-,T]}(\xi(\mathbf{z}))\Big]\Big] \\
&= (m_n(\mathbf{z}) - T)\Big[\mathbb{E}_n\Big[\mathbf{1}_{[T^-,T]}(\xi(\mathbf{z}))\Big] - \mathbb{E}_n\Big[\mathbf{1}_{[T,T^+]}(\xi(\mathbf{z}))\Big]\Big] \\
&= (m_n(\mathbf{z}) - T)\left[2\,\phi\Big(\frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \phi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right] \qquad (3.41)
\end{aligned}
$$

For the calculation of ③, the use of Lemma 2 is necessary, using the notations of it.

$$
\begin{aligned}
③ :&= -\Big[\mathbb{E}_n\Big[(m_n(\mathbf{z}) - \xi(\mathbf{z}))\mathbf{1}_{[T^-,T]}(\xi(\mathbf{z}))\Big] - \mathbb{E}_n\Big[(m_n(\mathbf{y}) - \xi(\mathbf{z}))\mathbf{1}_{[T,T^+]}(\xi(\mathbf{z}))\Big]\Big] \\
&= \sigma_n(\mathbf{z})\left[\mathbb{E}_n\left[\Big(\frac{\xi(\mathbf{z}) - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\mathbf{1}_{\left[\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}, \frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\right]}\Big(\frac{\xi(\mathbf{z}) - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right]\right. \\
&\qquad \left. - \mathbb{E}_n\left[\Big(\frac{\xi(\mathbf{z}) - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\mathbf{1}_{\left[\frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}, \frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\right]}\Big(\frac{\xi(\mathbf{z}) - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right]\right] \\
&= \sigma_n(\mathbf{z})\left[\mathbb{E}\left[N\,\mathbf{1}_{\left[\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}, \frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\right]}(N)\right] - \mathbb{E}\left[N\,\mathbf{1}_{\left[\frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}, \frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\right]}(N)\right]\right] \\
&= -\sigma_n(\mathbf{z})\left[2\,\varphi\Big(\frac{T - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \varphi\Big(\frac{T^- - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big) - \varphi\Big(\frac{T^+ - m_n(\mathbf{z})}{\sigma_n(\mathbf{z})}\Big)\right] \qquad (3.42)
\end{aligned}
$$

The expected result is then obtained by re-injecting the expressions of ①, ②, and ③ obtained in Equations (3.40) to (3.42) in Equation (3.39). □

Remark 3.
• In this proof of Proposition 2, the fact that $m_n(\mathbf{z})$, $\sigma_n(\mathbf{z})$, $\epsilon_{\mathbf{x}}(\mathbf{z})$ and $T^{\pm}$ are constant with respect to $\xi(\mathbf{z})$ is implicitly used, in particular to output $m_n(\mathbf{z})$, $\sigma_n(\mathbf{z})$ and $\epsilon_{\mathbf{x}}(\mathbf{z})$ of the conditional expectation $\mathbb{E}_n$, but also for the renormalization of $\xi(\mathbf{z})$ and the transition to the density probability and cumulative distribution functions of the standard normal distribution.

D Robustness of estimators with respect to the GP stationarity assumption.

A certain instability of the approximation error was observed at the beginning of the enrichment in Figures 3.4 and 3.3 in the case of estimator $\hat{\Gamma}_2$ corresponding to Vorob'ev expectation, in comparison to $\hat{\Gamma}_1$ naive estimator. This robustness of naive estimator $\hat{\Gamma}_1$ compared to $\hat{\Gamma}_2$ can be explained by the fact that naive estimator corresponds to the median of Vorob'ev (Appendix A, equation (3.28)). Indeed, even if the notions of expectation and median of Vorob'ev are not similar to the classical ones, the property of minimizing the first absolute central moment of the median is preserved when extending the notion of median to the framework of Vorob'ev random sets ([Molchanov, 2005] page 178). It is also possible to "read" the lack of robustness on Equation (3.29) of Appendix A which is recalled below:

$$
Q_{\alpha^\star} \overset{\text{a.s.}}{=} \Big\{\mathrm{x} \in \mathbb{X},\, m_n(\mathrm{x}) \leqslant T - \phi^{-1}(\alpha^\star)\,\sigma_n(\mathrm{x})\Big\}.
$$

Knowing the strong dependence of the σ_n term on the stationarity condition of the process, it is straightforward that the term $\phi^{-1}(\alpha^\star)\,\sigma_n(\mathbf{x})$ plays an important role in the non-robustness

of $\hat{\Gamma}_2 = Q_{\alpha^\star}$ estimator, compared to naive estimator $\hat{\Gamma}_1$ where this term $\phi^{-1}(\alpha^\star)\,\sigma_n(\mathbf{x})$ does not appear.

To illustrate this robustness issue, we define the Loggruy function in dimension 2 as follows:

$$\forall\,\mathbf{x}\in\mathbb{R}^2,\ \mathrm{Loggruy}(\mathbf{x}) = 10\log_{10}\left(3.6+10^4\left(\sum_{i=1}^{2}(\mathbf{x}_i-\mathbf{a}_i)^2-r^2\right)\times\cdots\times\left(\sum_{i=1}^{2}(\mathbf{x}_i-\mathbf{e}_i)^2-r^2\right)\right) \tag{3.43}$$

with $(\mathbf{a}_i)_i = \begin{pmatrix}0.153\\0.939\end{pmatrix}$; $(\mathbf{b}_i)_i = \begin{pmatrix}0.854\\0.814\end{pmatrix}$; $(\mathbf{c}_i)_i = \begin{pmatrix}0.510\\0.621\end{pmatrix}$; $(\mathbf{d}_i)_i = \begin{pmatrix}0.207\\0.386\end{pmatrix}$; $(\mathbf{e}_i)_i = \begin{pmatrix}0.815\\0.146\end{pmatrix}$ and $r := 0.07$.

The threshold chosen is $T = 10\log_{10}(3.6)$ and the corresponding $\Gamma^\star$ excursion set (Equation (3.1)) is composed of 5 disconnected components. This function is particularly interesting in our context, since it has strong gradients at the edges of the domain and weaker gradients in the middle, where the different zones of the $\Gamma^\star$ excursion set are located (Figure 3.8*c*). This means that the stationarity assumption of the kriging meta-model cannot be verified.

Figure 3.8 represents in the case of an enrichment of 100 and 200 points of SUR Vorob'ev, from an initial DoE of size 10, the contour lines of the coverage probability p_n and the kriging mean m_n. Estimators $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are also represented and compared to $\Gamma^\star$. An irregularity is observed for the contour lines of p_n either after 100 or 200 iterations. But, m_n is relatively accurate even after 100 iterations. $\hat{\Gamma}_2$ estimator is then less efficient than $\hat{\Gamma}_1$ estimator.

In summary, $\hat{\Gamma}_2$ estimator related to Vorob'ev theory is more dependent on the stationarity hypothesis than $\hat{\Gamma}_1$ naive estimator, and this is essentially explained by the construction of Vorob'ev expectation which is more sensitive to the kriging standard deviation (Equation (3.29)). Consequently, the approximation error $\mathrm{Err}\big(\hat{\Gamma}_2(\chi_{j,n})\big)$ is more dependent on the stationarity hypothesis than the naive approximation error $\mathrm{Err}\big(\hat{\Gamma}_1(\chi_{j,n})\big)$.

3.8 Additions to the article

A Influence of the κ parameter

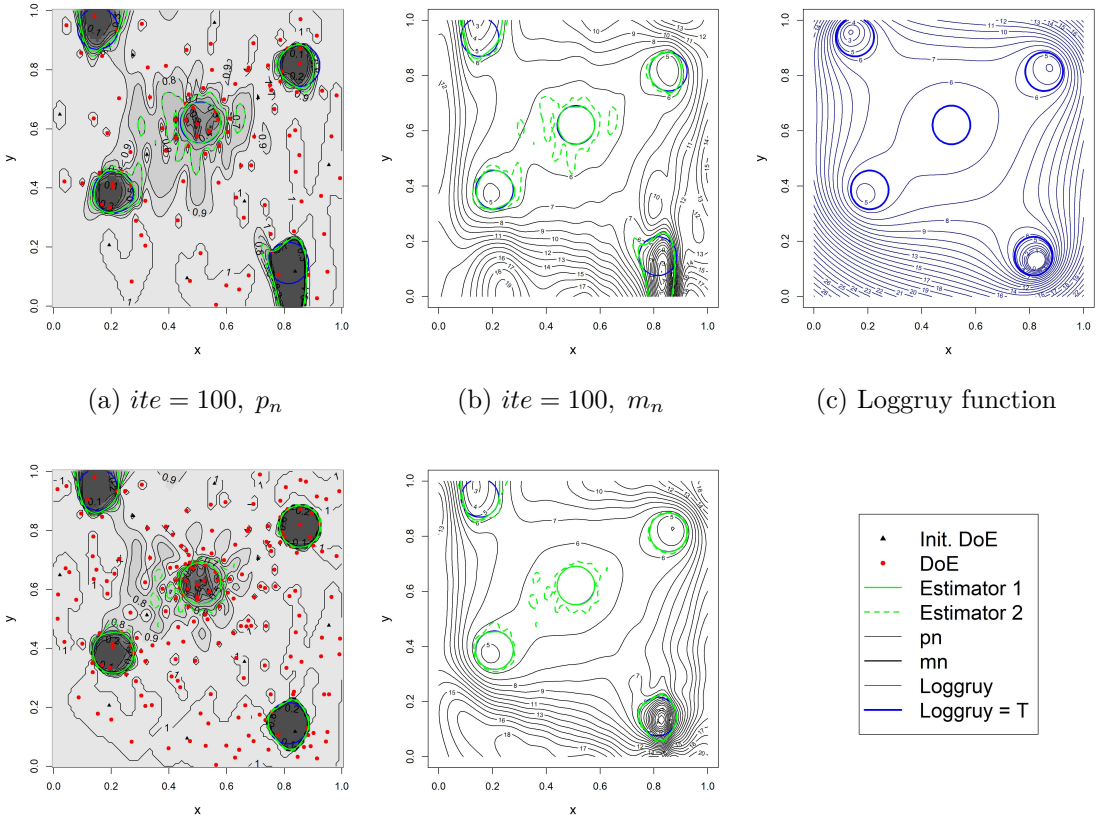
The impact of the κ parameter on the performance of SUR Bichon criterion is demonstrated using the Branin-rescaled test function. The tests previously conducted with the Branin-rescaled function (Section 3.4) are repeated with κ taking values from the set $0.1, 0.5, 1, 2, 10$ for SUR Bichon criterion. As shown in Figure 3.3, the results align with the theoretical expectation that κ should not be too large or too small to strike a balance between exploration and exploitation. The value $\kappa = 1$, which has been used so far, appears to be quite appropriate.

Additionally, it is observed that it is preferable to decrease κ rather than increase it. This is consistent with the formulation of SUR Bichon criterion, where the term multiplied by κ primarily controls the exploration aspect, while the other term balances both exploration and exploitation due to the expectation (see Equation (3.17)).

B Use of Importance Sampling for evaluating the SUR Bichon criterion

Importance sampling (see, for example, [Tokdar and Kass, 2010]) is a statistical method used to estimate properties of a target distribution by sampling from a different, more convenient distribution and reweighting the samples to account for the discrepancy between the two distributions. This approach is particularly useful for exploring rare events or low-probability regions by focusing computational resources on areas that contribute most to the estimation.

The use of an importance sampling method was considered instead of Sobol' sequences for evaluating the integral involved in the explicit formulation of SUR Bichon criterion (see Section



(a) $ite = 100,\ p_n$ (b) $ite = 100,\ m_n$ (c) Loggruy function

(d) $ite = 200,\ p_n$ (e) $ite = 200,\ m_n$

Figure 3.8: At left and center, representation of the two $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ estimators after 100 iterations (first line) and 200 iterations (second line) for SUR Vorob'ev, and with the contour lines of p_n (first column) and m_n (second column), in comparison with the true excursion set (top right), for Loggruy function inversion ($d = 2$) with $T = 10\,log_{10}(3.6)$, for a particular initial DoE of size 10.

κ / $(\times 10^{-3})$	0.1	0.5	1	2	10	SUR Vorob'ev
Mean	1.67	1.69	1.74	1.72	2.24	3.94
Median	1.59	1.6	1.65	1.69	2.14	1.64
$Q_{0.05}$	1	1	0.97	1.1	1.25	0.97
$Q_{0.95}$	2.55	2.55	2.73	2.5	3.65	3.98
SD	0.49	0.52	0.57	0.45	0.72	10.65
IQR	0.675	0.675	0.74	0.58	0.94	0.66

Table 3.3: Characteristic values of the empirical distributions of the comparison measure $\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}\Delta\Gamma^{\star})$ (with a grid 200×200) for the different criteria after 20 iterations, in the case of the inversion of the Branin-rescaled function ($d = 2$) with $T = 10$, for 100 different initial plans of size 10 of type LHS Maximin, for $\kappa \in \{0.1, 0.5, 1, 2, 10\}$ and with $n.points = 10\,000$.

3.4.1), similar to the approach implemented in the KrigInv package from [Chevalier et al., 2014b]. However, the results showed that while this method slightly reduces computational cost, it significantly limits the exploratory potential that makes SUR Bichon criterion particularly appealing.

Summary:

The aim of this chapter is to introduce, in the context of excursion set estimation via GPR, a new DoE enrichment criterion, based on the Bichon criterion. The idea is to propose a more efficient version of the Bichon criterion by adapting a SUR version of this criterion, while retaining the interesting exploratory character of the Bichon criterion. This exploratory character allows us to explore the design space sufficiently, in the case of a complex excursion set, e.g. made up of several connected components. An explicit formulation of the new criterion is proposed, making it easier to implement. Numerical simulations on 2 and $6d$ analytic examples, and comparisons with the Bichon and SUR Vorob'ev criteria, demonstrate the performance and robustness of the new SUR Bichon criterion. Also, the SUR Bichon criterion proves resistant to the stationarity assumption of the underlying Gaussian process.

The next chapter presents, within the framework of vector black box models, new methodologies for the simultaneous estimation of partial excursion sets on each output component.

Chapter 4

Excursion set estimation on vector valued models

Outlines

The aim of this chapter is to study the estimation of excursion sets using Gaussian processes in the context of a vector-valued black box function. In the context of simultaneous estimation of each partial excursion set defined for each model output component, we propose three experimental design enrichment criteria, all three inspired by [Bichon et al., 2008]. The first two are derived from the framework of scalar Gaussian process regression, while the third one is based on a multi-output Gaussian process surrogate model. The three criteria are compared on two analytical examples with input space dimension 2 and 6, and output space dimension 2.

Contents

4.1 The framework for estimating excursion sets **67**
4.1.1 Reminders on Gaussian process regression and scalar Bichon criterion 67
4.1.2 Excursion sets in the case of several outputs 68
4.2 Two natural extensions of Bichon criterion **70**
4.2.1 Alternating Scalar Bichon criterion 70
4.2.2 Pareto Scalar Bichon criterion 70
4.3 Vector output extension to Bichon criterion **72**
4.3.1 Reminders on multi-output Gaussian process regression 72
4.3.2 The proposed Vector Bichon criterion 73
4.4 Numerical experiments **74**
4.4.1 Implementation choices 75
4.4.2 Performance tests based on the 2D Branin function 76
4.4.3 Performance tests based on the 4D Hartmann function 84
4.5 Conclusion **88**
4.6 Appendices **90**
A Proof of the integral formulation for Vector Bichon criterion . . . 90
B Estimation of $H : (\alpha_1, \alpha_2, \rho) \mapsto \int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t$ 92

4.1 The framework for estimating excursion sets

For this chapter to be self contained, we recall first basics on Gaussian process regression and Scalar Bichon criterion reviewed in Chapters 1 and 2.

4.1.1 Reminders on Gaussian process regression and scalar Bichon criterion

Gaussian process regression

In the context of estimating an excursion set, given $\mathbb{X}$ a compact set of $\mathbb{R}^d$ and $g : \mathbb{X} \to \mathbb{R}$ a black box function, our objective is to estimate the set

$$\Gamma^\star := \Big\{\mathbf{x} \in \mathbb{X}, g(\mathbf{x}) \leq T\Big\}, \tag{4.1}$$

with T a fixed threshold, while limiting the number of costly g evaluations. To tackle this problem, a sequential Design of Experiments (DoE) enrichment strategy is used, where we recall that the DoE designates the set of evaluation points where g is evaluated. The first step is to define a small initial DoE and to build a first surrogate model based on the corresponding g evaluations. This DoE and the associated surrogate model are then sequentially updated using a well chosen acquisition criterion. Finally, a last update of the substitution model is performed to obtain an estimator of $\Gamma^\star$.

Gaussian Process Regression (GPR) is a popular surrogate model: the black box function g is considered as a realization of a Gaussian process (GP) ξ, which is written as

$$\xi(\mathbf{x}) := m(\mathbf{x}) + \mathbf{Z}(\mathbf{x}), \quad \forall\, \mathbf{x} \in \mathbb{X} \tag{4.2}$$

with m the trend of ξ (deterministic part) and $\mathbf{Z}$ a centered stationary GP of known covariance kernel $k : \mathbb{X}^2 \to \mathbb{R}$ (stochastic part). The choice of the covariance kernel k is crucial, since it determines the regularity of the predictor (see Section 1.3.1 in Chapter 1). To limit the difficulty of estimating the m function, it is common practice to set m as a linear combination of given basis functions $(f_i)_{i=1}^l$ with coefficients $\boldsymbol{\beta} := \{\beta_i\}_{i=1}^l$ to be estimated.

Let $\chi_n := (\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)})$ be the initial experimental design, $g(\chi_n)$ and $\xi(\chi_n)$ denote the vectors of evaluations of g and ξ on χ_n and notation $\mathcal{E}_n$ corresponds to the event $\xi(\chi_n) = g(\chi_n)$. $K := (k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}))_{1\leq i,j\leq n}$ is the covariance matrix of $\xi(\chi_n)$ and $\mathbf{k}(\mathbf{x})$ the cross-covariance matrix between $\xi(\chi_n)$ and $\xi(\mathbf{x})$ defined by $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x}, \mathbf{x}^{(1)}), ..., k(\mathbf{x}, \mathbf{x}^{(n)}))^\top$ for any $\mathbf{x}$ in $\mathbb{X}$. We denote $\mathbf{f}(\mathbf{x}) := (f_1(\mathbf{x}), ..., f_l(\mathbf{x}))^\top$ the evaluation vector of $\mathbf{f}$ on $\mathbf{x}$ defining the trend and $\mathbb{F} \in \mathbb{R}^{n\times l}$ the matrix with $\mathbf{f}(\mathbf{x}^{(i)})^\top$ as i^{th} row. When $\boldsymbol{\beta}$ is unknown and estimated by the maximum likelihood estimator (MLE)

$$\widehat{\boldsymbol{\beta}} := \left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\mathbb{F}^\top K^{-1} g(\chi_n),$$

the mean, variance and covariance of the conditionned process $\xi \,|\, \mathcal{E}_n$, respectively denoted m_n, σ_n^2, and k_n are given by

$$m_n(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \widehat{\boldsymbol{\beta}} + \mathbf{k}(\mathbf{x})^\top K^{-1}\big(g(\chi_n) - \mathbb{F}\widehat{\boldsymbol{\beta}}\big), \tag{4.3}$$

$$\begin{aligned}\sigma_n^2(\mathbf{x}) = \sigma^2(\mathbf{x}) - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}) + \\ \big(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\big)\left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\big(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\big)^\top,\end{aligned} \tag{4.4}$$

$$\begin{aligned}k_n(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}') + \\ \big(\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F}\big)\left(\mathbb{F}^\top K^{-1}\mathbb{F}\right)^{-1}\big(\mathbf{f}(\mathbf{x}')^\top - \mathbf{k}(\mathbf{x}')^\top K^{-1}\mathbb{F}\big)^\top.\end{aligned} \tag{4.5}$$

In the sequential DoE enrichment strategy, the choice of acquisition function is essential and depends on the objective on the black box function: global knowledge, optimization or estimation of an excursion set, etc. For more details on the various possible acquisition criteria, see Section 1.6 of Chapter 1 (for global knowledge and optimization) and Chapter 2 (for estimation of an excursion set). Below, we recall Bichon criterion, designed for estimating an excursion set and often used for its exploratory properties, while maintaining good performances in terms of exploitation.

Scalar Bichon criterion

Bichon criterion, originally presented in [Bichon et al., 2008], is a goal-oriented DoE enrichment criterion. This criterion is an adaptation of Expected Improvement (EI) from [Jones et al., 1998], introduced in the context of global optimization, to the context of excursion set estimation. This adaptation of EI criterion is made by considering both the variability of the surrogate model and the potential improvement in the knowledge of the excursion set boundary.

The new points selected according to Bichon criterion are such that the Expected Feasibility Function (EFF) is maximized, i.e.,

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}}\ \mathrm{EFF}(\mathbf{x}), \tag{4.6}$$

where EFF is defined by

$$\mathrm{EFF}(\mathbf{x}) := \mathbb{E}\Big[\big(\epsilon(\mathbf{x}) - |T - \xi(\mathbf{x})|\big)^+ \,\Big|\, \mathcal{E}_n\Big], \tag{4.7}$$

with $(\,.\,)^+ := \max(\,.\,, 0)$ and $\epsilon(\mathbf{x})$ a fixed function.

The function EFF represents, on average, the distance of the surrogate model to the limits of the interval $[T - \varepsilon(\mathbf{x}), T + \varepsilon(\mathbf{x})]$ only if the surrogate model belongs to this interval and is 0 otherwise. In practice, the width of the interval $\epsilon(\mathbf{x})$ is chosen to be proportional to the posterior standard deviation of the surrogate model $\sigma_n(\mathbf{x})$, leading in particular to a null value of the criterion for points already in the DoE. Criterion maximization in Equation (4.6) aims to select points close to the boundary of the excursion set approximated through GPR, or points associated with high values of $\epsilon(\mathbf{x})$. This is a compromise between exploitation and exploration.

For a statistical interpretation of Bichon criterion and an explicit formulation, see Section 2.1.4 of Chapter 2.

4.1.2 Excursion sets in the case of several outputs

Vector-valued black box models are widely used in industry to optimize and calibrate complex systems. They take various parameters as inputs and provide several critical components as outputs, such as mechanical stress and acoustic response in automotive, or mechanical stress and production cost in industrial manufacturing. These models ensure a close match between simulated results and experimental data, and can be leveraged for improving system performance and efficiency. It is sometimes crucial to check, for a given set of input parameters, which constraints on the outputs are respected, to ensure that the system meets the desired specifications and performance limits.

Simulator data are assumed to be isotopic, which means that the various simulator output components are evaluated simultaneously for a given input point. This means that a point must be chosen as DoE enrichment, in order to use the information provided by all simulator output components.

In this framework, the vector-valued black box model is denoted $\boldsymbol{g} := (g_1, \ldots, g_p)^\top$ and for a fixed $\boldsymbol{T} := (T_1, \ldots, T_p)^\top$ vector of thresholds, the partial excursion sets are defined by

$$\Gamma_i^\star := \{\mathbf{x} \in \mathbb{X},\, g_i(\mathbf{x}) \leq T_i\}, \tag{4.8}$$

and the global excursion set $\boldsymbol{\Gamma}^\star$ is defined as the intersection of all partial excursion sets:

$$\boldsymbol{\Gamma}^\star := \{\mathbf{x} \in \mathbb{X},\, \boldsymbol{g}(\mathbf{x}) \leq \boldsymbol{T}\} = \bigcap_{i=1}^{p} \Gamma_i^\star. \tag{4.9}$$

For the estimation of the global excursion set $\boldsymbol{\Gamma}^\star$, extensions of SUR Excursion measure variance and Integrated Bernoulli variance criteria have been developed in [Fossum et al., 2021] (see Section 2.2.4). These criteria use a multi-output Gaussian process surrogate model, and a generalization of the notion of coverage probability.

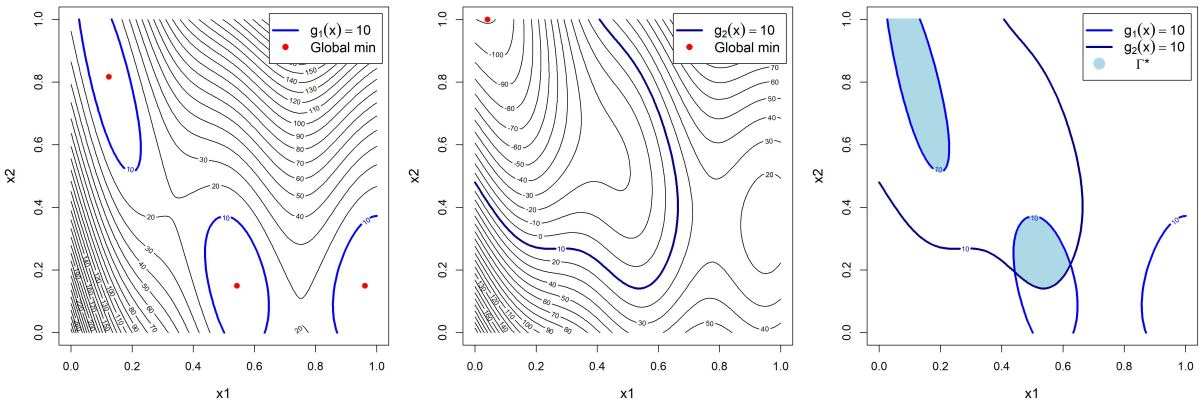
In this work, we aim to simultaneously estimate each of the partial excursion sets $\Gamma_i^\star$. This problem is quite different from that of estimating the global excursion set, since it requires discovering not only the boundary of the global excursion set, but also each of the boundaries of partial excursion sets. To illustrate this difference, we define the 2D function

$$g_1(\mathbf{x}) := \Big(\bar{\mathbf{x}}_2 - \frac{5\bar{\mathbf{x}}_1^2}{4\pi^2} + \frac{5\bar{\mathbf{x}}_1}{\pi} - 6\Big)^2 + 10\Big(1 - \frac{1}{8\pi}\Big)\cos(\bar{\mathbf{x}}_1) + 10 \tag{4.10}$$

and

$$g_2(\mathbf{x}) := \Big(\bar{\mathbf{x}}_2 - \frac{3\bar{\mathbf{x}}_1^2}{4\pi^2} + \frac{4\bar{\mathbf{x}}_1}{\pi} - 6\Big)^2 + 10\Big(1 - \frac{1}{8\pi}\Big)\cos(\bar{\mathbf{x}}_1) + 2\bar{\mathbf{x}}_1 - 9\bar{\mathbf{x}}_2 + 32 \tag{4.11}$$

with $\bar{\mathbf{x}}_1 := 15\mathbf{x}_1 - 5$ and $\bar{\mathbf{x}}_2 := 15\mathbf{x}_2$. Figure 4.1 shows for this 2D function the level sets for each component (left and middle), and excursion sets corresponding respectively to $T_1 = T_2 = 10$ (right).



(a) g_1 and $\Gamma_1^\star$ (b) g_2 and $\Gamma_2^\star$ (c) $\boldsymbol{\Gamma}^\star$

Figure 4.1: Left and middle: level sets for each component of the 2D function, and the partial excursion set $\Gamma_1^\star$ for threshold $T_1 = 10$ (left), $\Gamma_2^\star$ for threshold $T_2 = 10$ (middle). On the right, comparison of these partial excursion sets with the global excursion set $\boldsymbol{\Gamma}^\star$.

To the best of our knowledge, there are no criteria suitable for simultaneous estimation of each of the partial excursion sets $\Gamma_i^\star$. We introduce three new enrichment strategies based on

Bichon criterion for this simultaneous estimation objective. The first two strategies use scalar criteria based on independent surrogate models. The third strategy introduces a new criterion based on a multi-output Gaussian process, leveraging correlation between the different output components. These three enrichment procedures are compared on simple analytical examples on input space dimension 2 and 6, each with two output components. The study of the three criteria as well as the implementation on the analytical examples are presented below. The application of this methodology to a real case study is postponed to Chapter 5, with a detailed description of the characteristics of the numerical simulator.

4.2 Two natural extensions of Bichon criterion

The two criteria discussed below naturally extend Bichon criterion in the context of a vector-valued black box function and an objective of simultaneous estimation of different partial excursion sets $\Gamma_i^\star$. Each output component g_i of the black box model $\boldsymbol{g}$ is modeled by a Gaussian process ξ_i, and we assume that all surrogate models ξ_i are independent.

Due to the isotopic nature of the $\boldsymbol{g}$ model data, the new point chosen by the algorithm can be used to update all of the surrogate models. Although it is theoretically possible to design separate sequential DoEs for each component, this would lead to a potential loss of cross-information and require twice as many iterations. Therefore, it is preferable to opt for a common sequential DoE, enabling the simultaneous use of information from both output components of the black box function at each evaluation.

For simplicity, scalar criteria proposed below are presented for a black-box model with two output components ($p = 2$), although they can be generalized for any number of components.

4.2.1 Alternating Scalar Bichon criterion

The first criterion we introduce, called Alternating Scalar Bichon criterion, consists in maximizing Bichon criterion of each output component with the associated ξ_i surrogate model and T_i threshold alternatively. Note

$$\mathbf{x}_i^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}}\ \mathrm{EFF}_i(\mathbf{x}), \forall\, i \in \{1,2\}, \tag{4.12}$$

with EFF_i Bichon criterion associated with the surrogate model ξ_i and threshold T_i (see Section 4.1.1). Alternating Scalar Bichon criterion is then expressed as follows:

$$\mathbf{x}^{(n+1)} := \begin{cases} \mathbf{x}_1^{(n+1)} & \text{if } n+1 \text{ is even} \\ \mathbf{x}_2^{(n+1)} & \text{otherwise.} \end{cases} \tag{4.13}$$

Alternating Scalar Bichon criterion places equal importance on exploring each partial excursion set associated to each output component, regardless of the complexity of their respective estimation. Although simple to implement, this criterion seems rather limited when the difference in complexity between the estimation of the two partial excursion sets is substantial.

4.2.2 Pareto Scalar Bichon criterion

The aim is to develop a DoE enrichment criterion that takes into account the difficulty of estimating each partial excursion set, in order to choose a point that offers an optimal compromise for simultaneously improving knowledge of both outputs near the thresholds.

To do this, let us start by recalling the notion of Pareto front, a central concept in the field of multi-objective optimization (see for example [Marler and Arora, 2004]). When optimizing a vector-valued function, the optimization of one component does not necessarily coincide with the optimization of another component. The Pareto front represents a set of points in the output space which are not dominated by any other point, i.e., for each point on the Pareto front, there is no other point that has better values in all output dimensions ([Marler and Arora, 2004]). Figure 4.2 illustrates this concept in the case of maximizing a vector-valued function with two output components.

Each point on the Pareto front represents an optimal compromise between the different output components. Several methods exist for selecting the 'best' point on this front, although the notion of 'best' point is relative to the objective set. These include the optimisation of scalar aggregation functions that combine multiple objectives ([Srinivas and Deb, 1994]). For example, a common approach is to minimise the Euclidean distance to the ideal point, whose coordinates are the maximum values of each output ([Vincent and Grantham, 1981]). Another method is the search for the "knee point", where a small improvement in one objective leads to a large deterioration in another, making it possible to find balanced compromises ([Deb et al., 2003]). Finally, interactive approaches allow decision-makers to express their preferences in an iterative process, adjusting priorities until a satisfactory point is reached ([Miettinen and Mäkelä, 2006]).

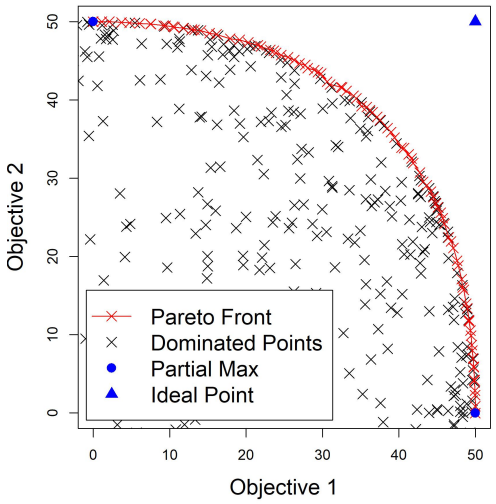


Figure 4.2: Example of a Pareto front.

The idea of the Pareto Scalar Bichon criterion is to construct a Pareto front for the vector-valued function whose i^{th} output component corresponds to Bichon criterion associated to surrogate model ξ_i and threshold T_i. Then, the enrichment point is the point in the input space corresponding to a point on Pareto front that minimizes the distance to the ideal point. This is possible because the output components are normalised on the basis of the initial DoE, which avoids unfairly favouring one or the other depending on their respective order of magnitude (see Section 4.4.1).

In summary, Pareto Scalar Bichon criterion is given by

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathcal{P}}{\operatorname{argmin}} \left\{ \sqrt{(\mathrm{EFF}_1(\mathbf{x}) - I_1)^2 + (\mathrm{EFF}_2(\mathbf{x}) - I_2)^2} \right\}. \tag{4.14}$$

with $(I_1, I_2) := \left(\max_{\mathbf{x}\in\mathcal{P}} \mathrm{EFF}_1(\mathbf{x}), \max_{\mathbf{x}\in\mathcal{P}} \mathrm{EFF}_2(\mathbf{x})\right)$ the ideal point and $\mathcal{P}$ the set of points in the input space corresponding to Pareto front.

If Bichon criterion of one of the two output components is relatively flat, this will lead to an enrichment oriented more towards the other component. Compared to Alternating Scalar Bichon criterion, Pareto Scalar Bichon criterion offers a compromise for the simultaneous estimation of both sets of partial excursions, taking into account the difficulty of estimating each of them.

4.3 Vector output extension to Bichon criterion

The aim of this section is to propose a new vector criterion based on joint modeling (with a MOGP) of the two outputs of the black box model.

4.3.1 Reminders on multi-output Gaussian process regression

Vector-valued Gaussian process regression, also known as multi-output Gaussian process (MOGP) regression is a generalization of GPR to build a surrogate of a deterministic vector-valued function $\boldsymbol{g} : \mathbb{X} \subset \mathbb{R}^d \mapsto \mathbb{R}^p$. MOGP is based on the fundamental belief that the output components are correlated in some manner. Therefore, a crucial aspect of MOGP is to take advantage of correlations between output components to mutualize the information acquired on each output and obtain more accurate predictions than if the output components were modeled independently. In the following, we are interested in a framework where prediction importance is the same for all model components (symmetric MOGP) and where the evaluation points are the same for each component of model output (isotopic data). This often occurs when the p output responses at a point $\mathbf{x}$ can be obtained through a single simulation (see for example [Liu et al., 2018] for more details on MOGP).

Let us assume that $\boldsymbol{g} := (g_1, \ldots, g_p)^\top$ is the realization of a vector-valued Gaussian process $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_p)^\top$. This process is characterized by its mean $M := (M_1, \ldots, M_p) : \mathbb{R}^d \to \mathbb{R}^p$ and covariance $K := (K_{i,j})_{1\leq i,j\leq p} : \mathbb{R}^d \times \mathbb{R}^d \to \mathcal{S}_p^+(\mathbb{R})$ functions (a priori, both assumed known here) defined for $(i, j) \in \{1, \ldots, p\}^2$ by

$$M_i(\mathbf{x}) := \mathbb{E}(\xi_i(\mathbf{x})) \quad \text{and} \quad K_{i,j}(\mathbf{x}, \mathbf{x}') := \mathrm{Cov}(\xi_i(\mathbf{x}), \xi_j(\mathbf{x}')), \tag{4.15}$$

with $\mathcal{S}^+(\mathbb{R})$ denoting real symmetric positive semi-definite matrices of size p. We note $\Sigma(\mathbf{x}) := K(\mathbf{x}, \mathbf{x})$ the covariance matrix of $\boldsymbol{\xi}(\mathbf{x})$. Let $\chi_n := (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})^\top$ denote the DoE at step n. We define

$$\boldsymbol{\xi}(\chi_n) := \left(\xi_1(\mathbf{x}^{(1)}), \ldots, \xi_1(\mathbf{x}^{(n)}), \ldots, \xi_p(\mathbf{x}^{(1)}), \ldots, \xi_p(\mathbf{x}^{(n)})\right)^\top \in \mathbb{R}^{pn\times 1}, \tag{4.16}$$

$$\boldsymbol{g}(\chi_n) := \left(g_1(\mathbf{x}^{(1)}), \ldots, g_1(\mathbf{x}^{(n)}), \ldots, g_p(\mathbf{x}^{(1)}), \ldots, g_p(\mathbf{x}^{(n)})\right)^\top \in \mathbb{R}^{pn\times 1}, \tag{4.17}$$

and $\boldsymbol{\mathcal{E}}_n$ the event $\boldsymbol{\xi}(\chi_n) = \boldsymbol{g}(\chi_n)$. The process $\boldsymbol{\xi}(\chi_n)$ is characterized by

$$\boldsymbol{M}(\chi_n) := \left(M_1(\mathbf{x}^{(1)}), \ldots, M_1(\mathbf{x}^{(n)}), \ldots, M_p(\mathbf{x}^{(1)}), \ldots, M_p(\mathbf{x}^{(n)})\right)^\top \in \mathbb{R}^{pn\times 1}. \tag{4.18}$$

and

$$\mathbf{K}_{\chi_n,\chi_n} := \begin{bmatrix} K_{11}(\chi_n, \chi_n) & \cdots & K_{1p}(\chi_n, \chi_n) \\ \vdots & \ddots & \vdots \\ K_{p1}(\chi_n, \chi_n) & \cdots & K_{pp}(\chi_n, \chi_n) \end{bmatrix} \in \mathbb{R}^{pn\times pn}, \tag{4.19}$$

with $K_{ii'}(\chi_n, \chi_n) = \mathbb{E}[(\xi_i(\chi_n) - \mathbb{E}[\xi_i(\chi_n)])(\xi_{i'}(\chi_n) - \mathbb{E}[\xi_{i'}(\chi_n)])^\top] \in \mathbb{R}^{n\times n}$ the cross-covariance matrix between $\xi_i(\chi_n)$ and $\xi_{i'}(\chi_n)$. In the same way, we note $\mathbf{K}_{\chi_n}(\mathbf{x}) \in \mathbb{R}^{pn\times p}$ the cross-covariance matrix between $\boldsymbol{\xi}(\chi_n)$ and $\boldsymbol{\xi}(\mathbf{x})$ defined from blocks

$$K_{ii'}(\mathbf{x}, \chi_n) := (K_{ii'}(\mathbf{x}, \mathbf{x}^{(1)}), \dots, K_{ii'}(\mathbf{x}, \mathbf{x}^{(n)}))^\top \in \mathbb{R}^{1\times n}. \tag{4.20}$$

Surrogate update formulas, which are extensions of Formulas (4.3), (4.4) and (4.3) to the vector case (with β known), are then given (see for example [Liu et al., 2018]) by:

$$M_n(\mathbf{x}) = M(\mathbf{x}) + \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1}\big(\boldsymbol{g}(\chi_n) - \boldsymbol{M}(\chi_n)\big), \tag{4.21}$$

$$\Sigma_n(\mathbf{x}) = \Sigma(\mathbf{x}) - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{K}_{\chi_n}(\mathbf{x}), \tag{4.22}$$

$$K_n(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}, \mathbf{x}') - \mathbf{K}_{\chi_n}(\mathbf{x})^\top \mathbf{K}_{\chi_n,\chi_n}^{-1} \mathbf{K}_{\chi_n}(\mathbf{x}'). \tag{4.23}$$

where M_n, Σ_n and K_n denote the mean, variance and covariance functions, respectively, of the process $\boldsymbol{\xi}_n := \boldsymbol{\xi} \,|\, \boldsymbol{\mathcal{E}}_n$.

In the case where the a priori trend is no longer known, but assumed to be a linear combination of basis functions with coefficients to be estimated, there are also update formulas ([Helterbrand and Cressie, 1994]). The sequential procedure in the vector case is similar to the scalar one (Chapter 1, Section 1.6 , Figure 1.5): an initial DoE is generated and an initial MOGP is fitted, then they are followed by a succession of MOGP updates and optimizations from a well-chosen enrichment criterion.

4.3.2 The proposed Vector Bichon criterion

In this section, we look at a generalization of Bichon criterion (Section 4.1.1) in the context of multi-output Gaussian process regression, again with the aim of estimating different partial excursion sets simultaneously.

Recall that $\Sigma_n(\mathbf{x})$ is the covariance matrix of $\boldsymbol{\xi}_n(\mathbf{x}) := (\xi_{n,1}(\mathbf{x}), \dots, \xi_{n,p}(\mathbf{x}))^\top = \boldsymbol{\xi}(\mathbf{x}) \,|\, \boldsymbol{\mathcal{E}}_n$. We extend scalar Bichon criterion (Equations (4.6) and (4.7)) to the vector setting as follows:

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}}\ \mathrm{VEFF}(\mathbf{x}) \tag{4.24}$$

with

$$\mathrm{VEFF}(\mathbf{x}) := \det\big(\Sigma_n(\mathbf{x})\big)^{\frac{1}{2p}}\, \mathbb{E}\left[\left(\kappa - \min_i \left(\frac{|T_i - \xi_i(\mathbf{x})|}{\sigma_{n,i}(\mathbf{x})}\right)\right)^+ \,\middle|\, \mathcal{E}_n\right] \tag{4.25}$$

and $\sigma_{n,i}(\mathbf{x}) := \sqrt{(\Sigma_n(\mathbf{x}))_{i,i}}$. Using the minimum allows us to select a point that improves our knowledge of at least one of the partial excursion sets we are looking for. The multiplicative term $\sigma_n(\mathbf{x})$ of Scalar Bichon criterion, which helps improving the exploratory power, has been replaced here by the determinant of the covariance matrix $\Sigma_n(\mathbf{x})$ raised to the power $1/(2p)$. This multiplicative term gives a smaller weight to points $\mathbf{x}$ for which the components of $\boldsymbol{\xi}_n(\mathbf{x})$ have strong correlation. The $1/(2p)$ power ensures homogeneity with the scalar case, allowing us to have a term in $\sigma_n(\mathbf{x})$ when all marginal variables are independent. It is important to note that the correlation of $\boldsymbol{\xi}$ plays a role not only in the multiplicative term with the determinant, but also in the expectation, since this expectation involves a minimum on random variables, thus structure depending on the correlation.

Initially, a criterion using the minimum over $B(\mathbf{x})^{-1}(\boldsymbol{T} - \boldsymbol{\xi}(\mathbf{x}))$ was considered, where $B(\mathbf{x})B(\mathbf{x})^\top$ represents the Cholesky decomposition of $\Sigma_n(\mathbf{x})$. This approach would decorrelate the variable $\boldsymbol{\xi}(\mathbf{x})$, simplifying the expectation calculations while accounting for output correlations within the $B(\mathbf{x})^{-1}$ term. However, this does not fully align with the objective,

which is to select an $\mathbf{x}$ such that at least one of the $\xi_i(\mathbf{x})$ is close to T_i, rather than ensuring that at least one of the $(B(\mathbf{x})^{-1}\boldsymbol{\xi}(\mathbf{x}))_i$ is close to $(B(\mathbf{x})^{-1}\boldsymbol{T})_i$.

In the following, we restrict ourselves to the case of a vector-valued black box model with two output components ($p = 2$). Proposition 1 below presents an explicit formulation of Vector Bichon criterion, facilitating its practical implementation. The proof of this proposition is given in Appendix A.

Proposition 1. Noting $F_{Y_\mathbf{x}}$ the cumulative distribution function of $Y_\mathbf{x} := \min_i \left(\frac{|T_i - \xi_i(\mathbf{x})|}{\sigma_{n,i}(\mathbf{x})} \right) \Big| \boldsymbol{\mathcal{E}}_n$, we have

$$\mathrm{VEFF}(\mathbf{x}) = \det\big(\Sigma_n(\mathbf{x})\big)^{\frac{1}{2p}} \int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t. \tag{4.26}$$

Furthermore, for every $\mathbf{x} \in \mathbb{X}$ and every $t \in \mathbb{R}$,

$$F_{Y_\mathbf{x}}(t) = \sum_{i=1}^{2} \Big(\phi(t+\alpha_i) - \phi(-t+\alpha_i) \Big) - \mathbb{P}\big((U_1, U_2) \in [\alpha_1 - t, \alpha_1 + t] \times [\alpha_2 - t, \alpha_2 + t]\big), \tag{4.27}$$

with $\alpha_i := \frac{T_i - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})}$, ϕ c.d.f of $\mathcal{N}(0,1)$, $(U_1, U_2) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$ and ρ the correlation coefficient between $\xi_{n,1}(\mathbf{x})$ and $\xi_{n,2}(\mathbf{x})$.

It should be pointed out that the values α_i and ρ, defined in Proposition 1, are functions of n and $\mathbf{x}$, although this dependence is not explicitly stated in the simplified notation. Note also that in Equation (4.27), $F_{Y_\mathbf{x}}$ depends only on α_i and ρ and the integral of $F_{Y_\mathbf{x}}$ in Equation (4.26) between 0 and κ can thus be calculated outside the DoE sequential enrichment loop (see Section 4.4.1 and Appendix B).

4.4 Numerical experiments

We present here the numerical experiments carried out on two analytical test functions in order to compare the different criteria proposed in Sections 4.2 and 4.3 We begin by detailing the choices made for the implementation of the acquisition criteria and the parameters linked to enrichment of the DoE. We then present the first results obtained on a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ based on the Branin function in 2D ([Picheny et al., 2013]). Finally, we present the results obtained on a function from $\mathbb{R}^4$ to $\mathbb{R}^2$ based on the Hartmann function in 4D ([Picheny et al., 2013]).

The performances of the different criteria are compared after n iterations by computing the partial relative approximation error

$$\mathrm{Err}_i := \frac{\mathbb{P}_\mathbb{X}(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mathbb{P}_\mathbb{X}(\Gamma_i^\star)} \tag{4.28}$$

on each output component, with $\hat{\Gamma}_i$ the naive estimator $\hat{\Gamma}_i := M_{n,i}^{-1}(]-\infty, T_i])$. Each partial relative error measures, for the i^{th} output component, the relative volume of the symmetric difference between estimator $\hat{\Gamma}_i$ and true partial excursion set $\Gamma_i^\star$ defined in (4.8). The sum of partial relative errors defined by

$$\mathrm{Err}_{\mathrm{sum}} := \sum_i \frac{\mathbb{P}_\mathbb{X}(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mathbb{P}_\mathbb{X}(\Gamma_i^\star)}, \tag{4.29}$$

is also computed to analyse performances of the various acquisition criteria.

4.4.1 Implementation choices

Criteria implementation

The implementation of Alternating and Pareto Scalar Bichon criteria are respectively based on Equations (4.13) and (4.14). For Scalar Pareto Bichon criterion, which requires scaled data, we subtract the mean of the values obtained on the initial DoE, then divide by the standard deviation of these values. The Pareto front is obtained using the NSGA-II algorithm in the *mco* package ([Trautmann et al., 2013]), with 150 generations of size 100.

The implementation of Vector Bichon criterion is based on Equations (4.26) and (4.27) of Proposition 1. As previously discussed, according to Equation (4.27), $F_{Y_{\mathbf{x}}}$ depends solely on α_i and ρ. We thus propose to approximate the function

$$H : \left\{ \begin{array}{ccc} \mathbb{R} \times \mathbb{R} \times [-1,1] & \longrightarrow & \mathbb{R} \\ (\alpha_1, \alpha_2, \rho) & \longmapsto & \int_0^\kappa F_{Y_{\mathbf{x}}}(t)\,\mathrm{d}t \end{array} \right.$$

outside the enrichment procedure. The function H is approximated with a Gauss-Legendre quadrature method (see Appendix B for details on interpolation and quadrature methods). Then at each iteration of the enrichment the approximation of H is evaluated at $(\alpha_1, \alpha_2, \rho)$ with $\alpha_i := \frac{T_i - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})}$ and ρ the correlation coefficient between $\xi_{n,1}(\mathbf{x})$ and $\xi_{n,2}(\mathbf{x})$.

Model selection

For both scalar Bichon criteria (Alternating and Pareto), the ξ_is are assumed to be GPs with constant mean and Matérn 5/2 covariance:

$$k(\mathbf{x}, \mathbf{x}') := \sigma_{\mathrm{c}}^2 \prod_{j=1}^{d} R_{\text{Matérn } 5/2}(h_j, \theta_j), \quad \forall\, (\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2, \tag{4.30}$$

with a vector of parameters θ belonging to $\mathbb{R}_+^{\star\, d}$, estimated by maximizing the likelihood at each iteration, $h_j = |\mathbf{x}_j - \mathbf{x}'_j|$ for $j \in \{1, \ldots, d\}$, σ_{c} a fixed parameter and

$$R_{\text{Matérn } 5/2}(h_j, \theta_j) := \left(1 + \frac{\sqrt{5}|h_j|}{\theta_j} + \frac{5h_j^2}{3\theta_j^2}\right) \exp\left(-\frac{\sqrt{5}|h_j|}{\theta_j}\right). \tag{4.31}$$

Moreover, ξ_1 and ξ_2 are assumed to be independant from each other.

For Vector Bichon criterion, we use a MOGP with an ICM-type separable kernel (see [Goovaerts, 1997]). Separable models simplify output correlations, improving parameter estimation and efficiency while offering flexibility in multi-output Gaussian processes, as demonstrated in [Alvarez et al., 2012] and [Liu et al., 2018]. They can be expressed as a product of a kernel function in the input space and a kernel function that encodes inter-output correlations. Specifically,

$$\left(K(\mathbf{x}, \mathbf{x}')_{i,j}\right)_{1 \leq i,j \leq p} := k(\mathbf{x}, \mathbf{x}')\, \mathbf{B}, \tag{4.32}$$

where k is a scalar kernel on $\mathbb{X}^2$ and $\mathbf{B}$ is a $p \times p$ symmetric positive definite matrix. Hyperparameters of the covariance structure can be estimated using maximum likelihood (see, e.g., [Bonilla et al., 2007]).

As outlined in [Alvarez et al., 2012] and [Liu et al., 2018], different structures for the matrix $\mathbf{B}$ can reduce the number of parameters to estimate (see Chapter 1, Section 1.3). The choice of $\mathbf{B}$ in our case follows [Pelamatti et al., 2024], which corresponds to a special case of

the "spherical parametrization" introduced by [Osborne, 2007]. For $p = 2$, the matrix $\mathbf{B}$ is parameterized as:

$$\mathbf{B} = \sigma^2_{kOut} \begin{pmatrix} 1 & \cos(\theta_{kOut}) \\ \cos(\theta_{kOut}) & 1 \end{pmatrix},$$

where $\sigma_{\text{kOut}} \in \mathbb{R}$ is a scaling factor common to all components (homoscedasticity), and $\theta_{kOut} \in [0, \pi]$ is the parameter controlling inter-output correlation under the spherical parameterization. The scalar kernel k is selected as in Equation (4.30), with $\sigma_c = 1$.

The different scalar and vector models presented above are implemented using the *kergp* package from [Deville et al., 2015]. Hyperparameter optimization is performed using a LN COBYLA algorithm with a multistart of 10, on the following search ranges: $\theta_j \in [0.1, 500]$, $\theta_{kOut} \in [0, \pi]$, $\sigma_c^2 \in [10^{-1}, 10^5]$ et $\sigma^2_{kOut} \in [10^{-1}, 10^5]$.

In the following, we will refer to the two scalar strategies associated with Alternating and Pareto Bichon criteria respectively as "Alternating Scal" and "Pareto Scal". Vector strategy associated with Vector Bichon criterion will be referred to as "Vect".

Other parameters

Criteria are optimized with the genetic algorithm Genoud (with $pop.size = 1000$) (package *rgenoud*, [Mebane Jr and Sekhon, 2011]). Initial DoEs are Latin Hypercube Sampling (LHS) designs optimized from Maximin distance ([Dupuy et al., 2015]), with size to be specified in the following.

Partial relative errors Err_i and their sum Err_{sum} (Equations (4.28) and (4.29)) are estimated from a grid of size 400×400 for the example based on Branin function, and from a Sobol' sequence (package *randtoolbox* [Dutang and Savicky, 2013]) of size 10 000 for the example based on Hartmann function.

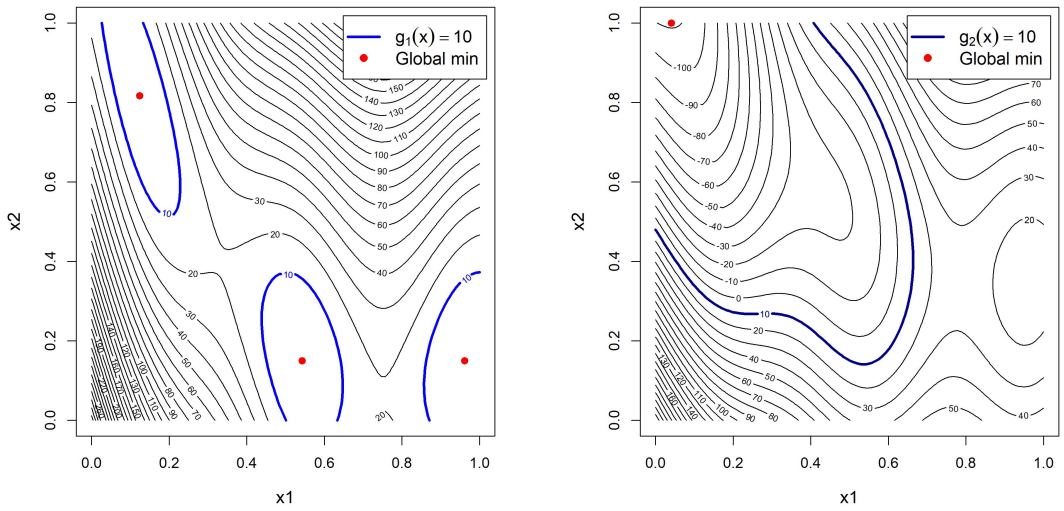
4.4.2 Performance tests based on the 2D Branin function

A model based on the 2D Branin function with two output components

In this Section, we evaluate performances of the proposed criteria on the 2D Branin function. The first output component is identical to Branin function (see [Roustant et al., 2012] and [Picheny et al., 2013]) and the second component is a modification of this function, as presented in [Liu et al., 2018]. This is the same test function as for Figure 4.1, whose expression is given by Equations (4.10) and (4.11). For this test case, we choose a threshold vector $\boldsymbol{T} := (10, 10)$. The partial excursion set associated with the first output component has three connected components, while that associated with the second component has only one. The level sets of the test function under consideration, as well as the partial excursion sets associated with $\boldsymbol{T}$, are plotted in Figure 4.3.

Standard Performance tests

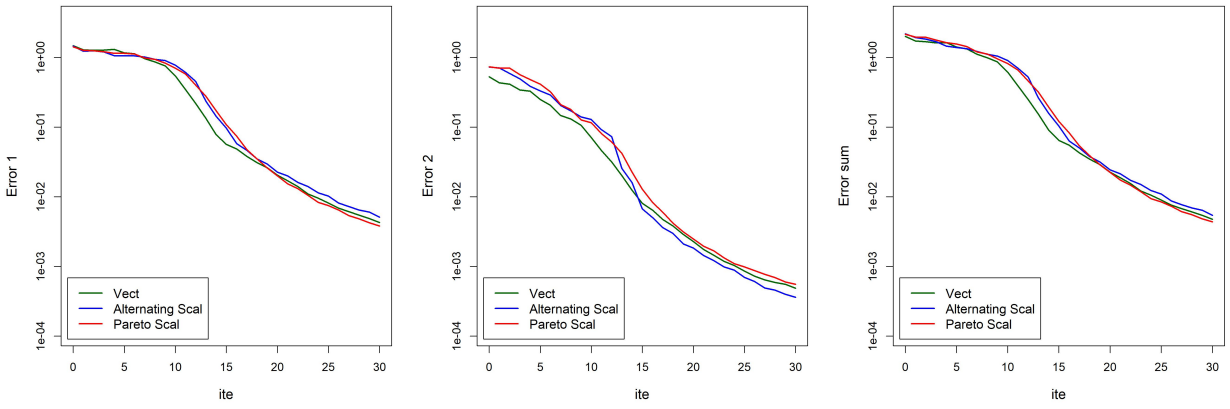
We study the enrichment of 40 initial LHS Maximin DoEs of size 5. The study is carried out over 30 iterations for the different DoE enrichment strategies, by studying the partial relative errors and their sum (Equations (4.28) and (4.29)). Figure 4.4 shows, in logarithmic scale, the average errors over the 40 initial DoEs, as a function of the number of iterations. The results show good performance for the three strategies studied, with errors decreasing significantly as the number of iterations grows. Note that for all three strategies, the relative partial error associated with the second component decreases faster than that associated with the first component, in line with the relative simplicity of the partial excursion set $\Gamma_2^\star$ compared to



(a) $g_1(\mathbf{x})$ (b) $g_2(\mathbf{x})$

Figure 4.3: Representation of the model based on the 2D Branin function with the first component on the left and the second one on the right and the contour of excursion sets in blue (with $\boldsymbol{T} = (10, 10)$).

$\Gamma_1^\star$ (see Figure 4.4). Despite an apparent slight advantage for Vector strategy at the start of enrichment, the graph shows few significant differences between the various enrichment strategies.



(a) Err_1 (b) Err_2 (c) $\mathrm{Err}_{\mathrm{sum}}$

Figure 4.4: Averaged partial relative errors and of their sum with respect to the number of iterations when estimating excursion sets of the enhanced 2D Branin test function with $\boldsymbol{T} = (10, 10)$. The three enrichment stategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Averages are evaluated over 40 LHS Maximin intial DoEs randomly chosen.

We plotted functional boxplots of partial relative errors and their sum. Functional box-

plots, as an extension of traditional boxplots, are used to visualize and summarize sets of curves or functional data, showing the distribution, median, quartiles and any outliers. We use the *fda* package with the Modified Band Depth (MBD) algorithm (see [Sun and Genton, 2011]).

Based on the center outward ordering induced by band depth for functional data, the main indicators of a functional boxplot are: the median curve, the envelope of the 50% central region ('Box') and the maximum non-outlying envelope ('Bar') based on the concept of band depth ([Sun and Genton, 2011]). Curves outside this envelope are considered outliers.

The functional boxplots of the partial relative errors and their sum for the three enrichment strategies are shown in Figure 4.5. These graphical representations confirm the differences in the estimation of partial excursion sets for the two output components, while showing no notable differences between the different strategies.

Standard "Data profiles"

In this section, we analyze the data on partial relative errors in a way that differs from functional boxplots. The approach is to examine, for a given iteration and a fixed threshold, the percentage of relative error curves that are definitively below that threshold from that iteration onwards. These graphs, which we call "Data profiles", provide a clear visualization of data conformity in relation to a specified threshold. This approach is particularly relevant in industrial contexts where it is crucial that errors remain below a defined tolerance threshold.

More precisely, let ite denote the iteration stage and C denote a threshold value, the partial "Data profile" associated with the i^{th} output component is given by :

$$\mathrm{DP}_i(\mathrm{ite}, C) := 100 \times \frac{\#\{j \in \{1, \ldots, \mathrm{N_{LHS}}\}, \forall\, k \geq \mathrm{ite}, \mathrm{Err}_i^{(j)}(k) < C\}}{\mathrm{N_{LHS}}}, \tag{4.33}$$

with $\mathrm{N_{LHS}}$ the number of repetitions, $\mathrm{Err}_i^{(j)}$ the error for the i^{th} component and the j^{th} initial DoE and for any set A, $\#\{A\}$ the cardinal of A.

The total "Data profile" corresponding to the percentage of errors definitely below threshold C for both output components after iteration ite is given by :

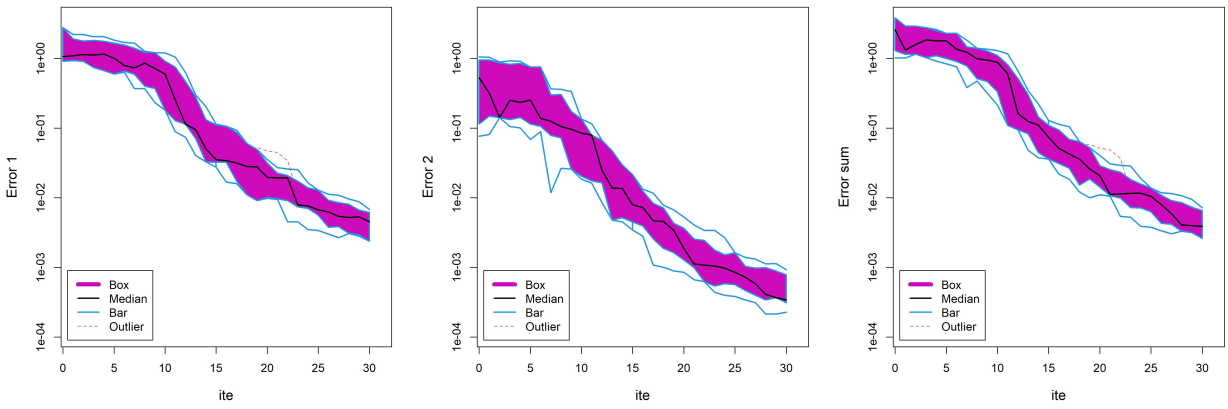
$$\mathrm{DP}_{\mathrm{tot}}(\mathrm{ite}, C) := 100 \times \frac{\#\{j \in \{1, \ldots, \mathrm{N_{LHS}}\}, \forall\, i, \forall\, k \geq \mathrm{ite}, \mathrm{Err}_i^{(j)}(k) < C\}}{\mathrm{N_{LHS}}}. \tag{4.34}$$

Figure 4.6 shows "Data profiles" (Equations (4.33) and (4.34)) associated with the partial relative errors, for thresholds $C = 20\%$, 10% and 5%. We observe a rapid growth of the "Data profiles" from 0 to 100%, which is consistent with the relative simplicity of the excursion set estimation problem under consideration (see Figure 4.4). We also observe that for the different threshold values, Vector strategy seems to perform slightly better than the other two, as its data profiles reach 100% more quickly.

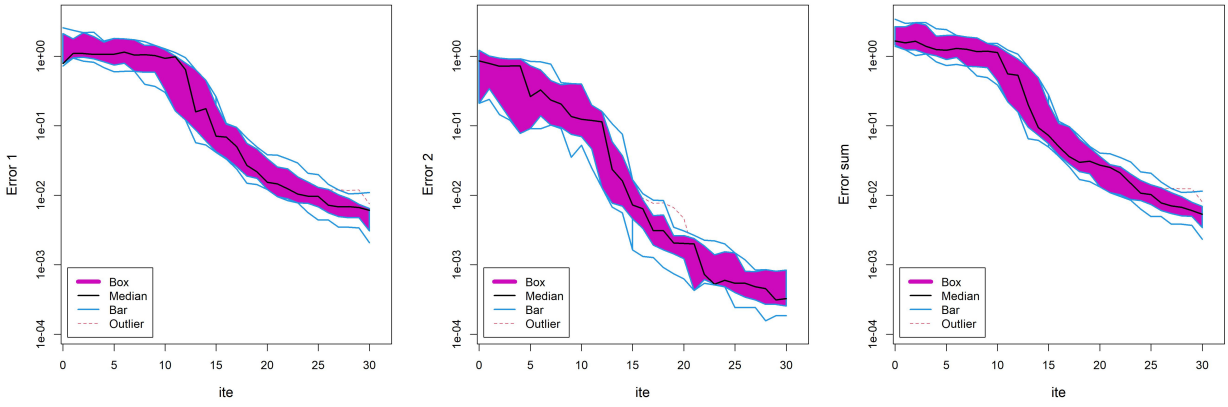
These results must be set against the higher computation times of Vector strategy compared with the two Scalar strategies (Figure 4.7). This Figure shows that the computation time for Vector strategy increases exponentially with the number of iterations. In the next section, we will plot data profiles with respect to total computation time (enrichment and $\boldsymbol{g}$ evaluations) rather than the number of iterations.

"Data profiles" as a function of total computing time

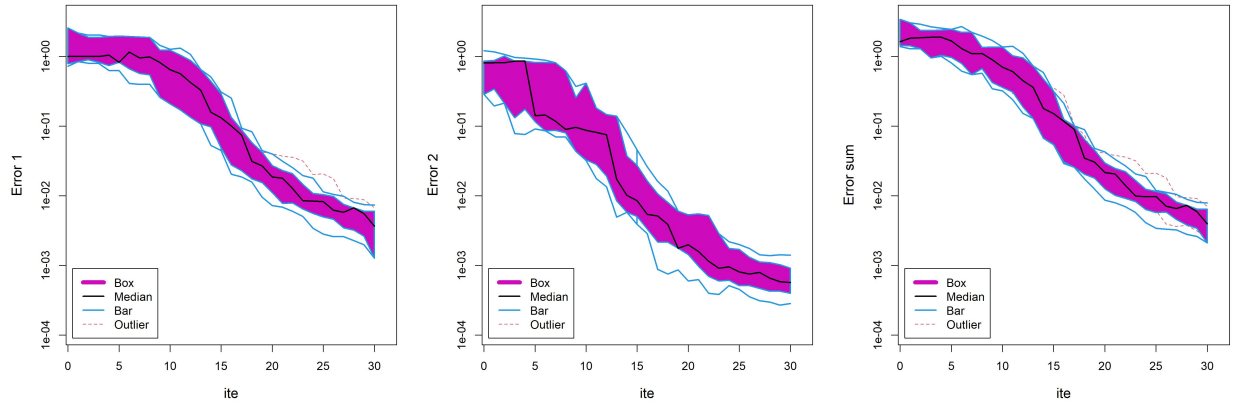
We now represent "Data profiles" as a function of total computation time, rather than the number of iterations. By total computation time, we mean the time required for enrichment (optimization of acquisition criteria and updating surrogate model) as well as the time required



(a) Vect, Err_1 (b) Vect, Err_2 (c) Vect, $\mathrm{Err}_{\mathrm{sum}}$

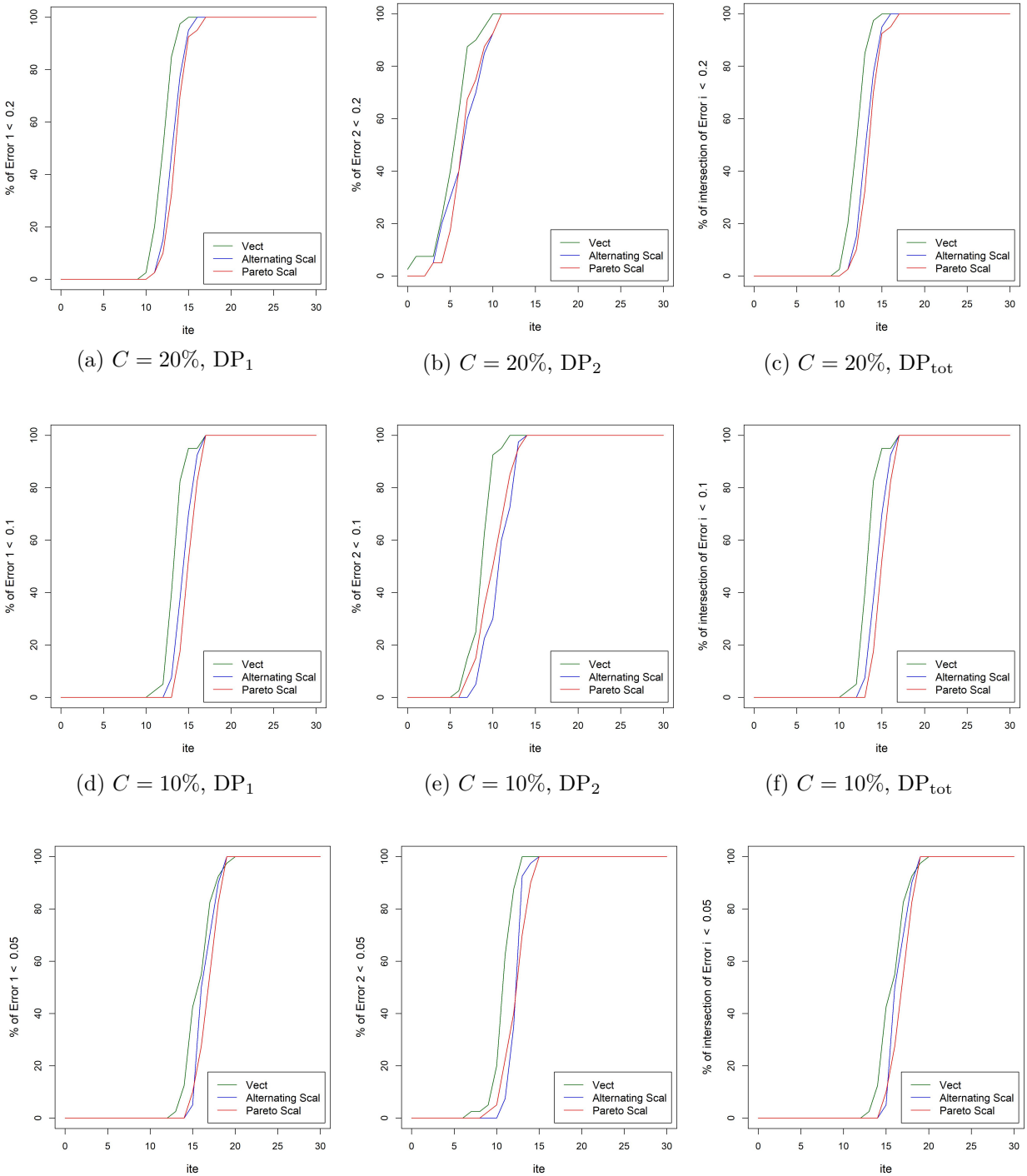


(d) Alternating Scal, Err_1 (e) Alternating Scal, Err_2 (f) Alternating Scal, $\mathrm{Err}_{\mathrm{sum}}$



(g) Pareto Scal, Err_1 (h) Pareto Scal, Err_2 (i) Pareto Scal, $\mathrm{Err}_{\mathrm{sum}}$

Figure 4.5: Functional boxplots of partial relative errors and of their sum, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin test function with $\boldsymbol{T} = (10, 10)$.

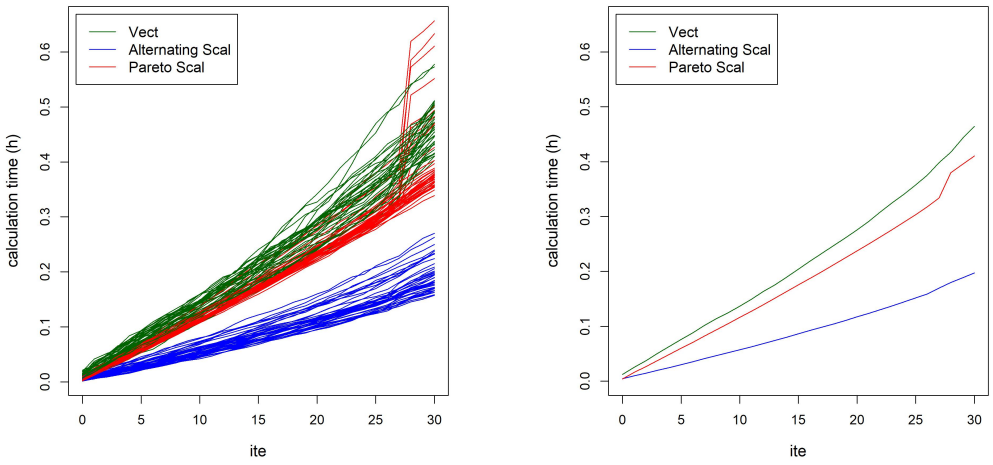


(a) $C = 20\%$, DP_1 (b) $C = 20\%$, DP_2 (c) $C = 20\%$, $\mathrm{DP}_{\mathrm{tot}}$

(d) $C = 10\%$, DP_1 (e) $C = 10\%$, DP_2 (f) $C = 10\%$, $\mathrm{DP}_{\mathrm{tot}}$

(g) $C = 5\%$, DP_1 (h) $C = 5\%$, DP_2 (i) $C = 5\%$, $\mathrm{DP}_{\mathrm{tot}}$

Figure 4.6: Standard "Data profiles" of partial relative errors with threshold set equal to 20%, 10% and 5%, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10)$.



(a) Line plots

(b) Means

Figure 4.7: Enrichment calculation time (assuming that the evaluation time of $\boldsymbol{g}$ is negligible) for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10)$.

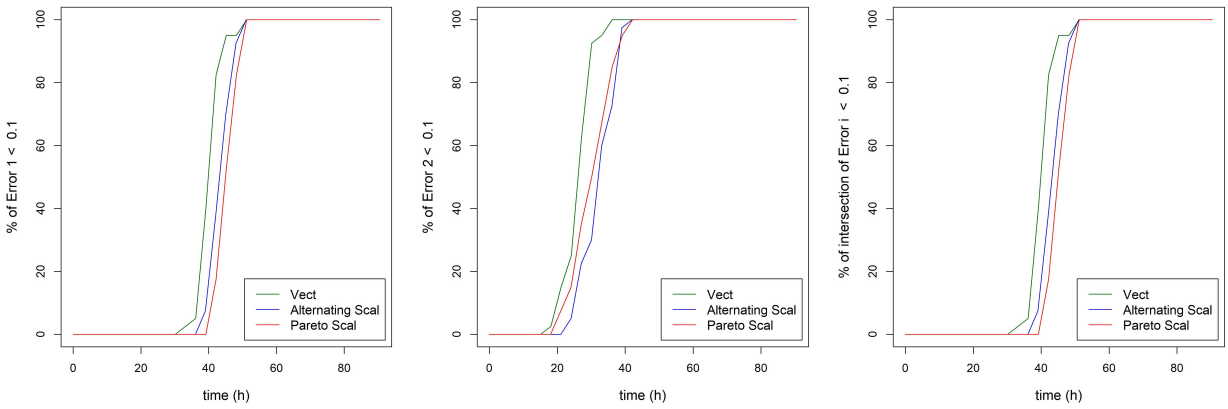
for evaluations of black box model $\boldsymbol{g}$. To do this, we set a threshold $C = 10\%$ and consider several evaluation times for $\boldsymbol{g}$: 3h, 10min and 1min. Data profiles are obtained in a similar way to Equations (4.33) and (4.34), replacing iteration by total computation time.

The results obtained (Figure 4.8) show that for long evaluation times of $\boldsymbol{g}$ (3h), curves are analogous to standard data profiles (Figure 4.7). However, when the evaluation time of $\boldsymbol{g}$ is shorter, Vector strategy becomes less advantageous and the strategy associated with Alternating Scalar criterion seems much more interesting. This is due to the significant computation time (a few minutes) required to update the vector surrogate model, which is negligible when the evaluation time of $\boldsymbol{g}$ is 3h. It should also be noted that the strategy associated with Pareto Scalar criterion is more expensive in computation time than that associated with Alternating Scalar criterion (Figure 4.7), which explains its inferior performance when the computation time of $\boldsymbol{g}$ is very short (1min).

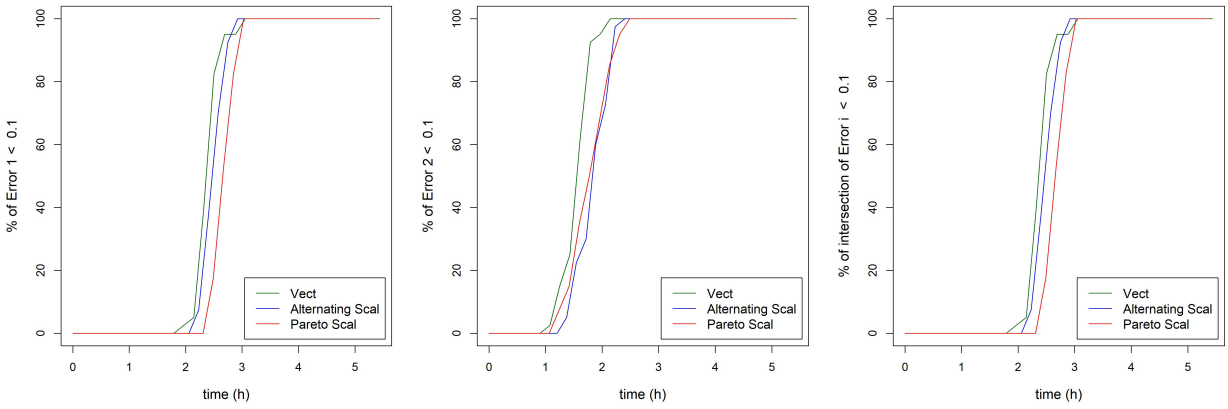
Failure of Alternating Scalar Bichon criterion for an inappropriate choice of threshold $\boldsymbol{T}$

Before moving on to the 4D test case, we look at a test case with an inappropriate choice of threshold $\boldsymbol{T}$ to show the limits of Alternating Scalar criterion. The test function is still based on the 2D Branin function, but we choose $\boldsymbol{T} = (10, 10\,000)$ instead of $\boldsymbol{T} = (10, 10)$. The partial excursion set associated with the first output component is the same as in the previous setting (Figure 4.3a), while that of the second component corresponds to the full design space $\mathbb{X}$.

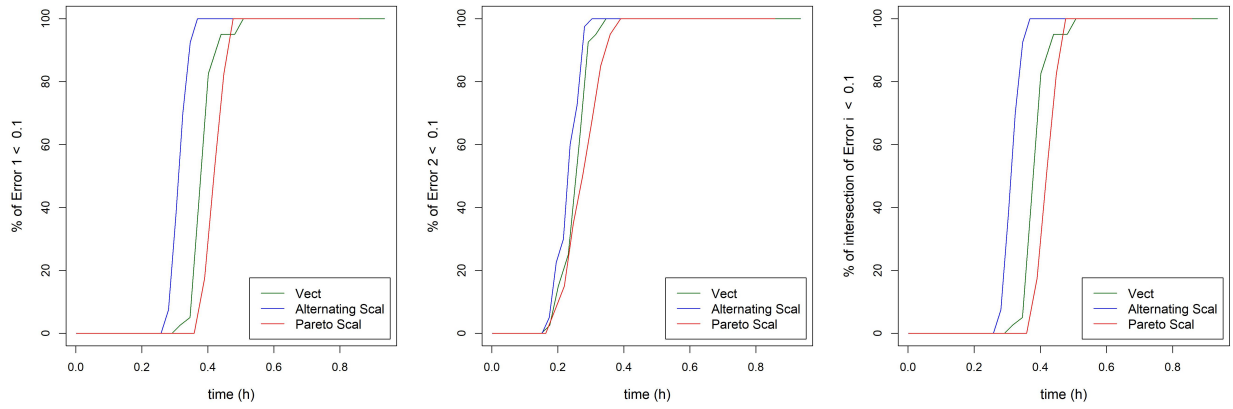
We are still considering the enrichment of 40 initial LHS Maximin DoEs of size 5, with 30 iterations. Figure 4.9, which plots means of partial relative errors and of their sum, shows that Alternating Scalar criterion performs much worse than the other two criteria, with jerky improvements. This is consistent with the fact that Alternating Scalar criterion go enriching the DoE with respect to the second component (one in two) while the excursion set associated to this component is already fully identified. Moreover, for all strategies, the error associated



(a) Eval $\boldsymbol{g} = 3\text{h}$, DP_1 (b) Eval $\boldsymbol{g} = 3\text{h}$, DP_2 (c) Eval $\boldsymbol{g} = 3\text{h}$, DP_{tot}



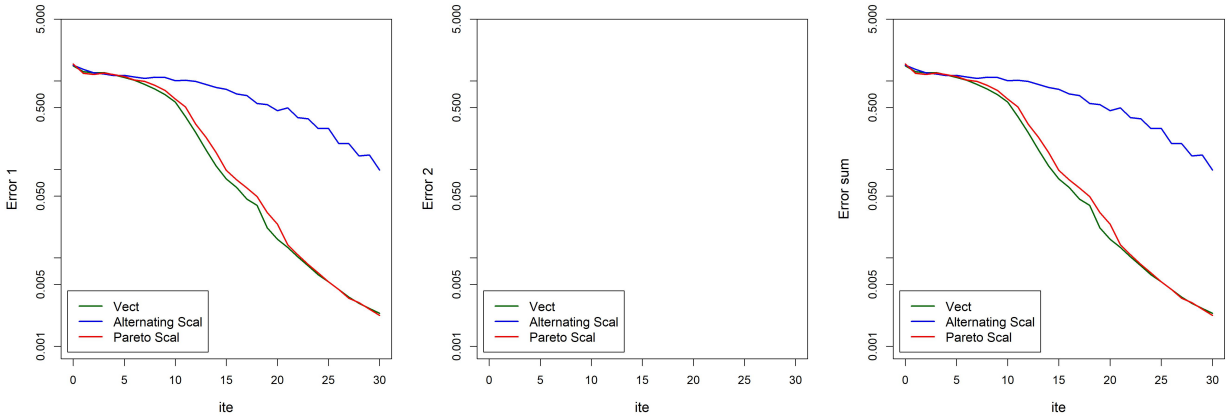
(d) Eval $\boldsymbol{g} = 10\text{min}$, DP_1 (e) Eval $\boldsymbol{g} = 10\text{min}$, DP_2 (f) Eval $\boldsymbol{g} = 10\text{min}$, DP_{tot}



(g) Eval $\boldsymbol{g} = 1\text{min}$, DP_1 (h) Eval $\boldsymbol{g} = 1\text{min}$, DP_2 (i) Eval $\boldsymbol{g} = 1\text{min}$, DP_{tot}

Figure 4.8: "Data profiles" as a function of total computing time of partial relative errors for evaluation times of $\boldsymbol{g}$ of 3h, 10min and 1min, with $C = 10\%$, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10)$.

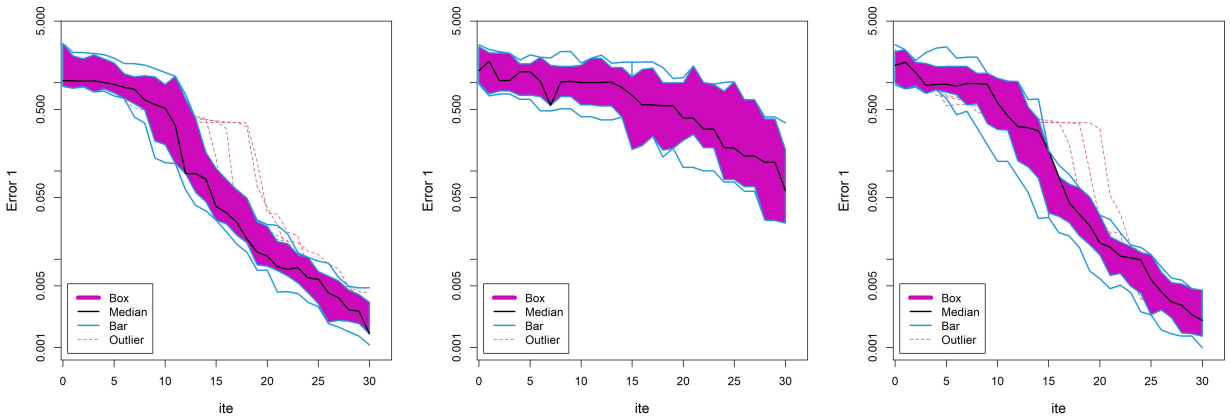
with the second output component is zero, since the partial excursion set $\Gamma_2^\star$ is very simple to identify.



(a) Err_1 (b) Err_2 (c) $\mathrm{Err}_{\mathrm{sum}}$

Figure 4.9: Averaged partial relative errors and of their sum with respect to the number of iterations when estimating excursion sets of the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10\,000)$. The three enrichment stategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Averages are evaluated over 40 LHS Maximin intial DoEs randomly chosen.

The functional boxplots corresponding to the error on the first component (Figure 4.10) confirm that Alternating Scalar Bichon criterion fails for this choice of $\boldsymbol{T}$. In contrast, Vector and Pareto Scalar strategies continue to perform well, despite the substantial difference in difficulty between the estimation of the two partial excursion sets. Thus, Alternating Scalar criterion is highly dependent on the balance between the difficulty of estimating different partial excursion sets, which is a real weakness compared to other strategies.



(a) Vect, Err_1 (b) Alternating Scal, Err_1 (c) Pareto Scal, Err_1

Figure 4.10: Functional boxplots of partial relative errors for the first component, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 5 with 30 iterations, for the model based on the 2D Branin function with $\boldsymbol{T} = (10, 10\,000)$.

4.4.3 Performance tests based on the 4D Hartmann function

A model based on the 4D Hartmann function with two output components

The test function used is based on the 4D Hartmann function, and is composed with two output components. The first component is the 4D Hartmann function ([Picheny et al., 2013]) defined by

$$\forall\, \mathbf{x} \in [0,1]^4,\ g_1(\mathbf{x}) = \frac{1}{0.839}\left[1.1 - \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{4} A_{ij}\left(x_j - P_{ij}\right)^2\right)\right], \tag{4.35}$$

with

$$(A_{ij}) = \begin{pmatrix} 10 & 3 & 17 & 3.50 \\ 0.05 & 10 & 17 & 0.1 \\ 3 & 3.5 & 1.7 & 10 \\ 17 & 8 & 0.05 & 10 \end{pmatrix}, \quad (P_{ij}) = 10^{-4} \begin{pmatrix} 1312 & 1696 & 5569 & 124 \\ 2329 & 4135 & 8307 & 3736 \\ 2348 & 1451 & 3522 & 2883 \\ 4047 & 8828 & 8732 & 5743 \end{pmatrix}$$

and $\alpha = (1.0, 1.2, 3.0, 3.2)^\top$. The second output component is obtained by modifying the 4D Hartmann function, adding noise from a random variable uniformly distributed over the interval $[-1, 1]$ to the coefficients α and A :

$$A' = A + 5A_{\text{noise}} \quad \text{et} \quad \alpha' = \alpha + \alpha_{\text{noise}}.$$

No modification of P is made to preserve a certain correlation between the two output components. Rounding the values of $5A_{\text{noise}}$ and α_{noise} to the hundredth and the unit respectively, we finally obtain

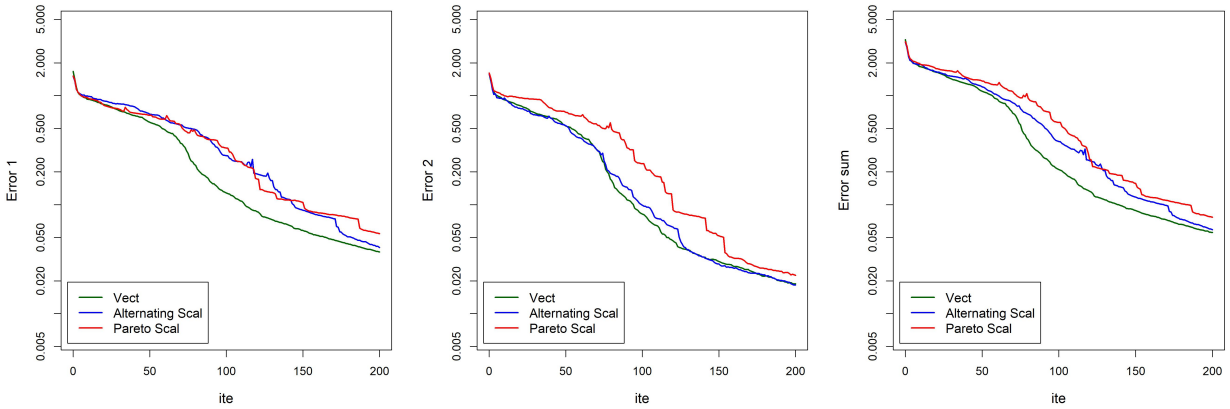
$$(A'_{ij}) = \begin{pmatrix} 7 & -2 & 14 & 0.50 \\ -0.95 & 9 & 19 & 5.1 \\ 7 & 1.5 & 1.7 & 10 \\ 16 & 13 & 4.05 & 5 \end{pmatrix} \quad \text{and} \quad \alpha' = (0.44, 0.25, 2.41, 2.63)^\top.$$

The choice of threshold vector $\boldsymbol{T} = (-1, -1.6)$ yields relative volumes around 9.5% (relatively to the volume of the total design space) for each partial excursion set, while the volume of the global excursion set represents 5.67% of the one of total design space. Clustering tests on the data have been carried out and show that each of the partial excursion sets appears to comprise a single connected component.

Standard Performance tests

In the context of the enrichment of 40 LHS Maximin initial DoEs of size 20, we compare over 200 iterations the different DoE enrichment strategies by studying the partial relative errors and their sum (Equations (4.28) and (4.29)). Figure 4.11 shows, in logarithmic scale, the averaged errors on the initial 40 DoEs, as a function of the number of iterations in the sequential enrichment strategy. The results show good performance for all three strategies, with errors decreasing significantly.

Alternating Scalar Bichon criterion performs well in estimating the partial excursion set associated with the second component. However, this increased performance on the second component seems to be detrimental to the estimation of the partial excursion set associated with the first one. Vector Bichon criterion better manages the trade-off between simultaneous learning of both components, by concentrating efforts where they are most needed, i.e., on



(a) Err_1 (b) Err_2 (c) $\mathrm{Err}_{\mathrm{sum}}$

Figure 4.11: Averaged partial relative errors and of their sum with respect to the number of iterations when estimating excursion sets of the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$. The three enrichment stategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 20 and with 200 enrichment iterations. Averages are evaluated over 40 LHS Maximin intial DoEs randomly chosen.

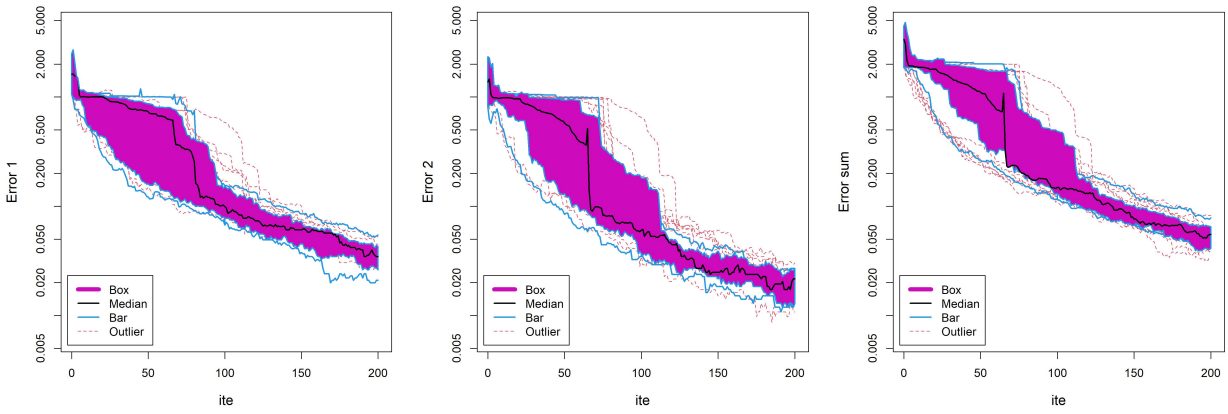
the first component. Pareto Scalar Bichon criterion performs slightly less than the other two criteria in estimating partial excursion sets, but is nonetheless relatively efficient.

We also plot functional boxplots of partial relative errors and their sum in Figure 4.12. For details on the interpretation of functional boxplots, we refer to Section 4.4.2. These graphical representations confirm the results obtained with the averages, showing a better compromise in the simultaneous search for the two partial excursion sets for Vector criterion (see Figure 4.12c versus Figures 4.12f and 4.12i). The functional boxplots also reveal a better robustness of Vector criterion, with less instability than the other strategies, which are characterized by "spikes" in functional boxplots.

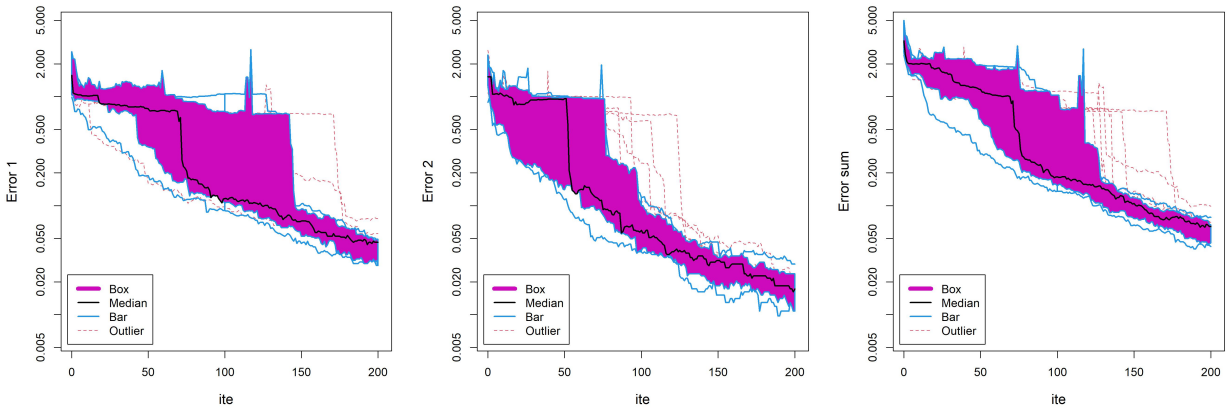
Standard "Data profiles"

In this section, we analyze "Data profiles" of the partial relative errors, in a similar way to what was done in Section 4.4.2 for the enhanced 2D Branin function (see Equations (4.33) and (4.34)). "Data profiles" show, as a function of the number of iterations and for a given threshold, the percentage of initial DoEs tested for which the relative partial error is definitely below a prescribed threshold. The total "Data profile" corresponds to the percentage of errors that are definitively below the threshold for both output components.

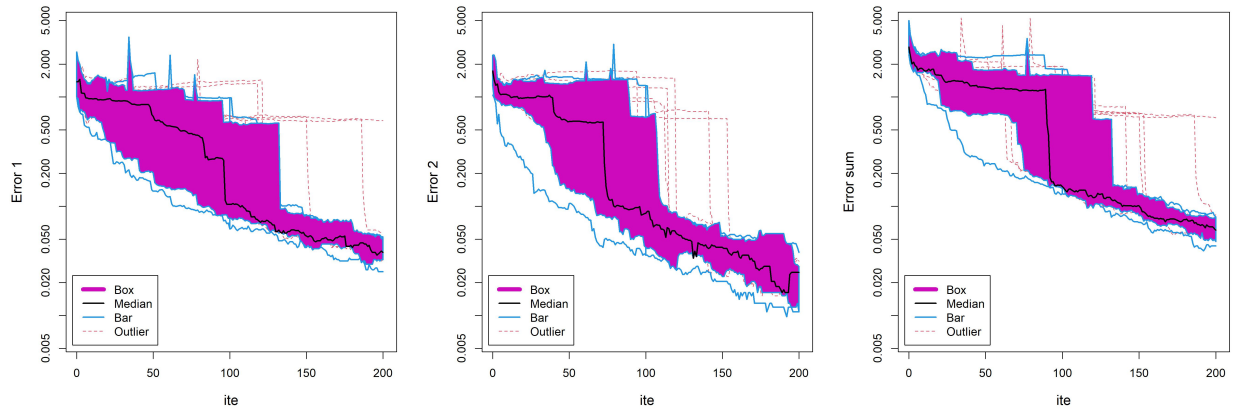
Figure 4.13 shows "Data profiles" associated with partial relative errors for thresholds values $C = 20\%$, 10% and 5%. We observe slower growth of the "Data profiles" from 0 to 100%, compared to that of the model based on the 2D Branin function (Figure 4.6). This is due to the increased dimension of design space $\mathbb{X}$, which renders the estimation of partial excursion sets more difficult. For the different thresholds, we observe that Vector strategy offers a better balance in estimating the two partial excursion sets. Pareto Scalar strategy also performs well, rivalling Vector strategy, particularly for thresholds set to 10% and 5%. Alternating Scalar strategy appears to perform less well overall than the other two, particularly with regards to first output component.



(a) Vect, Err_1 (b) Vect, Err_2 (c) Vect, $\mathrm{Err}_{\mathrm{sum}}$

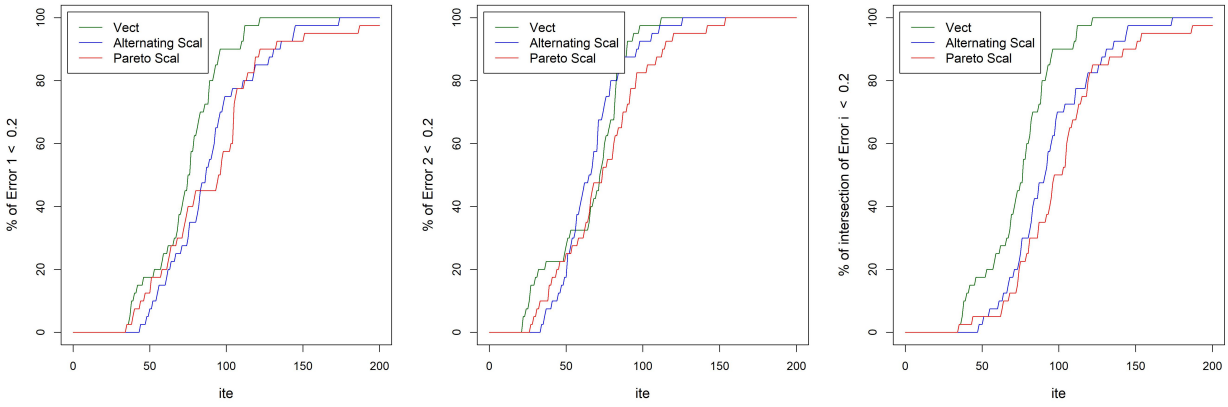


(d) Alternating Scal, Err_1 (e) Alternating Scal, Err_2 (f) Alternating Scal, $\mathrm{Err}_{\mathrm{sum}}$

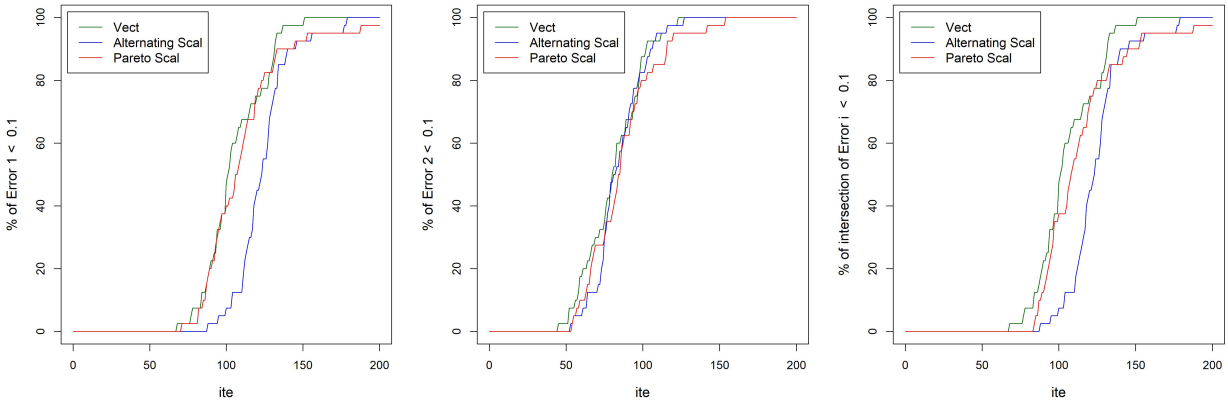


(g) Pareto Scal, Err_1 (h) Pareto Scal, Err_2 (i) Pareto Scal, $\mathrm{Err}_{\mathrm{sum}}$

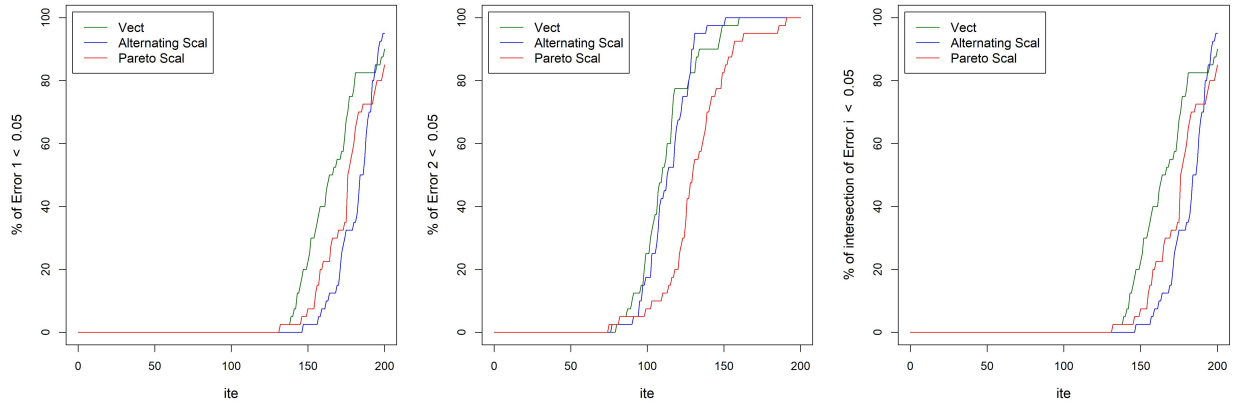
Figure 4.12: Functional boxplots of partial relative errors and of their sum, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$.



(a) $C = 20\%$, DP_1 (b) $C = 20\%$, DP_2 (c) $C = 20\%$, $\mathrm{DP}_{\mathrm{tot}}$



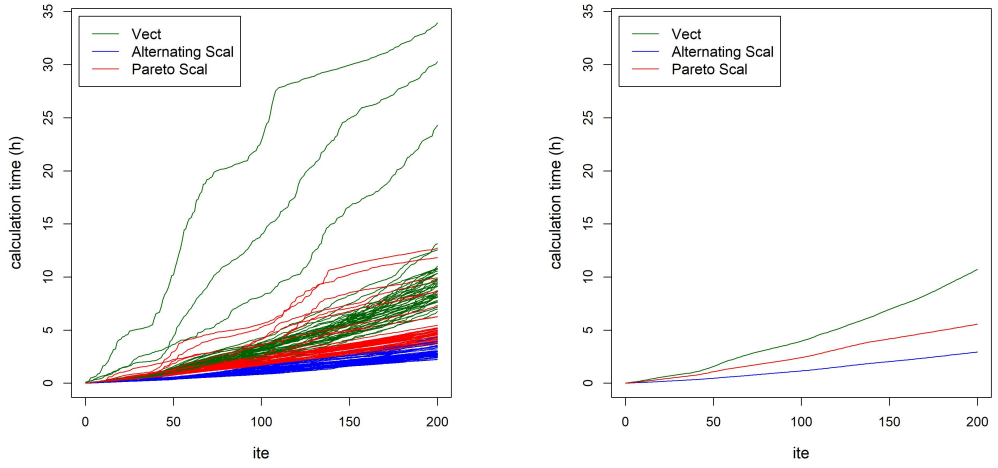
(d) $C = 10\%$, DP_1 (e) $C = 10\%$, DP_2 (f) $C = 10\%$, $\mathrm{DP}_{\mathrm{tot}}$



(g) $C = 5\%$, DP_1 (h) $C = 5\%$, DP_2 (i) $C = 5\%$, $\mathrm{DP}_{\mathrm{tot}}$

Figure 4.13: Standard "Data profiles" of partial relative errors with $C = 20\%$, 10% and 5%, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$.

These results must be set against the higher computation times of Vector strategy compared with Alternating and Pareto Scalar strategies (Figure 4.14). In the following section we plot data profiles in terms of total computation time (enrichment and $\boldsymbol{g}$ evaluations) rather than in terms of number of iterations.



(a) Line plots (b) Means

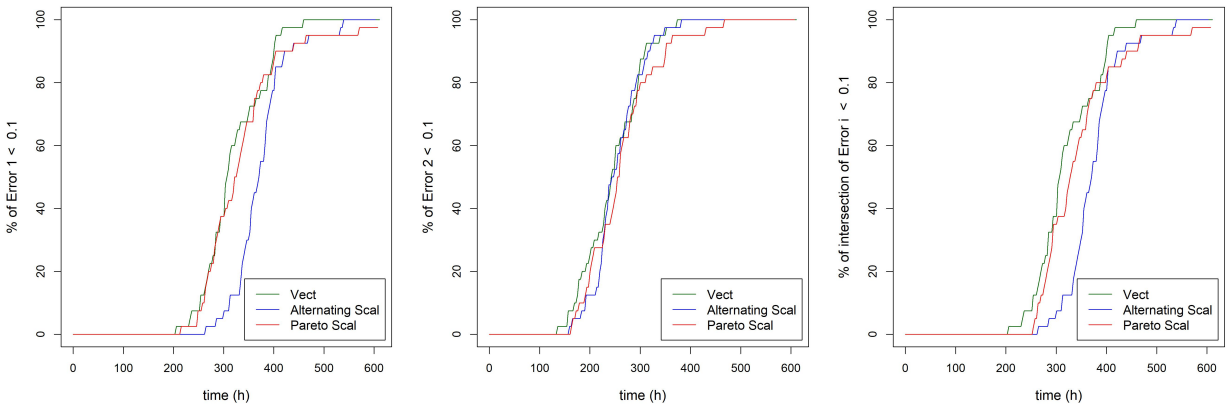
Figure 4.14: Enrichment calculation time (assuming that the evaluation time of $\boldsymbol{g}$ is negligible) for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the model based on the 4D Hartmann function with $\boldsymbol{T} = (-1.6, -1)$.

"Data profiles" as a function of total computing time

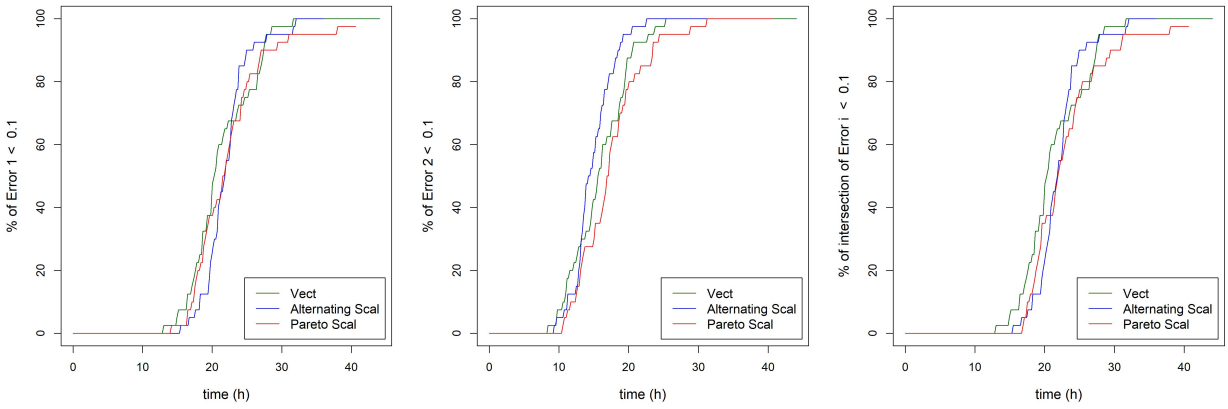
We fix the threshold to $C = 10\%$ and plot the data profiles as a function of total computation time (time required for enrichment and evaluations of $\boldsymbol{g}$) for different evaluation times of $\boldsymbol{g}$: 3h, 10min and 1min. The results obtained (Figure 4.15) show that for large evaluation times (3h), the curves are analogous to standard data profiles (Figure 4.14). However, when the evaluation time of $\boldsymbol{g}$ is shorter, Alternating Scalar strategy becomes more attractive than the other two strategies, due to its low computation time for sequential DoE enrichment. The long computation time required to update the vector surrogate model of order of a few minutes, is no longer negligible when the evaluation time of $\boldsymbol{g}$ is small (1min). Finally, Pareto Scalar and Vector strategies show comparable performance when the evaluation time of $\boldsymbol{g}$ is small (1min), with a slight advantage for Vector strategy.

4.5 Conclusion

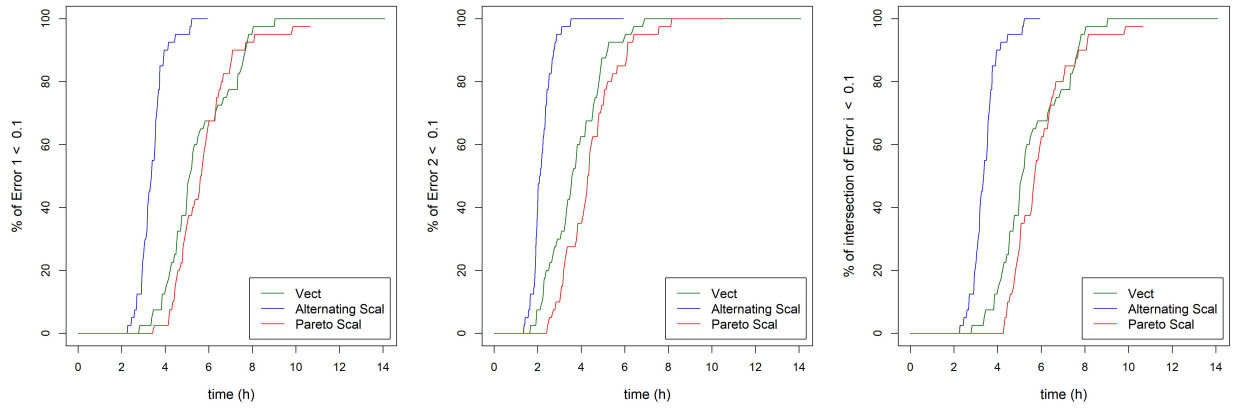
To simultaneously estimate the partial excursion sets of a vector-valued black box function with isotopic data (simultaneous evaluations of all output components), we have proposed three sequential DoE enrichment strategies. Two of these strategies use a scalar surrogate model for each output component with a suitable criterion for choosing a common enrichment point, while the third strategy uses a vector surrogate model (MOGP) with a criterion that takes into account the correlation between outputs. The performance of the three criteria was tested on 2 and 4 dimensional test functions, each with two output components.



(a) Eval $\boldsymbol{g}$ = 3h, DP_1 (b) Eval $\boldsymbol{g}$ = 3h, DP_2 (c) Eval $\boldsymbol{g}$ = 3h, $\mathrm{DP}_{\mathrm{tot}}$



(d) Eval $\boldsymbol{g}$ = 10min, DP_1 (e) Eval $\boldsymbol{g}$ = 10min, DP_2 (f) Eval $\boldsymbol{g}$ = 10min, $\mathrm{DP}_{\mathrm{tot}}$



(g) Eval $\boldsymbol{g}$ = 1min, DP_1 (h) Eval $\boldsymbol{g}$ = 1min, DP_2 (i) Eval $\boldsymbol{g}$ = 1min, $\mathrm{DP}_{\mathrm{tot}}$

Figure 4.15: "Data profiles" as a function of total computing time of partial relative errors for evaluation times of $\boldsymbol{g}$ of 3h, 10min and 1min, with $C = 10\%$, for the different criteria, in the case of enrichment of 40 LHS Maximin initial DoEs of size 20 with 200 iterations, for the enhanced 4D Hartmann test function with $\boldsymbol{T} = (-1.6, -1)$.

All three strategies showed a significant reduction in relative partial errors over the course of iterations. Analysis of these errors via functional boxplots revealed, especially for the 4-dimensional case, that Vector strategy offers a better compromise for the simultaneous search of the two excursion sets, with increased robustness. Also, we have highlighted the limitations of Alternating Scalar strategy with an unsuitable threshold, where it completely failed to estimate partial excursion sets. Vector and Pareto Scalar strategies were robust and efficient even in these difficult cases.

Data profiles, which show the percentage of curves with errors remaining below a given threshold confirmed these results. Vector strategy performed better in these examples with a good compromise between the two components but due to much higher computation time for updating vector surrogate model, it was above all interesting for functions $\boldsymbol{g}$ with high evaluation times (3h). For very low evaluation times of $\boldsymbol{g}$ (1min), Alternating Scalar strategy was more advantageous, due to its low computation time for enrichment (updating the substitution model and optimizing the enrichment criterion). In this last setting, Pareto Scalar and Vector strategies showed relatively similar performance .

Finally, although the results of Vector strategy are promising, the MOGP surrogate model we considered has limitations due to the separable nature of the correlation structure and induced property of autokrigeability (see Chapter 1, Section 1.7.3), which simplifies calculations but can reduce flexibility and accuracy by not necessarily capturing complex interactions between outputs.

4.6 Appendices

A Proof of the integral formulation for Vector Bichon criterion

Lemma 3. Let $\kappa > 0$ and Y be a non negative real random variable. Then we have

$$\mathbb{E}\big[(\kappa - Y)^+\big] = \int_0^\kappa F_Y(t)\,\mathrm{d}t, \tag{4.36}$$

where $(.)^+$ denotes the positive part function and F_Y the cumulative distribution function of Y.

Proof.

$$\begin{aligned}
\mathbb{E}\big[(\kappa - Y)^+\big] &= \mathbb{E}\big[(\kappa - Y)\mathbf{1}_{\kappa - Y \geq 0}\big] \\
&= \kappa\,\mathbb{E}[\mathbf{1}_{Y\leq\kappa}] - \mathbb{E}[Y\mathbf{1}_{Y\leq\kappa}] \\
&= \kappa\,F_Y(\kappa) - \int_0^{+\infty} \mathbb{P}[Y\mathbf{1}_{Y\leq\kappa} \geq t]\,\mathrm{d}t \qquad \text{(because } Y\mathbf{1}_{Y\leq\kappa} \geq 0\text{)} \\
&= \kappa\,F_Y(\kappa) - \int_0^{\kappa} \mathbb{P}[Y\mathbf{1}_{Y\leq\kappa} \geq t]\,\mathrm{d}t \\
&= \kappa\,F_Y(\kappa) - \int_0^{\kappa} \mathbb{P}[t \leq Y \leq \kappa]\,\mathrm{d}t \\
&= \kappa\,F_Y(\kappa) - \int_0^{\kappa} \big(F_Y(\kappa) - F_Y(t)\big)\,\mathrm{d}t \\
&= \int_0^{\kappa} F_Y(t)\,\mathrm{d}t.
\end{aligned}$$

□

Proposition 1. Noting $F_{Y_{\mathbf{x}}}$ the cumulative distribution function of $Y_{\mathbf{x}} := \min_i \left(\frac{|T_i - \xi_i(\mathbf{x})|}{\sigma_{n,i}(\mathbf{x})} \right) \Big| \boldsymbol{\mathcal{E}}_n$, we have

$$\mathrm{VEFF}(\mathbf{x}) = \det\big(\Sigma_n(\mathbf{x})\big)^{\frac{1}{2p}} \int_0^{\kappa} F_{Y_{\mathbf{x}}}(t)\,\mathrm{d}t. \tag{4.37}$$

Furthermore, for every $\mathbf{x} \in \mathbb{X}$ and every $t \in \mathbb{R}$,

$$F_{Y_{\mathbf{x}}}(t) = \sum_{i=1}^{2} \Big(\phi(t+\alpha_i) - \phi(-t+\alpha_i) \Big) - \mathbb{P}\big((U_1,U_2) \in [\alpha_1 - t, \alpha_1 + t] \times [\alpha_2 - t, \alpha_2 + t]\big), \tag{4.38}$$

with $\alpha_i := \frac{T_i - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})}$, ϕ c.d.f of $\mathcal{N}(0,1)$, $(U_1,U_2) \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \Big)$ and ρ the correlation coefficient between $\xi_{n,1}(\mathbf{x})$ and $\xi_{n,2}(\mathbf{x})$.

Proof. The integral formulation of Equation (4.37) follows directly from Lemma 3 applied with $Y = Y_{\mathbf{x}}$ (see initial formulation of VEFF in Equation (4.25)).

To calculate $F_{Y_{\mathbf{x}}}$, we note $\mathbb{P}_n := \mathbb{P}(\,.\,|\,\boldsymbol{\mathcal{E}}_n)$. Let $t \in \mathbb{R}$,

$$\begin{aligned} F_{Y_{\mathbf{x}}}(t) &= \mathbb{P}_n\left[\min_i \left(\frac{|T_i - \xi_i(\mathbf{x})|}{\sigma_{n,i}(\mathbf{x})} \right) \le t \right] \\ &= \underbrace{\mathbb{P}_n\left(\frac{|T_1 - \xi_1(\mathbf{x})|}{\sigma_{n,1}(\mathbf{x})} \le t \right) + \mathbb{P}_n\left(\frac{|T_2 - \xi_2(\mathbf{x})|}{\sigma_{n,2}(\mathbf{x})} \le t \right)}_{①} \\ &\quad \underbrace{- \mathbb{P}_n\left(\left\{ \frac{|T_1 - \xi_1(\mathbf{x})|}{\sigma_{n,1}(\mathbf{x})} \le t \right\} \bigcap \left\{ \frac{|T_2 - \xi_2(\mathbf{x})|}{\sigma_{n,2}(\mathbf{x})} \le t \right\} \right)}_{②}. \end{aligned} \tag{4.39}$$

The marginal distributions of $\boldsymbol{\xi}(\mathbf{x}) := (\xi_1(\mathbf{x}), \xi_2(\mathbf{x}))^\top$ verify

$$\xi_i(\mathbf{x}) \sim \mathcal{N}(M_{n,i}(\mathbf{x}), \sigma_{n,i}(\mathbf{x})^2), \forall\, i \in \{1,2\},$$

so for i fixed,

$$\begin{aligned} \mathbb{P}_n\left(\frac{|T_i - \xi_i(\mathbf{x})|}{\sigma_{n,i}(\mathbf{x})} \le t \right) &= \mathbb{P}_n\left(-t \le \frac{\xi_i(\mathbf{x}) - T_i}{\sigma_{n,i}(\mathbf{x})} \le t \right) \\ &= \mathbb{P}_n\left(-t + \frac{T_i - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})} \le \frac{\xi_i(\mathbf{x}) - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})} \le t + \frac{T_i - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})} \right) \\ &= \phi(t+\alpha_i) - \phi(-t+\alpha_i), \end{aligned}$$

with $\alpha_i := \frac{T_i - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})}$ and ϕ c.d.f. of $\mathcal{N}(0,1)$. Thus,

$$① = \sum_{i=1}^{2} \Big(\phi(t+\alpha_i) - \phi(-t+\alpha_i) \Big). \tag{4.40}$$

Next,

$$② = \mathbb{P}_n\left(\left\{ \frac{|T_1 - \xi_1(\mathbf{x})|}{\sigma_{n,1}(\mathbf{x})} \le t \right\} \bigcap \left\{ \frac{|T_2 - \xi_2(\mathbf{x})|}{\sigma_{n,2}(\mathbf{x})} \le t \right\} \right)$$

$$
\begin{aligned}
&= \mathbb{P}_n\Big((\xi_1(\mathbf{x}), \xi_2(\mathbf{x})) \in [T_1 \pm t\,\sigma_{n,1}(\mathbf{x})] \times [T_2 \pm t\,\sigma_{n,2}(\mathbf{x})]\Big) \\
&= \mathbb{P}_n\left(\left(\frac{\xi_1(\mathbf{x}) - M_{n,1}(\mathbf{x})}{\sigma_{n,1}(\mathbf{x})}, \frac{\xi_2(\mathbf{x}) - M_{n,2}(\mathbf{x})}{\sigma_{n,2}(\mathbf{x})}\right) \in [\alpha_1 \pm t] \times [\alpha_2 \pm t]\right) \\
&= \mathbb{P}\big((U_1, U_2) \in [\alpha_1 \pm t] \times [\alpha_2 \pm t]\big), \qquad (4.41)
\end{aligned}
$$

with $(U_1, U_2) \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\Big)$ and ρ the correlation coefficient between $\xi_1(\mathbf{x})$ and $\xi_2(\mathbf{x})$.

By reinjecting ① and ② from Equations (4.40) and (4.41) into Equation (4.39), we obtain the expected result. □

B Estimation of $H : (\alpha_1, \alpha_2, \rho) \mapsto \int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t$

In this appendix, we present the way in which, for a fixed $\mathbf{x}$, the quantity $\int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t$, required to implement Vector Bichon criterion, is estimated (see Proposition 1, Equation (4.26)). We begin by recalling that, according to Equation (4.27), the quantity $F_{Y_\mathbf{x}}$ (and therefore its integral between 0 and κ) depend only on α_i and ρ. Here, we present the interpolation method used to estimate the function

$$
H : \left\{ \begin{array}{ccc} \mathbb{R} \times \mathbb{R} \times [-1, 1] & \longrightarrow & \mathbb{R} \\ (\alpha_1, \alpha_2, \rho) & \longmapsto & \int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t \end{array} \right. ,
$$

outside the DoE enrichment loop. The first part presents the estimation of H for a triplet $(\alpha_1, \alpha_2, \rho)$ fixed in $\mathbb{R}^2 \times [0, 1]$ using a Gauss-Legendre quadrature method (used for estimation of H at the interpolation points). The second part studies the various symmetries of the H function, with a view to reducing the computation time required for interpolation. The third part concerns the interpolation method for estimating H on $[-50, 50]^2 \times [-1, 1]$, while the interpolation of H outside $[-50, 50]^2 \times [-1, 1]$ is deferred to the last part.

Estimation of H for $(\alpha_1, \alpha_2, \rho)$ fixed using Gaussian quadrature method

To estimate H for $(\alpha_1, \alpha_2, \rho)$ fixed, we start by recalling that according to Proposition 1,

$$
H(\alpha_1, \alpha_2, \rho) := \int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t \qquad (4.42)
$$

and for every $\mathbf{x} \in \mathbb{X}$ and every $t \in \mathbb{R}$,

$$
F_{Y_\mathbf{x}}(t) = \sum_{i=1}^{2} \Big(\phi(t + \alpha_i) - \phi(-t + \alpha_i)\Big) - \mathbb{P}\big((U_1, U_2) \in [\alpha_1 \pm t] \times [\alpha_2 \pm t]\big), \qquad (4.43)
$$

with ϕ c.d.f of $\mathcal{N}(0, 1)$ and $(U_1, U_2) \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\Big)$.

The calculation of $F_{Y_\mathbf{x}}(t)$ in (4.43) involves an orthant probability, in the particular case of the bivariate centered normal distribution (for a recent review on orthant probabilities see for example [Azzimonti and Ginsbourger, 2018]). This orthant probability is obtained using the *pvtnorm* function in the *mvtnorm* package in ***R*** computer language. The calculation method used is *'TVPACK'* from [Genz, 2004], which is particularly suited to the 2 and 3 dimensions.

For $(\alpha_1, \alpha_2, \rho)$ fixed, the estimation of $H(\alpha_1, \alpha_2, \rho) := \int_0^\kappa F_{Y_\mathbf{x}}(t)\,\mathrm{d}t$ is done using a Gauss-Legendre quadrature method (see [Abramowitz and Stegun, 1965]) between 0 and κ with the package *gaussquad* and $n_{\text{quad}} = 10$ quadrature points. Using 10 quadrature points leads to a

relative error of 10^{-16} (compared to a reference with 100 quadrature points), and we do not do any better with more quadrature points due to limitations on the estimation of $F_{Y_{\mathbf{x}}}$. The results of estimating H as a function of α_1 and α_2 and for different values of ρ are shown in Figure 4.16 (with $\kappa = 1$). The function exhibits a certain regularity as a function of α_1, α_2 and ρ. Also, when ρ is close to 0 the maximum of the function H seems to be located around $(\alpha_1, \alpha_2) = (0, 0)$, whereas when $|\rho| > 0.5$ it appears that there are two maxima, symmetrically positioned relatively to point $(0, 0)$. This is consistent since, for example, if $\rho \simeq 1$ the random variables are highly correlated and so a α_1 and a α_2 small in $|.|$ but of opposite signs provide more information than $\alpha_1 = \alpha_2 = 0$.

Study of the symmetries of the H function

The H function has several symmetries, which is beneficial in terms of reducing the computation time for the interpolation presented next. Let $m \in \mathbb{R}^2$ and $\Sigma \in \mathbb{R}^{2\times 2}$, let $\mathcal{N}(m, \Sigma)$ be the bivariate normal distribution with mean m and covariance matrix Σ. By definition of H,

$$H(\alpha_1, \alpha_2, \alpha) = \int_0^{\kappa} \mathbb{P}\big(\min(|X_1|, |X_2|) \leq t\big)\, \mathrm{d}t, \tag{4.44}$$

with $(X_1, X_2) \sim \mathcal{N}\Big(\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\Big)$. With these notations, we can see that $(X_1, -X_2) \sim \mathcal{N}\Big(\begin{pmatrix} \alpha_1 \\ -\alpha_2 \end{pmatrix}, \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}\Big)$ and that $(-X_1, -X_2) \sim \mathcal{N}\Big(\begin{pmatrix} -\alpha_1 \\ -\alpha_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\Big)$, which allows us to demonstrate the following symmetries on the function H:

$$\forall\, \rho \in [-1, 1], \forall\, (\alpha_1, \alpha_2) \in \mathbb{R}^2, H(\alpha_1, \alpha_2, \rho) = H(\alpha_1, -\alpha_2, -\rho) \tag{4.45}$$

and

$$\forall\, \rho \in [-1, 1], \forall\, (\alpha_1, \alpha_2) \in \mathbb{R}^2, H(\alpha_1, \alpha_2, \rho) = H(-\alpha_1, -\alpha_2, \rho). \tag{4.46}$$

Equation (4.45) represents, for a fixed α_1, a central symmetry with respect to point $(\alpha_2, \rho) = (0, 0)$, which restricts our study to $\rho \in [0, 1]$. Furthermore, Equation (4.46), for a fixed ρ, exhibits central symmetry with respect to the point $(\alpha_1, \alpha_2) := (0, 0)$, thus restricting our study to $\alpha_1 \in \mathbb{R}^+$.

In summary, the symmetries of Equations (4.45) and (4.46) make it possible to restrict the interpolation of H function to the set $\mathbb{R}^+ \times \mathbb{R} \times [0, 1]$. In numerical terms, it makes sense to leave a margin equivalent to the interpolation step size at the boundaries ($\alpha_1 = 0$ and $\rho = 0$) to ensure consistent regularity of the interpolation function. We also note the presence of two axial symmetries with respect to the lines $\alpha_2 = \alpha_1$ and $\alpha_2 = -\alpha_1$ (see Figure 4.16). However, these symmetries are not taken into account to reduce the computation time required for interpolation, as their implementation would be too complex.

Interpolation of H on $[-50, 50]^2 \times [-1, 1]$

Recall that our objective is to estimate the function $H : (\alpha_1, \alpha_2, \rho) \mapsto \int_0^{\kappa} F_{Y_{\mathbf{x}}}(t)\, \mathrm{d}t$ over its entire domain, outside the DoE enrichment loop, in order to reduce the computation time needed to evaluate Vector Bichon criterion. With this in mind, and given the observed regularity of H (see Figure 4.16), we opt for an interpolation method, using the *ipol* routine from the *chebpol* package ([Gaure, 2013]), while exploiting the symmetries of H previously highlighted. To begin with, we focus on interpolating the function H over the compact subset $[-50, 50]^2 \times [-1, 1]$.

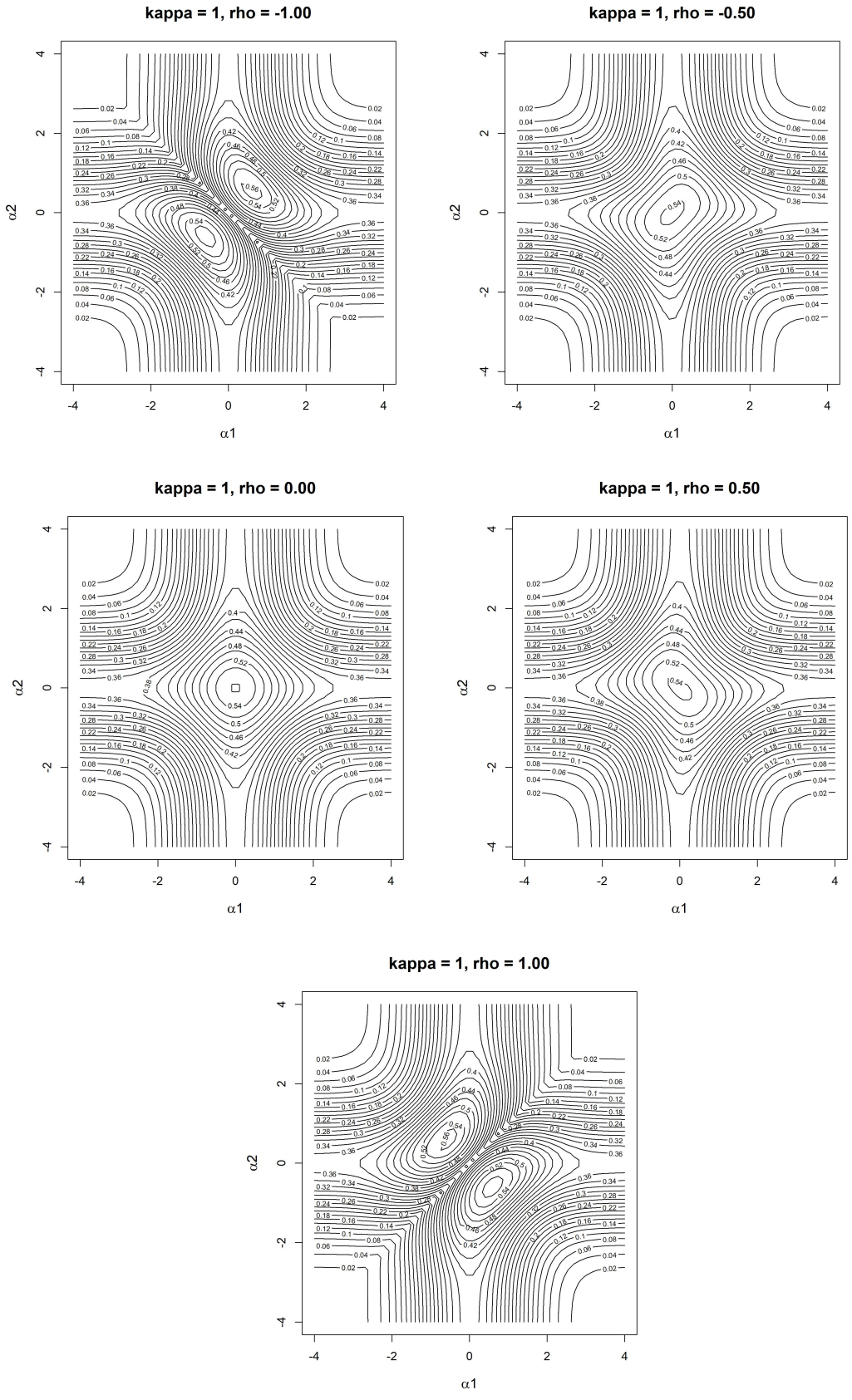


Figure 4.16: Representation of the level sets of H (with $\kappa = 1$) on a grid of size 50×50 for $\alpha := (\alpha_1, \alpha_2) \in [-4, 4]$, and for ρ values of -1, -0.5, 0, 0.5 and 1.

The interpolation methods compared are: multilinear interpolation (see, for example, [Wagner, 2008]), the Floater-Hormann method ([Floater and Hormann, 2007]) which relies on the construction of pole-free barycentric rational interpolants, and the Stalker method, a spline-based method proposed in the *chebpol* package that preserves monotonicity and extrema. The Stalker method preserves any existing monotonicity or, failing that, limits overshoot, i.e., the difference between the extrema values of the interpolation points and the extrema values of the interpolation function (see [Gaure, 2013]).

The type of grid used for interpolation plays a crucial role in the quality of the interpolation process. The distribution of grid points along the ρ component is chosen on a regular basis. For the α_1 and α_2 components, we compare a regular grid with an irregular grid obtained by a cubic transformation on each component (see the example in Figure 4.17 for $(\alpha_1, \alpha_2) \in [-1, 1]^2$). The use of an irregular grid based on a cubic transformation aims to concentrate more points in the region of interest of the H function. This region of interest corresponds to high values of H, i.e., values of (α_1, α_2) not too far to the axes $\alpha_1 = 0$ and $\alpha_2 = 0$ (see Figure 4.16).

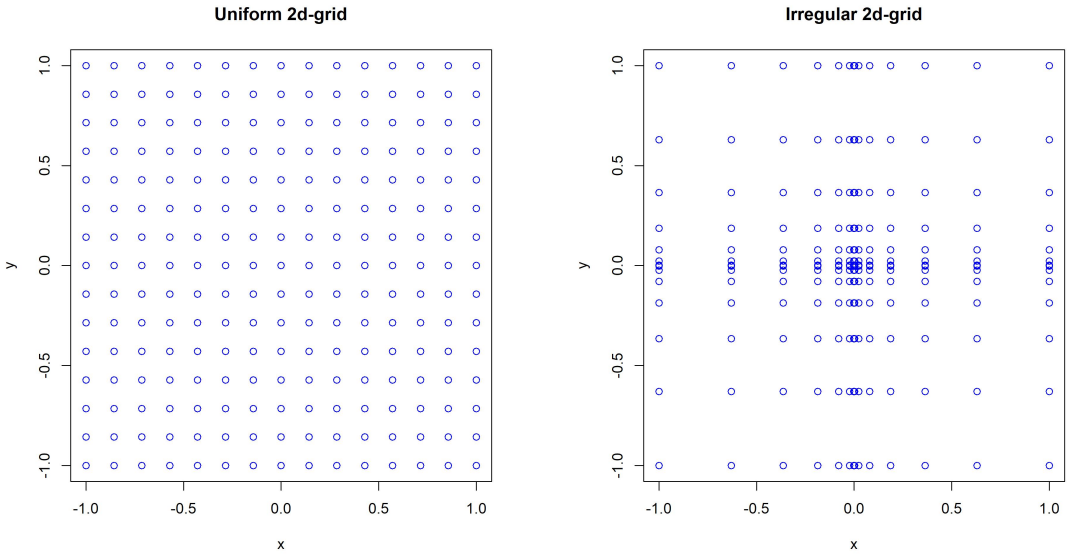


Figure 4.17: Comparison between a regular grid and an irregular grid resulting from a cubic transformation, on $[-1, 1]^2$, with 15 points per dimension.

The various interpolation methods are evaluated on $[-50, 50]^2 \times [-1, 1]$, using the two proposed grids (regular and cubic irregular), with 200 interpolation points per dimension. The results in term of interpolation errors are computed over two different sets of points: 10^5 random points selected in the area of interest $[-5, 5]^2 \times [-1, 1]$ (cf. Table 4.1) and 10^5 random points selected in the global interpolation area ($[-50, 50]^2 \times [-1, 1]$ (cf. Table 4.2). These results indicate that the Stalker (cubic) and Floater-Hormann (regular) methods perform best. However, the Floater-Hormann method turns out to be excessively expensive when evaluating the interpolation function at a specific point, as shown in Table 4.3. Consequently, the interpolation method chosen for the H function is the Stalker (cubic) method.

Interpolation of H outside $[-50, 50] \times [-1, 1]$

By default, the *chebpol* package creates the interpolation function even beyond the interpolation domain by multilinearly extending beyond the domain boundary. Table 4.4 summarizes

Method \ Error	Mean	Median	$q_{0.05}$	$q_{0.95}$	Maximum
Multilinear (regular)	$2.75e^{-3}$	$2.20e^{-3}$	$7.66e^{-5}$	$7.34e^{-3}$	$4.15e^{-2}$
Floater-Hormann (regular)	$3.65e^{-5}$	$\underline{5.27e^{-8}}$	$\underline{9.00e^{-10}}$	$9.88e^{-5}$	$2.22e^{-2}$
Stalker (regular)	$8.60e^{-4}$	$3.54e^{-4}$	$1.44e^{-5}$	$4.30e^{-3}$	$2.90e^{-2}$
Multilinear (cubic)	$2.17e^{-4}$	$1.40e^{-4}$	$8.61e^{-6}$	$6.25e^{-4}$	$\underline{3.81e^{-3}}$
Floater-Hormann (cubic)	$7.82e^{15}$	$2.94e^{13}$	$1.03e^{10}$	$4.34e^{15}$	$2.00e^{19}$
Stalker (cubic)	$\underline{1.03e^{-5}}$	$7.47e^{-6}$	$4.95e^{-7}$	$\underline{2.70e^{-5}}$	$5.93e^{-3}$

Table 4.1: Absolute errors of different interpolation methods chosen with 200 interpolation points per dimension and for 10^5 points chosen randomly on $[-5, 5]^2 \times [-1, 1]$.

Method \ Error	Mean	Median	$q_{0.05}$	$q_{0.95}$	Maximum
Multilinear (regular)	$3.34e^{-4}$	0	0	$2.74e^{-3}$	$2.68e^{-2}$
Floater-Hormann (regular)	$\underline{1.30e^{-6}}$	$3.00e^{-10}$	0	$\underline{1.35e^{-6}}$	$7.80e^{-3}$
Stalker (regular)	$1.63e^{-4}$	0	0	$5.02e^{-4}$	$1.89e^{-2}$
Multilinear (cubic)	$2.62e^{-5}$	0	0	$1.98e^{-4}$	$1.35e^{-3}$
Floater-Hormann (cubic)	$1.33e^{21}$	$1.53e^{18}$	$5.37e^{14}$	$3.57e^{20}$	$9.83e^{24}$
Stalker (cubic)	$1.36e^{-6}$	0	0	$1.08e^{-5}$	$\underline{1.24e^{-4}}$

Table 4.2: Absolute errors of different interpolation methods chosen with 200 interpolation points per dimension and for 10^5 points chosen randomly on $[-50, 50]^2 \times [-1, 1]$.

Grid \ Method	Multilinear	Floater-Hormann	Stalker	True evaluation
regular	$2.29e^{-2}$s	3min 12s	$8.95e^{-2}$s	3min 29s
cubic	$6.95e^{-3}$s	2min 55s	$1.86e^{-2}$s	3min 29s

Table 4.3: Comparison of interpolation function evaluation computation times for 10^4 randomly selected points, between the different interpolation methods selected.

the interpolation errors made on a few points outside the interpolation domain or on its boundary, allowing us to compare the case of a point outside the domain with that of its projection onto the domain. Since all the results show a very low error, especially when the values of the H function are high, we retain the default multilinear interpolation method to predict outside the domain.

Point $(\alpha_1, \alpha_2, \rho)$	True value of H	Stalker	Error
$(-50, 0, 0)$	$3.687e^{-1}$	$3.687e^{-1}$	$2.084e^{-10}$
$(-100, 0, 0)$	$3.687e^{-1}$	$3.687e^{-1}$	$2.084e^{-10}$
$(-100, 0, 0.5)$	$3.687e^{-1}$	$3.687e^{-1}$	$2.084e^{-10}$
$(-10000, 0, 0)$	$3.687e^{-1}$	$3.687e^{-1}$	$2.084e^{-10}$
$(-10000, 0, 0.5)$	$3.687e^{-1}$	$3.687e^{-1}$	$2.084e^{-10}$
$(-100, 0.1, 0)$	$3.672e^{-1}$	$3.672e^{-1}$	$3.053e^{-7}$
$(-100, 0.1, 0.5)$	$3.672e^{-1}$	$3.672e^{-1}$	$3.053e^{-7}$
$(-100, 3, 0.5)$	$7.869e^{-3}$	$7.734e^{-3}$	$1.357e^{-4}$
$(-50, 3, 0.5)$	$7.869e^{-3}$	$7.734e^{-3}$	$1.357e^{-4}$

Table 4.4: Values and absolute errors of the Stalker method (with cubic grid) for points, mainly chosen outside or on the boundary of the interpolation domain, in the context of interpolation of H on $[-50, 50]^2 \times [-1, 1]$ with 200 interpolation points per dimension.

Summary:

This chapter explores the estimation of excursion sets via Gaussian processes for a vector black-box function. Three strategies to enrich the DoE are proposed: the first two are based on scalar surrogate models for each output component, and the third one is based on a vector surrogate model (MOGP) taking the correlation between outputs into account. Tested on functions in dimensions 2 and 4 with two output components, all three strategies show a substantial reduction in relative errors as a function of the number of iterations. Vector strategy offers better performances on the presented examples, but is more costly in terms of computation time of the enrichment criterion. Alternating Scalar strategy is more advantageous for short computation times but fails to properly estimate partial excursion sets in the case of an inappropriate choice of threshold in relation to the values of the black box function. The MOGP surrogate model used is, however, limited by its separable correlation structure, which reduces flexibility and accuracy.

The following chapter presents the application of these different methodologies to our industrial test case. This application, proposed by IFP Énergies Nouvelles, concerns the pre-calibration of a wind turbine numerical model.

Chapter 5

Multi-output excursion set estimation applied to the pre-calibration of a wind turbine numerical model

Outlines

The aim of this chapter is to apply the methodology proposed in Chapter 4 to the pre-calibration of a wind turbine numerical model. The numerical simulator outputs are the vibration frequencies and deformation modes of the wind turbine in response to a wind load. In practice, observations of the modes and frequencies are used to pre-calibrate the numerical model. This pre-calibration is expressed as a multi-output excursion set estimation. Details of the numerical simulator and the formalization of the simulator pre-calibration problem are presented first. Next, the various enrichment strategies introduced in Chapter 4 are applied to the numerical wind turbine simulator, seen as a vector-valued black box function.

Contents

5.1 Introduction **99**
5.1.1 Excited deformation modes in wave theory 99
5.1.2 The case of a wind turbine structure 99
5.2 The wind turbine numerical model **101**
5.2.1 Presentation of the DTU 10MW Reference Wind Turbine 101
5.2.2 Operational Modal Analysis (OMA) 103
5.2.3 Model overview 105
5.3 Mode matching **107**
5.3.1 Matching example and mode matching algorithm 107
5.3.2 Checking the mode matching 109
5.4 Pre-calibration as a multiple excursion set estimation problem **110**
5.5 Pre-calibration using only the two main modes (Formulation 1) **112**
5.6 Pre-calibration using frequencies and modes (Formulation 2) **115**
5.7 Conclusion **118**
5.8 Appendix **120**
A Checking mode matching with respect to vibration frequencies . . 120

5.1 Introduction

5.1.1 Excited deformation modes in wave theory

In the wave theory of physics and engineering, the notion of an oscillation eigenmode in dynamical systems with several degrees of freedom is a fundamental concept, providing an essential tool for studying the vibratory behavior of complex structures (see for example [Blevins and Plunkett, 1980], [Nayfeh and Mook, 2008] and [Kneubühl, 2013]). For linearized complex dynamical systems, eigenmodes characterize specific oscillation states of the system, i.e., specific configurations in which the system can oscillate after being disturbed in the vicinity of its equilibrium state. The associated natural frequencies of vibration are called the eigenfrequencies. Classic examples of dynamical systems are mechanical systems, acoustic systems, molecular systems and electrical systems.

In the case of mechanical systems, the study of eigenmodes of deformation is fundamental to understanding the overall behavior and structure of objects. In fact, each object has its own set of deformation modes that vary according to its geometry, materials and the stresses to which it is subjected. Analysis of these eigenmodes enables the system to be decomposed as a set of independent harmonic oscillators, simplifying the understanding of vibratory phenomena. As a reminder, a harmonic oscillator is an oscillator model characterized by a sinusoidal time evolution of fixed amplitude, where the frequency is determined solely by the properties of the system. An example of eigenmodes of deformation in the case of a circular diaphragm fixed at its circumference is shown in Figure 5.1. Other classic examples include a vibrating string in 2D space, fixed at both ends.

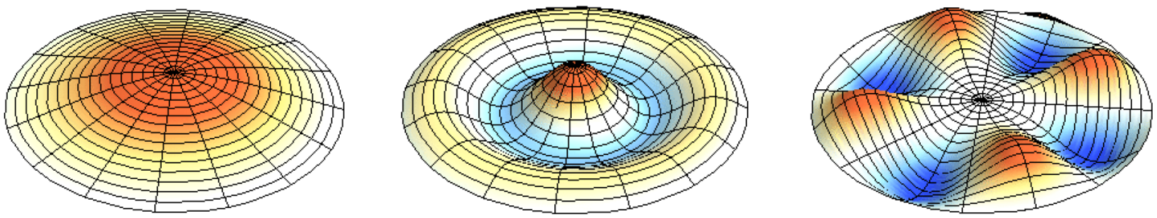


Figure 5.1: Vibrational modes of a circular membrane (CC BY-SA-ND 4.0; [Russell, 2018]).

When a dynamical system is subjected to external excitation, it may resonate with its own vibration frequencies associated with its various eigenmodes. This resonance is of paramount importance in engineering, particularly civil engineering, where the determination of eigenfrequencies is crucial to avoid structural damage caused by excessive vibrations and to design stronger, safer structures. The collapse of the Tacoma Narrows Bridge in the USA is one of the most famous civil engineering accidents in history (see Figure 5.2).

Figure 5.2: Tacoma Narrows bridge collapse, Nov. 1940.

Beyond their practical uses, eigenmodes are of significant theoretical importance, forming the basis of key concepts in physics such as eigenstates in quantum mechanics or the notion of the photon in electromagnetism. They provide an unified conceptual framework for understanding oscillation and vibration phenomena in a wide range of complex systems.

5.1.2 The case of a wind turbine structure

The application motivating our study is proposed by IFP Énergies Nouvelles and concerns the pre-calibration of a wind turbine numerical simulator. The OpenFAST ([Release, 2020]) wind turbine numerical simulator can calculate the vibration frequencies and deformation modes of the mechanical structure of the DTU 10-MW reference wind turbine (DTU 10MW RWT), both at standstill and in response to wind loading. The numerical simulator is considered a black box model, taking system parameters (such as stiffness coefficients of certain materials) as inputs and returning vibration frequencies and deformation eigenmodes as outputs. In this Chapter, inputs are noted Θ and outputs $\lambda_i(\Theta)$ for frequencies and $\text{Mod}_i(\Theta)$ for modes, with $i \in \{1, \ldots, p\}$ and p the number of modes. It should be recalled that studying the vibration signature (frequencies and modes) of a wind turbine is of practical interest, as it can be used to identify structural faults as mass or stiffness defects, premature wear, and misalignment of components (see [Cadoret, 2023]).

Figure 5.3: Illustration of offshore wind turbines.

Our aim here is to pre-calibrate the numerical model (Figure 5.4) by determining a set of feasible input parameters Θ for the simulator. These parameters should ensure that vibration frequencies $\lambda_i(\Theta)$ and deformation modes $\text{Mod}_i(\Theta)$ calculated by the simulator are sufficiently close, according to a certain measure of dissimilarity and within fixed thresholds, to the observed frequencies $\lambda_i^\star$ and modes $\text{Mod}_i^\star$, derived from experimental data based on the theory of Operational Modal Analysis (OMA).

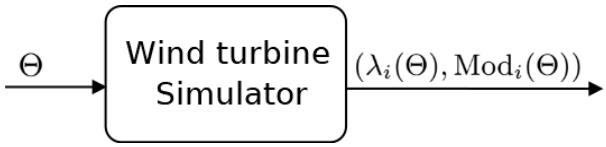


Figure 5.4: Schematic diagram of the wind turbine simulator.

Bayesian History Matching (BHM) aims to restrict the Θ parameter space by eliminating implausible sets, using a comparison between simulated and observed data. This approach improves the accuracy of simulator predictions by effectively reducing the parameter space to be explored (see [Kennedy and O'Hagan, 2001], [Bower et al., 2010] and [Andrianakis et al., 2015] for more details on BHM). The method used in this thesis to pre-calibrate the numerical wind turbine simulator can be seen as an alternative to BHM with an emulator based on Gaussian process regression, using OMA experimental data.

The black box model representing the simulator is presented in Section 5.2 , where we detail the wind turbine under study and OMA theory. Section 5.3 deals with mode matching, an essential step before solving the calibration problem, aimed at matching simulated modes with reference ones. The pre-calibration problem is presented in Section 5.4 . Two simplified formulations of the pre-calibration problem are addressed in Sections 5.5 and 5.6 , using the new methodology based on GPR introduced in Chapter 4. The first formulation uses only the two main modes, while the second uses both frequencies and modes. Finally, Section 5.7 provides the conclusion, and Section 5.8 includes an appendix related to mode matching (Section 5.3).

5.2 The wind turbine numerical model

In this section, we begin by describing the wind turbine itself: the DTU 10 MW RWT. Next, we discuss the Operational Modal Analysis (OMA) approach to determining reference frequencies $\lambda_i^\star$ and modes $\text{Mod}_i^\star$. Finally, the numerical simulator calculating simulated frequencies $\lambda_i(\Theta)$ and modes $\text{Mod}_i(\Theta)$, presented as a black-box function, is introduced.

5.2.1 Presentation of the DTU 10MW Reference Wind Turbine

Since the late 1970s, wind turbines have steadily increased in size, presenting challenges due to the linear dependency of turbine mass on the cube of the rotor radius. To address this challenge, the Light Rotor project, a collaboration between the DTU Wind Energy research team and Vestas Wind Systems, aimed to optimize blade design considering aerodynamic, structural, and aero-servo-elastic factors. As part of this initiative, a 10 MW reference rotor, known as the DTU 10 MW Reference Wind Turbine (DTU 10MW RWT), was developed to assess the performance of new designs (see [Bak et al., 2013]). The design process involves a comprehensive aerodynamic, structural and aero-servoelastic study of the turbine. Below, we outline some key features of this wind turbine. Further details are available in [Bak et al., 2013].

Key features

The DTU 10MW RWT wind turbine, designed for offshore installation, is a traditional three-bladed wind turbine with a rated output of 10 MW. The turbine features a rotor diameter of 178.3m, a hub height of 119.0m, and a hub diameter of 5.6m. Its general geometry is illustrated in Figure 5.5. The structural definition of the DTU 10MW RWT, excluding the blades, was derived by scaling up the reference NREL 5MW wind turbine from [Jonkman et al., 2009]. Modifications from the NREL 5MW include: adaptation to the offshore wind climate, conversion of the drivetrain from high to medium speed, a reduced hub height, and prebending of the blades to ensure clearance from the tower.

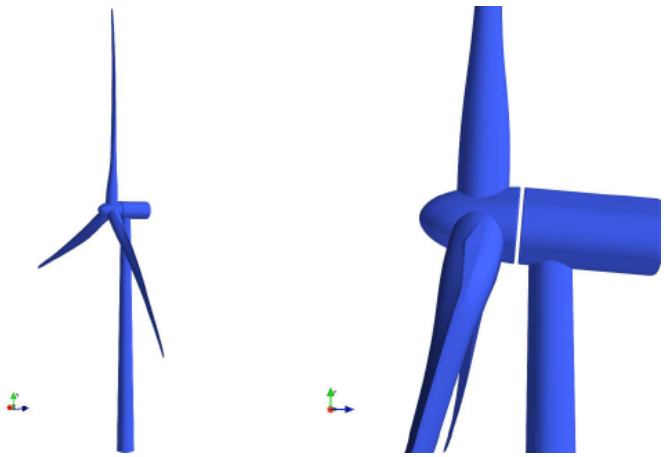


Figure 5.5: Plots of the DTU 10 MW RWT geometry from [Bak et al., 2013].

Aerodynamic design

The aerodynamic design of the DTU 10MW RWT is based on Blade Element Momentum (BEM) theory, which combines plane element theory and moment theory to calculate local forces on propellers or wind turbines. This overcomes the difficulties associated with calculating induced rotor speeds. The theory uses a simplified Rankine-Froude actuator-disk model (see [Rankine, 1865] and [Froude, 1889]) to estimate wind turbine efficiency, taking angular momentum into account. The rotor design is also evaluated using computational fluid dynamics (CFD) (see for example [Anderson and Wendt, 1995]).

Airfoil selection was influenced by the rotor's lightweight, which requires relatively thick airfoils. FFA-W3 profiles were chosen for their availability and thickness, and Gurney flaps were added to improve aerodynamic performance (Figure 5.6). The rotor's aerodynamic characteristics were calculated using EllipSys2D, with 3D corrections and manual adjustments. Rotor design was carried out using HAWTOPT software ([Fuglsang and Thomsen, 2001]), based on BEM theory and numerical optimization.

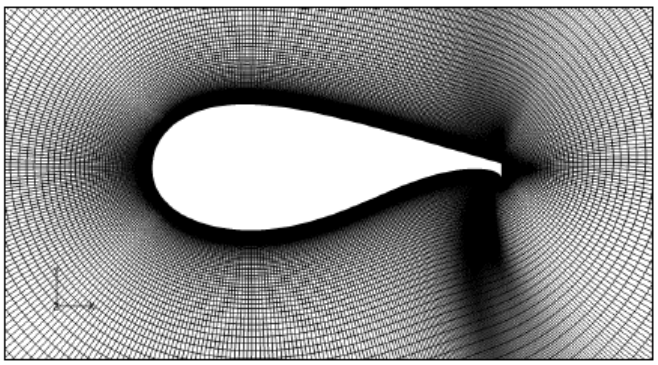


Figure 5.6: 2D mesh around the FFA-W3-336 airfoil fitted with a Gurney flap from [Bak et al., 2013].

Structural design

The blades are made of fiberglass-reinforced composites with balsa as the core material. The elastic properties of the multidirectional laminates were calculated using Classical Lamination Theory (CLT) ([Jones, 2018]). The internal structure of the blades was modeled using ABAQUS software ([Manual, 2012]), with approximately 35 000 elements. Each blade was divided into 11 circular regions and 100 radial regions to define the laminated composite (Figure 5.7). The stiffness and mass properties of the cross-sections were calculated using BECAS software ([Blasques, 2012]). Linear buckling analysis showed blade resistance to buckling modes. The tower is made of S355 steel, with a diameter ranging from 8.3m at the base to 5.5m at the top. The tower's buckling analysis showed a buckling load 3.15 times greater than the given force, guaranteeing sufficient safety.

Aero-Servo-Elastic design

The DTU 10MW RWT was designed to withstand anticipated loads and operating conditions, with detailed analyses ensuring structural safety and aeroelastic performance ([Bak et al., 2013]). Turbine components were determined through a scaling procedure, with the nacelle and hub located 119m above ground level. The drivetrain uses a 50 : 1 ratio, ensuring a compact,

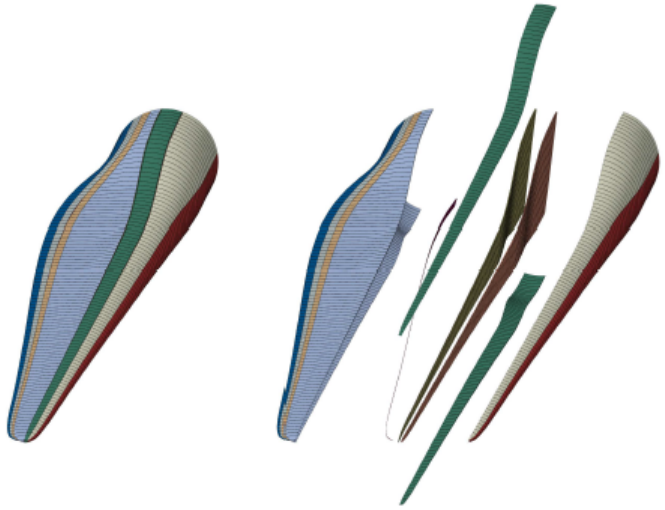


Figure 5.7: Representation of the blade's composite stratification, divided into 11 circumferential and 100 radial regions from [Bak et al., 2013].

lightweight design. The turbine is designed for offshore conditions conforming to IEC 61400-1 Ed.3 standard, with load simulations carried out using HAWC2 software ([Larsen and Hansen, 2009]). Vibration and stability issues have been resolved by adjusting natural blade frequencies and aerodynamic damping.

5.2.2 Operational Modal Analysis (OMA)

Operational Modal Analysis (OMA) (see for example [Brincker and Ventura, 2015]) is a method based on observed data used to identify the vibratory characteristics of mechanical structures, such as wind turbines. The result of OMA is then exploited to calibrate the numerical simulator by reducing the gap between the frequencies and modes obtained from the simulator and those obtained from OMA, as is the focus of our application ([Cadoret, 2023]).

In classical modal analysis (see for example [Fu and He, 2001]), the structure under study is subjected to a known excitation, and the measurement of inputs and responses enables the identification of a linear transfer function expressed as a function of the eigenmodes to be determined. This type of modal analysis is only suitable for studies where the environment is controllable, which is not the case for a wind turbine subjected to wind loading. Operational modal analysis (OMA) is a more advanced technique, where the structure is subjected to an unknown wind excitation. This excitation is modeled through a stochastic process with known mean and covariance. In this approach, measurements of responses alone enable the identification of a transfer function expressed in terms of eigenmodes. A final identification phase yields the desired eigenmodes (see Figure 5.8).

Standard mechanical vibration problem

A mechanical vibration problem is generally a second-order linear problem of the form

$$\begin{cases} M_s\ddot{\xi} + C_s\dot{\xi} + K_s\xi = L_\rho\rho \\ y = O_a\ddot{\xi} + O_v\dot{\xi} + O_m\xi + L_\mu\mu \end{cases}, \tag{5.1}$$

where ξ denotes the displacement vector of the structure, ρ the stochastic excitation and μ the measurement noise. The matrices M_s, C_s and K_s are the mass, damping and stiffness matrices

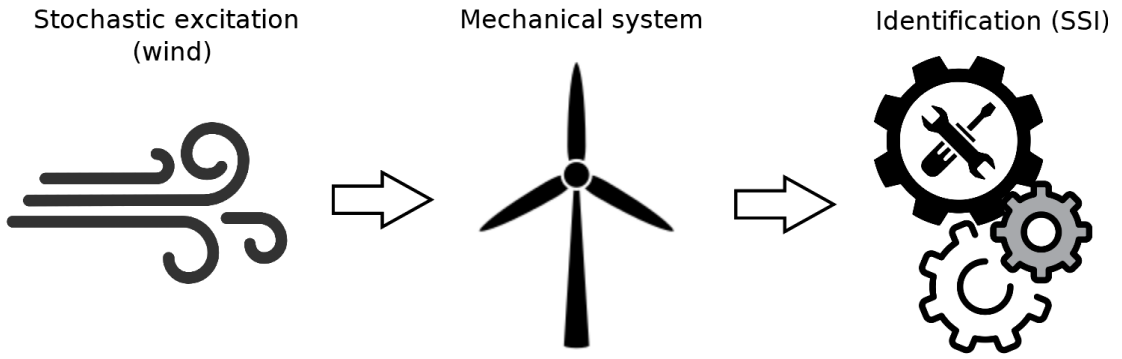


Figure 5.8: OMA summary diagram.

of the system, and it is further assumed that M_s and K_s are positive definite symmetric matrices. The outputs y of the system are assumed to be a linear combination of motion, velocity and acceleration via the matrices O_a, O_v and O_m. By setting $\boldsymbol{x} := (\xi^\top, \dot{\xi}^\top)^\top$, it is possible to reduce ([Tisseur and Meerbergen, 2001]) to a first-order linear problem of the form

$$\begin{cases} \dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{v} \\ \boldsymbol{y} = C\boldsymbol{x} + D\boldsymbol{w} \end{cases}, \tag{5.2}$$

with

$$A := \begin{pmatrix} 0 & I \\ -M_s^{-1}K_s & -M_s^{-1}C_s \end{pmatrix},\ B := \begin{pmatrix} 0 \\ {M_s}^{-1}L_\rho \end{pmatrix},$$

$$C := \begin{pmatrix} O_m - O_a M_s^{-1}K_s & O_v - O_a M_s^{-1}C_s \end{pmatrix},\ D := \begin{pmatrix} O_a M_s^{-1}L_\rho & L_\mu \end{pmatrix},$$

as well as $\boldsymbol{v} := \rho$ and $\boldsymbol{w} := (\rho^\top\ \mu^\top)^\top$.

From discrete measurements $\boldsymbol{y}_k$ sampled at times $t_k = k\Delta t$ with $k \geq 0$ integer, we consider the simplified discrete system assuming the excitation $\boldsymbol{v}$ constant on $[k\Delta t, (k+1)\Delta t]$:

$$\begin{cases} \boldsymbol{x}_{k+1} = \tilde{A}_k\boldsymbol{x}_k + \tilde{B}_k\boldsymbol{v}_k \\ \boldsymbol{y}_k = \tilde{C}_k\boldsymbol{x}_k + \tilde{D}_k\boldsymbol{w}_k \end{cases}, \tag{5.3}$$

with $\tilde{A}_k := \exp(\Delta t A)$, $\tilde{B}_k := (A - I)A^{-1}B$, $\tilde{C}_k := C$ and $\tilde{D}_k := D$. Since the system is time-invariant (none of the 4 matrices $\tilde{A}_k$, $\tilde{B}_k$, $\tilde{C}_k$ and $\tilde{D}_k$ really depends on k), we can express the statistic $R_n := \mathbb{E}[\boldsymbol{y}_{k+n}\boldsymbol{y}_k^\top]$ as a function of $\tilde{A}$:

$$R_n := \tilde{C}\tilde{A}^{n-1}G \tag{5.4}$$

with $G := \mathbb{E}[\boldsymbol{x}_{k+1}\boldsymbol{y}_k^\top]$ (see [Cadoret, 2023] for more details on deriving (5.3) and (5.4)).

OMA principle

The aim of OMA is to identify a system, as described by Equation (5.3), that best matches the $\boldsymbol{y}_k$ series of measurements of the mechanical system under study. The OMA principle is based on the fact that, when excited by white noise, the correlation functions R_n of the responses play the same role as the impulse responses. This makes it possible to apply classical system identification techniques by replacing temporal responses with observation correlation functions ([Cadoret, 2023]).

Stochastic Subspace Identification (SSI) algorithms (see for example [Reynders, 2012] and [Van Overschee and De Moor, 2012]) are commonly used to estimate the modal characteristics of structures from the resulting transfer function R_n. Based on simple linear algebra concepts, these algorithms represent an efficient and stable set of methods.

In the OMA framework, the SSI-Cov algorithm is applied to the Hankel matrix $\hat{R}_n$ of the unbiased estimators of the correlation matrices R_n obtained from the $\boldsymbol{y}_k$. More precisely, the singular value decomposition of the $\hat{R}_n$ Hankel matrix is used to estimate the A_k state matrix. From the relationship between A_k and A ($A_k := \exp(\Delta t A)$), the eigendecomposition of the A matrix, corresponding to the system's vibration frequencies and deformation modes, can be determined from the eigendecomposition of the $\hat{A}_k$ estimate. For more technical details on estimating the eigendecomposition of A, please refer to [Döhler and Mevel, 2013]. Variants of the SSI algorithm offer different approaches to estimating and organizing correlation matrices, influencing the factorization of subspace matrices (see for example [Peeters and De Roeck, 1999] and [Reynders et al., 2008]).

The SSI-Cov method estimates a state system based on correlations but requires the order (i.e., size n of R_n matrix) of the system to be specified. The classic approach is to estimate the system for different model orders, then identify the structures that remain invariant with the change in order. This is frequently done using a stabilization diagram, which displays the quantities identified as a function of model order.

To summarize, for an OMA analysis, a stochastic excitation is applied to the mechanical system to obtain a transfer function (quantity R_n in Equation (5.4)), which is then used by the SSI method to identify frequencies and eigenmodes.

Advantages and limitations of OMA

The main advantages of OMA are manifold. It allows the identification of the wind turbine's vibration signature, providing a discrete set of vibration modes that can be used to calibrate a simulator. OMA also enables the modal properties of the structure to be tracked over time and can reconstruct modal deformations when the structure is equipped with a sufficient number of sensors.

However, OMA is based on important assumptions, such as the linearity and time invariance of mechanical systems. These assumptions are not always respected in real-life situations, especially for operating wind turbines. This is why we have studied the simplified case of a wind turbine at standstill. OMA also relies on the stationarity and ergodicity of the excitation signal corresponding to the wind, which is not always verified. However, despite these limitations, OMA remains an effective method for monitoring wind turbine vibration, predicting maintenance operations, and calibrating numerical models.

5.2.3 Model overview

The black box model

The numerical wind simulator we are studying is regarded as a black-box model. We only have access to its inputs and outputs, without any knowledge or control over the internal mechanisms or equations of the model.

It takes as input the parameters of the system, corresponding to the stiffness coefficients of certain materials, and returns as output the vibration frequencies and deformation modes of the mechanical structure (see Figure 5.9). The simulator is implemented in the OpenFAST software ([Release, 2020]), which enables prior linearization of the mechanical system before the calculation of system vibration frequencies and deformation modes.

Figure 5.9: Schematic diagram of the black box wind model.

Data are assumed to be isotopic, meaning that one simulation produces the various output components simultaneously at each evaluation point. This assumption means that a common evaluation point must be chosen for experimental design enrichment and contributes to all output components.

Model inputs

The inputs to the black box model represent multiplicative stiffness coefficients at the level of various structural elements and are represented by $\Theta \in \mathbb{X} \subset \mathbb{R}^3$. More precisely, the cross-section of the reference blade (Figure 5.7) is decomposed along two main axes (one vertical and one horizontal) and the first two components of Θ correspond to material stiffness coefficients along each of these main directions. The third component of Θ corresponds to a stiffness coefficient at the tower level.

By default, the multiplicative stiffness coefficients are set to 1 and the information provided by the mechanical experts at IFP Énergies Nouvelles enables us to define a precise $\mathbb{X}$ design space on which to restrict our study. The range of possible multiplicative stiffness coefficients is given by $\Theta \in \mathbb{X} := [0.8, 1.2] \times [0.6, 1.4]^2$.

Model outputs

The outputs of the black box model correspond to the vibration frequencies $\lambda_i(\Theta)$ and associated deformation modes $\mathrm{Mod}_i(\Theta)$, defined for i ranging from 1 to 26. These eigenvalues and eigenvectors are conjugated in pairs due to the quadratic formalization of the problem (see [Tisseur and Meerbergen, 2001]), which reduces the number of unique vibration frequencies and deformation modes to 13.

The real part of the vibration frequencies provided by the simulator (eigenvalues of the quadratic problem) represents the resonance frequency, which indicates the speed of oscillation. The imaginary part of the vibration frequencies reflects the damping, describing how the energy decreases over time. These two aspects capture both oscillatory behavior and energy dissipation in the system.

Observed data

Normally, the reference $\lambda_i^\star$ frequencies and $\mathrm{Mod}_i^\star$ modes to which the simulated frequencies and modes are compared are obtained from the OMA analysis presented in Section 5.2.2. However, for our methodological purpose, we did not use real wind turbine data as required for the OMA approach. Instead, we emulated the OMA reference modes and frequencies by using the simulator response for fixed parameters Θ with reference $\Theta^\star := (1, 1, 1)$. This illustrates the problem of calibrating a numerical wind turbine simulator, even though the reference frequencies and modes should normally be obtained more rigorously from OMA analysis. The reference modes $\lambda_i^\star$ and $\mathrm{Mod}_i^\star$ are thus defined for all i as follows:

$$\lambda_i^\star := \lambda_i(\Theta^\star) \quad \text{and} \quad \mathrm{Mod}_i^\star := \mathrm{Mod}_i(\Theta^\star). \tag{5.5}$$

In physics, the modulus of the complex vibration frequency represents the overall magnitude of the oscillation and can be interpreted as a measure integrating both oscillation speed and energy loss. Table 5.1 shows the 13 moduli of the reference vibration frequencies. This table allows us to identify the main modes and their influence on the overall system in relation to the other modes.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
$\lvert\lambda_i^\star\rvert$	<u>15.44</u>	<u>15.49</u>	11.99	11.83	11.01	6.41	6.49	5.82	1.62	1.61	4.18	4.01	4.09

Table 5.1: Summary table of $|\lambda_i^\star|$ for $i \in \{1, \dots, 13\}$, in percentages relative to $\sum |\lambda_i^\star|$.

The formalization of the simulator pre-calibration problem involves defining an order between the frequencies and modes obtained from the simulator, before comparing them with the references. The matching of simulated frequencies and modes with reference frequencies and modes is detailed in the next section.

5.3 Mode matching

Both formulations we propose in Section 5.4 to address the pre-calibration step of wind turbine simulator are built on a dissimilarity measure based on ordered frequencies and modes. Mode matching is a preliminary step ensuring label matching between frequencies and modes generated by the simulator and those of the reference case. In this section, we propose an algorithm designed for mode matching.

5.3.1 Matching example and mode matching algorithm

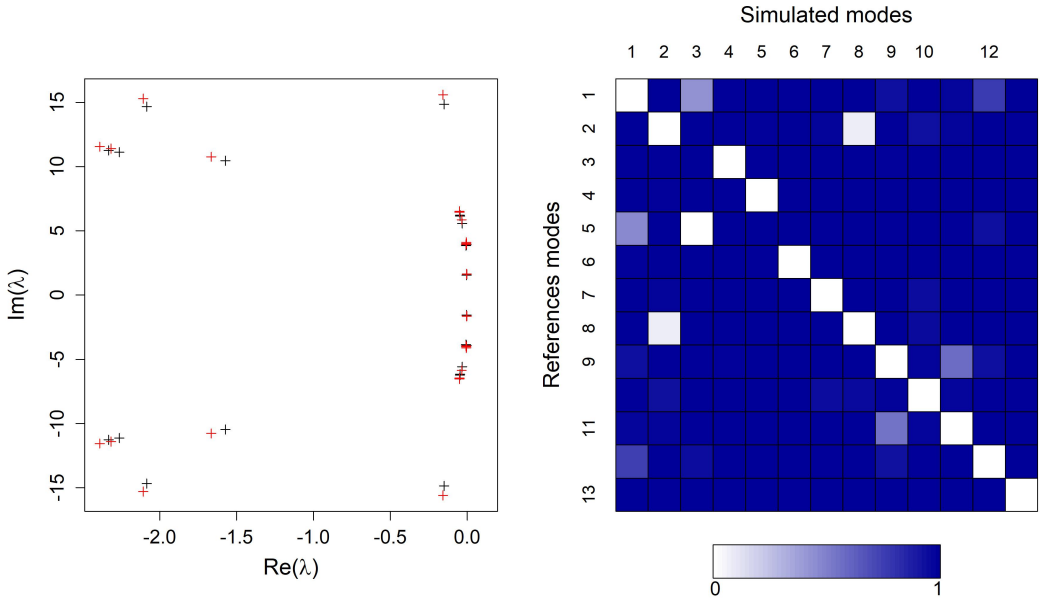
To recover the right pairing between the reference modes and the mode obtained with a given Θ, the idea is to compute the distances between each of the 13 reference modes and each of the 13 simulated modes and to bring together the closest ones. To do that we use the following dissimilarity measure :

$$\text{Meas}_{i,j} := \left(1 - \frac{| < \text{Mod}_i^\star, \text{Mod}_j(\Theta) > |^2}{\|\text{Mod}_i^\star\|^2 \|\text{Mod}_j(\Theta)\|^2}\right), \tag{5.6}$$

with $< .,. >$ being the canonical Hermitian scalar product. This measure allows to assess the degree of dissimilarity between $\text{Mod}_i^\star$ and $\text{Mod}_j(\Theta)$ (evaluated at 0 in the case of perfect collinearity and 1 in the case of orthogonality). We put all the pairwise distances in a 13×13 matrix. It can be seen that the lowest distances are mainly located on the diagonal, but not all of them. This indicates a better match than the natural one for several modes.

An example is given in Figure 5.10b, where the simulated modes are obtained with $\Theta = (1.05, 1.1, 1.1)$. For this example we first notice that, if we remove rows 2 and 8 and columns 2 and 8 in the matrix in Figure 5.10b, the corresponding submatrix has only one white cell per row and column. From this submatrix we can thus deduce a permutation of $\{1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13\}$, namely the circular permutation (3 4 5), which allows matching simulated modes with the ones obtained from OMA. Now if we focus on the values $\text{Meas}_{i,j}$ for i, $j \in \{2, 8\}$, we remark that $\text{Meas}_{2,8}$ is about 300 times higher than $\text{Meas}_{2,2}$ or $\text{Meas}_{8,8}$, meaning that for these two labels there is no reason to apply a permutation.

We now introduce a sequential algorithm based on Equation (5.6) to automatically recover the best matching. This algorithm begins by pairing the modes with the smallest dissimilarity



(a) Frequencies (b) Modes

Figure 5.10: On the left, comparison of simulated (red) and reference (black) frequencies for $\Theta = (1.05, 1.1, 1.1)$. On the right, representation of the table of dissimilarity measure values between simulated and reference modes.

and remove the corresponding row and column from the matrix. It then iterates until all pairs are obtained.

Algorithm 1: Mode matching

Input : Matrix $Meas := (\text{Meas}_{i,j})_{i,j}$
Output: Matrix *Res* for mode matching

$k \leftarrow 1$, *Res* $\leftarrow$ NULL;
while $k \leq 13$ **do**
  Select the indices i_k and j_k corresponding to the smallest value of *Meas*;
  All elements in row i_k or column j_k of *Meas* are replaced by 1;
  Add a line in *Res* with (i_k, j_k);
  $k \leftarrow k + 1$;
end
Sort rows of matrix *Res* by first column;
return *Res*

Algorithm 1 sequentially matches mode labels to correct the default labels assigned by the simulator. It begins with straightforward matches and proceeds to handle more complex ones. The algorithm outputs a permutation that maps reference mode labels to simulated mode labels. For instance, using $\Theta = (1.05, 1.1, 1.1)$, Algorithm 1 generates the following permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ 1 & 2 & 4 & 5 & 3 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \end{pmatrix},$$

indicating how mode labels from the reference case are matched to the simulated case. This permutation corresponds to the (3 4 5) cycle identified earlier (Figure 5.10).

5.3.2 Checking the mode matching

We then perform more general tests using a Maximin LHS design of size 100 for Θ values. For each row and each column of the matrix $\left(\text{Meas}_{i,j}\right)_{i,j}$, we study the ratio between the second smallest value and the smallest value, calling the minimum of these ratios over all rows and columns the minimum matching ratio. This minimum matching ratio indicates the potential difficulty for a sequential matching algorithm to accurately match mode labels. A minimum matching ratio close to 1 could cause problems for the algorithm in determining the best match between two elements in the same row or column. Conversely, a high minimum matching ratio indicates that the algorithm is unlikely to make mismatches. In the previous example, this ratio was 369. The boxplot of minimum matching ratios for Θ values from a LHS Maximin design of size 100 defined on $\mathbb{X} = [0.8, 1.2] \times [0.6, 1.4]^2$ (Figure 5.11), show relatively high values, with a median minimum matching ratio above 10.

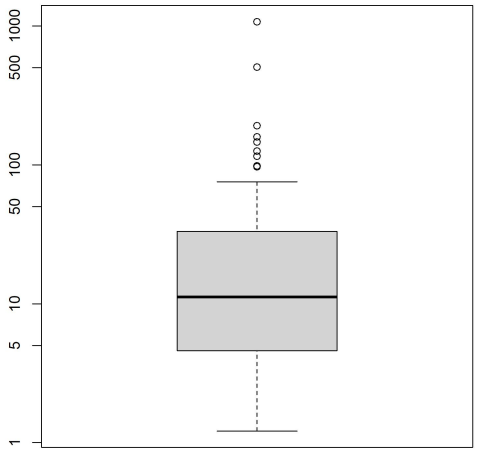


Figure 5.11: Boxplots (in logarithmic scale) of the minimum matching ratio for Θ values from a Maximin LHS design of size 100 on $[0.8, 1.2] \times [0.6, 1.4]^2$.

We checked that the case corresponding to the lowest minimum matching ratio on Figure 5.11 ($\Theta := (1.1673, 0.9646, 0.6424)$), was not problematic for Algorithm 1. This is because, even when two very low values are close together in the same row of $\left(\text{Meas}_{i,j}\right)_{i,j}$, Algorithm 1 initially eliminates the most obvious matches. The more complex cases become simpler by process of elimination, as fewer choices remain. This highlights the effectiveness of Algorithm 1 in globally optimizing the mode matching on the dissimilarity table, rather than performing row-by-row or column-by-column matching.

The results above demonstrate the effective performance of mode matching algorithm (Algorithm 1), showing robustness in handling challenging matching scenarios. Appendix A further explores the consistency of the permutation provided by this algorithm concerning vibration frequencies. Hereafter, we assume the permutation σ_{algo} provided by mode matching algorithm is applied to the simulated frequencies and modes. Thus, we denote $\lambda_i(\Theta)$ and $\text{Mod}_i(\Theta)$ instead of $\lambda_{\sigma_{\text{algo}}(i)}(\Theta)$ and $\text{Mod}_{\sigma_{\text{algo}}(i)}(\Theta)$, representing the simulated frequencies and modes aligned with the references $\lambda_i^\star$ and $\text{Mod}_i^\star$.

5.4 Pre-calibration as a multiple excursion set estimation problem

We begin by recalling that we use a black box simulator, which takes as input the stiffness parameters Θ of the mechanical system and returns as output the vibration frequencies $\lambda_i(\Theta)$ and the associated deformation modes $\mathrm{Mod}_i(\Theta)$, for i ranging from 1 to 26. As these frequencies and modes are conjugated in pairs, we limit our study to i ranging from 1 to 13. The reference frequencies $\lambda_i^\star$ and modes $\mathrm{Mod}_i^\star$ should normally be obtained by OMA analysis (see Section 5.2.2). Since we do not have access to these data, we simulate these reference frequencies and modes with the simulator using $\Theta^\star := (1, 1, 1)$. To simplify the problem and visualize results, we limit ourselves to two simulator input parameters, $\mathbb{X} := [0.8, 1.2] \times [0.6, 1.4]$, and set the third component of Θ to 1.

The pre-calibration problem initially supplied by IFP Énergies Nouvelles consists of identifying the input parameters Θ of the simulator so that the simulated modes and frequencies are not too far from the reference modes and frequencies. To define the notion of distance between simulated and reference modes, we need a measure of dissimilarity between modes. The notion proposed by the experts is comparable to that used for mode matching and is defined by:

$$\mathrm{Meas}_i(\Theta) := \left(1 - \frac{| < \mathrm{Mod}_i^\star, \mathrm{Mod}_i(\Theta) > |^2}{\|\mathrm{Mod}_i^\star\|^2 \|\mathrm{Mod}_i(\Theta)\|^2}\right), \tag{5.7}$$

with $i \in \{1, \dots 13\}$ and $< ., . >$ representing the canonical Hermitian scalar product. This measure evaluates the degree of proximity between two vectors, indicating their level of collinearity (evaluated as 0 in the case of perfect collinearity and 1 in the case of orthogonality). This dissimilarity measure, combined with thresholds $(T_1, \dots, T_{13})$, enables us to define what we mean by "not too far appart" for simulated modes compared to reference modes.

We do the same thing for vibration frequencies, considering the relative error defined by:

$$\mathrm{Meas}_i^\lambda(\Theta) := \frac{|\lambda_i(\Theta) - \lambda_i^\star|}{|\lambda_i^\star|}, \tag{5.8}$$

as a measure of dissimilarity. This dissimilarity measure combined with thresholds $(T_1^\lambda, \dots, T_{13}^\lambda)$ allows us to define the notion of "not too far appart" for simulated frequencies compared with reference frequencies.

In summary, we want to estimate partial excursion sets associated with each mode and each frequency. Specifically, for each $i \in 1, \dots, 13$, we aim to find $\Gamma_i^\star$ and $\Gamma_{\lambda,i}^\star$, which are the sets of input parameters Θ such that the distances $\mathrm{Meas}_i(\Theta)$ and $\mathrm{Meas}_i^\lambda(\Theta)$ do not exceed the thresholds T_i and T_i^λ, respectively. Mathematically, it gives:

$$\begin{cases} \Gamma_1^\star := \Big\{\Theta \in \mathbb{X},\, g_1(\Theta) := \ln\big(\mathrm{Meas}_1(\Theta)\big) \le T_1\Big\}, \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots \\ \Gamma_{13}^\star := \Big\{\Theta \in \mathbb{X},\, g_{13}(\Theta) := \ln\big(\mathrm{Meas}_{13}(\Theta)\big) \le T_{13}\Big\}, \\ \Gamma_{\lambda,1}^\star := \Big\{\Theta \in \mathbb{X},\, g_1^\lambda(\Theta) := \ln\big(\mathrm{Meas}_1^\lambda(\Theta)\big) \le T_1^\lambda\Big\}, \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots \\ \Gamma_{\lambda,13}^\star := \Big\{\Theta \in \mathbb{X},\, g_{13}^\lambda(\Theta) := \ln\big(\mathrm{Meas}_{13}^\lambda(\Theta)\big) \le T_{13}^\lambda\Big\}, \end{cases} \tag{5.9}$$

where $\mathbb{X} := [0.8, 1.2] \times [0.6, 1.4]$ is the (compact) design space, constructed included $\theta^\star := (1, 1)$ and ln denotes the natural logarithm. The use of the logarithm removes the positivity con-

straints on Meas_i and Meas_i^{λ}, which is not compatible with Gaussian process regression.

As it is not possible to consider all the modes and frequencies of the structure, we focus on the two principal modes (Figure 5.12), referred to as Formulation 1. These two main modes are determined by examining the modulus of their associated vibration frequencies (Table 5.1 in Section 5.2.3). This limitation is necessary since the methods presented in chapter 4 are restricted to the simultaneous estimation of only two partial excursion sets.

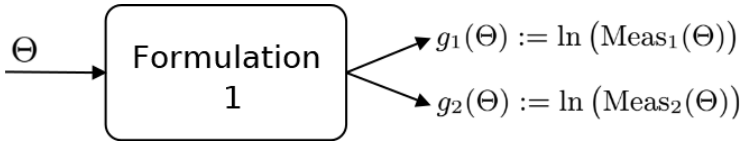
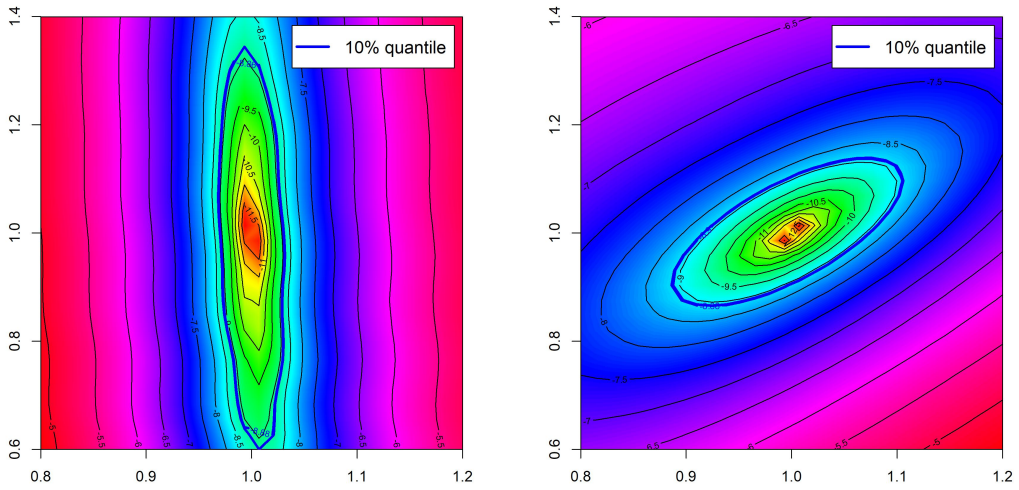


Figure 5.12: Schematic diagram of Formulation 1 of the pre-calibration problem using only the two main modes.

The problem is therefore reduced to estimating the two excursion sets:

$$\begin{cases} \Gamma_1^\star := \left\{\Theta \in \mathbb{X},\ g_1(\Theta) := \ln\left(\text{Meas}_1(\Theta)\right) \leq T_1\right\}, \\ \Gamma_2^\star := \left\{\Theta \in \mathbb{X},\ g_2(\Theta) := \ln\left(\text{Meas}_2(\Theta)\right) \leq T_2\right\}. \end{cases} \tag{5.10}$$

The short computation times of the black box $\boldsymbol{g}$, around 5 seconds, allow us to represent its level sets on a 30×30 grid (Figure 5.13). The dissimilarity measure for the first mode shows good identification of $\Theta_1^\star$ but not $\Theta_2^\star$, whereas the dissimilarity measure for the second mode allows good identification of $\Theta_1^\star - \Theta_2^\star$ but not $\Theta_1^\star + \Theta_2^\star$.



(a) Measure on Mod_1 (b) Measure on Mod_2

Figure 5.13: Representation of the logarithm of dissimilarity measures $\text{Meas}_1(\Theta)$ (left) and $\text{Meas}_2(\Theta)$ (right), on a grid of size 30 by 30 for the first two components of Θ on $[0.8, 1.2] \times [0.6, 1.4]$.

The low simulation time of the black box $\boldsymbol{g}$ limits the practical interest of strategies using a sequential DoE. However, this application enables us to illustrate the methodologies of

Chapter 4, and to compare their performances using true evaluations (Figure 5.13) instead of performing tests on an analytical test function. Simulation times are generally much higher in most industrial applications, justifying the interest of the approaches presented in Chapter 4.

As we have no precise indication of how to define the vector of thresholds $\boldsymbol{T}$, we choose to determine it from a quantile of order 10% defined using the 30×30 grid (see Figure 5.13), i.e., $\boldsymbol{T} = (-8.868, -8.891)$. Solving the pre-calibration problem using this formulation with the two main modes is detailed in Section 5.5.

According to table 5.1 in Section 5.2.3, information regarding the two principal modes comprises approximately one-third of the total information. We therefore want to include more information and not limit ourselves to the two principal modes. We propose a second formulation of the simulator pre-calibration problem that involves using information from both deformation modes and vibration frequencies, as both types of data are provided simultaneously by the simulator. The idea is to aggregate all measures associated with frequencies on one hand, and all those associated with modes on the other (Figure 5.14), referred to as Formulation 2.

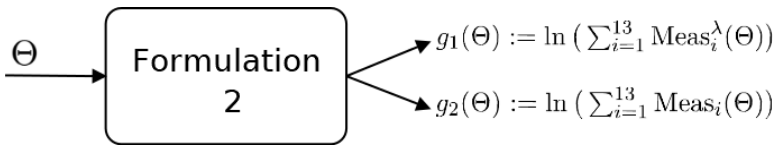


Figure 5.14: Schematic diagram of Formulation 2 of the pre-calibration problem using modes and frequencies.

The estimation problem we aim to address can be formulated as follows:

$$\begin{cases} \Gamma_1^\star := \left\{\Theta \in \mathbb{X},\, g_1(\Theta) := \ln\left(\sum_{i=1}^{13} \text{Meas}_i^{\lambda}(\Theta)\right) \leq T_1\right\}, \\ \Gamma_2^\star := \left\{\Theta \in \mathbb{X},\, g_2(\Theta) := \ln\left(\sum_{i=1}^{13} \text{Meas}_i(\Theta)\right) \leq T_2\right\}, \end{cases} \tag{5.11}$$

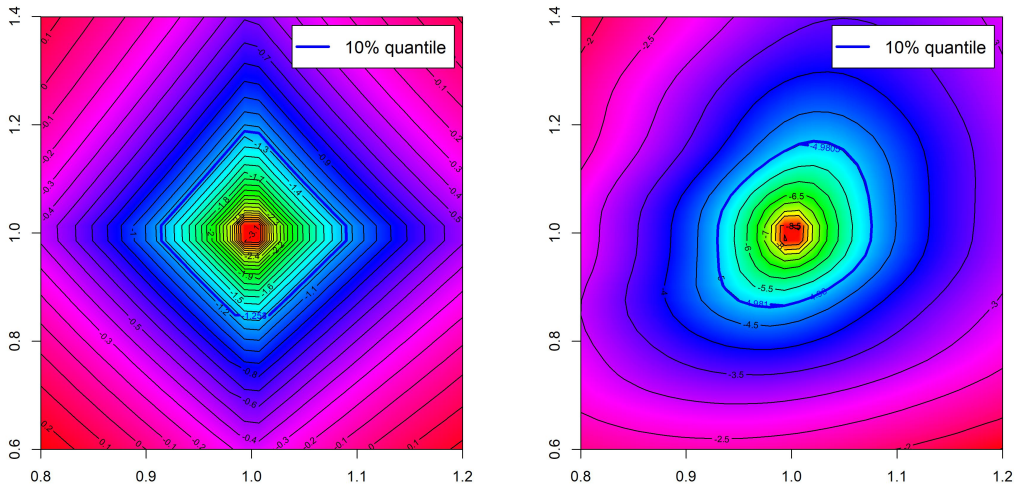
where $\boldsymbol{T} := (T_1, T_2)$ is a fixed vector of thresholds, and $\mathbb{X} := [0.8, 1.2] \times [0.6, 1.4]$.

The level sets of the new black box function considered are represented on a grid of size 30×30 (Figure 5.15). Unlike Formulation 1, the two components of the black box function each identify the two components of Θ. Once again, without precise guidance for determining the threshold vector $\boldsymbol{T}$, we determine it from a quantile of order 10% defined using the 30×30 grid (see Figure 5.13), i.e., $\boldsymbol{T} = (-1.254, -4.981)$. The two excursion sets thus proposed are easier to determine than those proposed in Formulation 1 (Figure 5.13), since the two sets of partial excursions resemble each other. Solving the pre-calibration problem using this formulation with frequencies and modes is detailed in Section 5.6.

5.5 Pre-calibration using only the two main modes (Formulation 1)

Enrichment is conducted over 30 iterations starting from initial LHS Maximin DoEs of size 5, employing the three methodologies ('Alternating Scal', 'Pareto Scal', and 'Vect') introduced in Chapter 4.

Implementation choices for surrogate models and enrichment criteria are similar to those in Section 4.4.1 of Chapter 4. The search bounds of the GP hyperparameters have been



(a) Measure on λ (b) Measure on Mod

Figure 5.15: Representation of the logarithm of the sum of dissimilarity measures $\text{Meas}_i^{\lambda}(\Theta)$ (left) and $\text{Meas}_i(\Theta)$ (right), on a grid of size 30 by 30 for the first two components of Θ on $[0.8, 1.2] \times [0.6, 1.4]$.

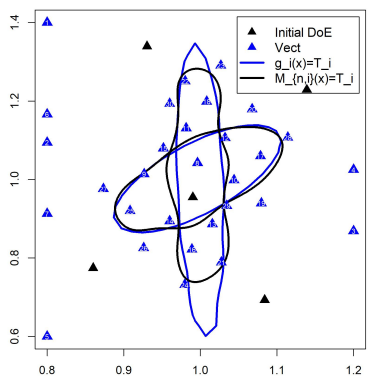
adjusted to the $\mathbb{X}$ design space by scaling the bounds selected in Chapter 4 according to the ranges of each component of $\mathbb{X}$: $\theta_1 \in [0.04, 250]$, $\theta_2 \in [0.08, 400]$. Partial relative errors Err_i and their sum Err_{sum} (Equations (4.28) and (4.29) from Chapter 4) are estimated from a grid of size 30 by comparing with the true evaluations of the black box function $\boldsymbol{g}$ (Figure 5.13).

For three different initial DoEs, sequential DoEs and partial excursion set estimates after 30 iterations are shown in Figure 5.16 for the different enrichment strategies. In these examples, Pareto Scalar strategy demonstrates the best estimates of partial excursion sets. Alternating Scalar strategy performs well overall, but exhibits limitations in the third example. Vector strategy, while less effective in estimating partial excursion sets, demonstrates good exploration capabilities across the design space.

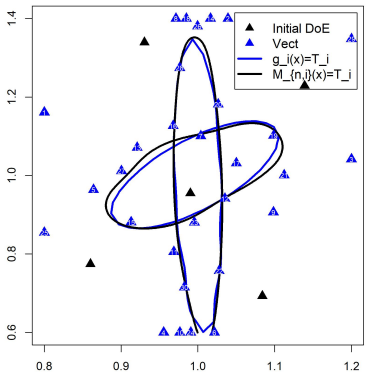
Line plots and means of partial relative errors and their sum are shown in Figure 5.17 for 10 different initial DoEs. The results show better performances for the scalar criteria, in particular for Pareto Scalar criterion. Plateaus observed in Figure 5.17a for relative errors below 0.1 highlight the limits of using a grid of size 30×30 to estimate partial relative errors.

We next plot standard Data profiles introduced in Sections 4.4.2 and 4.4.3 of Chapter 4. Due to the low precision of partial relative errors for values that are too low, relatively high thresholds are chosen for the Data profiles: $C = 100\%$, 50% and 20%. Data profiles obtained (Figure 5.18) confirm the better performance of Alternating and Pareto Scalar strategies compared with Vector strategy.

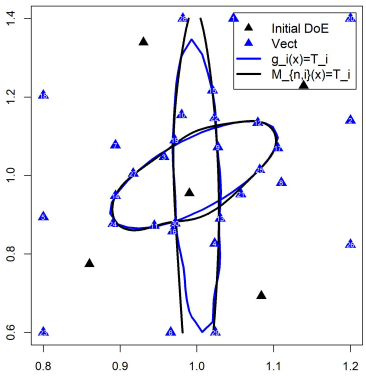
It is also interesting to compare the strategies in terms of range (length scale) hyperparameters estimation (Figure 5.19). For SOGP models (scalar strategies), $\theta_1^{(j)}$ and $\theta_2^{(j)}$ represent the ranges concerning the two input parameters for the j^{th} output component. These plots highlight significant differences in range between the two output components for SOGP models using scalar strategies, consistent with variations of the black box function shown in Figure 5.13. These range differences between output components cannot be captured by the MOGP model (Vector strategy) as the separable covariance model imposes a common range



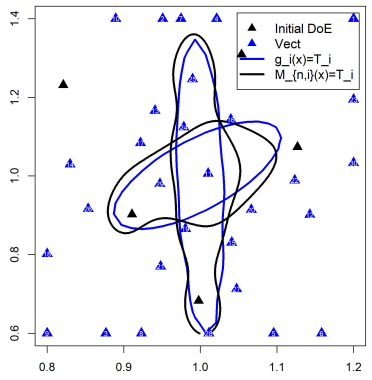
(a) Vect, LHS 1



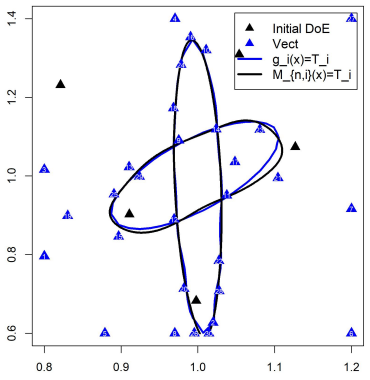
(b) Alternating Scal, LHS 1



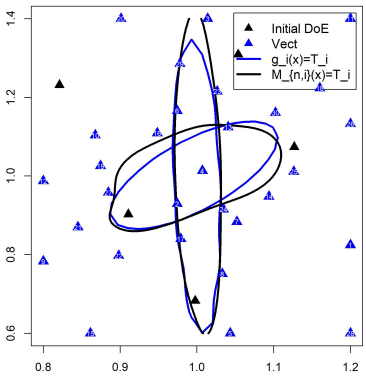
(c) Pareto Scal, LHS 1



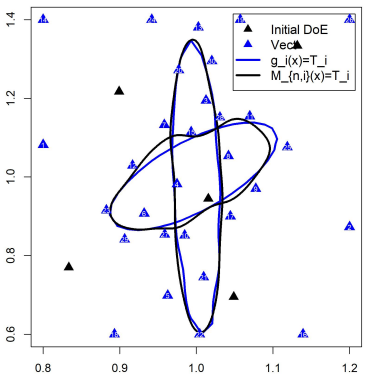
(d) Vect, LHS 2



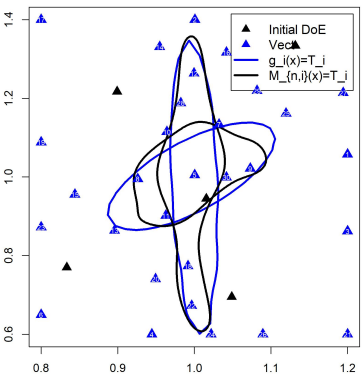
(e) Alternating Scal, LHS 2



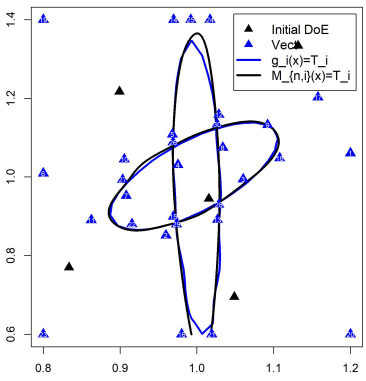
(f) Pareto Scal, LHS 2



(g) Vect, LHS 3

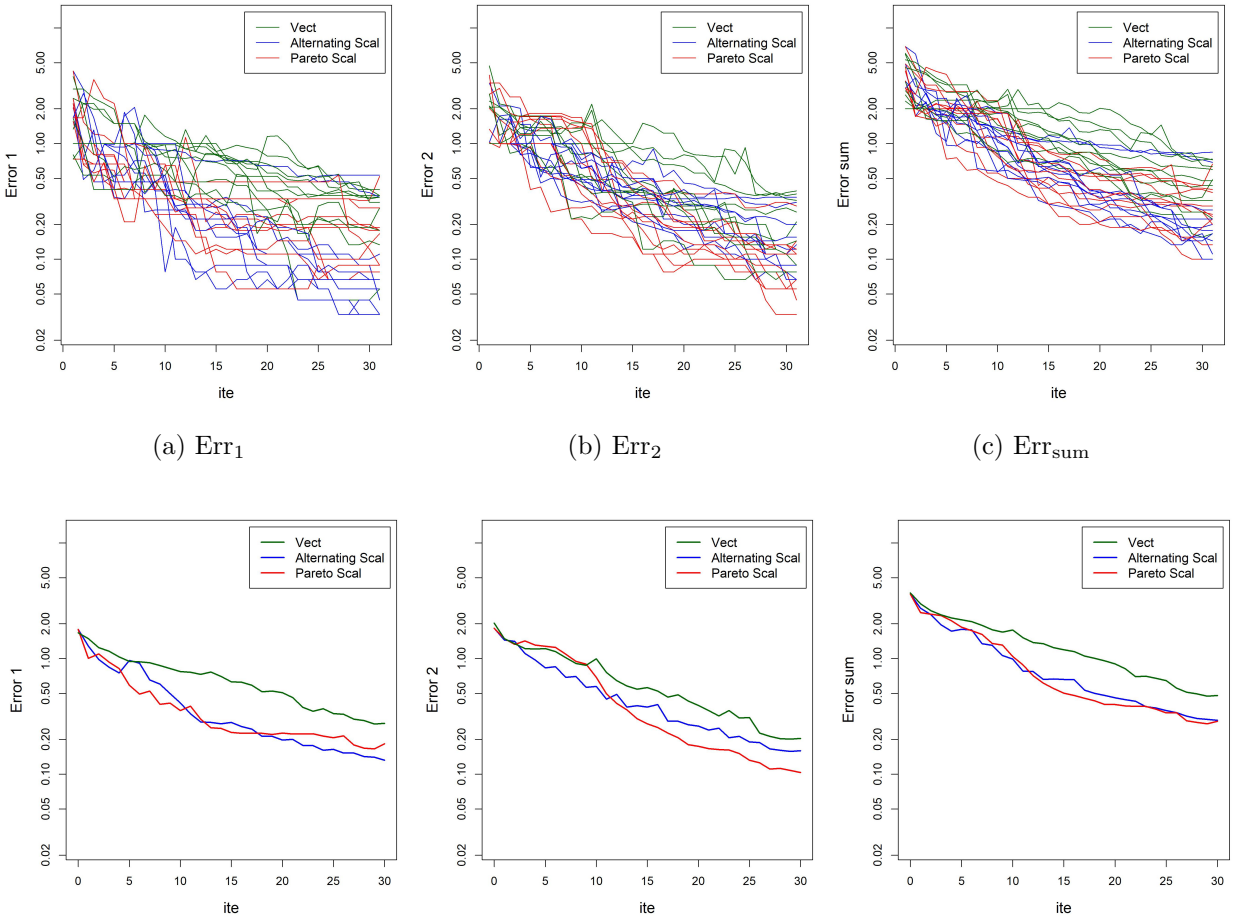


(h) Alternating Scal, LHS 3



(i) Pareto Scal, LHS 3

Figure 5.16: Representation of the enrichment of 3 initial LHS Maximin DoEs of size 5 after 30 iterations for the different enrichment criteria, for the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$. Boundaries of partial excursion sets are overlaid, calculated from a 30×30 grid.



(a) Err_1 (b) Err_2 (c) $\mathrm{Err}_{\mathrm{sum}}$

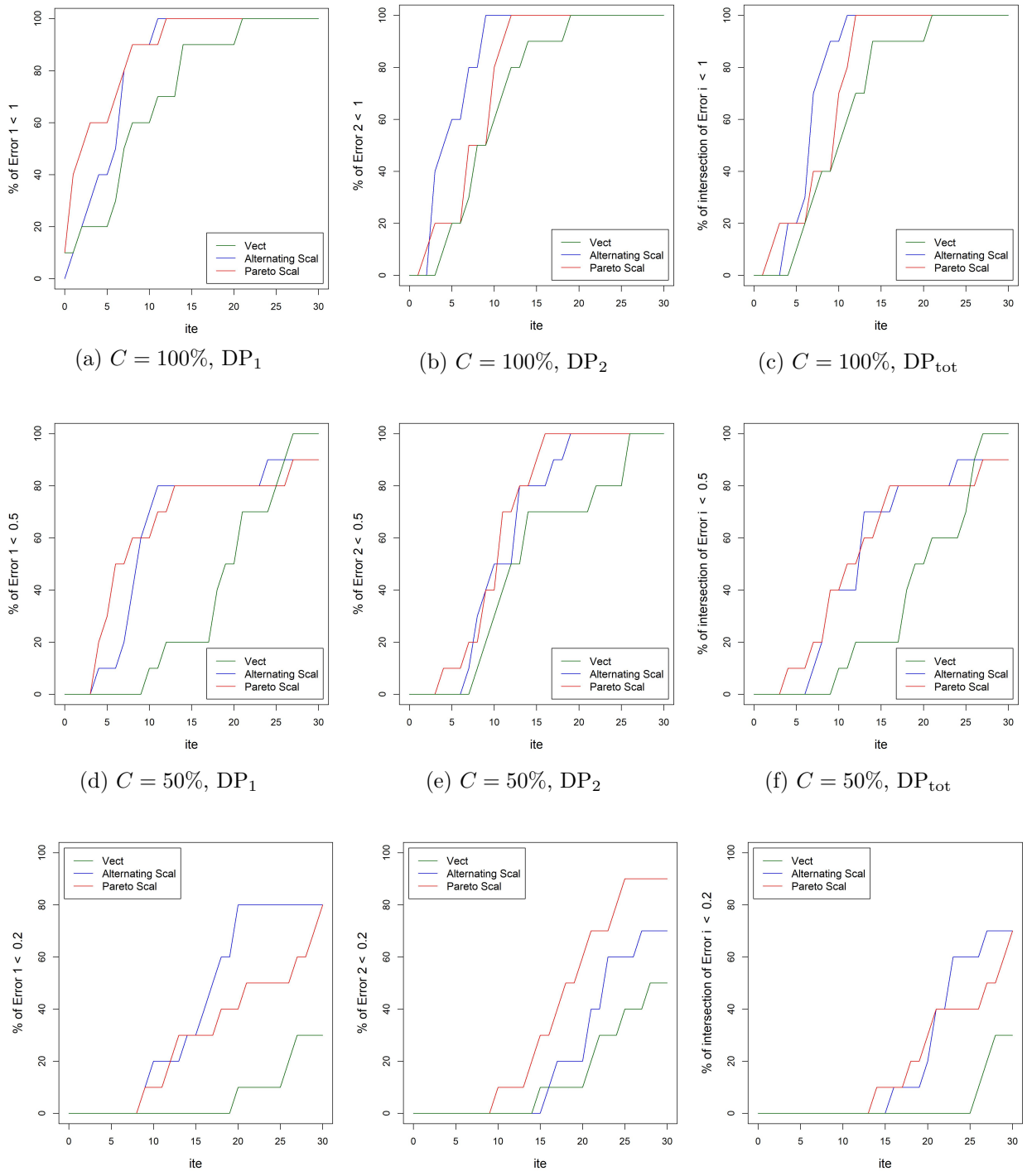
(d) Err_1 (means) (e) Err_2 (means) (f) $\mathrm{Err}_{\mathrm{sum}}$ (means)

Figure 5.17: Line plots and means of partial relative errors and of their sum with the number of iterations when estimating excursion sets of the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$. The three enrichment strategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Means are evaluated over 10 LHS Maximin intial DoE randomly chosen.

on both output components. This specific constraint of MOGP explains the relatively poor performance of Vector strategy in this context, underlining the limitations of such a separable MOGP model. It is also relevant to note that the θ_{kOut} parameter of the MOGP model (Figure 5.19b) is close to $\pi/2$, indicating a low correlation between the variables.

5.6 Pre-calibration using frequencies and modes (Formulation 2)

The same tests are conducted as previously. Figure 5.20 shows sequential DoEs and partial excursion set estimates after 30 iterations for three different initial DoEs, using the different enrichment strategies. In these examples, the accuracy of partial excursion set estimation appears to be higher than that observed in Formulation 1 (see Figure 5.16). Vector strategy

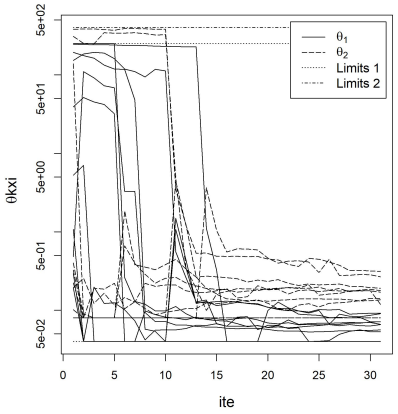


(a) $C = 100\%$, DP_1 (b) $C = 100\%$, DP_2 (c) $C = 100\%$, $\mathrm{DP}_{\mathrm{tot}}$

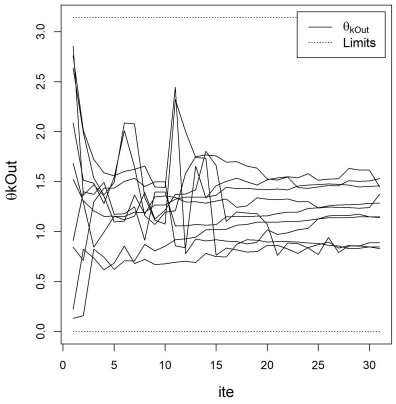
(d) $C = 50\%$, DP_1 (e) $C = 50\%$, DP_2 (f) $C = 50\%$, $\mathrm{DP}_{\mathrm{tot}}$

(g) $C = 20\%$, DP_1 (h) $C = 20\%$, DP_2 (i) $C = 20\%$, $\mathrm{DP}_{\mathrm{tot}}$

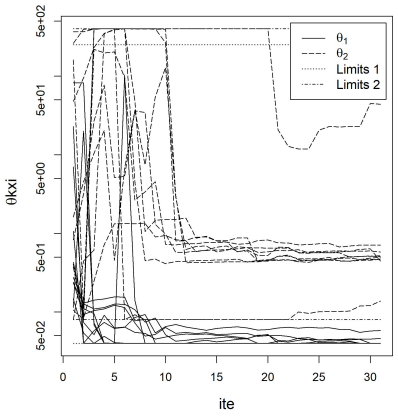
Figure 5.18: Standard "Data profiles" (DPs) of partial relative errors for the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$. DPs are represented for the different criteria, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations.



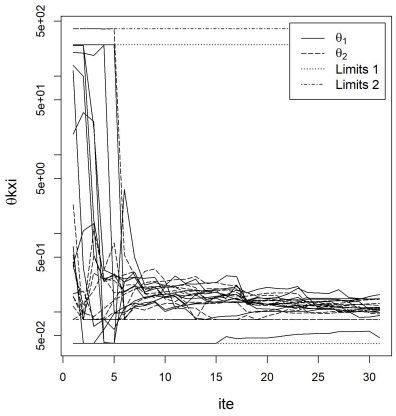
(a) Vect, θ_1 and θ_2



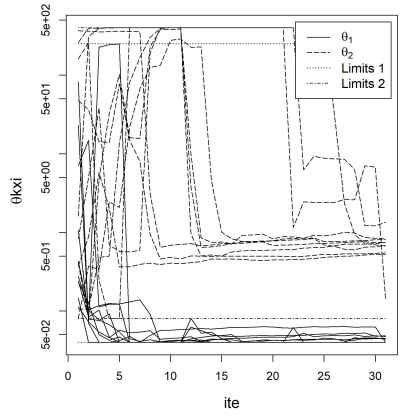
(b) Vect, θ_{kOut}



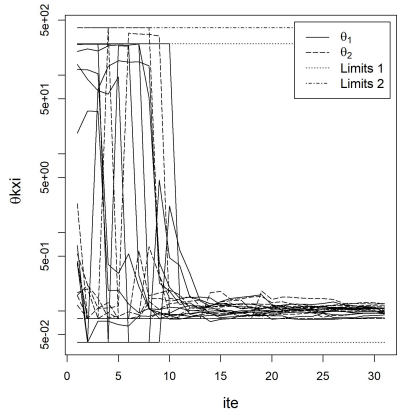
(c) Alternating Scal, $\theta_1^{(1)}$ and $\theta_2^{(1)}$



(d) Alternating Scal, $\theta_1^{(2)}$ and $\theta_2^{(2)}$



(e) Pareto Scal, $\theta_1^{(1)}$ and $\theta_2^{(1)}$



(f) Pareto Scal, $\theta_1^{(2)}$ and $\theta_2^{(2)}$

Figure 5.19: Representation of $\theta_i^{(1)}$ and $\theta_i^{(2)}$ hyperparameters of SOGP models associated with Alternating and Pareto Scalar criteria, and of θ_i and θ_{kOut} hyperparameters of MOGP model associated with Vector criterion, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations, for the pre-calibration using only the two main modes (Formulation 1) with $\boldsymbol{T} = (-8.868, -8.891)$.

again demonstrates the best ability to explore the design space (Figure 5.20).

Figure 5.21 presents line plots and means of partial relative errors and their sum for 10 different initial DoEs. The results demonstrate the superior performance of Pareto Scalar and Vector strategies compared to Alternating Scalar strategy. As observed in Formulation 1 of the pre-calibration problem, plateaus visible in Figure 5.21b, when relative errors are less than 0.1, highlight the limitations of estimating partial relative errors using a grid of size 30×30.

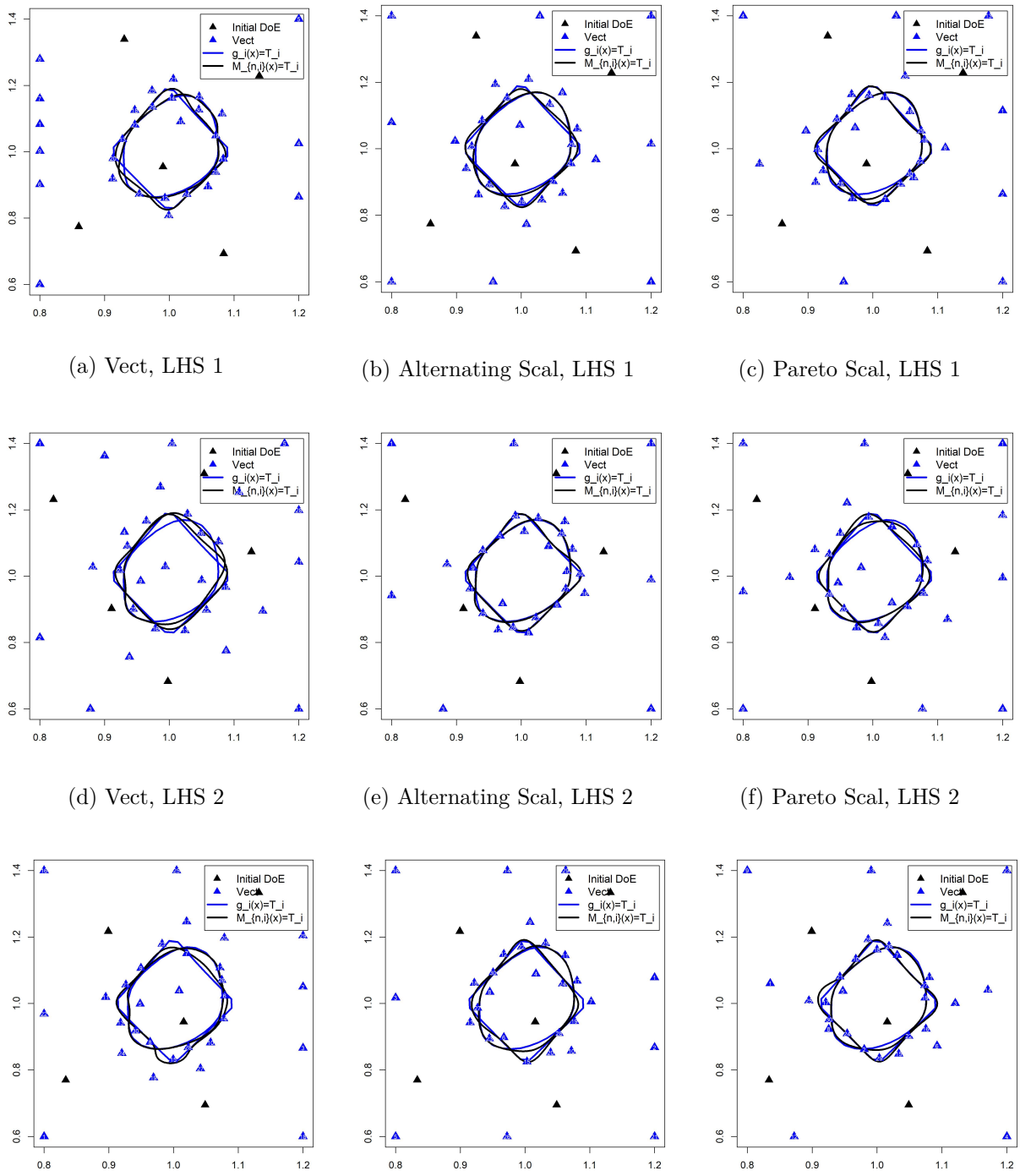
We once again plot the standard Data profiles presented in Sections 4.4.2 and 4.4.3 of chapter 4. In a similar way to Formulation 1, we use high thresholds for Data profiles, specifically $C = 100\%$, 50% and 20%, due to the limited precision of partial relative errors for very low values. The Data profiles obtained (Figure 5.22) do not show a significantly better performance of Pareto Scalar and Vector strategies compared to Alternating Scalar strategy. Upon analyzing the Data profiles, the performance appears to be relatively similar across the different strategies. Compared with Formulation 1, the results obtained for Vector strategy on Formulation 2 of the pre-calibration problem are much better. This is due to the greater correlation between the two output components (see Figure 5.15 in Section 5.4).

It is also interesting to compare the strategies in terms of range (length scale) hyperparameters estimation (Figure 5.23). For SOGP models, $\theta_1^{(j)}$ and $\theta_2^{(j)}$ denote the ranges concerning the two input parameters for the j^{th} output component. In contrast to Formulation 1, these plots reveal negligible differences in range between the two output components for SOGP models associated with the two scalar strategies, consistent with the variations observed in the black box function (Figure 5.15). This minor disparity in range between the output components in Formulation 2 explains the better performance of Vector strategy compared to Formulation 1, as the MOGP model used necessitates a common range hyperparameter for both output components (separable model). It is also worth noting that the θ_{kOut} parameter of MOGP model (Figure 5.23b) is closer to $\pi/2$ than in Formulation 1, indicating a stronger estimated correlation between the output components.

5.7 Conclusion

The various sequential DoE enrichment strategies presented in Chapter 4 are illustrated on two formulations of a wind turbine simulator pre-calibration problem. It is recalled that the strategies developed in chapter 4 are designed for the simultaneous estimation of partial excursion sets within the framework of a couple-valued black box function, and for isotopic data (simultaneous evaluation of all simulator output components).

Based on the analysis, the sequential enrichment strategies outlined in Chapter 4 show promising potential for the simultaneous estimation of partial excursion sets. They also highlight limitations of the MOGP model used in Vector strategy, which employs a separable ICM model with a shared range parameter for both simulator output components. However, in scenarios involving substantial correlation between the two outputs, the Vector strategy yields results comparable to other scalar strategies and exhibits a more exploratory nature. Unlike Chapter 4, the impact of this exploratory aspect on reducing partial relative errors compared to other strategies could not be emphasized due to the simplicity of the proposed partial excursion sets and the associated low accuracies of estimation errors.

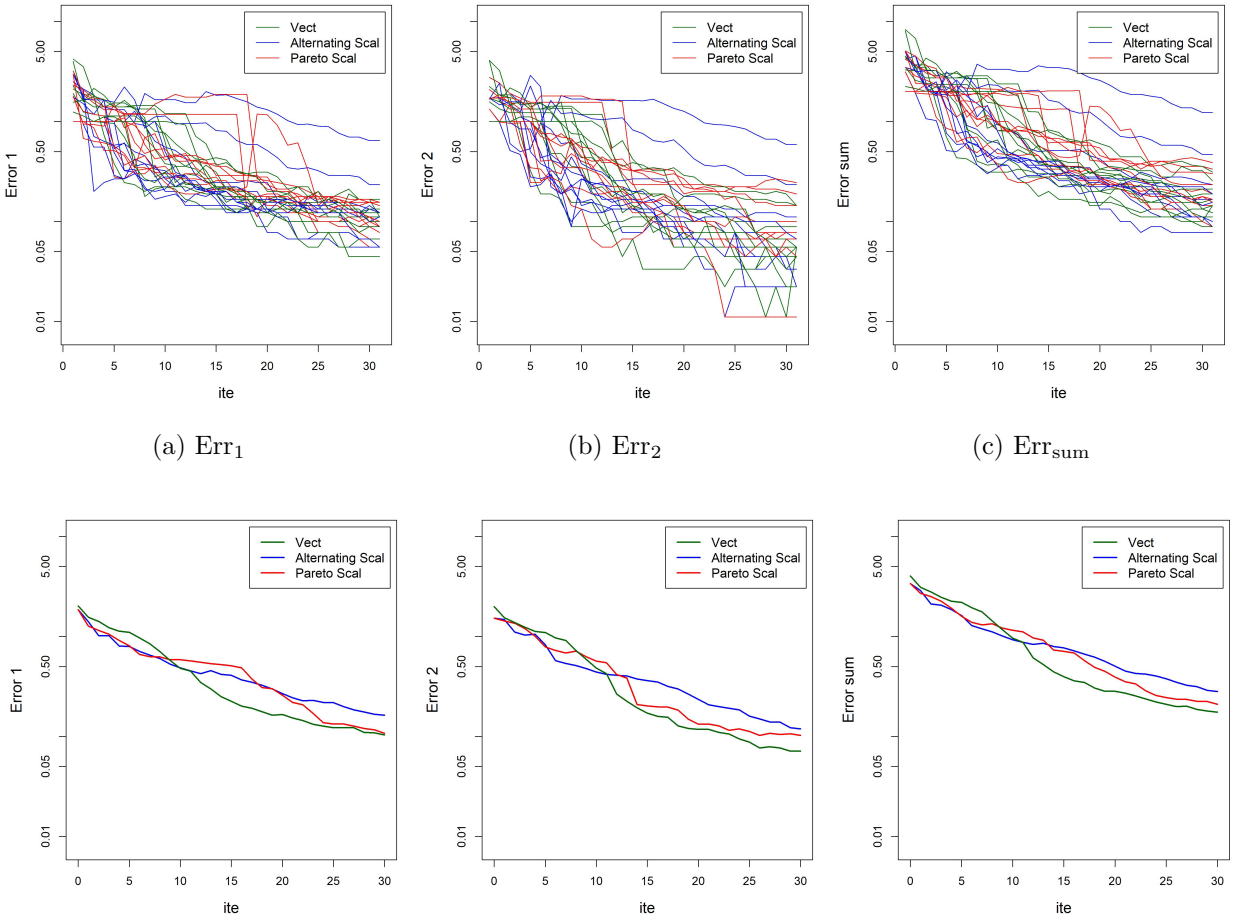


(a) Vect, LHS 1 (b) Alternating Scal, LHS 1 (c) Pareto Scal, LHS 1

(d) Vect, LHS 2 (e) Alternating Scal, LHS 2 (f) Pareto Scal, LHS 2

(g) Vect, LHS 3 (h) Alternating Scal, LHS 3 (i) Pareto Scal, LHS 3

Figure 5.20: Representation of the enrichment of 3 initial LHS Maximin DoEs of size 5 after 30 iterations for the different enrichment criteria, for the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$. Boundaries of partial excursion sets are overlaid, calculated from a 30×30 grid.



(a) Err_1 (b) Err_2 (c) $\mathrm{Err}_{\mathrm{sum}}$

(d) Err_1 (means) (e) Err_2 (means) (f) $\mathrm{Err}_{\mathrm{sum}}$ (means)

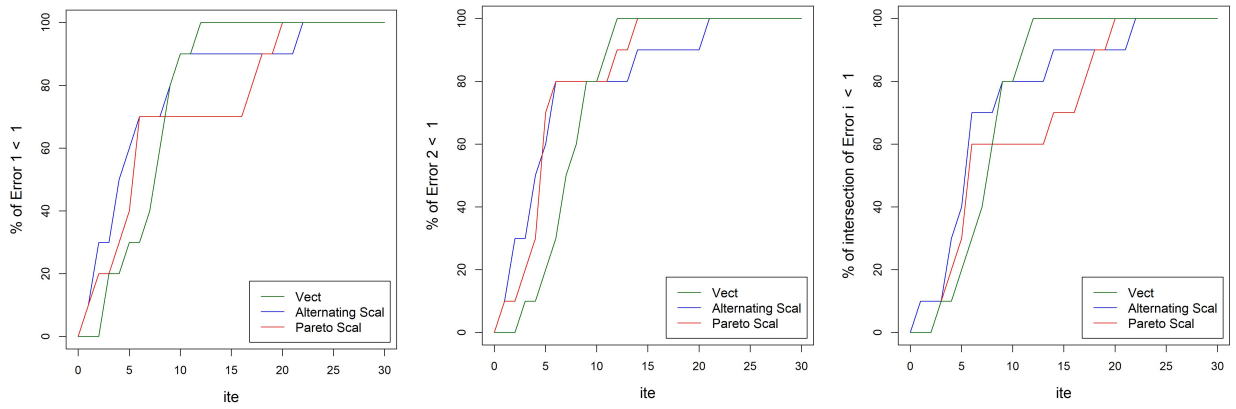
Figure 5.21: Line plots and means of partial relative errors and of their sum with the number of iterations when estimating excursion sets of the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$. The three enrichment strategies Vect, Alternating Scal and Pareto Scal, are performed from an initial DoE of size 5 and with 30 enrichment iterations. Means are evaluated over 10 LHS Maximin intial DoE randomly chosen.

5.8 Appendix

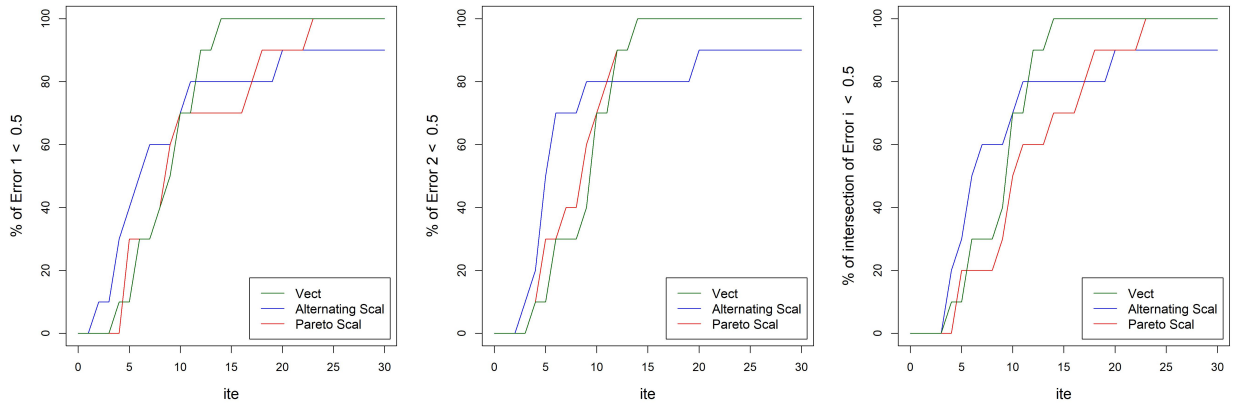
A Checking mode matching with respect to vibration frequencies

In this section, we assess the consistency of mode matching algorithm (Algorithm 1 from Section 5.3.1) with respect to vibration frequencies. We want to verify whether the frequencies are correctly paired after mode matching. We evaluate the performance of Algorithm 1 for two different values of Θ: $\Theta = (1.05, 1.1, 1.1)$ and $(1.673, 0.9646, 0.6424)$. For each of these cases, we compute the sum of relative errors associated with the vibration frequencies λ across 10 000 random permutations σ on $\{1, \ldots, 13\}$:

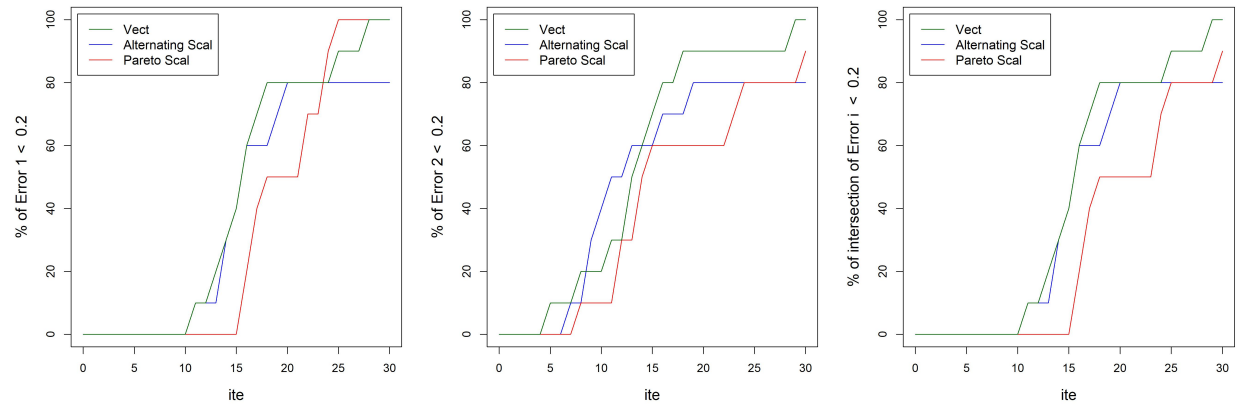
$$\sum_{i=1}^{13} \frac{|\lambda_{\sigma(i)}(\Theta) - \lambda_i^\star|}{|\lambda_i^\star|}. \tag{5.12}$$



(a) $C = 100\%$, DP_1 (b) $C = 100\%$, DP_2 (c) $C = 100\%$, $\mathrm{DP}_{\mathrm{tot}}$

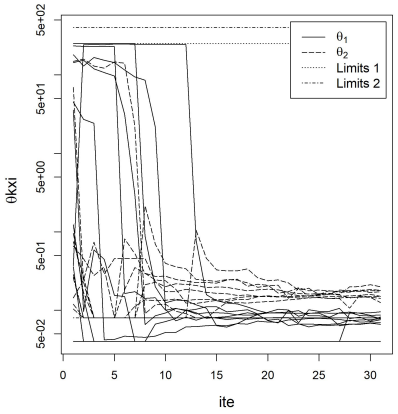


(d) $C = 50\%$, DP_1 (e) $C = 50\%$, DP_2 (f) $C = 50\%$, $\mathrm{DP}_{\mathrm{tot}}$

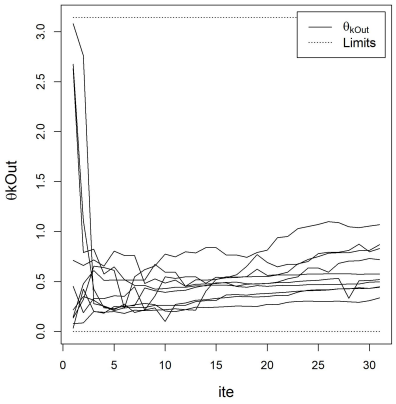


(g) $C = 20\%$, DP_1 (h) $C = 20\%$, DP_2 (i) $C = 20\%$, $\mathrm{DP}_{\mathrm{tot}}$

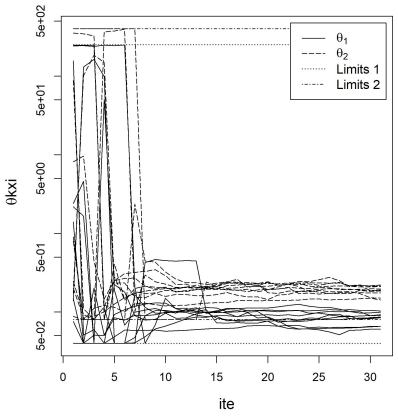
Figure 5.22: Standard "Data profiles" (DPs) of partial relative errors for the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$. DPs are represented for the different criteria, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations.



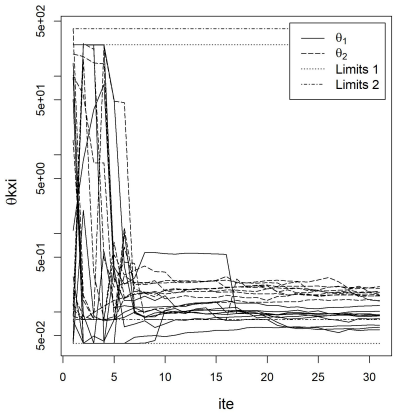
(a) Vect, θ_1 and θ_2



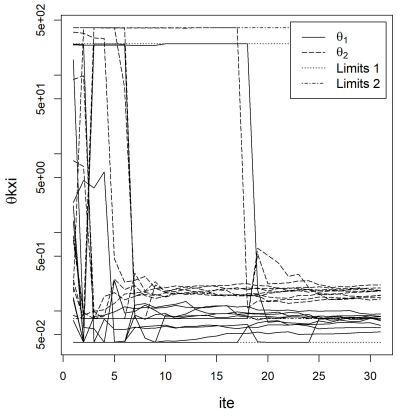
(b) Vect, θ_{kOut}



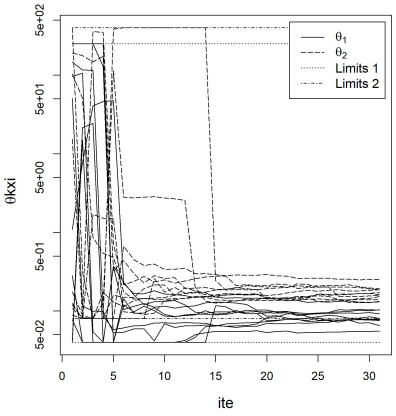
(c) Alternating Scal, $\theta_1^{(1)}$ and $\theta_2^{(1)}$



(d) Alternating Scal, $\theta_1^{(2)}$ and $\theta_2^{(2)}$



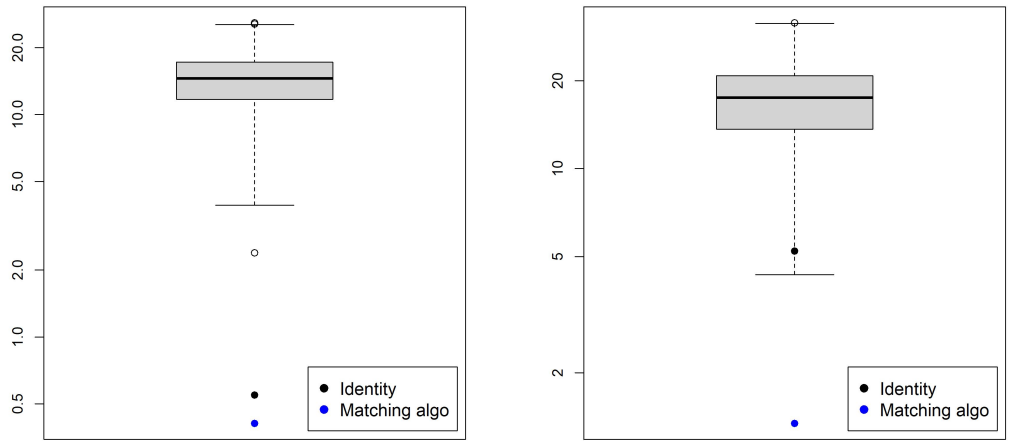
(e) Pareto Scal, $\theta_1^{(1)}$ and $\theta_2^{(1)}$



(f) Pareto Scal, $\theta_1^{(2)}$ and $\theta_2^{(2)}$

Figure 5.23: Representation of $\theta_i^{(1)}$ and $\theta_i^{(2)}$ hyperparameters of SOGP models associated with Alternating and Pareto Scalar criteria, and of θ_i and θ_{kOut} hyperparameters of MOGP model associated with Vector criterion, in the case of enrichment of 10 LHS Maximin initial DoEs of size 5 with 30 iterations, for the pre-calibration using frequencies and modes (Formulation 2) with $\boldsymbol{T} = (-1.254, -4.981)$.

We then compare the values obtained with different choices for permutation σ: we try 10 000 random permutations, the identity and finally the permutation provided by Algorithm 1. The results are shown in Figure 5.24 (see also Table 5.2). These results highlight the efficiency of Algorithm 1, at least for $\Theta = (1.05, 1.1, 1.1)$ and $\Theta = (1.673, 0.9646, 0.6424)$.



(a) $\Theta = (1.05, 1.1, 1.1)$ (b) $\Theta = (1.673, 0.9646, 0.6424)$

Figure 5.24: Boxplot (with logarithmic scale) of the distribution of the values of Equation (5.12) for 10 000 random permutations on $\{1, \dots 13\}$ and comparison with the same measure for the identity permutation and for the permutation resulting from mode matching algorithm (Algorithm 1).

Θ	Mean	Median	Min	Max	$\sigma = i_d$	$\sigma = \sigma_{\text{algo}}$
$(1.05, 1.1, 1.1)$	15.68	15.63	2.65	28.35	0.641	0.496
$(1.673, 0.9646, 0.6424)$	14.85	14.86	3.56	25.70	4.056	1.216

Table 5.2: Summary table of the distribution of the values of Equation (5.12) for 10 000 random permutations on $\{1, \dots 13\}$ and comparison with the same measure for the identity permutation and for the permutation resulting from mode matching algorithm (Algorithm 1).

We then repeat the experiment for 100 different values of Θ (sampled from a Maximin LHS). For each value of Θ we compute the sum in (5.12) for 10000 random permutations, but also for σ set to the identity and σ set to the permutation provided by Algorithm 1. Let us summarize the results from these experiments. For 23 values of Θ in the LHS, at least one of the 10000 random permutations beats the identity, while for any value of Θ in the LHS, the permutation provided by Algorithm 1 is better than any random permutation.

Summary:

This chapter implements the methodology presented in Chapter 4 for pre-calibrating a numerical wind turbine simulator capable of simulating vibration frequencies and deformation modes under wind loading. The objective is to determine a set of input parameters Θ for the simulator such that the simulated vibration frequencies $\lambda_i(\Theta)$ and deformation modes $\mathrm{Mod}_i(\Theta)$ closely match the reference values $\lambda_i^\star$ and $\mathrm{Mod}_i^\star$ obtained from experimental data using operational modal analysis (OMA).

Sequential DoE enrichment strategies are applied to two formulations of the pre-calibration problem: the first focuses on the two main deformation modes, while the second is more general, encompassing all deformation modes on one hand, and all vibration frequencies on the other hand. The results demonstrate that these strategies help in the estimation of partial excursion sets. However, we observe limitations probably due to the lack of flexibility for the covariance structure of the MOGP model we employed.

Conclusion/Perspectives

Cette thèse a permis d'explorer et de développer des méthodes avancées pour la conception séquentielle de plans d'expériences (DoEs), visant à estimer des ensembles d'excursion d'une fonction boîte noire à sortie scalaire ou vectorielle. Les différentes méthodes d'enrichissement proposées s'appuient sur un modèle de substitution, c'est-à-dire une approximation du véritable modèle, déterminé à partir d'un nombre restreint d'évaluations du modèle et moins coûteux à évaluer que ce dernier. Nous avons utilisé un modèle de substitution de type régression par processus gaussiens (GPR) pour le cas scalaire et son extension aux fonctions vectorielles, connue sous le nom de processus gaussiens multi-sorties (MOGP). L'un des principaux avantages de ces modèles est qu'ils offrent non seulement des prédictions ponctuelles, mais également une estimation de l'incertitude qui leur est associée.

Contributions principales

Dans le cadre de l'estimation des ensembles d'excursion pour des modèles boîtes noires à sortie scalaire, nous avons développé une version Stepwise Uncertainty Reduction (SUR) du critère Bichon, critère couramment utilisé avec la régression par processus gaussiens (GPR). Les stratégies SUR enrichissent le DoE en anticipant l'impact de l'ajout de nouveaux points pour réduire une incertitude résiduelle spécifique.

L'objectif était de créer une version plus efficace du critère Bichon et de la comparer à des critères classiques tels que les critères Bichon et SUR Vorob'ev. Une formulation explicite du critère a été développée pour faciliter sa mise en œuvre pratique. Les simulations numériques sur des fonctions analytiques en 2 et 6 dimensions ont montré que ce nouveau critère offre un bon comportement exploratoire dans différentes zones de l'espace de design. Grâce à son caractère exploratoire, il permet de détecter efficacement des ensembles d'excursion complexes avec plusieurs composantes connexes, tout en maintenant des performances équivalentes au critère SUR Vorob'ev en termes d'exploitation. En résumé, le critère SUR Bichon combine l'exploration et la robustesse du critère Bichon avec l'exploitation et les performances des stratégies SUR.

Dans le cadre de modèles boîtes noires à sorties vectorielles, nous avons développé trois stratégies pour l'estimation simultanée des ensembles d'excursions partiels associés à chaque composante de sortie. Ces stratégies s'appliquent à des données isotopiques, où toutes les composantes de sortie sont évaluées simultanément.

Les deux premières stratégies utilisent un modèle de substitution scalaire pour chaque composante de sortie avec un critère adapté (Alterné ou Pareto) pour le choix d'un point d'enrichissement commun. La troisième stratégie utilise un modèle de substitution vectoriel (MOGP) avec un critère prenant en compte la corrélation entre les sorties. Une formulation explicite de ce critère vectoriel a été développée pour permettre son implémentation pratique. La performance de ces trois critères a été testée sur des fonctions de test à 2 et 4 dimensions,

chacune ayant deux composantes de sortie. Toutes les stratégies montrent une réduction significative des erreurs relatives partielles au fil des itérations.

L'analyse de ces erreurs via des boxplots fonctionnels révèle que, surtout pour le cas 4-dimensionnel, la stratégie vectorielle offre un meilleur compromis pour la recherche simultanée des deux ensembles d'excursion, avec une robustesse accrue. La stratégie scalaire alternée a échoué dans le cas d'un choix de seuil inadapté, alors que les stratégies vectorielle et scalaire Pareto se sont avérées robustes et efficaces même dans ce cas difficile. Les profils de données (Data profiles) confirment ces résultats, montrant que la stratégie vectorielle est plus performante, bien que son temps de calcul plus élevé la rende particulièrement intéressante pour des fonctions avec des temps d'évaluation élevés. Pour des temps d'évaluation très courts, la stratégie scalaire alternée est plus avantageuse en raison de son faible temps de calcul pour l'enrichissement.

Application industrielle

Les différentes stratégies d'enrichissement séquentiel du DoE présentées au chapitre 4 sont illustrées sur un cas d'application industriel de pré-calibration d'un simulateur d'éolienne. Ce simulateur prend en entrée des paramètres du système, tels que les coefficients de rigidité de certains matériaux, et renvoie en sortie les fréquences de vibration et les modes de déformation de la structure mécanique en réponse à des charges de vent. L'éolienne étudiée est la DTU 10-MW de référence. L'objectif est d'identifier des jeux de paramètres admissibles en entrée du simulateur, pour que les fréquences de vibration et les modes de déformation en sortie correspondent aux données expérimentales obtenues par l'Operational Modal Analysis (OMA), méthode d'identification des caractéristiques vibratoires de structures mécaniques.

Les stratégies d'enrichissement séquentielles développées au Chapitre 4 ont été appliquées à deux formulations spécifiques du problème de pré-calibration. La première se concentre sur l'estimation simultanée des deux ensembles d'excursion partiels associés aux deux principaux modes de déformation, tandis que la seconde englobe tous les modes de déformation d'un côté et toutes les fréquences de vibration de l'autre. Les résultats montrent que les nouvelles méthodes offrent de bons résultats pratiques pour l'estimation des ensembles d'excursions partiels. Cependant, des limites du modèle MOGP utilisé pour la stratégie vectorielle ont été observées, notamment en raison de l'utilisation d'un paramètre de portée commun pour les deux composantes de sortie. Malgré cela, en cas de corrélation significative entre les deux sorties, la stratégie vectorielle produit des résultats comparables aux autres stratégies scalaires, avec un caractère exploratoire plus développé. L'impact de ce caractère exploratoire sur la réduction des erreurs relatives partielles n'a toutefois pas pu être mis en évidence en raison de la simplicité des ensembles d'excursion partiels proposés et des faibles précisions des erreurs relatives partielles.

Ces méthodologies, validées dans un contexte industriel, ouvrent la voie à des applications dans divers domaines nécessitant des modèles précis et efficaces pour des systèmes complexes à sorties multiples. Elles représentent une avancée significative vers l'optimisation des processus de conception de plans d'expériences pour l'estimation simultanée d'ensembles d'excursions partiels dans des environnements de simulation coûteuse à sortie vectorielle.

Limites de l'étude et perspectives de recherche

Plusieurs limites ont été identifiées dans notre étude. Elles sont présentées ci-dessous, accompagnées de perspectives de recherche.

Dans notre étude, les prédictions issues de la régression par processus gaussiens reposent sur l'hypothèse que le modèle est une réalisation d'un processus gaussien stationnaire. Cette hypothèse n'est cependant pas toujours vérifiée en pratique, ce qui peut restreindre la flexibilité et la précision des modèles. Pour surmonter ce problème, il est possible de relaxer cette hypothèse en utilisant des processus gaussiens non stationnaires (voir par exemple [Paciorek and Schervish, 2003] et [Heinonen et al., 2016]). Cette approche pourrait améliorer le modèle de substitution, particulièrement dans les contextes où les relations entre les variables d'entrée et de sortie évoluent de manière fortement non stationnaire.

Dans le Chapitre 4, la stratégie scalaire Alternée a montré des limites significatives, particulièrement lorsque l'une des valeurs du vecteur de seuil $\boldsymbol{T}$ est fortement inadaptée par rapport aux valeurs du modèle boîte noire de la composante de sortie correspondante. Dans de telles situations, cette stratégie échoue à estimer précisément les ensembles d'excursions partiels en attribuant un poids égal aux deux composantes, alors que l'ensemble d'excursion partiel associé à l'une d'entre elles est déjà complètement déterminé. Actuellement, nous ne voyons pas de perspectives claires d'amélioration pour cette problématique, à moins de vérifier attentivement le vecteur de seuils par rapport aux valeurs du modèle sur le DoE initial, ou d'utiliser les deux autres stratégies disponibles.

Ensuite, bien que la structure de corrélation séparable du modèle MOGP simplifie les calculs, elle réduit la flexibilité et la précision en ne capturant pas les interactions complexes entre les sorties (voir Section 1.7.3 du Chapitre 1 sur l'autokrigeabilité). Cette limite peut entraîner une perte d'efficacité dans certains scénarios où ces interactions jouent un rôle crucial. Nous avons également observé que l'utilisation d'un paramètre de portée commun pour les deux composantes de sortie est limitante lorsque les paramètres de portée estimés pour chaque composante sont très différents (voir Section 5.5 du Chapitre 5). Pour surmonter ces limites du modèle séparable, il serait pertinent d'explorer des structures de corrélation plus complexes et adaptatives qui pourraient mieux modéliser les interactions entre les différentes composantes de sortie, augmentant ainsi la précision et l'efficacité du modèle (voir Section 1.3 du Chapitre 1). Une étude sur des stratégies combinées, par exemple en utilisant un critère vectoriel associé à des modèles de GPR indépendants pour chaque composante, permettrait de déterminer avec précision l'impact du critère et du modèle sur la stratégie vectorielle proposée. Une autre approche pour résoudre les problèmes liés à l'autokrigeabilité du modèle MOGP consisterait à utiliser des données hétérotopiques, où l'évaluation du modèle n'est pas systématiquement réalisée pour toutes les composantes de sortie. Cependant, l'utilisation de cette méthode dépend du contexte d'application spécifique.

Une autre limite rencontrée est le temps de calcul élevé associé à la stratégie Vectorielle. Cette approche nécessite un temps de calcul considérablement plus long pour la mise à jour du modèle de substitution (MOGP) et l'optimisation du critère d'enrichissement, la rendant moins adaptée aux fonctions boîtes noires avec des temps d'évaluation courts. Pour surmonter cette limite, il serait bénéfique de développer des algorithmes plus efficaces pour la mise à jour du modèle de substitution vectoriel et pour l'optimisation du critère d'enrichissement. Cela permettrait de réduire les temps de calcul et améliorer les performances de la stratégie vectorielle, dans le cas de fonctions boîtes noires avec des temps d'évaluation courts.

Les trois stratégies ont été évaluées sur des cas à deux composantes de sortie, mais elles sont théoriquement généralisables à un nombre de sorties supérieur. Cependant, cette généralisation nécessite des efforts supplémentaires pour formuler des approches simples à implémenter, particulièrement pour la stratégie Vectorielle. Une voie d'amélioration serait donc de développer des formulations et des algorithmes qui facilitent cette extension, permettant ainsi l'application des stratégies à des problèmes plus complexes comportant $p > 2$

composantes de sortie.

Dans le Chapitre 5, nous avons discuté des limites précedemment citées associées au modèle MOGP utilisé. Du point de vue de l'application, il est également important de noter que nous n'avons pas eu accès aux données OMA réelles et que nous avons utilisé des fréquences de vibration et des modes de déformation simulés pour illustrer les méthodologies développées dans le Chapitre 4. En pratique, il serait essentiel d'utiliser les données OMA réelles fournies par les experts éoliens. De plus, nous nous sommes limités à seulement deux paramètres d'entrée du simulateur, alors qu'en pratique, jusqu'à six paramètres peuvent être considérés pour la calibration du simulateur, ce qui accroît la complexité de l'estimation des ensembles d'excursions partiels.

Par ailleurs, l'application de ces méthodes à d'autres secteurs industriels tels que la santé, l'automobile ou l'aérospatiale pourrait ouvrir de nouvelles perspectives et défis, renforçant ainsi l'impact de ce travail.

Conclusion/Perspectives

This thesis explored and developed advanced methods for sequential Design of Experiments (DoEs) aimed at estimating excursion sets of a black-box function with scalar or vector output. The various enrichment methods proposed are based on a surrogate model, i.e., an approximation of the true model, determined from a limited number of model evaluations and less costly to evaluate than the latter. We used Gaussian Process Regression (GPR) surrogate models for the scalar case and their extension to vector functions, known as Multi-Output Gaussian Processes (MOGP). One of the main advantages of these models is that they offer not only point predictions but also an estimate of the uncertainty associated with them.

Main contributions

In the context of estimating excursion sets for black box models with scalar outputs, we have developed a Stepwise Uncertainty Reduction (SUR) version of Bichon criterion, commonly used with Gaussian process regression (GPR). SUR strategies enrich the DoE by anticipating the impact of adding new points to reduce a specific residual uncertainty.

The aim was to create a more efficient version of Bichon criterion and compare it with classical criteria such as Bichon and SUR Vorob'ev criteria. An explicit formulation of the criterion was developed to facilitate its practical implementation. Numerical simulations on analytical functions in 2 and 6 dimensions have shown that this new criterion offers good exploratory behavior in different zones of the design space. Thanks to its exploratory nature, it can efficiently detect complex excursion sets with several connected components, while maintaining performance equivalent to SUR Vorob'ev criterion in terms of exploitation. In summary, SUR Bichon criterion combines the exploration and robustness of Bichon criterion with the exploitation and performance of SUR strategies.

Within the framework of black box models with vector outputs, we have developed three strategies for the simultaneous estimation of partial excursion sets associated with each output component. These strategies apply to isotopic data, where all output components are evaluated simultaneously.

The first two strategies use a scalar surrogate model for each output component with a suitable criterion (Alternating or Pareto) for choosing a common enrichment point. The third strategy uses a vector surrogate model (MOGP) with a criterion that takes into account the correlation between outputs. An explicit formulation of this vector criterion has been developed for practical implementation. The performance of these three criteria was tested on 2 and 4 dimensional test functions, each with two output components. All strategies show a significant reduction in relative partial errors with each iteration.

Analysis of these errors via functional boxplots reveals that, especially for the 4-dimensional case, Vector strategy offers a better compromise for the simultaneous search of both excursion sets, with increased robustness. Alternating scalar strategy failed in the case of an inappropri-

ate choice of threshold, while Vector and Pareto scalar strategies proved robust and effective even in this difficult case. Data profiles confirm these results, showing that Vector strategy performs better, although its higher computation time makes it particularly interesting for functions with high evaluation times. For very short evaluation times, Alternating scalar strategy is more advantageous due to its low computation time for enrichment.

Industrial application

The various sequential DoE enrichment strategies presented in Chapter 4 are illustrated through an industrial case study involving the pre-calibration of a wind turbine simulator. The simulator takes system parameters such as material stiffness coefficients as input and outputs vibration frequencies and deformation modes of the mechanical structure in response to wind loads. The wind turbine under study is the DTU 10-MW reference model. The objective is to identify acceptable parameter sets for the simulator input so that the output vibration frequencies and deformation modes closely match the experimental data obtained by Operational Modal Analysis (OMA), a method used to identify the vibratory characteristics of mechanical structures.

The sequential enrichment strategies developed in Chapter 4 have been applied to two distinct formulations of the pre-calibration problem. The first formulation focuses on the simultaneous estimation of two partial excursion sets associated with the two main deformation modes, while the second encompasses all deformation modes on one side and all vibration frequencies on the other. The results show that the new methods offer good practical results for estimating partial excursion sets. However, limitations of the MOGP model used for Vector strategy were observed, notably due to the use of a common range parameter for both output components. Despite this limitation, in cases where significant correlation exists between the two outputs, Vector strategy produced results comparable to the other scalar strategies, with a more developed exploratory character. Nevertheless, the impact of this exploratory approach on reducing partial relative errors could not be conclusively demonstrated, primarily due to the simplicity of the proposed partial excursion sets and the inherent imprecision associated with partial relative errors.

These methodologies, validated in an industrial context, pave the way for applications in various fields requiring accurate and efficient models for complex systems with multiple outputs. They represent a significant advancement in optimizing Design of Experiments (DoE) processes for the simultaneous estimation of partial excursion sets in costly vector-output simulation environments.

Study limitations and research perspectives

Several limitations were identified in our study. They are presented below, along with research perspectives.

In our study, predictions from Gaussian process regression rely on the assumption that the model is a realization of a stationary Gaussian process. However, this assumption is not always met in practice, which can restrict the flexibility and accuracy of models. To overcome this problem, it is possible to relax this assumption by using non-stationary Gaussian processes (see for example [Paciorek and Schervish, 2003] and [Heinonen et al., 2016]). This approach could improve the surrogate model, especially in scenarios where the relationships between input and output variables change in a highly non-stationary manner.

In Chapter 4, Alternating scalar strategy showed significant limitations, particularly when one of the values of the threshold vector $\boldsymbol{T}$ is unsuitable with respect to the model values of the corresponding output component. In such situations, this strategy fails to accurately estimate partial excursion sets by assigning equal weights to both components, when the partial excursion set associated with one of them is already fully determined. At present, we see no clear prospects of improvement for this problem, unless we carefully check the threshold vector against the model values on the initial DoE, or use the other two available strategies.

Although the separable correlation structure of the MOGP model simplifies calculations, it reduces flexibility and accuracy by not capturing complex interactions between outputs (see Section 1.7.3 of Chapter 1 on autokrigeability). This limitation can lead to a loss of efficiency in certain scenarios where these interactions play a crucial role. We have also observed that using a common range parameter for both output components restricts model adaptability when the estimated range parameters for each component differ significantly (see Section 5.5 of Chapter 5). To overcome these limitations of the separable model, exploring more intricate and adaptive correlation structures could be beneficial. Such structures could better capture interactions between different output components, thereby enhancing both the accuracy and efficiency of the model (see Section 1.3 of Chapter 1). A study of combined strategies, for example using a vector criterion combined with independent GPR models for each component, would enable us to accurately determine the impact of the criterion and model on the proposed vector strategy. An alternative approach to solving the problems associated with MOGP model autokrigeability would be to use heterotopic data, where model evaluation is not systematically performed for all output components. However, the use of this method depends on the specific application context.

Another limitation encountered is the high computation time associated with Vector strategy. This approach requires a considerably longer computation time for updating the surrogate model (MOGP) and optimizing the enrichment criterion, making it less suitable for black box functions with short evaluation times. To overcome this limitation, it would be beneficial to develop more efficient algorithms for updating the MOGP surrogate model and optimizing the enrichment criterion. This would reduce computation times and improve the performance of Vector strategy, in the case of black box functions with short evaluation times.

All three strategies have been evaluated in scenarios involving two output components, but they can theoretically be extended to a higher number of outputs. However, this extension requires additional effort to formulate approaches that are simple to implement, particularly for Vector strategy. One way forward would therefore be to develop formulations and algorithms that facilitate this extension, enabling the strategies to be applied to more complex problems with $p > 2$ output components.

In Chapter 5, we discussed the aforementioned limitations associated with the MOGP model used. From an application point of view, it is also noteworthy that we did not have access to real OMA data. instead, we used simulated vibration frequencies and deformation modes to illustrate the methodologies developed in Chapter 4. In practice, incorporating actual OMA data provided by wind experts would be crucial. Additionally, we restricted ourselves to only two simulator input parameters, whereas practical applications often involve up to six parameters for simulator calibration, thereby increasing the complexity of estimating partial excursion sets.

Moreover, applying these methodologies to other industrial sectors such as healthcare, automotive, or aerospace could introduce new perspectives and challenges, thereby enhancing the impact of this research.

Bibliography

[Abramowitz and Stegun, 1965] Abramowitz, M. and Stegun, I. A. (1965). Handbook of mathematical functions with formulas, graphs, and mathematical tables. *National Bureau of Standards Applied Mathematics Series. e*, 55:953.

[Abtini, 2018] Abtini, M. (2018). *Plans prédictifs à taille fixe et séquentiels pour le krigeage*. PhD thesis, Ecole Centrale Lyon.

[Allen, 1974] Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *technometrics*, 16(1):125–127.

[Alvarez and Lawrence, 2011] Alvarez, M. A. and Lawrence, N. D. (2011). Computationally efficient convolved multiple output gaussian processes. *The Journal of Machine Learning Research*, 12:1459–1500.

[Alvarez et al., 2012] Alvarez, M. A., Rosasco, L., Lawrence, N. D., et al. (2012). Kernels for vector-valued functions: A review. *Foundations and Trends® in Machine Learning*, 4(3):195–266.

[Anderson and Wendt, 1995] Anderson, J. D. and Wendt, J. (1995). *Computational fluid dynamics*, volume 206. Springer.

[Andrianakis et al., 2015] Andrianakis, I., Vernon, I. R., McCreesh, N., McKinley, T. J., Oakley, J. E., Nsubuga, R. N., Goldstein, M., and White, R. G. (2015). Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on hiv in uganda. *PLoS computational biology*, 11(1):e1003968.

[Arnaud and Emery, 2000] Arnaud, M. and Emery, X. (2000). *Estimation et interpolation spatiale: méthodes déterministes et méthodes géostatistiques*. Hermes.

[Azzimonti and Ginsbourger, 2018] Azzimonti, D. and Ginsbourger, D. (2018). Estimating orthant probabilities of high-dimensional gaussian vectors with an application to set estimation. *Journal of Computational and Graphical Statistics*, 27(2):255–267.

[Azzimonti, 2016] Azzimonti, D. F. (2016). *Contributions to Bayesian set estimation relying on random field priors*. PhD thesis, Philosophisch-naturwissenschaftliche Fakultät der Universität Bern.

[Baillargeon, 2005] Baillargeon, S. (2005). *Le krigeage: revue de la théorie et application à l'interpolation spatiale de données de précipitations*. PhD thesis, Université Laval Laval.

[Bak et al., 2013] Bak, C., Zahle, F., Bitsche, R., Kim, T., Yde, A., Henriksen, L., Andersen, P. B., Natarajan, A., and Hansen, M. H. (2013). Design and performance of a 10 mw wind turbine. *Wind Energy*, 124.

[Bect et al., 2019] Bect, J., Bachoc, F., and Ginsbourger, D. (2019). A supermartingale approach to gaussian process based sequential design of experiments. *Bernoulli*, 25(4A):2883–2919.

[Bect et al., 2012] Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793.

[Bichon et al., 2008] Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA journal*, 46(10):2459–2468.

[Blasques, 2012] Blasques, J. P. (2012). User's manual for becas. a cross section analysis tool for anisotropic and inhomogeneous beam sections of arbitrary geometry.

[Blevins and Plunkett, 1980] Blevins, R. D. and Plunkett, R. (1980). Formulas for natural frequency and mode shape. *Journal of Applied Mechanics*, 47(2):461.

[Bonilla et al., 2007] Bonilla, E. V., Chai, K., and Williams, C. (2007). Multi-task gaussian process prediction. *Advances in neural information processing systems*, 20.

[Bower et al., 2010] Bower, R. G., Goldstein, M., and Vernon, I. (2010). Galaxy formation: a bayesian uncertainty analysis.

[Breiman, 1992] Breiman, L. (1992). *Probability*. SIAM.

[Brincker and Ventura, 2015] Brincker, R. and Ventura, C. (2015). *Introduction to operational modal analysis*. John Wiley & Sons.

[Bryan et al., 2005] Bryan, B., Nichol, R. C., Genovese, C. R., Schneider, J., Miller, C. J., and Wasserman, L. (2005). Active learning for identifying function threshold boundaries. *Advances in neural information processing systems*, 18.

[Cadoret, 2023] Cadoret, A. (2023). *Analyse modale opérationnelle pour le suivi de santé structurelle des éoliennes*. PhD thesis, Université de Rennes.

[Chevalier, 2013] Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, Universität Bern.

[Chevalier et al., 2014a] Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014a). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.

[Chevalier et al., 2014b] Chevalier, C., Picheny, V., and Ginsbourger, D. (2014b). Kriginv: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging. *Computational statistics & data analysis*, 71:1021–1034.

[Cressie, 1990] Cressie, N. (1990). The origins of kriging. *Mathematical geology*, 22:239–252.

[Dagnelie, 1992] Dagnelie, P. (1992). *Statistique théorique et appliquée: Les bases théoriques*. Les Presses Agronomiques de Gembloux.

[Damblin et al., 2013] Damblin, G., Couplet, M., and Iooss, B. (2013). Numerical studies of space-filling designs: optimization of latin hypercube samples and subprojection properties. *Journal of Simulation*, 7(4):276–289.

[Deb et al., 2003] Deb, K., Mohan, M., and Mishra, S. (2003). A fast multi-objective evolutionary algorithm for finding well-spread pareto-optimal solutions. *KanGAL report*, 2003002:1–18.

[Deville et al., 2015] Deville, Y., Ginsbourger, D., Contributors, O. R., Durrande, N., and Roustant, M. O. (2015). Package 'kergp'.

[Döhler and Mevel, 2013] Döhler, M. and Mevel, L. (2013). Efficient multi-order uncertainty computation for stochastic subspace identification. *Mechanical Systems and Signal Processing*, 38(2):346–366.

[Duhamel et al., 2023] Duhamel, C., Helbert, C., Munoz Zuniga, M., Prieur, C., and Sinoquet, D. (2023). A sur version of the bichon criterion for excursion set estimation. *Statistics and Computing*, 33(2):41.

[Dupuy et al., 2015] Dupuy, D., Helbert, C., and Franco, J. (2015). Dicedesign and diceeval: Two r packages for design and analysis of computer experiments. *Journal of Statistical Software*, 65(11).

[Dutang and Savicky, 2013] Dutang, C. and Savicky, P. (2013). randtoolbox: Generating and testing random numbers. *R package*, 67.

[Echard et al., 2011] Echard, B., Gayton, N., and Lemaire, M. (2011). AK-MCS: an active learning reliability method combining kriging and Monte Carlo simulation. *Structural Safety*, 33(2):145–154.

[El Amri, 2019] El Amri, M. R. (2019). *Analyse d'incertitudes et de robustesse pour les modèles à entrées et sorties fonctionnelles.* Theses, Université Grenoble Alpes.

[El Amri et al., 2020] El Amri, M. R., Helbert, C., Lepreux, O., Zuniga, M. M., Prieur, C., and Sinoquet, D. (2020). Data-driven stochastic inversion via functional quantization. *Statistics and Computing*, 30:525–541.

[Floater and Hormann, 2007] Floater, M. S. and Hormann, K. (2007). Barycentric rational interpolation with no poles and high rates of approximation. *Numerische Mathematik*, 107:315–331.

[Fossum et al., 2021] Fossum, T. O., Travelletti, C., Eidsvik, J., Ginsbourger, D., and Rajan, K. (2021). Learning excursion sets of vector-valued gaussian random fields for autonomous ocean sampling. *The annals of applied statistics*, 15(2):597–618.

[Fricker et al., 2013] Fricker, T. E., Oakley, J. E., and Urban, N. M. (2013). Multivariate gaussian process emulators with nonseparable covariance structures. *Technometrics*, 55(1):47–56.

[Froude, 1889] Froude, R. E. (1889). On the part played in propulsion by differences of fluid pressure. *Trans. Inst. Naval Architects*, 30:390.

[Fu and He, 2001] Fu, Z.-F. and He, J. (2001). *Modal analysis.* Elsevier.

[Fuglsang and Thomsen, 2001] Fuglsang, P. and Thomsen, K. (2001). Site-specific design optimization of 1.5–2.0 mw wind turbines. *J. Sol. Energy Eng.*, 123(4):296–303.

[Gauchy, 2022] Gauchy, C. (2022). *Uncertainty quantification methodology for seismic fragility curves of mechanical structures: Application to a piping system of a nuclear power plant.* PhD thesis, Institut Polytechnique de Paris.

[Gaure, 2013] Gaure, S. (2013). Details of chebpol.

[Genz, 2004] Genz, A. (2004). Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and computing*, 14:251–260.

[Ginsbourger, 2017] Ginsbourger, D. (2017). Sequential design of computer experiments. *Wiley StatsRef*, 99:1–11.

[Ginsbourger et al., 2009] Ginsbourger, D., Dupuy, D., Badea, A., Carraro, L., and Roustant, O. (2009). A note on the choice and the estimation of kriging models for the analysis of deterministic computer experiments. *Applied Stochastic Models in Business and Industry*, 25(2):115–131.

[Goovaerts, 1997] Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press, USA.

[Gorbach et al., 2017] Gorbach, N. S., Bian, A. A., Fischer, B., Bauer, S., and Buhmann, J. M. (2017). Model selection for gaussian process regression. In *Pattern Recognition: 39th German Conference, GCPR 2017, Basel, Switzerland, September 12–15, 2017, Proceedings 39*, pages 306–318. Springer.

[Goulard and Voltz, 1992] Goulard, M. and Voltz, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24:269–286.

[Halton, 1960] Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90.

[Heinonen et al., 2016] Heinonen, M., Mannerström, H., Rousu, J., Kaski, S., and Lähdesmäki, H. (2016). Non-stationary gaussian process regression with hamiltonian monte carlo. In *Artificial Intelligence and Statistics*, pages 732–740. PMLR.

[Helbert et al., 2009] Helbert, C., Dupuy, D., and Carraro, L. (2009). Assessment of uncertainty in computer experiments from universal to bayesian kriging. *Applied Stochastic Models in Business and Industry*, 25(2):99–113.

[Helterbrand and Cressie, 1994] Helterbrand, J. D. and Cressie, N. (1994). Universal cokriging under intrinsic coregionalization. *Mathematical Geology*, 26:205–226.

[Horn and Johnson, 2012] Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

[Jin et al., 2002] Jin, R., Chen, W., and Sudjianto, A. (2002). On sequential sampling for global metamodeling in engineering design. In *International design engineering technical conferences and computers and information in engineering conference*, volume 36223, pages 539–548.

[Jobson, 1991] Jobson, J. (1991). Multiple linear regression. *Applied multivariate data analysis: Regression and experimental design*, pages 219–398.

[Johnson et al., 1990] Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148.

[Jones, 2001] Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, 21(4):345–383.

[Jones et al., 1998] Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492.

[Jones, 2018] Jones, R. M. (2018). *Mechanics of composite materials*. CRC press.

[Jonkman et al., 2009] Jonkman, J., Butterfield, S., Musial, W., and Scott, G. (2009). Definition of a 5-mw reference wind turbine for offshore system development. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

[Kennedy and O'Hagan, 2001] Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

[Kneubühl, 2013] Kneubühl, F. K. (2013). *Oscillations and waves*. Springer Science & Business Media.

[Krige, 1951] Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139.

[Kushner, 1964] Kushner, H. J. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise.

[Larsen and Hansen, 2009] Larsen, T. J. and Hansen, A. M. (2009). How 2 hawc2, the user's manual. *RisøReport, Risø*, 702.

[Lemieux, 2009] Lemieux, C. (2009). Quasi–monte carlo constructions. *Monte Carlo and Quasi-Monte Carlo Sampling*, pages 1–61.

[Lindgren, 2012] Lindgren, G. (2012). *Stationary stochastic processes: theory and applications*. CRC Press.

[Liu et al., 2018] Liu, H., Cai, J., and Ong, Y.-S. (2018). Remarks on multi-output gaussian process regression. *Knowledge-Based Systems*, 144:102–121.

[Manual, 2012] Manual, A. S. U. (2012). Abaqus 6.11. *http://130.149*, 89(2080):v6.

[Marler and Arora, 2004] Marler, R. T. and Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 26:369–395.

[Matheron, 1963] Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

[McKay et al., 1979] McKay, M., Beckman, R., and Conover, W. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

[Mebane Jr and Sekhon, 2011] Mebane Jr, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: the rgenoud package for r. *Journal of Statistical Software*, 42(1):1–26.

[Miettinen and Mäkelä, 2006] Miettinen, K. and Mäkelä, M. M. (2006). Synchronous approach in interactive multiobjective optimization. *European Journal of Operational Research*, 170(3):909–922.

[Molchanov, 2005] Molchanov, I. S. (2005). *Theory of random sets*, volume 87. Springer.

[Moustapha et al., 2021] Moustapha, M., Marelli, S., and Sudret, B. (2021). A generalized framework for active learning reliability: survey and benchmark. *arXiv preprint arXiv:2106.01713*.

[Nayfeh and Mook, 2008] Nayfeh, A. H. and Mook, D. T. (2008). *Nonlinear oscillations*. John Wiley & Sons.

[O'Hagan, 1978] O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(1):1–24.

[Osborne, 2007] Osborne, M. (2007). Gaussian processes for prediction.

[Paciorek and Schervish, 2003] Paciorek, C. and Schervish, M. (2003). Nonstationary covariance functions for gaussian process regression. *Advances in neural information processing systems*, 16.

[Paciorek, 2003] Paciorek, C. J. (2003). *Nonstationary Gaussian processes for regression and spatial modelling*. PhD thesis, Carnegie Mellon University.

[Peeters and De Roeck, 1999] Peeters, B. and De Roeck, G. (1999). Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical systems and signal processing*, 13(6):855–878.

[Pelamatti et al., 2024] Pelamatti, J., Le Riche, R., Helbert, C., and Blanchet-Scalliet, C. (2024). Coupling and selecting constraints in bayesian optimization under uncertainties. *Optimization and Engineering*, 25(1):373–412.

[Picheny et al., 2010] Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N.-H. (2010). Adaptive designs of experiments for accurate approximation of a target region.

[Picheny et al., 2013] Picheny, V., Wagner, T., and Ginsbourger, D. (2013). A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626.

[Pronzato and Müller, 2012] Pronzato, L. and Müller, W. G. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701.

[Ranjan et al., 2008] Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541.

[Rankine, 1865] Rankine, W. J. M. (1865). On the mechanical principles of the action of propellers. *Transactions of the Institution of Naval Architects*, 6.

[Rasmussen et al., 2006] Rasmussen, C. E., Williams, C. K., et al. (2006). *Gaussian processes for machine learning*, volume 1. Springer.

[Release, 2020] Release, N. O. D. (2020). v2. 3.0. *National Renewable Energy Laboratory: Golden, CO, USA*.

[Reynders, 2012] Reynders, E. (2012). System identification methods for (operational) modal analysis: review and comparison. *Archives of Computational Methods in Engineering*, 19:51–124.

[Reynders et al., 2008] Reynders, E., Pintelon, R., and De Roeck, G. (2008). Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mechanical systems and signal processing*, 22(4):948–969.

[Rossi, 2018] Rossi, R. J. (2018). *Mathematical statistics: an introduction to likelihood based inference*. John Wiley & Sons.

[Roustant et al., 2012] Roustant, O., Ginsbourger, D., and Deville, Y. (2012). Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization.

[Russell, 2018] Russell, D. (2018). Vibrational modes of a circular membrane.

[Sacks and Schiller, 1988] Sacks, J. and Schiller, S. (1988). Spatial designs. *Statistical decision theory and related topics IV*, pages 385–399.

[Sacks et al., 1989] Sacks, J., Schiller, S. B., and Welch, W. J. (1989). Designs for computer experiments. *Technometrics*, pages 41–47.

[Sobol', 1967] Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802.

[Srinivas and Deb, 1994] Srinivas, N. and Deb, K. (1994). Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3):221–248.

[Stein, 1999] Stein, M. L. (1999). Interpolation of spatial data. *Springer Series in Statistics*.

[Stigler, 2007] Stigler, S. M. (2007). The epic story of maximum likelihood. *Statistical Science*, pages 598–620.

[Stone, 1974] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2):111–133.

[Sun and Genton, 2011] Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of computational and graphical statistics*, 20(2):316–334.

[Teh et al., 2005] Teh, Y. W., Seeger, M., and Jordan, M. I. (2005). Semiparametric latent factor models. In *International Workshop on Artificial Intelligence and Statistics*, pages 333–340. PMLR.

[Tisseur and Meerbergen, 2001] Tisseur, F. and Meerbergen, K. (2001). The quadratic eigenvalue problem. *SIAM review*, 43(2):235–286.

[Tokdar and Kass, 2010] Tokdar, S. T. and Kass, R. E. (2010). Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):54–60.

[Trautmann et al., 2013] Trautmann, H., Steuer, D., Mersmann, O., and Mersmann, M. O. (2013). Package 'mco'.

[Van Overschee and De Moor, 2012] Van Overschee, P. and De Moor, B. (2012). *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science & Business Media.

[Vazquez and Bect, 2009] Vazquez, E. and Bect, J. (2009). A sequential bayesian algorithm to estimate a probability of failure. *IFAC Proceedings Volumes*, 42(10):546–550.

[Ver Hoef and Barry, 1998] Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69(2):275–294.

[Vincent and Grantham, 1981] Vincent, T. and Grantham, W. (1981). *Optimality in Parametric Systems*. John Wiley and Sons, New York.

[Wackernagel, 2003] Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.

[Wagner, 2008] Wagner, R. (2008). Multi-linear interpolation. *Beach Cities Robotics*.

[Wahba, 1990] Wahba, G. (1990). *Spline models for observational data*. SIAM.