Calage bayésien sous incertitudes des outils de calcul scientifique couplés : application en simulation numérique du combustible

Oumar Balde

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Université de Toulouse

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Calage bayésien sous incertitudes des outils de calcul scientifique couplés : application en simulation numérique du combustible

---

Thèse présentée et soutenue, le 26 avril 2024 par

Oumar BALDE

École doctorale
EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse

Spécialité
Mathématiques et Applications

Unité de recherche
Institut de Mathématiques de Toulouse (IMT), UMR 5219

Thèse dirigée par
Amandine MARREL

Composition du jury
M. Emmanuel VAZQUEZ, Président, Université Paris Saclay
Mme Clémentine PRIEUR, Rapporteure, Université Grenoble Alpes
M. Guillaume PERRIN, Rapporteur, Université Gustave Eiffel
Mme Claire CANNAMELA, Examinatrice, CEA DAM Île-de-France
M. Fabrice GAMBOA, Examinateur, Université Toulouse III - Paul Sabatier
Mme Amandine MARREL, Directrice de thèse, CEA CADARACHE

Membres invités
M. Guillaume DAMBLIN, Commissariat à l'énergie atomiques et aux énergies alternatives (CEA)
M. Antoine BOULORÉ, CEA Cadarache

À la mémoire d'Oussou Baldé (Pouthiel), mon frère,
À la mémoire de Lomby Baldé, ma grande-mère,
À mes parents.

Table des matières

Liste des figures

Listes des tableaux

Remerciements

Par ces quelques mots, je tiens à exprimer ma profonde reconnaissance envers toutes les personnes qui ont contribué, directement ou indirectement, à la réussite de mon parcours doctoral. En premier lieu, je souhaite adresser mes plus sincères remerciements à ma directrice de thèse, Amandine MARREL, ainsi qu'à mes deux encadrants, Guillaume DAMBLIN et Antoine BOULORÉ. Votre soutien indéfectible et votre disponibilité sans faille ont été d'une aide précieuse pendant ces trois années. Je suis extrêmement reconnaissant pour le temps et les efforts que vous m'avez accordés, ainsi que pour vos conseils avisés. Votre accompagnement permanent m'a permis d'acquérir des connaissances tant sur le plan scientifique qu'humain. Enfin, je vous remercie d'avoir créé un environnement propice à mon épanouissement et à ma progression sereine dans ma quête du savoir.

Je tiens également à exprimer toute ma gratitude à l'égard mes deux rapporteurs de thèse, Clémentine PRIEUR et Guillaume PERRIN, pour avoir accepté d'évaluer mon travail. Vos rapports détaillés ainsi que vos remarques et suggestions pertinentes ont été d'une grande valeur. C'est un véritable honneur pour moi d'avoir eu la chance de bénéficier de votre expertise en tant que rapporteurs de ma thèse. J'adresse aussi mes remerciements aux autres membres du jury : Claire CANNAMELA, Emmanuel VAZQUEZ et Fabrice GAMBOA, d'avoir accepté d'examiner mon travail.

Dans le même élan, un grand merci à toi Gabriel SARAZIN d'avoir accepté de travailler avec nous. J'ai beaucoup appris de ta rigueur scientifique et rédactionnelle, et c'est agréable de collaborer avec toi. Un grand merci à toi aussi Loïc GIRALDI pour ton aide précieuse durant les deux premières années de ma thèse.

Toute ma reconnaissance à l'ensemble du personnel du SGLS que j'ai eu le privilège de rencontrer tout au long de mon parcours doctoral. Je tiens particulièrement à remercier chaleureusement mes collègues de bureau : Anthony, Souroure, Léopold et Antonin. Votre présence et nos échanges ont constitué des moments précieux pour moi. Nos conversations ont été une source de réconfort, particulièrement vers la fin de ma thèse. Merci pour ces instants de détente et de partage qui ont contribué à rendre cette période plus supportable. À tous, je vous souhaite le meilleur pour l'avenir et beaucoup de succès dans vos projets à venir.

Dans le même esprit, je tiens également à remercier tous les doctorants du bâtiment 451 (SGLS et LMSF). Je pense notamment à Sébastien, Mathieu, Anthony, Antonin, Sanae, Capucine, Nathalie (merci pour ta bonne humeur constante), Loïc, Mathis, Jiayi, Gabriel, Andrew, Elie, William et Etienne. Je vous souhaite à tous plein de succès dans votre quête du précieux sésame qu'est le titre de docteur. J'en profite également pour remercier chaleureusement ceux qui l'ont déjà obtenu : Clément et Antonin.

Enfin, je souhaite exprimer ma profonde gratitude envers mes proches. À mes chers parents, je vous remercie du fond du cœur pour votre amour indéfectible et votre soutien constant. Je n'aurai jamais réussi sans vous. À mon cousin Cheickou Sylla et à ma famille de Nangis, mes mots ne sauraient véritablement exprimer toute l'étendue de ma reconnaissance pour tout ce que vous avez fait pour moi. Merci infiniment !

Chapter 1

Introduction (français)

Sommaire

1.1 Calage de paramètres incertains en simulation numérique

1.1.1 Contexte général de la simulation combustible pour les réacteurs nucléaires

Au cours des dernières décennies, les modèles numériques (appelés également codes de calcul ou simulateurs) sont devenus des outils incontournables pour l'analyse, la compréhension, la modélisation et la prévision de certains systèmes physiques complexes tels que les centrales nucléaires. Ces codes de calcul permettent d'étudier ces systèmes physiques lorsque les expérimentations réelles sont limitées, voire impossibles pour des raisons économiques et/ou éthiques. **Les modèles numériques peuvent s'avérer coûteux en temps de calcul CPU (quelques heures à plusieurs jours pour une simulation) et prennent généralement un grand nombre de paramètres en entrée**. Ces paramètres dont dépendent les modélisations physique et numérique sont souvent incertains. Cela découle essentiellement d'une méconnaissance du phénomène étudié, et d'hypothèses simplificatrices introduisant des incertitudes de paramètres ou encore des erreurs de mesure. Par conséquent, **les sorties du code de calcul sont également entachées d'incertitudes**. Il devient alors nécessaire de quantifier précisément ces dernières avant toute utilisation dans un contexte industriel tel que l'évaluation d'une occurrence de défaillance, le calcul de marges de sécurité, la prise de décision, etc. **La quantification des incertitudes paramétriques est un volet très important pour la plupart des domaines industriels tel que le domaine du combustible nucléaire où les paramètres des modèles sont totalement ou partiellement inconnus** [Bouloré, 2019; Delipei, 2019]. Dans ce domaine, les codes de calcul sont utilisés pour étudier le comportement des éléments combustibles en réacteur en situation de fonctionnement normal, incidentel et accidentel. En conditions accidentelles par exemple, ces codes servent à évaluer la dégradation subie par les éléments combustibles et leur potentielle dispersion dans les différents composants du réacteur. Ces codes sont également utilisés pour optimiser la conception des combustibles nucléaires vis-à-vis de certains critères de fonctionnement.

La recherche et le développement sur les combustibles, menée au CEA en collaboration avec EDF et Framatome, a conduit à l'élaboration de la plateforme de simulation PLEIADES, illustrée par la Figure 1.1. Chaque filière combustible (réacteur à eau pressurisée, rapide refroidi au sodium, rapide refroidi au gaz, etc.) fait l'objet d'une application dédiée sur la plateforme PLEIADES. Son objectif principal est d'une part de prédire le comportement des combustibles nucléaires en réacteur en fonction de leurs caractéristiques initiales et de leur évolution sous irradiation, afin d'assurer le maintien de la première barrière de sécurité (la gaine) selon le mode d'utilisation. D'autre part, elle permet d'effectuer des expériences numériques en explorant des domaines de fonctionnement difficiles à reproduire dans la réalité ou de simuler des situations accidentelles de type LOCA (*Loss of Coolant Accident* pour Accident de Perte de Réfrigérant Primaire) et RIA (*Reactivity Insertion Accident* pour Accident d'Insertion de Réactivité). Ces expérimentations numériques guident les chercheurs vers une conception optimisée et plus innovante de la forme du combustible. Les codes de calcul de la filière des Réacteurs à Eau Pressurisée (REP) de cette plateforme sont **ALCYONE** [Delipei, 2019; Michel et al., 2021] et CYRANO3 respectivement développés par le CEA et EDF. **Le simulateur ALCYONE simule ainsi le comportement des combustibles dans un réacteur de type REP. Il est constitué d'un ensemble de modèles physiques qui permettent de calculer le comportement mécanique (déformations subies par le combustible) et thermique (évolution de la température au sein du coeur) ainsi que le comportement des produits de fission**. Une boucle de convergence globale assure la convergence de l'ensemble du chaînage des modèles à chaque pas de temps de calcul.

Comme nous l'avons évoqué précédemment, les modèles numériques, mis en oeuvre pour représenter fidèlement la réalité physique, comportent souvent un grand nombre de paramètres incertains [Campbell, 2006] (données d'entrée, paramètres de modélisation physiques ou conceptuels, etc.). Dans le cas de la simulation du comportement du combustible nucléaire, les sources d'incertitude sont nombreuses et de nature variée [Bouloré, 2019]. Par conséquent,

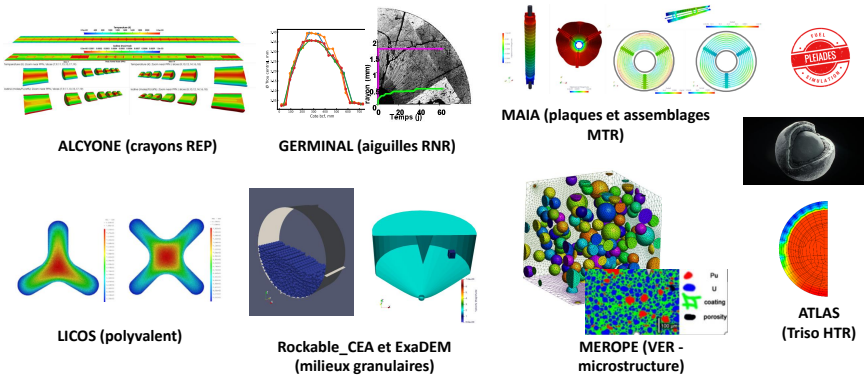


Figure 1.1: Plateforme PLEIADES [Marelle et al., 2016; Michel et al., 2021].

la quantification la plus précise possible de toutes les sources d'incertitude et d'erreurs susceptibles d'influencer la capacité prédictive du modèle numérique est une étape importante pour les études en simulation du combustible. La démarche générale de traitement de toutes les incertitudes en simulation numérique a fait l'objet de nombreux travaux depuis bientôt une vingtaine d'années [De Rocquigny et al., 2008; Santner et al., 2018; Ghanem et al., 2017; Baccou et al., 2020]. Dans cette thèse, notre attention se concentre sur deux étapes de cette démarche : **le calage de modèle** qui constitue l'axe principal des travaux et **l'analyse de sensibilité en support à ce calage**. L'analyse de sensibilité va servir à sélectionner les paramètres influents à prendre en compte dans le processus de calage de modèle. Nous détaillons les objectifs et principes de ces deux étapes dans ce qui suit.

Le calage de modèle est un processus visant à quantifier les incertitudes paramétriques d'un modèle numérique en se basant sur les données expérimentales disponibles du système physique étudié [Campbell, 2006]. Ce processus d'ajustement a pour but de quantifier et de réduire au mieux l'incertitude de prédiction du modèle afin d'améliorer l'adéquation des sorties à la réalité avec la précision la plus fiable possible. Il existe deux types de calage de modèle : le calage déterministe et **le calage statistique, en particulier bayésien**. Dans le calage déterministe, l'objectif est de trouver le vecteur de paramètres optimal qui minimise l'écart entre les données expérimentales et les simulations du modèle numérique [Tuo and Wu, 2015; Wong et al., 2017]. **Le calage bayésien repose sur une approche probabiliste où l'incertitude paramétrique est quantifiée par des distributions de probabilité** [Kennedy and O'Hagan, 2001; Campbell, 2006; Gu and Wang, 2018]. Cette approche est particulièrement intéressante lorsque nous avons à disposition des données expérimentales en nombre limité ou entachées d'incertitude car elle permet d'utiliser toutes les informations à notre disposition (connaissances *a priori*, données expérimentales et données simulées) pour produire une estimation robuste et informative des incertitudes résiduelles sur les paramètres. Pour ce faire, le calage bayésien se base sur trois étapes. Premièrement, la distribution dite distribution *a priori* quantifie les connaissances sur les paramètres incertains avant prise en compte des données expérimentales. Ces informations *a priori* sont souvent obtenues par avis d'experts et permettent d'établir une telle distribution soit par le principe du maximum d'entropie, soit par le choix d'une distribution paramétrique qui s'ajuste au mieux aux données issues de ces informations *a priori* [Garthwaite et al., 2005; Ghosh et al., 2006; Oakley and O'Hagan, 2007; O'Hagan, 2019]. En l'absence d'informations *a priori*, le statisticien s'attachera à construire une distribution aussi non informative que possible [Kass and Wasserman, 1996; Gelman et al., 2014; Reich and Ghosh, 2019; Murphy, 2022]. Deuxièmement, la disposition d'une équation probabiliste reliant le vecteur des données expérimentales disponibles aux sorties du code permet d'établir la fonction de vraisemblance des paramètres. La vraisemblance mesure

le degré de plausibilité d'un ensemble de valeurs prises par les paramètres incertains au vue des données expérimentales. Enfin, à l'aide de **la formule de Bayes**, une distribution dite distribution *a posteriori* des paramètres incertains est calculée. La distribution *a posteriori* quantifie l'incertitude résiduelle après l'utilisation des informations apportées par les données expérimentales. Cette distribution s'interprète comme une mise à jour de la distribution *a priori* et à partir de laquelle une valeur optimale des paramètres telle que le maximum *a posteriori* (MAP) peut être déduite.

L'analyse de sensibilité, autre étape fondamentale de la démarche de prise en compte des incertitudes, vise à quantifier comment la variabilité des entrées d'un modèle numérique influence la variabilité de la sortie ou de la quantité d'intérêt. Les grands objectifs de l'analyse de sensibilité sont de classer les variables d'entrée par ordre d'influence [Hamby, 1993, 1994, 1995] ou d'effectuer un criblage des entrées [Morris, 1991; Morris and Mitchell, 1995; Campolongo et al., 2007, 2011] (c'est-à-dire éliminer les entrées non influentes). Pour répondre à ces objectifs, il existe deux types d'analyse de sensibilité : l'analyse de sensibilité locale (LSA pour *Local Sensitivity Analysis*) et l'analyse de sensibilité globale (GSA pour *Global Sensitivity Analysis*). La LSA vise à quantifier la variabilité de la sortie lorsque de faibles perturbations sont appliquées aux entrées autour d'une valeur de référence (également appelée valeur nominale) [Morris, 1991; Morris and Mitchell, 1995]. La GSA cherche à prendre en compte la distribution de probabilité des variables d'entrée sur l'ensemble de leur domaine de variation [Saltelli et al., 2004, 2008; Da Veiga et al., 2021]. Dans cette thèse, l'analyse de sensibilité est utilisée en support au calage bayésien afin d'effectuer un criblage préalable. Plus précisément, on considérera la GSA afin de prendre en compte la distribution *a priori* des paramètres.

1.1.2 Calage bayésien de paramètres dans les modèles numériques

Un système physique d'intérêt, représenté par $r(\boldsymbol{x}) \in \mathscr{D}_{\boldsymbol{Y}} \subset \mathbb{R}^q$ $(q \geq 1)$, peut être considéré comme une fonction mathématique caractérisant la relation possible entre le vecteur d'entrée $\boldsymbol{x}$ et la sortie $r(\boldsymbol{x})$. Le vecteur d'entrée, $\boldsymbol{x} \in \mathscr{D}_{\boldsymbol{X}} \subset \mathbb{R}^d$, contient principalement des variables appelées variables de contrôle [Santner et al., 2018]. Ces variables de contrôle représentent généralement les conditions expérimentales et les propriétés géométriques du système simulé. Pour des configurations $\{\boldsymbol{x}_i\}_{i=1}^{n_{\text{exp}}}$, les réalisations expérimentales $\boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_{n_{\text{exp}}})^t$ du système physique $r(\boldsymbol{x})$ sont en général reliées à celui-ci par l'équation probabiliste suivante :

$$\boldsymbol{z}_i = r(\boldsymbol{x}_i) + \epsilon_i, \ 1 \leq i \leq n_{\text{exp}}, \tag{1.1}$$

où ϵ_i est la réalisation de l'incertitude expérimentale qui est essentiellement due à des imprécisions de mesure. Cette incertitude expérimentale est souvent considérée comme une réalisation d'une distribution normale multivariée de moyenne nulle et de matrice de covariance $\Sigma_{\epsilon_i} = \text{diag}(\sigma^2_{\epsilon_{i,1}}, \cdots, \sigma^2_{\epsilon_{i,q}})$ connue.

Un modèle numérique déterministe visant à représenter fidèlement la réalité physique $r(\boldsymbol{x})$ peut être envisagé comme une fonction mathématique paramétrée par des paramètres de calage $\boldsymbol{\theta} \in \mathscr{D}_{\boldsymbol{\Theta}} \subset \mathbb{R}^p$ $(p \geq 1)$ incertains le plus souvent, qui décrit aussi la relation entre les variables de contrôle $\boldsymbol{x}$ et la sortie notée $y_{\boldsymbol{\theta}}(\boldsymbol{x})$. Si $\boldsymbol{\theta}$ et/ou $\boldsymbol{x}$ sont entachés d'incertitude, alors $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ le sera lui aussi. Supposons dans un premier temps qu'il existe une valeur optimale des paramètres $\boldsymbol{\theta}$ notée $\boldsymbol{\theta}_{\text{opt}}$ telle que le modèle numérique $y_{\boldsymbol{\theta}_{\text{opt}}}(\boldsymbol{x})$ reproduise parfaitement le système physique étudié $r(\boldsymbol{x})$. Ceci implique que, les données expérimentales $\{\boldsymbol{z}_i\}_{i=1}^{n_{\text{exp}}}$ sont des réalisations du modèle numérique paramétré par $\boldsymbol{\theta}_{\text{opt}}$:

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta}_{\text{opt}}}(\boldsymbol{x}_i) + \epsilon_i, \ 1 \leq i \leq n_{\text{exp}}. \tag{1.2}$$

L'équation (1.2) stipule que le modèle numérique reproduit parfaitement le phénomène physique étudié pour un ensemble de paramètres optimaux $\boldsymbol{\theta}_{\text{opt}}$. Cependant, en pratique, les modèles parfaits sont rares, et sous $\boldsymbol{\theta}_{\text{opt}}$, le modèle $y_{\boldsymbol{\theta}_{\text{opt}}}(\boldsymbol{x})$ ne peut fournir qu'une approximation de

la réalité physique. Si l'écart entre cette approximation et la réalité est significatif et donc qu'il n'est pas raisonnable de considérer cet écart comme indissociable de l'incertitude expérimentale, un terme supplémentaire non observable dépendant souvent uniquement de $\boldsymbol{x}$, noté $b(\boldsymbol{x})$ [Kennedy and O'Hagan, 2001], est introduit pour le représenter. Il est défini par :

$$b(\boldsymbol{x}) := r(\boldsymbol{x}) - y_{\boldsymbol{\theta}_{\text{opt}}}(\boldsymbol{x}). \tag{1.3}$$

Par conséquent, l'équation probabiliste générale reliant les données expérimentales aux simulations numériques est donnée par :

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta}_{\text{opt}}}(\boldsymbol{x}_i) + b(\boldsymbol{x}_i) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}. \tag{1.4}$$

Cependant, en pratique, nous ne connaissons jamais les paramètres optimaux $\boldsymbol{\theta}_{\text{opt}}$ à l'avance et l'objectif du calage de modèle est de fournir un estimateur $\widehat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}_{\text{opt}}$ en se basant sur l'équation suivante :

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta}}(\boldsymbol{x}_i) + b(\boldsymbol{x}_i) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}, \tag{1.5}$$

En ce qui concerne le calage déterministe, la quantité $\widehat{\boldsymbol{\theta}}$ est obtenue en minimisant autant que possible la norme quadratique de la discrépance de modèle (ou erreur de modèle) $b(\boldsymbol{x})$ à partir des données expérimentales et des simulations $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ [Tuo and Wu, 2015; Wong et al., 2017] :

$$\widehat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta} \in \mathscr{D}_{\boldsymbol{\Theta}}}{\operatorname{argmin}} \frac{1}{n_{\text{exp}}} \sum_{i=1}^{n_{\text{exp}}} \left(\boldsymbol{z}_i - y_{\boldsymbol{\theta}}(\boldsymbol{x}_i)\right)^t \Sigma_{\epsilon_i}^{-1} \left(\boldsymbol{z}_i - y_{\boldsymbol{\theta}}(\boldsymbol{x}_i)\right). \tag{1.6}$$

En général, nous disposons pour ce faire d'un nombre limité de données expérimentales et le code $y_{\boldsymbol{\theta}}(\cdot)$ est souvent coûteux en temps de calcul CPU ce qui empêche la résolution du problème d'optimisation (1.6) en un temps de raisonnable. Cependant, pour pallier le problème du coût d'évaluation de $y_{\boldsymbol{\theta}}(\boldsymbol{x})$, il est possible de le résoudre en remplaçant $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ par un métamodèle $\widehat{y}_{\boldsymbol{\theta}}(\boldsymbol{x})$ [1] [Wong et al., 2017; Lee and Park, 2020]. C'est ce qui est également fait dans la littérature dédiée au calage bayésien où le modèle numérique et la discrépance de modèle sont souvent remplacés par des métamodèles.

Par ailleurs, lorsque nous avons des données expérimentales limitées et un code de calcul coûteux en temps de calcul CPU, le calage bayésien est souvent adopté [Kennedy and O'Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007; Bachoc et al., 2014; Gu and Wang, 2018]. Il permet de quantifier les incertitudes associées à $\boldsymbol{\theta}$ par des distributions de probabilité. Ce calage sous incertitudes permet de quantifier et de réduire l'incertitude dans les prédictions [Campbell, 2006; Trucano et al., 2006]. Pour ce faire, le processus de calage bayésien, comme décrit précédemment, combine une distribution *a priori* $\pi(\boldsymbol{\theta})$ quantifiant l'incertitude initiale associée à $\boldsymbol{\theta}$ et la vraisemblance des données $\boldsymbol{z}$ notée $\mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta})$ à travers la formule de Bayes pour obtenir la distribution *a posteriori* $\pi_{\text{full}}(\boldsymbol{\theta} \mid \boldsymbol{z})$:

$$\pi_{\text{full}}(\boldsymbol{\theta} \mid \boldsymbol{z}) \propto \mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{1.7}$$

La distribution *a posteriori* est une correction de la distribution *a priori* à l'aide des informations apportées par $\boldsymbol{z}$ via la vraisemblance. La vraisemblance $\mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta})$ mesure la plausibilité d'un ensemble de valeurs prises par $\boldsymbol{\theta}$ compte tenu des données expérimentales observées $\boldsymbol{z}$ et elle se calcule à l'aide de l'équation (1.5). Les premiers travaux de calage de modèle sous incertitudes utilisant l'équation probabiliste (1.5) sont ceux de Kennedy and O'Hagan [2001].

[1] Le principe de la métamodélisation repose sur la substitution d'un modèle numérique $y_\theta(\boldsymbol{x})$ coûteux en temps de calcul CPU par une fonction mathématique $\widehat{y}_\theta(\boldsymbol{x})$ permettant de prédire la sortie $y_\theta(\boldsymbol{x})$ avec une grande précision (ou une précision souhaitée), et ce, de manière instantanée.

Méthode de Kennedy et O'Hagan et ses variantes

La méthode présentée par Kennedy and O'Hagan [2001] est le papier de référence du calage bayésien des modèles numériques coûteux en temps de calcul CPU avec l'utilisation de métamodèles de type processus gaussien [Currin et al., 1991; Rasmussen et al., 2006; Gramacy, 2020]. L'idée principale de cette méthode consiste d'abord à substituer le modèle $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ par un métamodèle processus gaussien dont les hyperparamètres ϕ_y [2] sont inconnus et de modéliser également la discrépance $b(\boldsymbol{x})$ par un autre processus gaussien indépendant et ayant aussi des hyperparamètres ϕ_b qui sont également inconnus. Ensuite, les auteurs proposent d'estimer les paramètres $\boldsymbol{\theta}$ et les hyperparamètres de ces deux processus gaussiens indépendants en deux temps. Cette estimation en deux temps sera plus tard appelée approche modulaire par Liu et al. [2009]. Enfin, un prédicteur $\widehat{r}(\boldsymbol{x})$ du système $r(\boldsymbol{x})$ est proposé tenant compte de la correction des réponses du modèle par la discrépance estimée. Nous donnons ci-dessous quelques détails de la méthode.

Le modèle numérique et la discrépance de modèle sont modélisés par deux processus gaussiens indépendants *a priori* :

$$y(\boldsymbol{x},\boldsymbol{\theta}) := y_{\boldsymbol{\theta}}(\boldsymbol{x}) \sim \mathscr{GP}\left(m_{\beta_y}(\boldsymbol{x},\boldsymbol{\theta}),\, \sigma_y^2 K_{\psi_y}((\boldsymbol{x},\boldsymbol{\theta}),(\boldsymbol{x}',\boldsymbol{\theta}')\right),\ \phi_y := (\beta_y,\sigma_y^2,\psi_y)^t, \tag{1.8}$$

$$b(\boldsymbol{x}) \sim \mathscr{GP}\left(m_{\beta_b}(\boldsymbol{x}),\, \sigma_b^2 K_{\psi_b}(\boldsymbol{x},\boldsymbol{x}')\right),\ \phi_b := (\beta_b,\sigma_b^2,\psi_b)^t. \tag{1.9}$$

L'apprentissage de ces deux métamodèles processus gaussiens s'appuie sur M simulations disponibles du modèle numérique

$$y\left(D_M\right) = \left(y_{\boldsymbol{\theta}^{(1)}}(\boldsymbol{x}^{(1)}),\cdots,y_{\boldsymbol{\theta}^{(M)}}(\boldsymbol{x}^{(M)})\right)^t$$

obtenues à partir d'un plan d'expériences numérique

$$D_M = \left\{\left(\boldsymbol{x}^{(1)},\boldsymbol{\theta}^{(1)}\right),\cdots,\left(\boldsymbol{x}^{(M)},\boldsymbol{\theta}^{(M)}\right)\right\},$$

où $\{\boldsymbol{x}^{(l)}\}_{1\leq l\leq M}$ sont des réalisations de $\boldsymbol{x}\in\mathscr{D}_{\boldsymbol{X}}$ et $\{\boldsymbol{\theta}^{(l)}\}_{1\leq l\leq M}$ sont des réalisations de $\boldsymbol{\theta}\in\mathscr{D}_{\boldsymbol{\Theta}}$. Suivant la formule de Bayes, l'estimation de $\boldsymbol{\theta}$ et l'estimation des hyperparamètres ϕ_y et ϕ_b s'obtiennent à partir de la distribution *a posteriori* jointe qui est donnée par :

$$\pi_{\text{full}}(\boldsymbol{\theta},\phi_y,\phi_b \mid \boldsymbol{d}) \propto \mathscr{L}(\boldsymbol{d} \mid \boldsymbol{\theta},\phi_y,\phi_b)\pi(\boldsymbol{\theta},\phi_y,\phi_b), \tag{1.10}$$

où $\boldsymbol{d} := (\boldsymbol{z}, y(D_M))^t$ se compose des données expérimentales et simulées à notre disposition. Cependant, comme l'expliquent Kennedy and O'Hagan [2001], la distribution $\pi(\boldsymbol{\theta},\phi_y,\phi_b \mid \boldsymbol{d})$ n'est pas explicite et son estimation nécessite l'utilisation de méthodes Monte Carlo par Chaîne de Markov (MCMC). Ces méthodes MCMC souffrant du fléau de la dimension, elles pourraient ne pas converger lorsque la dimension de $(\boldsymbol{\theta},\phi_y,\phi_b)$ est grande. De plus, une distribution *a priori* appropriée sur le couple (ϕ_y,ϕ_b) doit être choisie ce qui n'est pas toujours facile en pratique [Paulo, 2005]. C'est pourquoi, les auteurs proposent une procédure d'estimation en deux étapes de $(\boldsymbol{\theta},\phi_y,\phi_b)$ en estimant (ϕ_y,ϕ_b) par $(\widehat{\phi}_y,\widehat{\phi}_b)$ soit par maximum de vraisemblance, soit par validation croisée [Bachoc, 2013], puis en estimant la distribution *a posteriori* $\pi(\boldsymbol{\theta} \mid \boldsymbol{d},\widehat{\phi}_y,\widehat{\phi}_b)$ par un algorithme MCMC tel que celui de Metropolis-Hastings [Chib and Greenberg, 1995; Robert et al., 1999; Andrieu and Thoms, 2008]. Après ces étapes d'estimation, les auteurs proposent un prédicteur de la quantité physique $r(\boldsymbol{x})$. Il est obtenu

[2] Un hyperparamètre est un paramètre dont la valeur permet d'ajuster un processus d'apprentissage tel que les processus gaussiens. Il permet de régler le processus d'apprentissage et la performance de ce processus d'apprentissage dépend grandement de la valeur prise par son (ou ses) hyperparamètre(s) [Hastie et al., 2009; Hutter et al., 2014].

en intégrant la moyenne $\mathbb{E}[r(\boldsymbol{x}) \mid \boldsymbol{d}, \boldsymbol{\theta}, \widehat{\phi}_y, \widehat{\phi}_b]$ de la distribution prédictive par rapport à la mesure de probabilité *a posteriori* $\pi(\boldsymbol{\theta} \mid \boldsymbol{d}, \widehat{\phi}_y, \widehat{\phi}_b)d\boldsymbol{\theta}$:

$$\widehat{r}(\boldsymbol{x}) = \int \mathbb{E}[r(\boldsymbol{x}) \mid \boldsymbol{d}, \boldsymbol{\theta}, \widehat{\phi}_y, \widehat{\phi}_b]\pi(\boldsymbol{\theta} \mid \boldsymbol{d}, \widehat{\phi}_y, \widehat{\phi}_b)d\boldsymbol{\theta}, \tag{1.11}$$

où l'expression de $\mathbb{E}[r(\boldsymbol{x}) \mid \boldsymbol{d}, \boldsymbol{\theta}, \widehat{\phi}_y, \widehat{\phi}_b]$ est donnée dans [Kennedy and O'Hagan, 2001; Kennedy and O'Hagan, 2001; Damblin, 2015; Leoni, 2022].

Cette méthode a été largement reprise dans de nombreux travaux [Bayarri et al., 2007; Higdon et al., 2008; Qian and Wu, 2008; Liu et al., 2009; Wu et al., 2018]. Certains, dont [Higdon et al., 2004; Arendt et al., 2012; Brynjarsdottir and O'Hagan, 2014; Bachoc et al., 2014; Tuo and Wu, 2015; Damblin et al., 2016; Plumlee, 2017; Gu and Wang, 2018; Carmassi et al., 2019; Kamary et al., 2019; Leoni, 2022] montrent qu'elle produit parfois des résultats non identifiables en $(\boldsymbol{\theta}, b)$. Cette non identifiabilité est souvent due au choix de modélisation pour $b(\boldsymbol{x})$. Ainsi, pour y remédier, plusieurs modifications ont été apportées [Higdon et al., 2004; Bayarri et al., 2007; Bachoc et al., 2014; Brynjarsdottir and O'Hagan, 2014; Gu and Wang, 2018; Leoni, 2022]. Parmi elles, nous avons la proposition de Higdon et al. [2004] qui consiste à faire un calage sans la discrépance de modèle dans un premier temps puis à ne considérer le terme $b(\boldsymbol{x})$ que lorsque les prédictions du modèle calé et les mesures expérimentales $\boldsymbol{z}$ sont inconsistantes entre elles. En d'autres termes, les auteurs recommandent de ne prendre en compte $b(\boldsymbol{x})$ que lorsque l'écart entre les prédictions du modèle calé et les données expérimentales semble être plus important que l'incertitude expérimentale. Cette approche permet de contourner *a priori* le problème d'identifiabilité, évoqué précédemment, causé par la prise en compte du terme $b(\boldsymbol{x})$ dès le début du processus de calage. Contrairement à Kennedy and O'Hagan [2001], les auteurs proposent d'estimer les paramètres $\boldsymbol{\theta}$ et les hyperparamètres des processus gaussiens conjointement. Cette estimation conjointe est appelée *Full Bayesian approach* (pour approche bayésienne complète ou approche pure bayésienne). Soulignons également les travaux de Damblin et al. [2016] qui reprennent les idées de Higdon et al. [2004] et proposent une méthode pour tester la significativité de b à travers le calcul du facteur de Bayes entre le modèle sans b et le modèle avec b. Les auteurs proposent alors une règle de décision en comparant la valeur du facteur de Bayes à 1 : si cette valeur est très supérieure à 1 alors il ne faut pas prendre en compte la discrépance de modèle dans l'étude.

Continuons avec une autre variante de l'approche de Higdon et al. [2004] proposée par Bachoc et al. [2014]. Elle consiste à linéariser la sortie du code $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ et à rajouter une discrépance de modèle à nouveau lorsque l'amplitude des écarts absolus entre les observations et les prédictions du modèle linéaire calé n'est pas similaire à l'amplitude des incertitudes expérimentales caractérisées par les variances $\sigma^2_{\epsilon_{i,t}}$ $(1 \leq i \leq n, 1 \leq t \leq q)$. La discrépance de modèle est alors modélisée par un processus gaussien centré dont les hyperparamètres sont estimés séparément de $\boldsymbol{\theta}$, par la méthode du maximum de vraisemblance restreint [Bachoc et al., 2014; Santner et al., 2018]. Ensuite, nous avons les approches telle que l'approche de Plumlee [2017] et l'approche de Gu and Wang [2018] qui prennent la discrépance de modèle d'emblée dans le processus de calage avec une distribution *a priori* sur $b(\boldsymbol{x})$ bien choisie. Plumlee [2017] propose de choisir un processus gaussien sur la discrépance $b_{\boldsymbol{\theta}}(\boldsymbol{x})$ [3] avec une structure de covariance *a priori* particulière. La distribution *a priori* $\pi\left(b_{\boldsymbol{\theta}}(\boldsymbol{x}) \mid \boldsymbol{\theta}\right)$ est construite sous l'hypothèse d'orthogonalité entre $b_{\boldsymbol{\theta}}(\boldsymbol{x})$ et le gradient de $y_{\boldsymbol{\theta}}(\boldsymbol{x})$. L'auteur propose d'estimer conjointement $\boldsymbol{\theta}$ et les hyperparamètres du processus gaussien ainsi que $b_{\boldsymbol{\theta}}(\boldsymbol{x})$ par un algorithme de Gibbs [Robert et al., 1999; Gelman et al., 2014], alors que l'approche proposée par Gu and Wang [2018] concilie le calage déterministe et le calage bayésien dans le but d'avoir le plus faible écart quadratique possible entre la réalité physique $r(\boldsymbol{x})$ et le modèle $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ dans une inférence bayésienne. Pour ce faire, cet écart est modélisé par une variable aléatoire $\boldsymbol{Z}$ et la discrépance de modèle par un processus gaussien de fonction moyenne nulle dont

[3] La dépendance de la discrépance de modèle b à $\boldsymbol{\theta}$ c'est-à-dire $(b_{\boldsymbol{\theta}}(\boldsymbol{x}))$ introduite par Plumlee [2017] n'est que par souci de clarté. En effet, dans la littérature, cette dépendance est souvent supprimée.

l'intégrale quadratique est la réalisation de $\boldsymbol{Z}$. De cette manière, les auteurs construisent une distribution *a priori* sur la discrépance de modèle à partir de ce processus gaussien contraint. Cela favorise ainsi l'ajustement du modèle $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ aux données expérimentales du système physique $r(\boldsymbol{x})$. Enfin, récemment, Leoni [2022] a proposé une nouvelle méthode adaptative appelée FMP (pour *Full Maximum a Posteriori*) pour l'estimation des hyperparamètres du processus gaussien modélisant la discrépance. L'idée centrale de cette méthode adaptative consiste à estimer les hyperparamètres du processus gaussien conditionnellement à chaque valeur prise par les paramètres de calage $\boldsymbol{\theta}$. Autrement dit, cette approche adapte la forme de la discrépance de modèle associée à chaque nouvelle prédiction du modèle numérique en ré-estimant les hyperparamètres du processus gaussien associée à la discrépance. Comme le montre Leoni [2022], la méthode FMP a de bonnes propriétés et s'avère plus efficace que la méthode de Kennedy and O'Hagan [2001] pour l'échantillonnage des distributions *a posteriori* complexes telles que celles présentant plusieurs modes. Cependant, la méthode FMP nécessite en contrepartie un grand nombre d'estimations des hyperparamètres.

En conclusion, dans la littérature dédiée au calage bayésien, il n'y a pas un consensus général pour la résolution du problème d'identifiabilité du couple $(\boldsymbol{\theta}, b)$ induit par la prise en compte d'emblée d'une discrépance de modèle dans le processus de calage. La prise ou non d'une discrépance de modèle est à étudier au cas par cas. C'est pour cette raison que dans la communauté du calage bayésien, certains considèrent que la discrépance de modèle peut, dans certains cas, être indistinguable de la réalisation de l'incertitude expérimentale ϵ_i [Damblin et al., 2018; Yi et al., 2019] et par conséquent, ils utilisent l'équation probabiliste sans le terme discrépance malgré les risques de surapprentissage (*overfitting*). Nous adoptons ce principe dans cette thèse et considérons un calage sans discrépance de modèle se basant sur l'équation probabiliste suivante :

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta}}(\boldsymbol{x}_i) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}. \tag{1.12}$$

Le calage de modèle nous permet d'estimer $\boldsymbol{\theta}_{\text{opt}}$ en fonction des données disponibles sous l'hypothèse que $\boldsymbol{\theta}$ est indépendant de $\boldsymbol{x}$. Si nous n'avons plus cette indépendance aux variables de contrôle, c'est-à-dire si certaines composantes de $\boldsymbol{\theta}$ sont fonction de $\boldsymbol{x}$, alors le calage de modèle devra tenir compte de cette dépendance fonctionnelle [Plumlee et al., 2016; Brown and Atamturktur, 2018]. On parle alors de **calage fonctionnel**. Dans ce qui suit, nous décrivons l'approche proposée par Brown and Atamturktur [2018] qui apporte une solution à la problématique du calage fonctionnel.

Méthode de Brown et Atamturktur

Le principe de cette méthode réside dans le calage des paramètres $\boldsymbol{\theta}$ en tant que fonction des variables de contrôle $\boldsymbol{x}$. Les auteurs démontrent que l'intégration de cette dépendance fonctionnelle peut parfois prévenir l'inadéquation du modèle calibré. En effet, ils montrent que négliger cette dépendance peut tromper le statisticien en le faisant croire qu'une correction par un terme de discrépance est nécessaire pour pallier l'écart entre le modèle calibré et la réalité physique, alors qu'il suffit simplement d'introduire une dépendance fonctionnelle entre $\boldsymbol{\theta}$ et $\boldsymbol{x}$ dans le processus de calage.

Dans le cadre de cette approche, $\boldsymbol{\theta}$ est remplacé par $\boldsymbol{\theta}(\boldsymbol{x}_i)$ dans l'équation (1.12) :

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta}(\boldsymbol{x}_i)}(\boldsymbol{x}_i) + \epsilon_i,\ i = 1, \cdots, n_{\text{exp}}.$$

Les hypothèses formulées par les auteurs sont alors les suivantes :

- $\boldsymbol{x}_i \in [0,\ 1]^d$, $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \vartheta_y^{-1})$, $\vartheta_y \sim \Gamma(a_y, b_y)$, a_y, b_y connues,

- $\boldsymbol{\theta}(\boldsymbol{x}_i) = (\boldsymbol{\theta}_1(\boldsymbol{x}_i)^t, \boldsymbol{\theta}_2^t)^t$ où $\boldsymbol{\theta}_1(\boldsymbol{x}_i) \in \mathscr{C}^0(\mathbb{R}^d, \mathbb{R}^{p_1})$ est le vecteur de paramètres fonctionnels et $\boldsymbol{\theta}_2$ est un vecteur de paramètres constants (de taille p_2 tel que $p_1 + p_2 = p$) et $\pi(\boldsymbol{\theta}_2) \propto 1$,

- $\boldsymbol{\theta}_1(\cdot)$ est *a priori* indépendant de $\boldsymbol{\theta}_2$ et chacune de ses composantes est une réalisation *a priori* d'un processus gaussien indépendant. La loi *a priori* sur $\boldsymbol{\theta}(\boldsymbol{x})$ s'écrit donc :

$$\pi(\boldsymbol{\theta}(\boldsymbol{x})) = \pi(\boldsymbol{\theta}_1(\boldsymbol{x}))\pi(\boldsymbol{\theta}_2) = \prod_{j=1}^{p_1} \pi(\boldsymbol{\theta}_{1j}(\boldsymbol{x}))\pi(\boldsymbol{\theta}_2), \tag{1.13}$$

$$\boldsymbol{\theta}_{1j}(\cdot) \overset{\text{indep}}{\sim} \mathscr{GP}\left(\mu_{\boldsymbol{\theta},j}, \vartheta_{\boldsymbol{\theta},j}^{-1} R_j(\cdot,\cdot)\right),\ \mu_{\boldsymbol{\theta},j} \in \mathbb{R},\ 1 \leq j \leq p_1, \tag{1.14}$$

$$R_j(\boldsymbol{x}, \boldsymbol{x}') = \exp\left\{-4 \sum_{k=1}^{d} \gamma_{\boldsymbol{\theta},j} \left(\boldsymbol{x}_k - \boldsymbol{x}'_k\right)^2\right\} \tag{1.15}$$

où $\gamma_{\boldsymbol{\theta},j}$ contrôle la régularité des trajectoires de $\boldsymbol{\theta}_{1j}(\cdot)$.

Nous présentons cette approche dans le cas où $p_1 = p_2 = 1$ ($p = 2$) et $d = 1$. La distribution *a priori* de la fonctionnelle $\boldsymbol{\theta}_1(\cdot)$ se déduit de l'hypothèse processus gaussien :

$$\boldsymbol{\theta}_1(\cdot) \sim \mathscr{GP}\left(\mu_{\boldsymbol{\theta}}, \vartheta_{\boldsymbol{\theta}}^{-1} R_{\rho_{\boldsymbol{\theta}}}(\cdot,\cdot)\right), \tag{1.16}$$

où $R_{\rho_{\boldsymbol{\theta}}}(\boldsymbol{x}, \boldsymbol{x}') = \rho_{\boldsymbol{\theta}}^{4(\boldsymbol{x}-\boldsymbol{x}')^2}$, $\rho_{\boldsymbol{\theta}} = \exp(-\gamma_{\boldsymbol{\theta}})$ et $\mu_{\boldsymbol{\theta}} \in \mathbb{R}$ connu.

Cette méthode est une approche bayésienne hiérarchique car les auteurs ont spécifié des distributions *a priori* sur les hyperparamètres $\rho_{\boldsymbol{\theta}}$, $\vartheta_{\boldsymbol{\theta}}$ qui sont données respectivement par :

$$\vartheta_{\boldsymbol{\theta}} \sim \Gamma(0.01, 0.01), \tag{1.17}$$

$$\rho_{\boldsymbol{\theta}} \sim Beta(1, \alpha_{\boldsymbol{\theta}}). \tag{1.18}$$

Les auteurs recommandent de prendre $\alpha_{\boldsymbol{\theta}} = 0.1$ ou 0.2. Ensuite, la distribution jointe *a posteriori* est obtenue en appliquant la formule de Bayes :

$$\begin{aligned}\pi(\boldsymbol{\theta}_1(\boldsymbol{x}), \boldsymbol{\theta}_2, \rho_{\boldsymbol{\theta}}, \vartheta_{\boldsymbol{\theta}}, \vartheta_y \mid \boldsymbol{z}) \propto\ & \vartheta_y^{\frac{n_{\text{exp}}}{2}+a_y-1} \exp\left\{-\frac{\vartheta_y}{2}\left(\boldsymbol{z} - y_{\boldsymbol{\theta}(\boldsymbol{x})}(\boldsymbol{x})\right)^t \left(\boldsymbol{z} - y_{\boldsymbol{\theta}(\boldsymbol{x})}(\boldsymbol{x})\right)\right\} \vartheta_{\boldsymbol{\theta}}^{\frac{n_{\text{exp}}}{2}-0.99} \\ & \times |R_{\rho_{\boldsymbol{\theta}}}|^{-\frac{1}{2}} \exp\left\{-\frac{\vartheta_{\boldsymbol{\theta}}}{2}\left(\boldsymbol{\theta}_1(\boldsymbol{x}) - \mu_{\boldsymbol{\theta}}\mathbf{1}_n\right)^t R_{\rho_{\boldsymbol{\theta}}}^{-1}\left(\boldsymbol{\theta}_1(\boldsymbol{x}) - \mu_{\boldsymbol{\theta}}\mathbf{1}_{n_{\text{exp}}}\right)\right\} \\ & \times \exp(-b_y\vartheta_y) \exp(-\alpha_{\boldsymbol{\theta}}\vartheta_{\boldsymbol{\theta}})(1-\rho_{\boldsymbol{\theta}})^{\alpha_{\boldsymbol{\theta}}-1}.\end{aligned} \tag{1.19}$$

L'estimation des hyperparamètres $(\rho_{\boldsymbol{\theta}}, \vartheta_{\boldsymbol{\theta}}, \vartheta_y)$ et des paramètres $(\boldsymbol{\theta}_1(\boldsymbol{x}), \boldsymbol{\theta}_2)$ se fait conjointement via l'équation (1.19). Les auteurs utilisent un échantillonneur de Gibbs avec une étape de Metropolis pour échantillonner la distribution jointe

$$\pi(\boldsymbol{\theta}_1(\boldsymbol{x}), \boldsymbol{\theta}_2, \rho_{\boldsymbol{\theta}}, \vartheta_{\boldsymbol{\theta}}, \vartheta_y \mid \boldsymbol{z}).$$

Pour plus de détails, se référer à [Brown and Atamturktur, 2018; Plumlee et al., 2016].

En général, la distribution *a posteriori* de $\boldsymbol{\theta}$ n'est connue qu'à une constante près. Son estimation nécessite souvent l'usage d'algorithmes MCMC [Robert et al., 1999] tels que les algorithmes de Metropolis-Hastings et de Gibbs [Chib and Greenberg, 1995; Andrieu and Thoms, 2008; Gelman et al., 2014] qui sont les plus utilisés en pratique. Pour bien explorer l'espace *a priori* des paramètres de calage, l'algorithme de Metropolis-Hastings qui est un algorithme itératif fait un appel intensif au code de calcul $y_{\boldsymbol{\theta}}(\boldsymbol{x})$, ce qui peut être problématique lorsque le code de calcul est coûteux en temps CPU. Pour surmonter cette limite, l'approche standard consiste, comme évoqué précédemment, à remplacer $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ par une fonction mathématique $\widehat{y}_{\boldsymbol{\theta}}(\boldsymbol{x})$ appelée métamodèle (ou modèle de substitution ou encore émulateur) [Sacks et al., 1989; Currin et al., 1991; Santner et al., 2018] et dont l'évaluation est instantanée. À noter que dans la littérature, l'un des métamodèles le plus couramment utilisé est le métamodèle processus gaussien [Rasmussen et al., 2006; Marrel et al., 2009; Gramacy, 2020]. Ce métamodèle est ensuite utilisé dans l'algorithme Metropolis-Hastings afin d'estimer la distribution *a posteriori* de $\boldsymbol{\theta}$ en un temps raisonnable. Ce modèle de substitution est souvent

construit sur l'espace produit $\mathscr{D}_{\boldsymbol{X}} \times \mathscr{D}_{\boldsymbol{\Theta}}$ pour être capable de prédire le simulateur $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ pour de nouvelles valeurs de $\boldsymbol{x}$ ou de $\boldsymbol{\theta}$ [Wu et al., 2018; Damblin et al., 2018; Santner et al., 2018]. Bien que cette construction dans l'espace joint soit toujours préférable, elle peut perdre de sa pertinence dans certains problèmes de calage où les configurations expérimentales $\{x_i\}_{i=1}^{n_{\text{exp}}}$ sont hautement structurées, comme en simulation thermohydraulique [Damblin and Gaillard, 2020]. Dans de telles situations, il est préférable de construire un métamodèle $\widehat{y}_i(\boldsymbol{\theta})$ pour chaque $y_i(\boldsymbol{\theta}) := y_{\boldsymbol{\theta}}(\boldsymbol{x}_i)$.

L'usage d'un métamodèle se substituant au code de calcul dans l'algorithme de Métropolis-Hastings permet de contourner le coût prohibitif de ce dernier. Cependant, lorsque le nombre de paramètres de calage est important (au-delà d'une dizaine), deux autres problèmes demeurent :

1. en raison du grand nombre d'hyperparamètres inconnus à estimer, la construction du métamodèle peut être ardue. Des recherches sont en cours pour proposer des solutions à cette problématique [Iooss and Marrel, 2019; Marrel et al., 2022; Zhou et al., 2022; Binois and Wycoff, 2022],

2. la convergence de l'algorithme de Metropolis-Hastings devient difficile également. Plus le nombre de paramètres est grand, plus l'algorithme de Metropolis-Hastings a du mal à bien mélanger la chaîne de Markov induite. Cela résulte d'un mauvais réglage des hyperparamètres de l'algorithme de Metropolis-Hastings [Haario et al., 2001; Andrieu and Thoms, 2008; Gelman et al., 2014] qui a pour conséquence la génération de distributions marginales fortement corrélées. On se retrouve ainsi face à un problème d'identifiabilité entre les composantes de $\boldsymbol{\theta}$ [Liu et al., 2009; Cole, 2020].

Une solution pour contourner ces deux problèmes, ci-dessus, consiste à utiliser d'une part, des méthodes de métamodélisation avancées [Marrel et al., 2022] et d'autre part, des algorithmes de Metropolis-Hastings adaptatifs [Andrieu and Thoms, 2008] ou MCMC hamiltoniens (HMC) [Neal et al., 2011]. Cependant, ces algorithmes adaptatifs et HMC sont complexes et difficiles à régler en pratique. Une solution alternative consiste alors à effectuer une réduction de la dimension des paramètres $\boldsymbol{\theta}$, avant calage, en utilisant des techniques d'analyse de sensibilité [Van Oijen et al., 2011; Zambrano-Bigiarini et al., 2013; Xu et al., 2016; Wu et al., 2017, 2019; Nagel et al., 2020; Dighe et al., 2022; Perret et al., 2022]. **Dans ce manuscrit, nous utiliserons cette solution mais en se basant uniquement sur une technique d'analyse de sensibilité globale afin de détecter l'indépendance entre chaque paramètre et la/les sortie(s) d'intérêt. Ainsi, un paramètre du modèle sera considéré comme un paramètre de calage influent lorsqu'une dépendance statistiquement significative entre ce paramètre et la/les sortie(s) d'intérêt aura été détectée.**

1.1.3 Analyse de sensibilité globale en support au calage de modèle bayésien

L'analyse de sensibilité globale, telle que définie précédemment, vise à évaluer de manière quantitative l'impact des incertitudes liées aux paramètres d'un modèle numérique, représentées par des distributions de probabilité, sur la variabilité de la sortie d'intérêt [Saltelli et al., 2004; Da Veiga et al., 2021]. Cette sortie d'intérêt peut englober l'ensemble de la sortie vectorielle du code de calcul, une composante de cette sortie ou même une fonction des paramètres telle que la fonction des moindres carrés. En pratique, l'analyse de sensibilité peut être utilisée pour simplifier les modèles numériques trop complexes en éliminant les entrées ayant peu ou pas d'influence sur la variabilité de la sortie d'intérêt. Cette simplification a pour avantage pratique de faciliter la construction de métamodèles [Marrel et al., 2022]. De plus, elle contribue à améliorer la convergence des algorithmes MCMC dans un contexte de calage bayésien, en ne prenant en compte que les paramètres du modèle qui ont une influence significative sur la variabilité de la sortie d'intérêt [Wu et al., 2019; Minunno et al., 2013].

Pour atteindre cet objectif de la réduction du nombre de paramètres, on réalise un criblage des paramètres du modèle capable de caractériser l'indépendance entre chaque composante

du vecteur de paramètres et la/les sortie(s) d'intérêt. Plus précisément, nous cherchons à détecter une indépendance statistique entre chaque composante de $\boldsymbol{\theta}$ et la/les sortie(s) d'intérêt à partir d'un jeu données de taille finie. Le choix de la mesure de sensibilité utilisée pour ce criblage dépendra, d'une part, de sa capacité à s'appliquer aussi bien à une sortie d'intérêt scalaire qu'à une sortie vectorielle et d'autre part, de son adaptabilité à la nature des données, au contexte de données fournies, ainsi qu'au nombre limité de simulations disponibles (environ une centaine ici).

Parmi le large spectre de mesures de sensibilité, les indices de Sobol' occupent une place prépondérante dans la littérature [Ratto et al., 2001; Ciuffo and Azevedo, 2014; Mathieu et al., 2018; Wu et al., 2019; Nagel et al., 2020; Wagner et al., 2020; Gou et al., 2020; Dighe et al., 2022]. Ces indices, introduits par Sobol [1993], sont largement utilisés en pratique dans diverses applications pratiques couvrant plusieurs domaines tels que l'ingénierie nucléaire [Bouloré et al., 2012; Iooss and Le Gratiet, 2019; Lefebvre et al., 2023], l'hydrogéologie [Marrel et al., 2011; Gou et al., 2020; Saveleva et al., 2021] et tant d'autres [Minunno et al., 2013; Wagner et al., 2020; Dighe et al., 2022]. Ces indices reposent sur la décomposition de la variance de la sortie d'intérêt où chaque élément représente la part d'influence d'une seule entrée ou d'un groupe d'entrées et s'interprètent en termes de pourcentage de variance expliquée pour chaque entrée ou groupe d'entrées par rapport à la variance totale de Y. Ces éléments sont obtenus grâce à la décomposition ANOVA (pour *ANalysis Of VariAnce*) [Hoeffding, 1992]. Plus précisément, pour $\boldsymbol{Y} = f(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_p) \in \mathbb{R}$[4] où les entrées $\boldsymbol{X}_i$ sont supposées mutuellement indépendantes, un indice de Sobol' pour un groupe d'entrées noté $\boldsymbol{X}_u$ est donné par:

$$\mathrm{S}_u = \frac{\mathbb{V}\left(\mathbb{E}\left[\boldsymbol{Y} \mid \boldsymbol{X}_u\right]\right)}{\mathbb{V}\left(\boldsymbol{Y}\right)} = \frac{\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{Y} \mid \boldsymbol{X}_u\right]^2\right] - \mathbb{E}\left[\boldsymbol{Y}\right]^2}{\mathbb{V}\left(Y\right)}. \tag{1.20}$$

Lorsque $\boldsymbol{X}_u = \boldsymbol{X}_j$ $(1 \leq j \leq p)$, on parle d'indice du premier ordre ou d'effet principal, qui quantifie la contribution individuelle de $\boldsymbol{X}_j$ sur la variabilité de $\boldsymbol{Y}$. Lorsque $\boldsymbol{X}_u = (\boldsymbol{X}_i, \boldsymbol{X}_j)$ avec $1 \leq i < j \leq p$, on parle d'indice d'ordre 2 qui mesure la contribution due au couple $(\boldsymbol{X}_i, \boldsymbol{X}_j)$ sur la variabilité de $\boldsymbol{Y}$ indépendemment de leurs effets principaux, et ainsi de suite pour les indices d'ordre supérieur. Pour tout $1 \leq j \leq p$, il est possible de définir un indice de Sobol' qui mesure toute la contribution de $\boldsymbol{X}_j$ sur la variabilité totale de $\boldsymbol{Y}$ incluant son effet principal ainsi que l'ensemble de ses possibles interactions avec les autres variables $\boldsymbol{X}_{-j} := (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_{j-1}, \boldsymbol{X}_{j+1}, \cdots, \boldsymbol{X}_p)$. Cet indice est appelé indice d'ordre total et est défini par

$$T_j = 1 - \mathrm{S}_{-j} = 1 - \frac{\mathbb{E}\left[\mathbb{E}\left[Y \mid \boldsymbol{X}_{-j}\right]^2\right] - \mathbb{E}\left[\boldsymbol{Y}\right]^2}{\mathbb{V}\left(\boldsymbol{Y}\right)}. \tag{1.21}$$

La nullité de l'indice total T_j est équivalente à l'indépendance entre les variables $\boldsymbol{X}_j$ et $\boldsymbol{Y}$.

En général, l'obtention des valeurs théoriques de ces indices de Sobol' (équations (1.20) et (1.21)) n'est pas réalisable. Pour cette raison, divers estimateurs dotés de bonnes propriétés mathématiques (consistance, vitesse de convergence explicite, normalité asymptotique, etc.) ont été proposés au cours des dernières décennies afin d'estimer les quantités $\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{Y} \mid \boldsymbol{X}_u\right]^2\right]$. Le défi pratique de l'estimation des indices de Sobol' réside en effet essentiellement dans l'estimation des quantités $\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{Y} \mid \boldsymbol{X}_u\right]^2\right]$, puisque $\mathbb{E}\left[\boldsymbol{Y}\right]$ et $\mathbb{V}\left(\boldsymbol{Y}\right)$ sont facilement estimables (en prenant simplement la moyenne et la variance empiriques). Ces estimateurs sont souvent construits à partir de plans d'expériences des entrées $\{\boldsymbol{X}_j\}_{1 \leq j \leq p}$ de type Monte Carlo ou quasi-Monte Carlo [Sobol, 2001]. Notons dans ce qui suit, la taille n de ces plans.

Au sein de ces diverses classes d'estimateurs, nous trouvons les premiers estimateurs de variances d'espérances conditionnelles proposés par McKay [1997] et qui sont basés sur une double boucle d'échantillonnage. Ce schéma d'échantillonnage nécessitant un très grand

[4] La fonction f est supposée être de carré intégrable.

nombre de simulations (plan de taille $n \geq 10^4$) pour avoir une estimation satisfaisante de chaque indice peut s'avérer impossible pour les codes de calcul coûteux en temps de calcul CPU. C'est pour cette raison que d'autres estimateurs ont été proposés tels que les estimateurs issus de plans d'expériences Monte Carlo spécifiques, connus sous le nom de *Pick-Freeze design of experiments* [Saltelli, 2002]. Les estimateurs *Pick-Freeze* présentent de bonnes propriétés mathématiques telles que la consistance et la normalité asymptotique [Janon et al., 2014; Gamboa et al., 2016]. Pour ce faire, le *Pick-Freeze* se base sur une réécriture des quantités $\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{Y} \mid \boldsymbol{X}_u\right]^2\right]$ par

$$\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{Y} \mid \boldsymbol{X}_u\right]^2\right] = \mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^u\right] \quad \text{avec} \quad \boldsymbol{Y}^u := f\left(\boldsymbol{X}_u, \boldsymbol{X}'_{-u}\right), \tag{1.22}$$

où X' est une copie indépendante de X. En utilisant l'égalité (1.22), établie par le lemme 2.2 dans [Janon et al., 2014], dans les équations (1.20) (avec $u = \{j\}$) et (1.21), nous obtenons :

$$\mathrm{S}_j = \frac{\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^j\right] - \mathbb{E}\left[\boldsymbol{Y}\right]^2}{\mathbb{V}\left(\boldsymbol{Y}\right)} = \frac{\mathrm{cov}\left(\boldsymbol{Y}, \boldsymbol{Y}^j\right)}{\mathbb{V}\left(\boldsymbol{Y}\right)}, \tag{1.23}$$

$$T_j = 1 - \frac{\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^{-j}\right] - \mathbb{E}\left[\boldsymbol{Y}\right]^2}{\mathbb{V}\left(\boldsymbol{Y}\right)} = 1 - \frac{\mathrm{cov}\left(\boldsymbol{Y}, \boldsymbol{Y}^{-j}\right)}{\mathbb{V}\left(\boldsymbol{Y}\right)}. \tag{1.24}$$

Le principe d'estimation *Pick-Freeze* pour les indices du premier ordre S_j consiste d'abord à générer un plan d'expériences $\boldsymbol{X}_A$ de taille n et à exécuter le code pour obtenir un vecteur de simulations de la sortie d'intérêt, noté $\boldsymbol{Y}_A \in \mathbb{R}^n$. Ensuite, un autre plan d'expériences $\boldsymbol{X}_B$, indépendant de $\boldsymbol{X}_A$, est généré sans en conduire les simulations correspondantes. En troisième étape, un troisième plan d'expériences $\boldsymbol{X}_C = \left(\boldsymbol{X}_A^j, \boldsymbol{X}_B^{-j}\right)$ est construit à partir des deux plans $\boldsymbol{X}_A$ et $\boldsymbol{X}_B$: la $j^{\text{ème}}$ colonne de $\boldsymbol{X}_C$ est la $j^{\text{ème}}$ colonne de $\boldsymbol{X}_A$ et les autres colonnes de $\boldsymbol{X}_C$ sont des colonnes de $\boldsymbol{X}_B$ à l'exception de sa $j^{\text{ème}}$ colonne. Enfin, le code est exécuté sur $\boldsymbol{X}_C$ pour obtenir $\boldsymbol{Y}_B^j \in \mathbb{R}^n$. L'indice S_j (équation (1.23)) est alors estimé à partir des vecteurs $\boldsymbol{Y}_A$ et $\boldsymbol{Y}_B^j$ en utilisant des estimateurs Monte Carlo classiques de covariance et de variance :

$$\widehat{\mathrm{S}}_j = \frac{\widehat{\mathrm{cov}}\left(\boldsymbol{Y}_A, \boldsymbol{Y}_B^j\right)}{\widehat{\mathbb{V}}\left(\boldsymbol{Y}_A\right)}. \tag{1.25}$$

En somme, la construction de l'estimateur de l'équation (1.25) a nécessité n appels au code pour obtenir $\boldsymbol{Y}_A$ et n appels pour obtenir également $\boldsymbol{Y}_B^j$. Le coût de simulation pour le calcul de $\widehat{S}_j$ est donc de $2n$. Par conséquent, le coût total d'estimation de tous les indices du premier ordre $\{\mathrm{S}_j\}_{j=1}^p$ s'élève à $n(p+1)$ [Saltelli, 2002]. Pour estimer les indices d'ordre total $\boldsymbol{S}_j^T$ (équation (1.24)), il suffit d'exécuter le code sur le plan $\boldsymbol{X}_B = \left(\boldsymbol{X}_B^j, \boldsymbol{X}_B^{-j}\right)$ pour obtenir un autre vecteur de simulations $\boldsymbol{Y}_B$ puis de calculer son estimateur comme suit :

$$\widehat{T}_j = 1 - \frac{\widehat{\mathrm{cov}}\left(\boldsymbol{Y}_B^j, \boldsymbol{Y}_B\right)}{\widehat{\mathbb{V}}\left(\boldsymbol{Y}_A\right)}. \tag{1.26}$$

Il est important de noter que l'astuce utilisée pour obtenir l'estimateur $\widehat{T}_j$ de l'équation (1.26) découle du fait que les plans $\boldsymbol{X}_C$ et $\boldsymbol{X}_B$ sont identiques à l'exception de la $j^{\text{ème}}$ colonne. En résumé, le coût total des appels nécessaires pour estimer tous les indices du premier ordre et d'ordre total est $N = n(p+2)$ [Saltelli, 2002]. Pour obtenir une précision satisfaisante des estimateurs

$$\left\{\widehat{S}_j\right\}_{1\leq j\leq p} \quad \text{et} \quad \left\{\widehat{T}_j\right\}_{1\leq j\leq p},$$

il est nécessaire d'avoir au moins une centaine de simulations [Iooss and Lemaître, 2015]. Par conséquent, plus le nombre de variables p (par exemple, $p \geq 10$) est élevé, plus le déploiement de la méthode *Pick-Freeze* devient quasiment impossible pour des codes de calcul coûteux en

temps de calcul CPU en raison de la dépendance linéaire du coût de simulation à la dimension p. Pour réduire ce coût élevé, l'approche standard consiste à calculer ces indices en utilisant des métamodèles [Sudret, 2008; Marrel et al., 2009; Crestaux et al., 2009] mais la construction d'un métamodèle prédictif s'avère de plus en plus complexe lorsque p augmente (en général à partir de $p > 10$).

Pour surmonter la dépendance linéaire par rapport à la dimension p dans le calcul des indices, de nouvelles variantes de la méthode *Pick-Freeze* ont été proposées. Ainsi, Tissot and Prieur [2015] ont proposé d'adapter la méthode du *Pick-Freeze* à deux plans hypercubes latins répliqués [5] conduisant ainsi à un budget de simulations $N = 2n$ nécessaire pour l'estimation de tous les indices du premier ordre et du second ordre. Ils montrent que les estimateurs des indices du premier ordre et du second ordre obtenus convergent presque sûrement vers les valeurs théoriques et vérifient également le théorème central limite sous condition que les variables aléatoires $\boldsymbol{Y}^4$ et $\boldsymbol{Y}^6$ soient respectivement intégrables (voir la proposition 3.2 dans [Tissot and Prieur, 2015]). Cependant, le coût de simulation pour estimer les indices d'ordre total reste inchangé. Autrement dit, il est toujours linéaire par rapport à la dimension p des entrées. Récemment, les auteurs de Gamboa et al. [2022], inspirés par les travaux de Chatterjee [2021], ont proposé une nouvelle méthode d'estimation des indices du premier ordre basée sur les statistiques d'ordre. Pour ce faire, ils ont proposé une méthodologie générale pour estimer, à partir uniquement d'un seul plan d'expériences et des simulations correspondantes $\left\{\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}\right\}_{i=1}^{n}$, les quantités définies par :

$$\mathbb{E}\left[\mathbb{E}\left[h_1(\boldsymbol{Y}) \mid \boldsymbol{X}_j\right] \mathbb{E}\left[h_2(\boldsymbol{Y}) \mid \boldsymbol{X}_j\right]\right], \tag{1.27}$$

où h_1 et h_2 sont deux fonctions mesurables bornées. Ils proposent ainsi l'estimateur suivant :

$$\frac{1}{n} \sum_{i=1}^{n} h_1\left(\boldsymbol{y}_j\right) h_2\left(\boldsymbol{y}_{\tau_n(i)}\right), \tag{1.28}$$

où τ_n est une permutation aléatoire qui n'a pas de points fixes et mesurable selon la tribu engendrée par $(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)})$. Cette permutation est définie dans [Gamboa et al., 2022; Chatterjee, 2021] par :

$$\tau_n(i) := \begin{cases} \sigma^{-1}\left(\sigma(i)+1\right) & \text{si } \sigma(i)+1 \leq n \\ \sigma^{-1}\left(1\right) & \text{si } \sigma(i) = n, \end{cases} \tag{1.29}$$

où $\sigma(i)$ est le rang de la statistique d'ordre $\boldsymbol{X}_j^{(i)}$ et $\tau_n(i)$ s'interprète comme l'antécédent de la statistique d'ordre qui se trouve juste après la statistique d'ordre $\boldsymbol{X}_j^{(i)}$ dans le classement. Ainsi, pour estimer tous les indices de Sobol' du premier ordre, il suffit de prendre

$$h_1(\boldsymbol{x}) = h_2(\boldsymbol{x}) = \boldsymbol{x}.$$

Finalement, tous les indices du premier ordre s'obtiennent à partir d'un seul échantillon ($N = n$) en considérant chaque paire $(\boldsymbol{X}_j, \boldsymbol{Y})$ associée à sa propre permutation basée sur les rangs des $\boldsymbol{X}_j$ et ils vérifient le théorème central limite [Gamboa et al., 2022]. Cette méthode basée sur le classement ne peut pas être utilisée pour estimer les indices d'ordre supérieur car nous n'avons pas une relation d'ordre pour des quantités vectorielles. Toutefois, il est possible de remplacer ce classement par une approche de plus de proches voisins [Broto et al., 2020] ou par une approche de noyaux de lissage [Da Veiga et al., 2023]. Pour plus de détails sur

[5] Une matrice $\boldsymbol{M}$ de taille $n \times p$ est dit plan hypercube latin lorsqu'elle constitue un plan d'expériences quasi-Monte Carlo où chaque colonne i de $\boldsymbol{M}$ est générée à partir d'une permutation $\tau_i \in \mathbb{S}_n$ [Santner et al., 2018]. Si $\boldsymbol{M}$ et $\boldsymbol{V}$ sont deux plans hypercube latins de taille n, alors on dit qu'ils sont répliqués lorsque qu'il existe une permutation $\tau_i \in \mathbb{S}_n$ telle que $\tau_i\left(\boldsymbol{V}_i\right) = \boldsymbol{M}_i$, où $\boldsymbol{M}_i$ et $\boldsymbol{V}_i$ représentent les colonnes respectives de $\boldsymbol{M}$ et $\boldsymbol{V}$ [McKay et al., 1999; Alex and Rakoto Joseph, 2008; Damblin and Ghione, 2021].

ces différentes méthodes, le lecteur intéressé peut se référer à [Saltelli et al., 2008; Da Veiga et al., 2021; Da Veiga, 2021; Da Veiga et al., 2023].

Hormis les indices de Sobol' qui sont basés sur la décomposition de la variance de la sortie d'intérêt $\boldsymbol{Y}$, des indices basés sur des mesures de dépendance [Da Veiga, 2015] ont aussi été développés tels que les indices basés sur des densités de probabilité [Borgonovo, 2007; Borgonovo and Plischke, 2016]. Ces indices de sensibilité mesurent la dépendance entre un groupe d'entrée $\boldsymbol{X}_u$ et la quantité $\boldsymbol{Y}$ à travers une mesure de dissimilarité mesurant l'écart entre la densité jointe $f_{\boldsymbol{Y},\boldsymbol{X}_u}(\boldsymbol{y},\boldsymbol{x}_u)$ et le produit des densités marginales $f_{\boldsymbol{Y}}(\boldsymbol{y})f_{\boldsymbol{X}_u}(\boldsymbol{x}_u)$. En pratique, ces densités de probabilité ne sont pas connues et le calcul des indices de Borgonovo nécessitent souvent une phase préliminaire d'estimation de densités de probabilité. Cependant, en présence de données de faible taille et en grande dimension, cette estimation pourrait ne pas être satisfaisante et par conséquent, les résultats fournis par ces indices pourraient ne pas être fiables. Enfin, récemment, de nouvelles mesures de sensibilité basées sur des noyaux ont vu le jour. Ces nouveaux indices de sensibilité s'affranchissent de la phase d'estimation de densités. Basés sur le Critère d'Indépendance Hilbert-Schmidt (HSIC pour *Hilbert-Schmidt Independence Criterion*) et présentés par Da Veiga [2015], ils caractérisent bien l'indépendance entre chaque entrée $\boldsymbol{X}_j$ et la sortie d'intérêt $\boldsymbol{Y}$ sous l'hypothèse de noyaux caractéristiques [6] [Fukumizu et al., 2008].

Le HSIC est une mesure statistique qui évalue le degré d'indépendance entre deux variables aléatoires $\boldsymbol{X}_j$ et $\boldsymbol{Y}$ [Gretton et al., 2005b]. Contrairement aux méthodes traditionnelles qui se limitent à détecter des dépendances linéaires telles que le tau (τ) de [Kendall, 1949] ou le rho (ρ) de [Spearman, 1904], le HSIC peut capturer des dépendances complexes et non-linéaires entre les variables. Cette capacité à explorer des relations plus profondes est rendue possible grâce à une approche exploitant des espaces mathématiques abstraits connus sous le nom d'espaces de Hilbert à noyau reproduisant (RKHS pour *Reproducing Kernel Hilbert Space*). L'idée de base consiste à associer à l'espace de départ (ici le domaine de définition de la variable aléatoire) un noyau et un espace fonctionnel muni d'un produit scalaire hérité de ce noyau. Ces éléments définissent ainsi un RKHS (théorème de Moore-Aronszajn [Aronszajn, 1950]). Les RKHS sont des espaces de Hilbert particuliers qui permettent de considérer des transformations non-linéaires sur les données. Plus précisément, à chaque élément de l'espace de départ, une fonction dans le RKHS lui est associée. Le noyau permet ensuite de mesurer les distances entre les fonctions dans le RKHS. Appliqués à des variables aléatoires, les projections dans le RKHS joint (un RKHS est associé à chacune des variables) permettent ainsi de calculer des mesures de dissimilarité entre les distributions de probabilité des variables (distance entre les représentants moyens des distributions de probabilité dans le RKHS joint). En résumé, l'utilisation des RKHS dans le contexte du HSIC permet de projeter les variables d'entrée dans un espace de dimension supérieure, où les dépendances linéaires et non-linéaires peuvent être capturées. Le HSIC offre donc un cadre d'analyse sophistiqué pour étudier l'indépendance statistique entre deux ensembles de données ou rechercher des motifs dans ces données [Gretton et al., 2006, 2007] en exploitant les propriétés des espaces de Hilbert à noyau reproduisant. Cette approche élargit les possibilités d'exploration des relations complexes entre les variables aléatoires, ce qui en fait un outil précieux dans les domaines de l'apprentissage automatique, et de la statistique [Da Veiga, 2015; De Lozzo and Marrel, 2016; El Amri and Marrel, 2024; Da Veiga, 2021]. Les détails mathématiques du HSIC ainsi que toute la procédure de test d'indépendance associée sont exposés au Chapitre 3.

Dans la section suivante, nous présentons en détail le cas test ayant motivé ces travaux de thèse. Nous présentons également les enjeux et objectifs du calage des paramètres de ce cas test.

[6] La définition de noyaux caractéristiques est donnée au Chapitre 3

Figure 1.2: Pastilles d'UO_2 [Guillet et al., 2008].

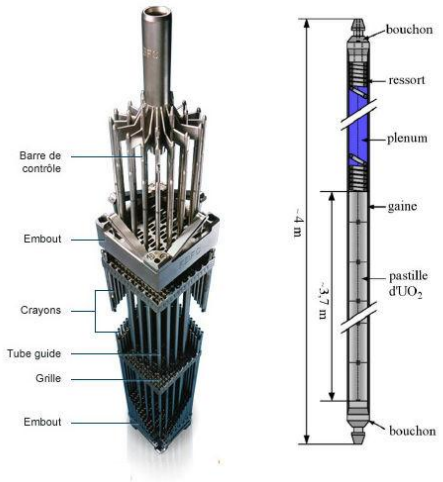


Figure 1.3: Assemblage des crayons combustibles [Guillet et al., 2008].

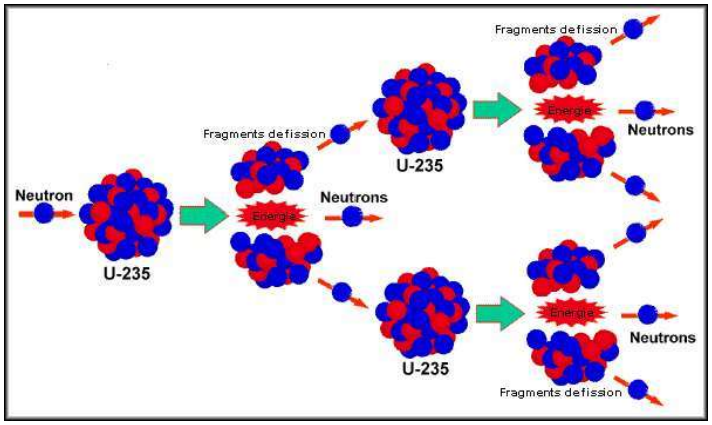


Figure 1.4: Réaction en chaîne de l'uranium 235 bombardé par un neutron [Guillet et al., 2008].

1.2 Description du cas applicatif

1.2.1 Présentation des crayons combustibles

Dans les cœurs des réacteurs nucléaires, le combustible nucléaire est un composé chimique contenant des matières fissiles comme l'uranium qui sont le siège de la fission nucléaire. Le combustible est la source de chaleur permettant de produire de l'énergie tout en entretenant la réaction en chaîne de fission nucléaire. Cette réaction en chaîne est illustrée par la Figure 1.4. Le combustible prend la forme d'un empilement gainé de pastilles cylindriques qui contiennent la matière fissile. Cet empilement est appelé crayon combustible [Guillet et al., 2008]. Notons que la gaine constitue la première barrière de confinement des matières radioactives produites lors de la réaction. Dans les réacteurs à eau pressurisée (REP), le cœur est constitué de plusieurs assemblages de plusieurs crayons combustibles gainés d'un alliage de zirconium (appelé aussi zircaloy) et contenant des pastilles d'oxyde d'uranium (UO_2). Les Figures 1.2 et 1.3 représentent respectivement les pastilles d'UO_2 et l'assemblage des crayons combustibles.

1.2.2 Simulateur ALCYONE

Le simulateur ALCYONE est une application multidimensionnelle [Guillet et al., 2008; Marelle et al., 2016; Delipei, 2019; Michel et al., 2021] de la plateforme PLEIADES pour la simulation du comportement des combustibles des REP. Elle est constituée d'un ensemble de modèles

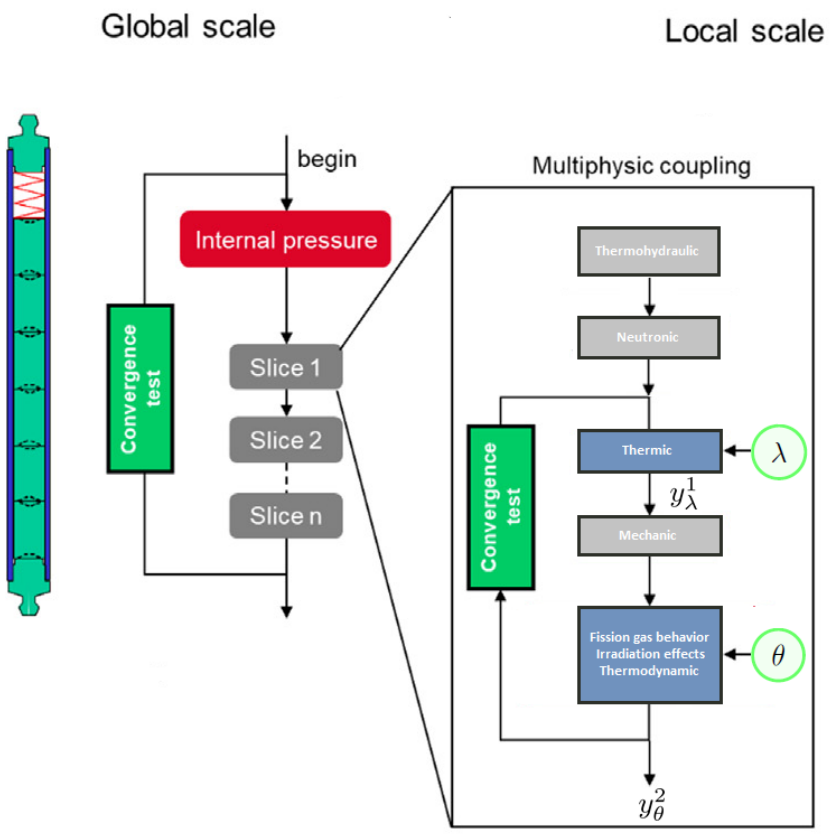


Figure 1.5: Chaînage des modèles thermique et de comportement des gaz de fission du simulateur ALCYONE : λ est la conductivité thermique (en entrée du modèle thermique) et $\boldsymbol{\theta}$ représente les paramètres du modèle de comportement des gaz de fission [Michel et al., 2021].

physiques dans le but de comprendre et simuler les comportements mécanique, thermique et chimique du combustible d'un REP. Le schéma de calcul multi-physique est illustré par la Figure 1.5 et une boucle itérative assure la convergence de l'ensemble des calculs en chaîne. Dans le cadre des travaux de thèse, on s'intéresse uniquement aux deux modèles suivants (représentés en bleu sur la Figure 1.5) : le **modèle thermique** et le **modèle de comportement des gaz de fission**. Chacun de ces modèles dépend d'un certain nombre de paramètres (paramètres physiques, numériques, propriétés matériaux, $\cdots$). Dans le cadre de ce travail, seule la conductivité thermique du combustible (notée $\lambda \in \mathscr{D}_{\boldsymbol{\Lambda}} \subset \mathbb{R}$) et les paramètres physiques du modèle de comportement des gaz de fission (notés $\boldsymbol{\theta} \in \mathscr{D}_{\boldsymbol{\Theta}}$) sont considérés.

1.2.3 Le modèle thermique

Le modèle thermique calcule l'évolution de la température au sein du cœur et l'évacuation avec précision de la chaleur induite vers le fluide caloporteur au cours de la réaction en chaîne. Pour cela, il résout l'équation de la chaleur produite dans la pastille combustible et donne en sortie le champ de température en tout point du maillage. Cette équation est la suivante :

$$\frac{\partial H}{\partial t} = -\overrightarrow{\nabla}.\overrightarrow{j} + \boldsymbol{q}, \tag{1.30}$$

où

- H est l'enthalpie volumique du matériau combustible,
- $\overrightarrow{j}$ est le flux de chaleur et $\boldsymbol{q}$ est la puissance volumique générée.

Notons que la quantité de chaleur produite $\boldsymbol{q}$ est reliée à la puissance d'irradiation $P(t)$. Cette puissance est connue pour chaque crayon combustible avec une incertitude épistémique de l'ordre de 5% [Bouloré et al., 2012]. Le calcul des flux de chaleur fait intervenir différents phénomènes physiques représentés sur la Figure 1.6. Le phénomène de conduction fait intervenir la conductivité du combustible λ qui est une propriété intrinsèque de l'UO_2 et qui évolue au cours du temps d'irradiation du combustible [Guillet et al., 2008; Michel et al., 2021], du fait des modifications physico-chimiques du combustible dues à la réaction en chaîne dans le réacteur. En régime normal permanent de fonctionnement du réacteur, et en représentation 1D radiale (schéma ALCYONE 1D), l'équation (1.30) peut s'écrire :

$$\boldsymbol{q} + \frac{1}{r_{\text{rod}}} \frac{d(\lambda(T, \tau_{\text{burnup}}) r_{\text{rod}} \frac{dT}{dr_{\text{rod}}})}{dr_{\text{rod}}} = 0, \tag{1.31}$$

où le taux de combustion τ_{burnup} (en MWj/tU) est calculé à partir de l'historique de puissance et est donné par :

$$\tau_{\text{burnup}}(t) = \frac{1}{\pi r_{\text{past}}^2 \rho} \frac{270}{238} 10^{-4} \int_0^t P(u) du, \tag{1.32}$$

avec

- r_{past} est le rayon initial de la pastille combustible (m),
- ρ est la masse volumique initiale du combustible (tonne/m^3),
- $P(t)$ est la puissance thermique fournie par unité de longueur (W/cm),
- t est la durée du matériau dans le coeur du réacteur REP (jours).

L'objectif du modèle thermique est de calculer le champ de température sur le rayon r_{past} de la pastille. La conductivité thermique de l'UO_2 est une grandeur importante pour l'évaluation du champ de température dans le combustible au cours de l'irradiation[7] quel que soit le régime utilisé (régime nominal, incidentel ou accidentel). C'est une propriété intrinsèque du matériau qui évolue avec la température et l'irradiation du matériau [Bouloré et al., 2023]. La loi modélisant la conductivité de l'UO_2 vierge utilisée dans l'application ALCYONE [Bouloré et al., 2023] est la suivante :

$$\lambda_0(T) = \frac{1 - p_o}{1 + 2p_o} \left(\frac{1}{a + bT} + \frac{c}{T^{2.05}} \exp\left(-\frac{d}{T}\right) \right), \tag{1.33}$$

où

- T est la température à laquelle est soumise le matériau,
- $p_o = 0.05$ est la porosité,
- $a = 0.035$, $b = 2,165.10^{-4}$, $c = 4.715.10^9$, $d = 16361$ sont des paramètres de modélisation calés dans de précédentes études [Bouloré et al., 2023].

En régime permanent de fonctionnement réacteur, la température ne dépasse pas $1500K$ et la relation (1.33) peut être simplifiée en :

$$\lambda_0(T) \approx \frac{(1 - p_o)}{(1 + 2p_o)} \frac{1}{(a + bT)}. \tag{1.34}$$

Les quantités $\lambda_0(T)$ et $\lambda(T)$ sont reliées par l'équation suivante :

$$\lambda(T, \tau_{\text{burnup}}) = \lambda_0(T) C_1(T, \tau_{\text{burnup}}) C_2(T, \tau_{\text{burnup}}) C_3(T, \tau_{\text{burnup}}), \tag{1.35}$$

où les coefficients $C_1(T, \tau_{\text{burnup}}), C_2(T, \tau_{\text{burnup}})$ et $C_3(T, \tau_{\text{burnup}})$ représentent respectivement la dégradation liée aux produits de fission dissous, le coefficient d'amélioration dû aux produits

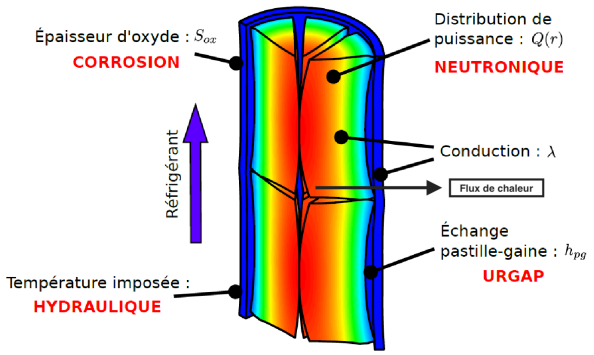


Figure 1.6: Phénomènes physiques intervenant dans le calcul thermique [Guillet et al., 2008].

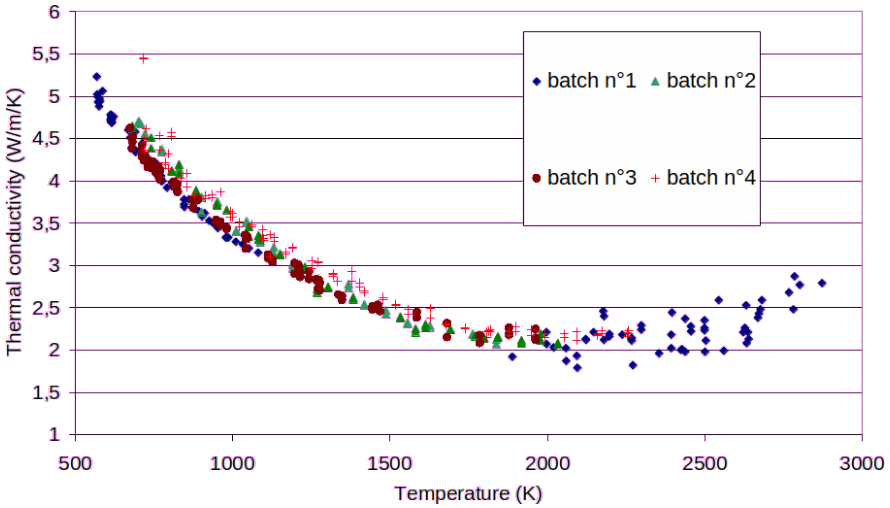


Figure 1.7: Mesures de la conductivité thermique de l'UO_2 vierge non irradié en fonction de la température calculée, pour quatre lots de mesures correspondant à des conditions expérimentales différentes.

de fission précipités et l'effet des défauts d'irradiation. Les expressions de ces coefficients sont données dans [Bouloré et al., 2023].

Les données expérimentales disponibles de l'UO_2 vierge non irradié comportent une incertitude expérimentale, illustrées par la Figure 1.7. Sur cette Figure sont représentées quatre lots de valeurs expérimentales $(T, \lambda(T, \tau_{\text{burnup}}))$ mesurées sur une plage de variation pour la température variant entre 500 et $3000K$. Chaque lot correspond à des conditions expérimentales différentes.

Une analyse d'incertitudes détaillée de ce modèle thermique a précédemment été réalisée dans plusieurs situations expérimentales [Bouloré et al., 2010]. Une analyse de sensibilité basée sur le calcul d'indices de Sobol' du premier ordre[8] (Figure 1.8) a notamment mis en lumière **la nécessité de bien quantifier l'incertitude de la conductivité thermique, celle-ci ayant un impact significatif sur l'incertitude des températures calculées a partir de la résolution de l'équation** (1.31) **à chaque pas de temps** [Bouloré et al., 2012]. L'application de la méthode probabiliste CIRCE [Damblin and Gaillard, 2020; Cocci et al., 2022] sur les données expérimentales a permis de montrer que la loi $\lambda_0(T)$ modélisant la conductivité

[7] L'irradiation est l'exposition d'un corps à un flux de rayonnements ionisants.

[8] Mesures de sensibilité probabilistes basées sur une décomposition de la variance de la sortie d'intérêt en fonction des incertitudes de chaque entrée (pour plus de détails, voir [Saltelli et al., 2009; Sobol, 1993]). À noter que les indices du 1er ordre quantifient, dans le cas de variables d'entrée indépendantes, l'influence de chaque entrée prise séparément, sans prendre en compte leurs éventuelles interactions.

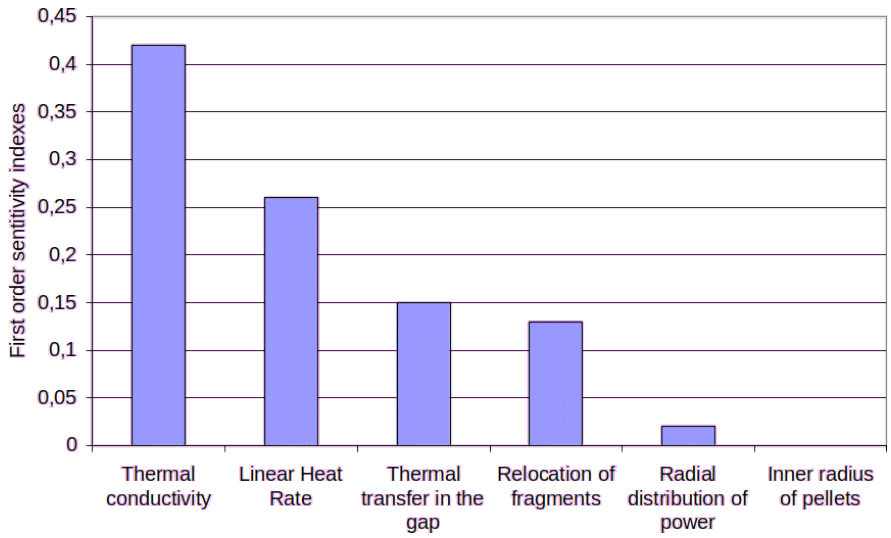


Figure 1.8: Indices de Sobol' du 1er ordre des paramètres du modèle thermique, pour la température calculée en sortie du modèle.

thermique pouvait s'écrire :

$$\lambda_0(T) = \mathbf{\Lambda}\frac{(1-p)}{(1+2p)}\frac{1}{(a+bT)}, \tag{1.36}$$

où $\mathbf{\Lambda}$ est une variable aléatoire gaussienne [Bouloré et al., 2023]. Dans la suite du document, quand on parle de la distribution de λ, il s'agit en fait par abus de langage de celle de $\mathbf{\Lambda}$. Ainsi, en propageant cette incertitude épistémique de la conductivité thermique de l'UO_2 dans le schéma de résolution du modèle thermique, nous obtenons en sortie une incertitude sur le champ de température calculé T. En somme, la résolution du modèle thermique consiste à calculer le champ de température dans le crayon combustible en tenant compte de deux grandeurs incertaines :

- la puissance d'irradiation $P(t)$ dont l'incertitude est de nature épistémique,
- la conductivité thermique du combustible λ dont l'incertitude est elle aussi de nature épistémique [Bouloré et al., 2023].

Dans le cadre des travaux de thèse, on considérera λ comme variable incertaine en entrée du modèle thermique pour une puissance d'irradiation donnée, et en régime de fonctionnement normal et permanent ($T < 1500K$).

1.2.4 Le modèle de comportement des gaz de fission

Le modèle de comportement des gaz de fission dans ALCYONE appelé CARACAS (pour CAlcul RApide du Comportement des gAz de fiSsion) représente de manière continue le comportement des produits de fission gazeux dans le matériau UO_2 en régime permanent et en régime transitoire [Jomard et al., 2014; Boulore et al., 2017; Michel et al., 2021; Marelle et al., 2016]. Tous les phénomènes physiques représentés dans ce modèle permettent d'avoir un meilleur état de répartition des gaz de fission dans le matériau avant une séquence accidentelle de type LOCA (*Loss of Coolant Accident* pour Accident de Perte de Réfrigérant Primaire) et RIA (*Reactivity Insertion Accident* pour Accident d'Insertion de Réactivité) et de calculer l'évolution de la pression interne dans le crayon combustible. Les gaz de fission produits par l'irradiation sont distribués sous forme de trois populations qui sont illustrées par la Figure 1.9 :

1. les gaz dissous dans le grain,

2. les gaz précipités dans des bulles intragranulaires,

3. les gaz précipités en bulles intergranulaires.

Les mécanismes physiques représentés à l'échelle du grain combustible sont :

- la diffusion sous forme atomique du gaz dans le grain de l'UO_2,
- la précipitation du gaz dans les bulles intra et interganulaires,
- la remise en solution du gaz des bulles dans le matériau sous l'effet des fragments de fission (remise en solution balistique),
- le mouvement des cavités sous gradient thermique (en régime transitoire).

Lorsque les bulles interganulaires du joint de grains UO_2 sont suffisamment couvrantes pour être interconnectées, nous obtenons le phénomène appelé relâchement des gaz de fission dans le volume libre du crayon combustible (Figure 1.10). Ce relâchement impacte directement la pression interne du crayon. La restructuration du combustible (structure à petits grains, milieu très poreux) se produit quand le combustible est irradié avec un fort taux de combustion. Cette restructuration a lieu en périphérie de la pastille d'où l'appellation "rim" en anglais (bord, extrémité, périphérie), parfois utilisée pour la qualifier. En situation accidentelle, cette restructuration a un fort impact sur le comportement des gaz de fission d'où sa prise en compte dans le modèle CARACAS. Les gaz initialement présents dans les bulles intragranulaires sont transférés progressivement dans les bulles de "rim" et la restructuration d'un grain UO_2 se fait de manière progressive en fonction du taux de combustion (appelé *burnup* ou *burn rate*) (Figure 1.11).

Le modèle de comportement de gaz de fission calcule ainsi différentes grandeurs en sortie :

- les déformations de gonflement (volume des bulles, gonflement solide, variations dimensionnelles, ...),
- les fractions de gaz relâchées dans le volume libre notées RGF (pour *released gas fraction*) par la suite,
- la distribution du gaz dans les différentes populations.

Tous ces phénomènes (Figures 1.9, 1.10 et 1.11) sont thermiquement activés. Autrement dit, toutes les grandeurs de sortie du modèle de comportement de gaz de fission dépendent de la température calculée par le modèle thermique présenté plus haut. Par conséquent, **ces grandeurs de sortie dépendent également de la conductivité thermique λ du combustible UO_2. Parmi ces grandeurs, dans le cadre de l'étude présente, seule la fraction de gaz relâchée notée RGF (RGF pour *Released Gas Fraction*) est considérée : elle constitue la quantité d'intérêt au centre des travaux de cette thèse.** Toutes les modélisations embarquées dans **le modèle CARACAS comportent des paramètres dont certains ne sont pas mesurables (paramètres de conception $\boldsymbol{\theta} \in \mathscr{D}_{\boldsymbol{\Theta}}$) et dont la valeur doit être ajustée afin que le modèle puisse représenter au mieux ces phénomènes**. Dans des études précédentes [Struzik and Boulore, 2010; Boulore et al., 2009], l'incertitude calculée sur le champ de température dans le combustible a été propagée directement dans le modèle de gaz de fission, aboutissant à une incertitude finale sur la RGF très importante. **Cette étude a ainsi montré la nécessité, dans le cadre de la quantification d'incertitudes de calcul, de faire dépendre les paramètres du modèle de gaz $\boldsymbol{\theta}$ à la conductivité thermique λ**. En effet, l'incertitude épistémique de la conductivité thermique n'a pas d'impact sur les mesures expérimentales de RGF. Autrement dit, si la conductivité change en entrée du modèle thermique, les paramètres $\boldsymbol{\theta}$ du modèle de comportement doivent être recalés en conséquence : on parle alors de **calage conditionnel**.

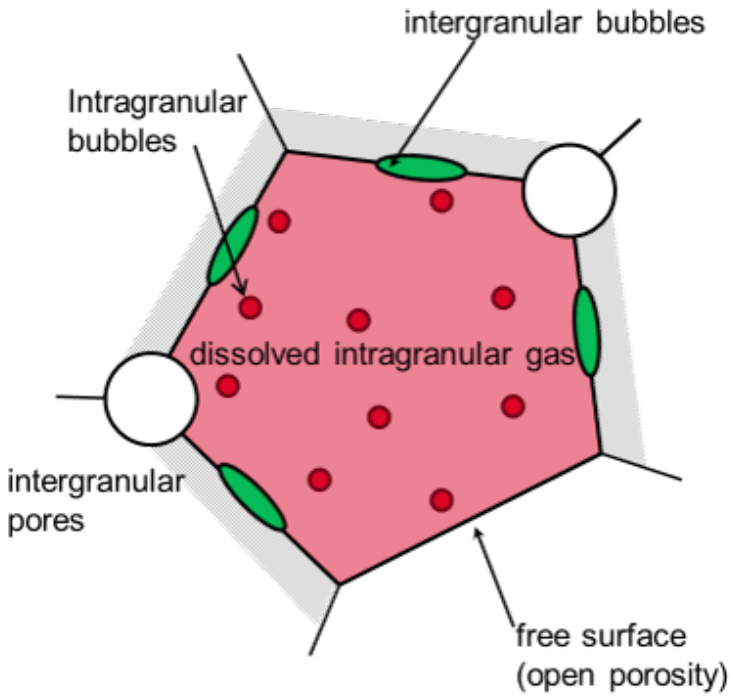


Figure 1.9: Les différentes populations de gaz de fission représentées dans CARACAS.

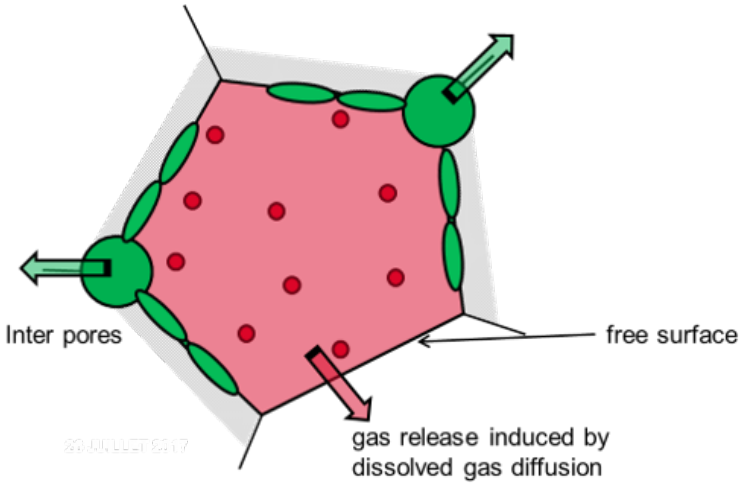


Figure 1.10: Relâchement des gaz de fission.

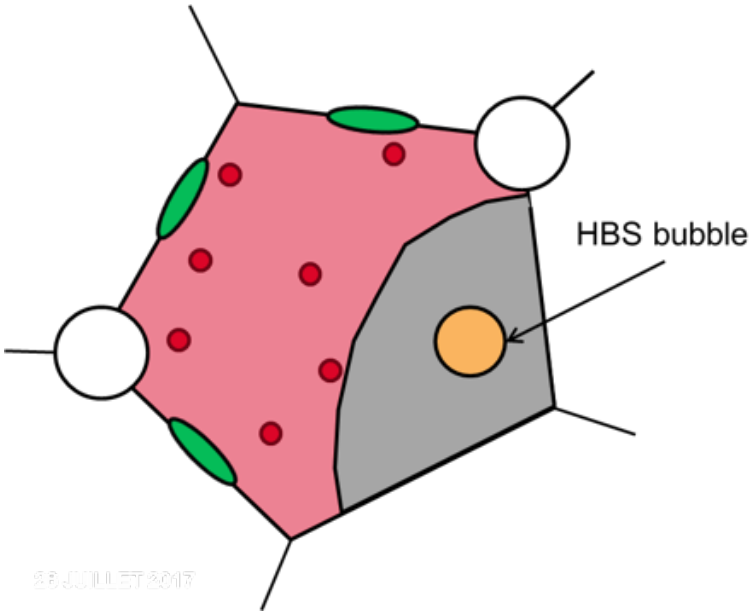


Figure 1.11: Modélisation de la restructuration du combustible (High Burnup Structure).

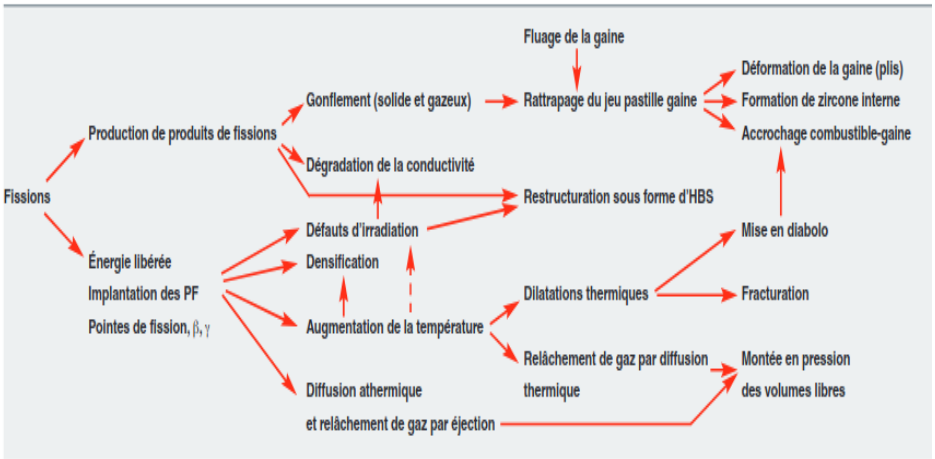


Figure 1.12: Résumé des différents phénomènes intervenant dans le combustible REP sous irradiation et leurs interactions [Guillet et al., 2008].

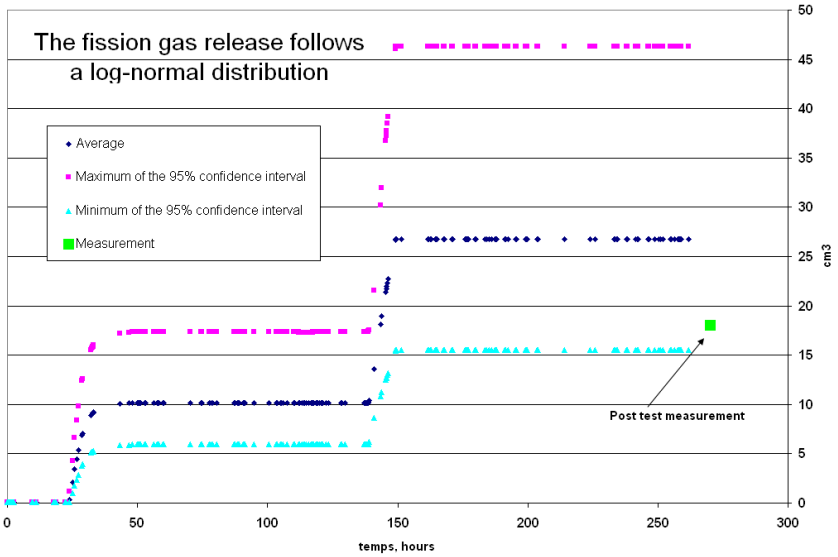


Figure 1.13: Incertitude de prédiction sur la RGF.

Notons que **lorsque l'on veut calculer la RGF (simulée) d'un crayon combustible, cela nécessite l'exécution de l'ensemble du code ALCYONE qui appelle dans sa boucle de convergence le modèle CARACAS dont les paramètres sont $\boldsymbol{\theta}$ (voir Figure 1.5). C'est pourquoi dans la suite de ce présent document, par abus de language, lorsque l'on mentionne le résulat du code CARACAS, il s'agit en fait du résultat de calcul ALCYONE-CARACAS.**

1.2.5 Enjeux et objectifs du calage conditionnel dans le contexte applicatif

L'objectif général du calage conditionnel est de **déterminer la distribution de $\boldsymbol{\theta}$ conditionnellement à l'incertitude de la conductivité thermique λ du combustible**. En effet, lorsque l'on propage l'incertitude de la conductivité thermique λ dans le schéma de calcul (Figure 1.5) indépendamment de l'incertitude de $\boldsymbol{\theta}$, l'incertitude sur la RGF prédite par CARACAS est doublée (Figure 1.13). Par conséquent, il est nécessaire de propager l'incertitude de λ et l'incertitude de $\boldsymbol{\theta}$ conjointement. Cette propagation conjointe devra tenir compte d'une **hypothèse physique** qui stipule que les mesures des RGF ne sont pas informatives sur l'incertitude de la conductivité thermique λ. Pour la simulation, cela induit que **quelle que soit la valeur de la conductivité thermique λ échantillonnée selon sa dsitribution de probabilité de $\boldsymbol{\Lambda}$, la prédiction RGF du code ALCYONE devrait rester inchangée**. Pour y parvenir, **l'effet potentiel de la conductivité thermique λ sur la RGF simulée doit être annihilé en recalant le paramètre $\boldsymbol{\theta}$ du modèle CARACAS conditionnellement à λ.** Il s'agit **d'un effet de compensation qui est attendu.** La Figure 1.14 montre un exercice de calage déterministe de $\boldsymbol{\theta}$ sur une base de calibration réduite, pour 3 valeurs différentes de λ. On obtient bien des prédictions globalement similaires pour la plupart des RGF des crayons de cette base réduite.

En conclusion, pour pouvoir propager les incertitudes du modèle thermique (ici uniquement l'incertitude de la conductivité thermique) dans le modèle de gaz de fission, il faut déterminer la distribution de $\boldsymbol{\theta}$ conditionnellement à la valeur de λ sur l'ensemble du domaine d'incertitude de λ (support de $\boldsymbol{\Lambda}$). Cela vise à représenter de manière aussi précise que possible les mesures expérimentales RGF tout en respectant l'hypothèse de compensation. Ce processus devra de plus prendre en compte l'incertitude sur ces mesures expérimentales qui est de l'ordre de 6%.

Dans ce contexte, le premier objectif de cette thèse est de proposer une méthodologie pour le calage conditionnel dans le cadre d'une inférence bayésienne. Pour ce faire, inspiré par

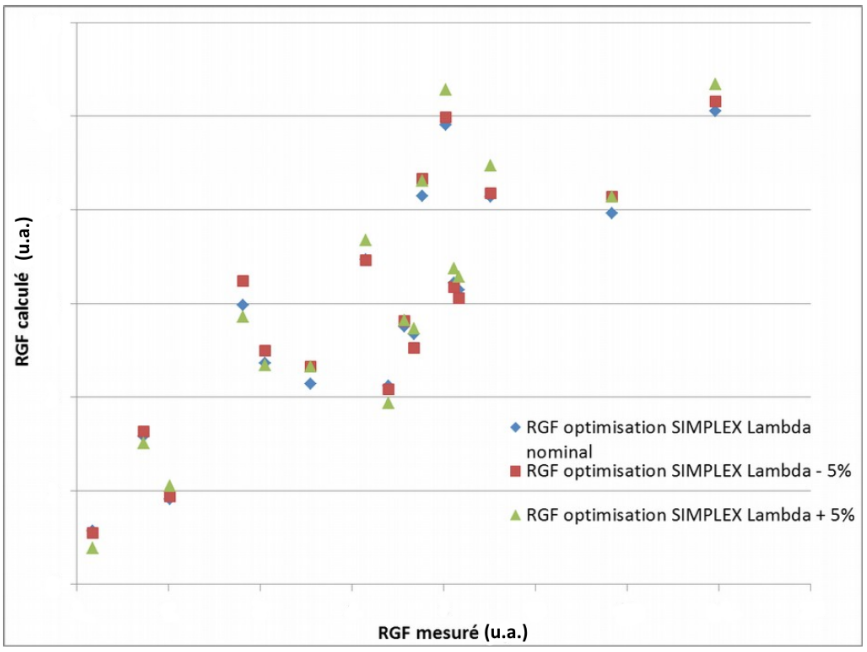


Figure 1.14: RGF mesurée versus RGF calculée pour 3 réalisations de λ.

les travaux de Brown and Atamturktur [2018], nous avons proposé **une méthode appelée GP-LinCC** (pour *Gaussian Process and Linearization-based Conditonal Calibration*) **qui apprend la relation fonctionnelle entre $\boldsymbol{\theta}$ et λ à partir de la distribution conditionnelle *a posteriori* $\pi(\boldsymbol{\theta} \mid \boldsymbol{z}, \lambda)$ dans le cas où la sortie du modèle CARACAS est supposée linéaire par rapport à la fonction de calage $\boldsymbol{\theta}(\lambda)$.** Cette étude fait l'objet du Chapitre 2. **Le second objectif est de proposer une méthode d'analyse de sensibilité a priori en amont du calage conditionnel bayésien.** Nous avons proposé une méthode d'analyse de sensibilité globale basée sur les indices de sensibilité HSIC [Da Veiga, 2015; Meynaoui, 2019; El Amri and Marrel, 2024] pour sélectionner les paramètres à caler. Ces travaux font l'objet du Chapitre 3.

1.2.6 Organisation du document

Dans le but de répondre à la problématique du calage conditionnel dans le contexte de l'application en simulation du combustible, le présent manuscrit est organisé comme suit. Le Chapitre 2 présente tout d'abord une nouvelle méthodologie bayésienne proposée pour répondre à la problématique du calage conditionnel bayésien. Comme cela sera illustré, cette nouvelle approche, appelée GP-LinCC, a montré de bonnes performances sur des exemples analytiques en petite dimension. Cependant son déploiement sur le cas applicatif ne peut pas se faire directement. Il nécessite un travail préliminaire de réduction de la dimension, c'est-à-dire du nombre de paramètres à caler. Pour cela, nous avons développé une méthode d'analyse de sensibilité globale adaptée à la problématique du calage conditionnel. Cette méthode d'analyse de sensibilité globale, basée sur les indices HSIC, fait l'objet du Chapitre 3. Enfin, dans le Chapitre 4, la méthode GP-LinCC est mise en œuvre sur le cas applicatif afin de quantifier les incertitudes des paramètres incertains du modèle de comportement des gaz de fission (appelé aussi modèle CARACAS) conditionnellement à l'incertitude de la conductivité thermique du modèle thermique. En conclusion, le Chapitre 5 présente un résumé des deux contributions développées dans ce document pour le calage bayésien des paramètres de modèles numériques chaînés. Des pistes d'améliorations de ces travaux sont également présentées.

Chapter 2

Conditional Bayesian calibration: GP-LinCC approach

Contents

Abstract

Nowadays, numerical models are widely used in most of engineering fields to simulate the behavior of complex systems, such as for example power plants or wind turbine in the energy sector. Those models are nevertheless affected by uncertainty of different nature (numerical, epistemic) which can affect the reliability of their predictions. We develop here a new method for quantifying conditional parameter uncertainty within a chain of two numerical models in the context of multi-physics simulation. More precisely, we aim to calibrate the parameters $\boldsymbol{\theta}$ of the second model of the chain conditionally on the value of parameters λ of the first model, while assuming the probability distribution of λ is known. This conditional calibration is carried out from the available experimental data of the second model. In doing so, we aim to quantify as well as possible the impact of the uncertainty of λ on the uncertainty of $\boldsymbol{\theta}$. To perform this conditional calibration, we set out a nonparametric Bayesian formalism to estimate the functional dependence between $\boldsymbol{\theta}$ and λ, denoted $\boldsymbol{\theta}(\lambda)$. First, each component of $\boldsymbol{\theta}(\lambda)$ is assumed to be the realization of a Gaussian process prior. Then, if the second model is written as a linear function of $\boldsymbol{\theta}(\lambda)$, the Bayesian machinery allows us to compute analytically the posterior predictive distribution of $\boldsymbol{\theta}(\lambda)$ for any set of realizations λ. The effectiveness of the proposed method is illustrated on several analytical examples.

Keywords. Conditional Bayesian calibration, nonparametric regression, Gaussian process, empirical Bayes, cut-off models.

2.1 Introduction

Over the last thirty years, numerical models have become important tools for modeling, understanding, analyzing and predicting the physical phenomena studied. In nuclear engineering, such models are essential because the physical experiments are often limited or even impossible for economic or ethical reasons (e.g., simulation of accidental transients for safety demonstration). The models, implemented to represent the physical reality faithfully, may depend on the specification of a large number of calibration parameters characterizing the studied phenomenon which are most often uncertain. Such parameters uncertainties are often due to a lack of knowledge of the phenomenon and, as a consequence, to some modeling assumptions. The process of quantifying parameter uncertainty according to the available experimental data is called model calibration. There are basically two types of calibration: deterministic calibration and calibration under uncertainties using Bayesian framework (Bayesian calibration) [Kennedy and O'Hagan, 2001; Campbell, 2006; Damblin, 2015; Santner et al., 2018; Wu et al., 2018; Willmann et al., 2022]. Bayesian calibration allows us to quantify the parameter uncertainty by probability distribution instead of a single optimal value as in deterministic calibration. Such an optimal value is generally obtained by minimizing a criterion which quantifies a difference between the simulated data and the available experimental data (e.g., least-square criterion) [Tuo and Wu, 2015; Wong et al., 2017].

This paper focuses on Bayesian calibration of models in the framework of multi-physics simulation. Multi-physics simulation refers to several numerical models of different physics which are connected to one another to simulate the entire phenomenon of interest. For example, in fuel simulation for nuclear power plants, the ALYCONE application is a multi-physics application [Michel et al., 2021]. It is composed of interdependent physical models that represent the mechanical, thermal and chemical behaviors of fuel rods in the core of pressurized water reactors. Among these different models, we will focus here on the thermal and the fission gas behavior models. The thermal model simulates the evolution of the temperature within the fuel rod during the fission reaction and provides as output, the associated temperature field. The fission gas behavior model, as a function of the thermal model, continuously represents the behavior of the fission products (fuel swelling and release of fission gas atoms) during the fission reaction. Figure 1.5 represents the chaining of these two models and their integration within the global calculation workflow. These models represented by blue boxes have as input parameters, respectively, the thermal conductivity $\lambda \in \mathscr{D}_{\boldsymbol{\Lambda}} \subseteq \mathbb{R}$ and the parameters of the fission gas behavior model denoted $\boldsymbol{\theta} \in \mathscr{D}_{\boldsymbol{\Theta}} \subset \mathbb{R}^p$ $(p \geq 1)$.

In the literature, two methods stand out for quantifying parametric uncertainties in interconnected models. First, we can independently calibrate $\boldsymbol{\theta}$ of Model 2 (fission gas behavior model in Figure 1.5) and the parameter λ of Model 1 (thermal model in Figure 1.5) using experimental data from each phenomenon separately. Alternatively, we can opt for joint calibration using the whole set of data as proposed in [Marque-Pucheu et al., 2016], then enabling us to derive of conditional distributions, like that of $\boldsymbol{\theta}$ given λ. In this work, our focus is only on the conditional calibration of $\boldsymbol{\theta}$ based on a well-known posterior distribution of λ, presumed accurate enough, thus eliminating the need for its reestimation. Therefore, the approaches presented in [Marque-Pucheu et al., 2016] are not suitable in our context. Indeed, we aim to estimate a conditional posterior probability density of $\boldsymbol{\theta}|\lambda$ using experimental measurements of Model 2 (fission gas behavior model in our application context). Unless specific modelling assumptions, such a posterior probability density is only known up to a constant, so Markov chain Monte Carlo (MCMC) methods are required for sampling it [Robert et al., 1999]. For the purpose of computing the probability distribution of $\boldsymbol{\theta}|\lambda$, a naive approach would be to run as many independent Markov chains as the number of λ samples of interest. This approach is computationally expensive and omits that, under some regularity conditions, the conditional distribution $\boldsymbol{\theta}|\lambda$ may give some informations about $\boldsymbol{\theta}|\lambda'$ if λ is not too far from λ'.

To overcome the above drawback, we propose a nonparameteric approach which directly learn the functional relation $\boldsymbol{\theta}(\lambda)$ and infers its probability distribution. This approach is inspired by the work of Brown and Atamturktur [2018] presented in another calibration context. The functional approach is based on the assumption that each component of $\boldsymbol{\theta}(\lambda)$ is represented a priori as a trajectory of an independent Gaussian process. Moreover, it is implemented in the special case where the output of Model 2 is assumed to be linear in $\boldsymbol{\theta}$ given λ. The proposed method eventually provides a posterior predictive probability distribution of $\boldsymbol{\theta}$ conditionally on any λ and its feasibility is studied on analytical examples.

This paper is divided into five sections. In Section 2.2, we present a short state of the art of Bayesian calibration of numerical models. Section 2.3 is devoted to possible methods for conditional density estimation. In Section 2.4, we detail our approach, called GP-LinCC for Gaussian Process and Linearization-based Conditional Calibration, and whose performances are illustrated in Section 2.5. Section 2.6 concludes the paper.

2.2 Bayesian calibration of numerical models

A physical system of interest $r(\boldsymbol{x}) \in \mathscr{D}_{\boldsymbol{Y}} \subset \mathbb{R}^q$ (here $q = 1$) can be seen as a function describing the relationship between the input x and the output of interest $r(\boldsymbol{x})$. The vector $\boldsymbol{x} \in \mathscr{D}_{\boldsymbol{X}} \subset \mathbb{R}^d$ is most often constituted of variables called control variables which typically represent experimental conditions and geometry of the physical system. Let $y_{\boldsymbol{\theta},\lambda}(\boldsymbol{x}) := y^2_{\boldsymbol{\theta}}(y^1_{\lambda}(\boldsymbol{x}))$ be a deterministic numerical model resulting from the chaining of two physical models (like in Figure 1.5) and supposed to be representative of $r(\boldsymbol{x})$. Inspired by the work of Kennedy and O'Hagan [2001], the experimental measurements $\boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_{n_{\text{exp}}})^t$ are related to the model outputs at the input values $\{\boldsymbol{x}_i\}_{1 \leq i \leq n_{\text{exp}}}$ by the following general probabilistic equation:

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta},\lambda}(\boldsymbol{x}_i) + b(\boldsymbol{x}_i) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}, \tag{2.1}$$

where ϵ_i is the experimental uncertainty often considered as a realization of a zero-mean Gaussian distribution and the variance $\sigma^2_{\epsilon_i}$ is known. The unobserved term $b(\boldsymbol{x})$, called model discrepancy in [Kennedy and O'Hagan, 2001], represents the gap between the numerical model $y_{\boldsymbol{\theta},\lambda}(\boldsymbol{x})$ and the physical system $r(\boldsymbol{x})$ when the model is run at the optimal value (or true) but unknown $(\boldsymbol{\theta}, \lambda)$ [1] of the uncertain parameters. The term $b(\boldsymbol{x})$ is often modeled

[1] Optimal value in the sense that the model run with $(\boldsymbol{\theta}, \lambda)$ gives the best possible prediction accuracy. Note that the best parameter can be different from the true parameter [Kennedy and O'Hagan, 2001; Wu et al., 2018].

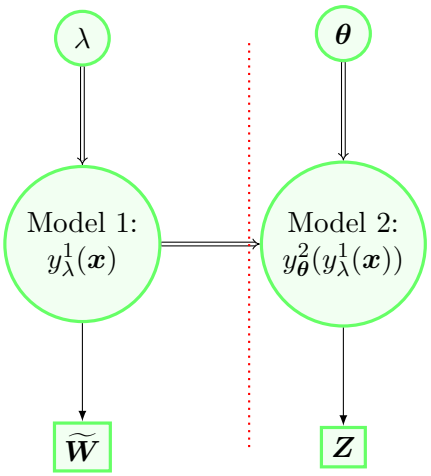


Figure 2.1: Graphical representation of a cut-off model for two chained physical models. Single and double arrows give respectively stochastic and logical dependences. Square nodes are used for observed data.

by a Gaussian process [Kennedy and O'Hagan, 2001; Bachoc et al., 2014]. In the literature dedicated to Bayesian calibration, the term $b(\boldsymbol{x})$ is sometimes neglected and thus assumed to be indistinguishable from the experimental uncertainty [Damblin et al., 2018]. Similarly, we will instead consider the simplified probabilistic equation:

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta},\lambda}(\boldsymbol{x}_i) + \epsilon_i,\ 1 \leq i \leq n. \tag{2.2}$$

The joint posterior distribution is obtained by using Bayes' formula:

$$\pi_{\text{full}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{z}) \propto \mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta}, \lambda)\pi(\boldsymbol{\theta}, \lambda), \tag{2.3}$$

where $\pi(\boldsymbol{\theta}, \lambda)$ is the prior distribution quantiying the uncertainty of $(\boldsymbol{\theta}, \lambda)$ before collecting the data $\boldsymbol{z}$, $\mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta}, \lambda)$ is the likelihood of the data $\boldsymbol{z}$ conditionally on the couple $(\boldsymbol{\theta}, \lambda)$ and the posterior distribution $\pi_{\text{full}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{z})$ quantifies the residual uncertainty of $(\boldsymbol{\theta}, \lambda)$ conditionally on $\boldsymbol{z}$.

Sometimes, as in fuel simulation, there may be other sources of data which can bring information on some components of $(\boldsymbol{\theta}, \lambda)$. In such cases, cut-off models can be used to properly partition the different sources of data that can provide information on the model parameters involved in the chaining [Liu et al., 2009; Plummer, 2015; Jacob et al., 2017]. Inspired by the work of Plummer [2015], Figure 2.1 presents a cut-off model for the two chained physical models where the direct measurements $\boldsymbol{w}$ (a realization of a some random variable $\widetilde{\boldsymbol{W}}$) bring information on λ whereas the data $\boldsymbol{z}$ (a realization of random variable $\boldsymbol{Z}$) are seen by experts as uninformative on this parameter. More precisely, in Figure 2.1, the graph is divided by a cut into two subgroups (left and rigth of the dotted red line). The posterior distribution of λ of the left subgroup is computed without considering the random variables of the right subgroup. Thus, the estimation of this posterior distribution is done only with the observation $\boldsymbol{w}$ of $\widetilde{\boldsymbol{W}}$ despite its possible dependence on $\boldsymbol{Z}$. On the opposite, when estimating the posterior distribution of $\boldsymbol{\theta}$ of the right subgroup, the terms of the left subgroup are taken into account. Then, we can write the probability distribution of the parameters $(\boldsymbol{\theta}, \lambda)$ conditionally on the complete data $(\widetilde{\boldsymbol{W}} = \boldsymbol{w}, \boldsymbol{Z} = \boldsymbol{z})$ as:

$$\pi_{\text{cut}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z}) = \pi(\boldsymbol{\theta} \mid \lambda, \boldsymbol{z})\pi(\lambda \mid \boldsymbol{w}), \tag{2.4}$$

where $\pi(\lambda \mid \boldsymbol{w}) \propto \mathscr{L}(\boldsymbol{w} \mid \lambda)\pi(\lambda)$ is the posterior distribution when $\boldsymbol{w}$ is observed only and $\pi(\boldsymbol{\theta}|\lambda, \boldsymbol{z})$ is the posterior distribution of $\boldsymbol{\theta}$ conditionally on λ. The distribution $\pi_{\text{cut}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z})$, called the cut distribution in [Plummer, 2015], is different from the true joint posterior

probability distribution $\pi_{\text{full}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z})$ which is written as:

$$\pi_{\text{full}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z}) = \pi(\boldsymbol{\theta}|\lambda, \boldsymbol{z})\pi(\lambda \mid \boldsymbol{w}, \boldsymbol{z}). \tag{2.5}$$

The two distributions are linked by the following relation:

$$\pi_{\text{cut}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z}) \propto \frac{\pi_{\text{full}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z})}{\pi(\boldsymbol{z}|\lambda)}. \tag{2.6}$$

Indeed, we have:

$$\pi_{\text{full}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z}) = \frac{\mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta}, \lambda)\pi(\boldsymbol{\theta} \mid \lambda) \times \mathscr{L}(\boldsymbol{w}|\lambda)\pi(\lambda)}{\pi(\boldsymbol{w}, \boldsymbol{z})} \tag{2.7}$$

$$= \frac{\pi(\boldsymbol{\theta} \mid \lambda, \boldsymbol{z})\pi(\boldsymbol{z}|\lambda) \times \pi(\lambda \mid \boldsymbol{w})\pi(\boldsymbol{w})}{\pi(\boldsymbol{w}, \boldsymbol{z})} \tag{2.8}$$

$$= \pi_{\text{cut}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{w}, \boldsymbol{z}) \times \frac{\pi(\boldsymbol{z} \mid \lambda)}{\pi(\boldsymbol{z} \mid \boldsymbol{w})}. \tag{2.9}$$

In the sequel, the posterior distribution $\pi(\lambda \mid \boldsymbol{w})$ is assumed to be well-known and thus our goal comes down to estimating only the conditional posterior distribution $\pi(\boldsymbol{\theta} \mid \lambda, \boldsymbol{z})$ in Equation (2.4). In the next section, we present some potential methods to achieve this goal.

2.3 Approaches for conditional density estimation

We aim to estimate the distribution of $\boldsymbol{\theta}$ conditionally on any λ varying in the support of $\pi(\lambda|w)$ and deduce an associated estimator for $\boldsymbol{\theta}(\lambda)$. At first glance, one could use a kernel density estimator (KDE) [Hastie et al., 2009; Murphy, 2022] to approximate the conditional posterior distribution $\pi(\boldsymbol{\theta}|\lambda, \boldsymbol{z})$. This density estimator would require a large number N_{mcmc} of samples $\{\boldsymbol{\theta}_i, \lambda_i\}_{i=1}^{N_{\text{mcmc}}}$ of $\pi_{\text{cut}}(\boldsymbol{\theta}, \lambda|\boldsymbol{w}, \boldsymbol{z})$ which could be generated by a Gibbs algorithm with possibly an inner Metropolis step (also called Metropolis-within-Gibbs) [Robert et al., 1999] as follows:

$$\boldsymbol{\theta}_i \sim \pi_{\text{cut}}(\boldsymbol{\theta}, \lambda_{i-1} \mid \boldsymbol{w}, \boldsymbol{z}) \propto \pi_{\text{full}}(\boldsymbol{\theta}, \lambda_{i-1} \mid \boldsymbol{w}, \boldsymbol{z}) \text{ (Equation (2.6))}, \tag{2.10}$$

$$\lambda_i \sim \pi_{\text{cut}}(\lambda, \boldsymbol{\theta}_i \mid \boldsymbol{w}, \boldsymbol{z}) \propto \frac{\pi_{\text{full}}(\boldsymbol{\theta}_i, \lambda \mid \boldsymbol{w}, \boldsymbol{z})}{\pi(\boldsymbol{z} \mid \lambda)},\ 1 \leq i \leq N_{\text{mcmc}}. \tag{2.11}$$

However, in the general case, the marginal likelihood in the denominator of Equation (2.11) is computationally intractable

$$\pi(\boldsymbol{z} \mid \lambda) = \int \mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta}, \lambda)\pi(\boldsymbol{\theta} \mid \lambda)d\boldsymbol{\theta}. \tag{2.12}$$

As a result, the above sampling scheme is unfeasible because we are not able to sample the λ_i according to Equation (2.11). Therefore, KDE-based appraoch could not be used with this sampling scheme to approximate $\pi(\boldsymbol{\theta} \mid \lambda, \boldsymbol{z})$. Another solution may be for each λ drawn in $\pi(\lambda \mid \boldsymbol{w})$, to use a Metropolis-Hastings algorithm (MH) [Chib and Greenberg, 1995] to approximate the associated conditional posterior distribution. More precisely, we would have the following MCMC sampling scheme:

$$\lambda_i \sim \pi(\lambda \mid \boldsymbol{w}), \tag{2.13}$$

$$\boldsymbol{\theta}_i \sim \pi(\boldsymbol{\theta} \mid \lambda_i, \boldsymbol{z}),\ 1 \leq i \leq N_{\text{mcmc}}. \tag{2.14}$$

However, if the numerical model is computationally expensive, this MH-based sampling scheme could not be repeated for a large number of different λ. This is why we can envisage to carry out this scheme by considering only a limited number m of well-chosen values of λ, then using them to build a regression model to extrapolate the conditional distribution to another values of λ. This is presented hereafter.

2.3.1 Moment-based estimation method

We begin by creating a numerical design for λ, $D_m := (\lambda_1, \cdots, \lambda_m)^t$ of size m by Latin hypercube sampling [McKay et al., 1979] to spread out the samples in the support of the distribution $\pi(\lambda \mid \boldsymbol{w})$. Then, we estimate independently each distribution $\pi(\boldsymbol{\theta} \mid \lambda_j, \boldsymbol{z})$ for $j = 1, \cdots, m$. Finally, we interpolate by a Gaussian process (GP) regression [Rasmussen et al., 2006] the first two moments of the conditional posterior distribution of $\boldsymbol{\theta}$ with respect to D_m in order to predict them at new realizations $\lambda^\star \notin D_m$. More precisely, the steps are the following:

(*i*) Creation of a numerical design D_m according to $\pi(\lambda \mid \boldsymbol{w})$,

(*ii*) Computation of the m conditional posterior distributions $\pi(\boldsymbol{\theta} \mid \lambda_j, \boldsymbol{z})$ by running a MH algorithm for each $1 \leq j \leq m$. Let $(\boldsymbol{\theta}_i^j)_{i=1}^N$ be the resulting set of samples associated with λ_j and generated from $\pi(\boldsymbol{\theta} \mid \lambda_j, \boldsymbol{z})$. The expectation and covariance matrix are then estimated by

$$\bar{\boldsymbol{\theta}}(\lambda_j) := \mathbb{E}(\boldsymbol{\theta} \mid \lambda_j, \boldsymbol{z}) \approx \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\theta}_i^j, \tag{2.15}$$

$$\mathrm{cov}(\boldsymbol{\theta} | \lambda_j, \boldsymbol{z}) \approx \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{\theta}_i^j - \bar{\boldsymbol{\theta}}(\lambda_j))(\boldsymbol{\theta}_i^j - \bar{\boldsymbol{\theta}}(\lambda_j))^t. \tag{2.16}$$

(*iii*) Fitting of a GP model on the quantities (2.15) and (2.16) respectively based on their estimated values on D_m. From the two first moments learnt by the two GP, a Gaussian assumption is finally used to build a predictive distribution associated with any new realization $\lambda^\star \notin D_m$.

The main difficulty of Step (*iii*) lies in preserving the positive semi-definite property of covariance matrices in the GP regression model. There are some solutions to satisfy this property:

- for $p = 1$, GP regression on the log variance can be considered,
- for $p > 1$, the solution proposed by Fiszeder and Orzeszko [2021] is to first carry out a regression on the Cholesky factors and then to use the inverse Cholesky decomposition to obtain predicted matrices ensuring the positive semi-definite property. However, this method is rather complex and omits any assessment of the prediction uncertainty.

To circumvent the use of an MH algorithm for estimating $\pi(\boldsymbol{\theta}(\lambda) \mid \boldsymbol{z}) := \pi(\boldsymbol{\theta} | \lambda, \boldsymbol{z})$ in Step (*ii*), we assume in the remainder of the paper that the output of the numerical model $y_{\boldsymbol{\theta},\lambda}(x)$ is linear in $\boldsymbol{\theta}$ conditionally on λ. Regarding the fuel application which constitutes the applicative framework, the available simulations of the fission gas behavior model show that this linearity assumption is reasonable. With this assumption, Steps (*i*) and (*ii*) are easier to implement because MH sampling is not required anymore. However, the problem of interpolating covariance matrices remains. This is why we propose in the next section a new method for the approximation of $\pi(\boldsymbol{\theta}(\lambda) | \boldsymbol{z})$ still under the framework of the linear assumption.

2.3.2 Proposed solution: method based on GP-prior and linear assumption (GP-LinCC method)

We aim to estimate the conditional posterior distribution $\pi(\boldsymbol{\theta} | \boldsymbol{z}, \lambda)$ for any realization λ. Our approach is to learn the relationship between $\boldsymbol{\theta}$ and λ through the function $\boldsymbol{\theta}(\lambda)$ (called the calibration function) using the Bayesian paradigm based on three assumptions: a Gaussian process prior, the linearity of the output of the numerical model in $\boldsymbol{\theta}(\lambda)$ and the compensation hypothesis that will be presented later (see Subsection 2.3.2.3).

2.3.2.1 Gaussian Process prior

Among the metamodels classically used for approximating computationally intensive numerical models, the GP regression is a powerful tool for nonparametric function metamodeling [Sacks et al., 1989; Marrel et al., 2008]. The theory of GP is extensively detailed in [Rasmussen et al., 2006] and more recently in [Gramacy, 2020]. Choosing it as a prior class of random functions (characterized by its mean and covariance functions) and conditioning on observed data yields a posterior GP which delivers a Gaussian predictive distribution for the model output at each prediction point with anaytical formulas for mean and covariance matrix [Currin et al., 1991] (see Appendix 2.7.1.4). GP regression has the property of providing a predictive distribution (resulting in a probabilistic metamodel) rather than a simple predictor, largely explains its popularity, particularly in risk and safety assessment applications. Indeed, an uncertainty and credibility intervals can be derived for any quantity of interest estimated from the GP predictive distribution: estimation of a quantile, probability [Marrel et al., 2022] or sensitivity indices [Marrel et al., 2009; Le Gratiet et al., 2014], performing an inversion problem [Chevalier et al., 2014], or construction of a sequential design of experiments [Bect et al., 2012; Gramacy, 2020]. These properties have also been successfully used to emulate several chained (also called nested) or coupled codes, as in [Marque-Pucheu et al., 2016, 2017] for calibration purpose.

In our case, we are quantifying the relationship between $\boldsymbol{\theta}$ and λ by learning the function $\boldsymbol{\theta}(\lambda) \in \mathbb{R}^p$. We suppose each component of $\boldsymbol{\theta}(\lambda)$ follows an independent GP which can be written as

$$\boldsymbol{\theta}_u(\lambda) \overset{\text{indep}}{\sim} \mathscr{GP}(m_{\beta_u}(\lambda),\, \sigma_u^2 K_{\psi_u}(\lambda,\, \lambda')),\ 1 \leq u \leq p, \tag{2.17}$$

where $m_{\beta_u}(\lambda)$ is the mean function (also called trend) of the u-th GP which has to be specified or estimated. A constant $m_{\beta_u}(\lambda) = \beta_u$ or a one-degree polynomial trend is usually considered and recommended in practice. But any linear regression model on a set of known basis functions could be used instead. For simplicity, we will assume in the sequel that the prior mean is a constant β_u. The covariance function $\sigma_u^2 K_{\psi_u}(\lambda,\, \lambda')$ controls the regularity and scale of the trajectories of the GP (ψ_u is the correlation length and σ_u^2 is the variance parameter). The covariance function $\sigma_u^2 K_{\psi_u}(\lambda,\, \lambda')$ is positive semi-definite and encodes the dependence structure of the u-th GP. One of the most popular choice of the covariance function is the Matérn 5/2 covariance function, as recommended in [Gu et al., 2018], among others. The Matérn 5/2 covariance function is given by:

$$\sigma_u^2 K_{\psi_u}(\lambda,\, \lambda') = \sigma_u^2 \left(1 + \sqrt{5}\frac{|\lambda - \lambda'|}{\psi_u} + \frac{5}{3}\frac{|\lambda - \lambda'|^2}{\psi_u}\right) e^{-\sqrt{5}\frac{|\lambda-\lambda'|}{\psi_u}}, \tag{2.18}$$

In the multidimensional case (i.e., $\lambda \in \mathscr{D}_{\boldsymbol{\Lambda}} \subseteq \mathbb{R}^s$, $s \geq 1$), we can use a tensorization of 1D-Matérn 5/2 covariance function.

2.3.2.2 Linear case

The interest for a linear framework is that the posterior distribution of $\boldsymbol{\theta}$ conditionally on λ can be computed explicitly without any use of MCMC algorithms. We will consider that for any λ, the output of the numerical model $y_{\boldsymbol{\theta},\lambda}(\boldsymbol{x}_i)$ is linear in $\boldsymbol{\theta}(\lambda)$. Thus, Equation (2.2) becomes:

$$\boldsymbol{z}_i = g_{\lambda,0}(\boldsymbol{x}_i) + g_{\lambda,1}(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}, \tag{2.19}$$

where $\epsilon_i \sim \mathscr{N}(0, \sigma_{\epsilon_i}^2 + \widehat{\sigma}_i^2)$ with $\widehat{\sigma}_i^2$ being an extra scale parameter modelling the possible linearization error. In practice, the coefficients $g_\lambda(\boldsymbol{x}_i) := (g_{\lambda,0}(\boldsymbol{x}_i),\, g_{\lambda,1}(\boldsymbol{x}_i))$ are often unknown and must be estimated by linearizing the code in $\boldsymbol{\theta}(\lambda)$ at a fixed $(\lambda, \boldsymbol{x}_i)$. To perform the linearization, we fit for each $\lambda_j \in D_m$ $(1 \leq j \leq m)$ and for each $\boldsymbol{x}_i$, a linear regression from a set of training samples. More precisely, for each λ_j, a n-size random sample of $\boldsymbol{\theta}$ denoted $\boldsymbol{\Theta}_n^{(j)}$

is generated and the corresponding simulations $y_{\boldsymbol{\Theta}_n^{(j)},\lambda_j}(\boldsymbol{x}_i)$ are run. In the end, a total of $m \times n_{\text{exp}} \times n$ simulations are performed. Each set of the n_{exp} regression coefficients of these m different linear regressions will constitute the estimates of the coefficients $\{g_{\lambda_j}(\boldsymbol{x}_i)\}_{1\leq i\leq n_{\text{exp}}}$ for each $\lambda_j \in D_m$ $(1 \leq j \leq m)$.

2.3.2.3 Compensation hypothesis

In the context of the fuel application considered and introduced in Section 2.1, the compensation hypothesis is considered by experts as a necessary condition to address the problem of conditional calibration in a consistent way. This hypothesis states that the experimental data $\boldsymbol{z}$ of the physical quantities of interest provide negligible information on the uncertainty of λ (compared to that brought by $\boldsymbol{w}$). And the same is desired for the corresponding simulated quantities $y^2_{\boldsymbol{\theta}}(y^1_{\lambda}(x))$ with $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n_{\text{exp}}})^t$ (Figure 2.1). This means that $\pi(\lambda \mid \boldsymbol{z}, \boldsymbol{w})$ must be as close as possible to $\pi(\lambda \mid \boldsymbol{w})$ in the sense of a dissimilarity measure (e.g. the Kullback-Leibler divergence). From Equations (2.4) and (2.5), this is equivalent to postulating that $\pi_{\text{cut}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{z}, \boldsymbol{w})$ is as close as possible to $\pi_{\text{full}}(\boldsymbol{\theta}, \lambda \mid \boldsymbol{z}, \boldsymbol{w})$, which happens when the ratio of the two marginal densities $\pi(\boldsymbol{z} \mid \boldsymbol{w})$ and $\pi(\boldsymbol{z}|\lambda)$ comes to 1 (see Equation (2.9)). Therefore, this hypothesis will be satisfied if the likelihood of the data $\boldsymbol{z}$ conditionally on the couple $(\boldsymbol{\theta}, \lambda)$ is non-identifiable. That is to say that for any $\lambda \neq \lambda'$, there would exist $\boldsymbol{\theta}(\lambda) \neq \boldsymbol{\theta}(\lambda')$ giving the same likelihood [Cole, 2020].

2.4 GP-LinCC method: estimation and prediction

2.4.1 GP-LinCC method resulting from Gaussian process prior and Linearization

The combination of a Gaussian prior given in Equation (2.17) with a Gaussian likelihood due to the linearity hypothesis leads to an explicit posterior Gaussian distribution: we speak in this case of a conjugate family [Reich and Ghosh, 2019]. Thus, a Gaussian predictive distribution of $\boldsymbol{\theta}$ conditionally on any λ of non-zero probability with respect to $\pi(\lambda \mid \boldsymbol{w})$ is obtained, as detailed in Subsection 2.4.3.

2.4.2 GP-LinCC method in practice

To apply GP-LinCC method in practice to calibrate $\boldsymbol{\theta}$ conditionally on λ, we need, in addition to the three basic assumptions described above (compensation, linearity of numerical model and GP prior), the following data:

- a set of experimental data noted $\boldsymbol{z}$ of size n_{exp},
- a set of simulations of the numerical model $\{y_{\boldsymbol{\theta},\lambda}(\boldsymbol{x}_i)\}_{1\leq i\leq n_{\text{exp}}}$ for different values of λ and $\boldsymbol{\theta}$ to estimate the coefficients $\{g_{\lambda}(\boldsymbol{x}_i)\}_{1\leq i\leq n_{\text{exp}}} = \{g_{\lambda 0}(\boldsymbol{x}_i), g_{\lambda,1}(\boldsymbol{x}_i)\}_{1\leq i\leq n_{\text{exp}}}$ of the linear model of Equation (2.19).

In the sequel, it is assumed that the coefficients of $\{g_{\lambda}(\boldsymbol{x}_i)\}_{1\leq i\leq n_{\text{exp}}}$ of the linear model of Equation (2.19) are well known and that $g_{\lambda,0}(\boldsymbol{x}_i) = 0$, $\forall\ 1 \leq i \leq n_{\text{exp}}$.

2.4.3 Estimation: calculation of the posterior distribution

Equation (2.19) is applied to the m realizations of λ to learn the relation between $\boldsymbol{\theta}$ and λ. Due to the compensation hypothesis presented above, we can then write m equations involving the vector of experimental data $\boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_{n_{\text{exp}}})^t \in \mathbb{R}^{n_{\text{exp}}}$:

$$\boldsymbol{z} = g_{\lambda_j}(\boldsymbol{x})\boldsymbol{\theta}(\lambda_j) + \epsilon,\ 1 \leq j \leq m \tag{2.20}$$

where $g_{\lambda_j}(\boldsymbol{x}) := (g_{\lambda_j}(\boldsymbol{x}_1)^t, \cdots, g_{\lambda_j}(\boldsymbol{x}_{n_{\text{exp}}})^t)^t \in \mathbb{R}^{n_{\text{exp}} \times p}$ and $\epsilon := (\epsilon_1, \cdots, \epsilon_{n_{\text{exp}}})^t$. By summarizing these m equations into a single matrix equation, we obtain:

$$(\boldsymbol{z}, \cdots, \boldsymbol{z}) = (g_{\lambda_1}(\boldsymbol{x})\boldsymbol{\theta}(\lambda_1), \cdots, g_{\lambda_m}(\boldsymbol{x})\boldsymbol{\theta}(\lambda_m)) + (\epsilon, \cdots, \epsilon). \tag{2.21}$$

Let $\mathbf{z} := (\boldsymbol{z}, \cdots, \boldsymbol{z}) \in \mathbb{R}^{n_{\text{exp}} \times m}$ be the matrix of the m copies of $\boldsymbol{z}$ and the associated macro parameter $\boldsymbol{\Theta}_m := (\boldsymbol{\theta}(\lambda_1), \cdots, \boldsymbol{\theta}(\lambda_m))^t \in \mathbb{R}^{m \times p}$. Each component of this macro parameter is a realization of a multivariate normal distribution according to Equation (2.17). Then, one way to infer the random matrix $\boldsymbol{\Theta}_m$ is to work on the vectorized form of $\boldsymbol{\Theta}_m$ noted $\vec{\boldsymbol{\Theta}}_m \in \mathbb{R}^{pm}$ [Barratt, 2018] (see Appendix 2.7.1.3). Therefore, the prior distribution on $\boldsymbol{\Theta}_m$ is written as:

$$\pi(\boldsymbol{\Theta}_m|\phi) \propto |\mathbf{K}_\phi|^{-\frac{1}{2}} \exp -\frac{1}{2}(\vec{\boldsymbol{\Theta}}_m - \vec{M}_\beta)^t \mathbf{K}_\phi^{-1}(\vec{\boldsymbol{\Theta}}_m - \vec{M}_\beta), \tag{2.22}$$

where

$$\vec{M}_\beta = (m_{\beta_1}(\lambda_1), \cdots, m_{\beta_p}(\lambda_1), \cdots, m_{\beta_1}(\lambda_m), \cdots, m_{\beta_p}(\lambda_m))^t \in \mathbb{R}^{pm},$$

$$\mathbf{K}_\phi = \{\text{cov}\,(\boldsymbol{\theta}(\lambda_j), \boldsymbol{\theta}(\lambda_{j'}))\}_{j,j'=1}^m, \text{ with } \text{cov}\,(\boldsymbol{\theta}(\lambda_j), \boldsymbol{\theta}(\lambda_{j'})) = \text{diag}\Big\{\sigma_l^2 K_{\psi_l}(\lambda_j, \lambda_{j'})\Big\}_{l=1}^p,$$

and $\phi := (\beta_l, \sigma_l^2, \psi_k)_{l=1}^p$. The expression of the likelihood of $\mathbf{z}$ conditionally on $\boldsymbol{\Theta}_m$ is given by:

$$\mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m) = \mathscr{L}(\boldsymbol{z}_1 = \boldsymbol{z}, \cdots, \boldsymbol{z}_m = \boldsymbol{z} \mid \boldsymbol{\theta}(\lambda_1), \cdots, \cdots \boldsymbol{\theta}(\lambda_m)) \tag{2.23}$$

$$\propto \prod_{j=1}^m \exp -\frac{1}{2}(\boldsymbol{z} - g_{\lambda_j}(\boldsymbol{x})^t\boldsymbol{\theta}(\lambda_j))^t \Sigma_\epsilon^{-1}(\boldsymbol{z} - g_{\lambda_j}(\boldsymbol{x})^t\boldsymbol{\theta}(\lambda_j)), \tag{2.24}$$

with $\Sigma_\epsilon := \text{diag}(\sigma_{\epsilon_1}^2 + \widehat{\sigma}_1^2, \cdots, \sigma_{\epsilon_{n_{\text{exp}}}}^2 + \widehat{\sigma}_{n_{\text{exp}}}^2)$. Finally, the posterior probability distribution is given by the Bayes' formula:

$$\pi(\boldsymbol{\Theta}_m \mid \mathbf{z}, \phi) \propto \mathscr{L}(\mathbf{z} \mid \boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi). \tag{2.25}$$

Theorem 1

The posterior probability distribution $\pi(\boldsymbol{\Theta}_m|\mathbf{z}, \phi)$ is a multivariate normal distribution with mean $\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z}, \phi]$ and covariance matrix $\boldsymbol{\Sigma}_\phi$ given below:

$$\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z}, \phi] = \boldsymbol{\Sigma}_\phi(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}) \in \mathbb{R}^{pm}, \tag{2.26}$$

$$\boldsymbol{\Sigma}_\phi = (\boldsymbol{\Delta}^{-1} + \mathbf{K}_\phi^{-1})^{-1} \in \mathbb{R}^{pm \times pm}, \tag{2.27}$$

where

$$G = (g_{\lambda_1}(\boldsymbol{x}), \cdots, g_{\lambda_m}(\boldsymbol{x})) \in \mathbb{R}^{n_{\text{exp}} \times pm},$$
$$\boldsymbol{\Delta}^{-1} = \text{diag}(g_{\lambda_1}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_1}(\boldsymbol{x}), \cdots, g_{\lambda_m}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_m}(\boldsymbol{x})).$$

See Appendix 2.7.2.1 for the proof of the theorem. From this posterior distribution, we can deduce a predictive distribution denoted $\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z}, \phi)$ associated with any new realization or set of realizations $\lambda^\star \notin D_m$.

2.4.4 Prediction: calculation of the predictive distribution

For any new set of realizations $\lambda^\star = (\lambda_1^\star, \cdots, \lambda_k^\star)^t$ belonging to the support of the distribution $\pi(\lambda|w)$, a predictive distribution of $\boldsymbol{\theta}(\lambda^\star)$ can be derived by integrating the conditional

Gaussian distribution $\pi(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{\Theta}_m, \phi)$ over the posterior probability measure $\pi(\boldsymbol{\Theta}_m|\mathbf{z}, \phi)d\boldsymbol{\Theta}_m$ given by Theorem 1:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z}, \phi) = \int \pi(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{\Theta}_m, \phi)\pi(\boldsymbol{\Theta}_m|\mathbf{z}, \phi)d\boldsymbol{\Theta}_m. \tag{2.28}$$

Theorem 2

The predictive distribution $\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z}, \phi)$ is a multivariate normal distribution with mean and covariance matrix respectively denoted by $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and $\Sigma_{p\text{pred}}(\lambda^\star, \lambda^{\star\prime})$ such that

$$\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star) := \vec{m}_\beta(\lambda^\star) + \mathbf{C}(\lambda^\star, D_m)\mathbf{K}_\phi^{-1}(\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z}, \phi] - \vec{M}_\beta), \tag{2.29}$$

$$\Sigma_{\text{pred}}(\lambda^\star, \lambda^{\star\prime}) := \boldsymbol{\Sigma}_{\text{cond}}(\lambda^\star, \lambda^{\star\prime}) + \mathbf{C}(\lambda^\star, D_m)\mathbf{K}_\phi^{-1}\boldsymbol{\Sigma}_\phi\mathbf{K}_\phi^{-1}\mathbf{C}(D_m, \lambda^{\star\prime}), \tag{2.30}$$

where

$$\vec{m}_\beta(\lambda^\star) = \left(m_{\beta_1}(\lambda_1^\star), \cdots, m_{\beta_p}(\lambda_1^\star), \cdots, m_{\beta_1}(\lambda_k^\star), \cdots, m_{\beta_p}(\lambda_k^\star)\right)^t \in \mathbb{R}^{pk},$$
$$\mathbf{C}(\lambda^\star, D_m) = \left\{\text{cov}\left(\boldsymbol{\theta}(\lambda_i^\star), \boldsymbol{\theta}(\lambda_j)\right)\right\}_{1\leq i\leq k, 1\leq j\leq m} \in \mathbb{R}^{pk\times pm},$$
$$\boldsymbol{\Sigma}_{\text{cond}}(\lambda^\star, \lambda^{\star\prime}) = \mathbf{C}(\lambda^\star, \lambda^{\star\prime}) - \mathbf{C}(\lambda^\star, D_m)\mathbf{K}_\phi^{-1}\mathbf{C}(D_m, \lambda^{\star\prime}) \in \mathbb{R}^{pk\times pk}.$$

The expression of $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ corresponds to the classical expression of the mean of a conditional GP where the unknown quantity $\boldsymbol{\Theta}_m$ is replaced by its posterior expectation $\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z}, \phi]$. The mean $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ is a predictor of $\boldsymbol{\theta}(\lambda^\star)$ which does not require the knowledge of $g_{\lambda^\star}(\boldsymbol{x}_i)$ $(1 \leq i \leq n_{\text{exp}})$ for any $\lambda^\star$. The predictive covariance matrix $\Sigma_{\text{pred}}(\lambda^\star, \lambda^{\star\prime})$ is composed of two terms. The first one $\boldsymbol{\Sigma}_{\text{cond}}(\lambda^\star, \lambda^{\star\prime})$ corresponds to the classical formula for the covariance matrix of a conditional GP when $\boldsymbol{\Theta}_m$ is known (we would have in this case $\Sigma_{\text{pred}}(\lambda^\star, \lambda^{\star\prime}) = \boldsymbol{\Sigma}_{\text{cond}}(\lambda^\star, \lambda^{\star\prime})$). The additional term is related to the lack of knowledge of $\boldsymbol{\Theta}_m$ conveyed by $\boldsymbol{\Sigma}_\phi$. Appendix 2.7.2.2 details the calculations of this predictive distribution.

All the previous formulas involve the hyperparameters ϕ which are never known in advance. We can estimate them either by likelihood maximization or by cross-validation [Bachoc, 2013]. This adds a preliminary step to enable us to evaluate the distributions of Theorems 1 and 2. In this paper, the hyperparameters ϕ are estimated by marginal likelihood maximization [Rasmussen et al., 2006]. This technique, kwown as empirical Bayes, consists in maximizing the marginal likelihood in ϕ obtained by integrating the likelihood $\mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)$ over $\boldsymbol{\Theta}_m$:

$$\widehat{\phi} := \underset{\phi}{\text{argmax}} \int \mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi)d\boldsymbol{\Theta}_m. \tag{2.31}$$

Finally, the accuracy of the predictive distribution in Theorem 2 can be evaluated by comparison with the distribution that we would obtain if using $g_{\lambda^\star}(\boldsymbol{x}_i)_{i=1}^n$ along with a Jeffreys prior on $\boldsymbol{\theta}(\lambda^\star)$. The expression of this distribution, called target distribution, is given by:

$$\pi_{\text{target}}(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{z}\mathbf{1}_k^t) \sim \mathscr{N}_{pk}(\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda^\star), \Sigma_{\text{target}}(\lambda^\star, \lambda^{\star\prime})). \tag{2.32}$$

It is obtained from Theorem 1 where the matrix $\mathbf{K}_\phi^{-1}$, the vector $\vec{M}_\beta$ and m are replaced respectively by the null matrix of $\mathbb{R}^{pk\times pk}$, the null vector of $\mathbb{R}^{pk}$ and k. Another target distribution might be the one using $g_{\lambda^\star}(\boldsymbol{x}_i)_{i=1}^n$ along with a Gaussian prior whose hyperparameters are given in Equation (2.31):

$$\pi_{\text{targetGP}}(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{z}\mathbf{1}_k^t, \hat{\phi}) \sim \mathscr{N}_{pk}(\bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda^\star), \Sigma_{\text{targetGP}}(\lambda^\star, \lambda^{\star\prime})). \tag{2.33}$$

It is derived from Theorem 1 with $m = k$ and $\phi = \hat{\phi}$. The comparison between the three distributions (predictive and the two targets) will be done via the MSE criterion described in the next section.

2.4.5 Accuracy criterion: MSE

The criterion we consider is the mean square error (MSE). It is calculated by integrating over $\pi(\lambda|w)$ the difference between the true value $\boldsymbol{\theta}(\lambda)$ and the mean predictor $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ (resp. $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda)$, $\bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda)$ for the target distributions):

$$\text{MSE} = \int (\boldsymbol{\theta}(\lambda) - \bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda))^t (\boldsymbol{\theta}(\lambda) - \bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)) \pi(\lambda|w) d\lambda. \tag{2.34}$$

In practice, it is approximated by a Monte Carlo estimator:

$$\frac{1}{N_\lambda} \sum_{j=1}^{N_\lambda} (\boldsymbol{\theta}(\lambda_j) - \bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda_j))^t (\boldsymbol{\theta}(\lambda_j) - \bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda_j)), \tag{2.35}$$

where $\{\lambda\}_j)_{j=1,\cdots,N_\lambda}$ are a N_λ-size sample of i.i.d. realizations of λ. The lower the MSE, the more accurate the predictor $\bar{\boldsymbol{\theta}}_{pred}(\lambda)$. In the numerical examples, we will consider an empirical MSE for each component $\boldsymbol{\theta}_u(\lambda)$ of $\boldsymbol{\theta}(\lambda)$:

$$\text{MSE}_u \approx \frac{1}{N_\lambda} \sum_{j=1}^{N_\lambda} (\boldsymbol{\theta}_u(\lambda_j) - \bar{\boldsymbol{\theta}}_{\text{pred}_u}(\lambda_j))^2,\ 1 \leq u \leq p \ \text{ (and resp. for } \bar{\boldsymbol{\theta}}_{\text{target}}(\lambda),\ \bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda)). \tag{2.36}$$

Afterwards, we can check to which extend the calibrated model $\{g_\lambda(\boldsymbol{x}_i)^t \bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)\}_{1 \leq i \leq n_{\text{exp}}}$ is constant in λ (in other words whether the compensation hypothesis is verified).

2.4.6 A test for checking the compensation hypothesis

The compensation hypothesis states that the predictions of the numerical model should be constant in λ. To inspect this hypothesis, we compute the predictive distribution of the numerical model associated to each λ:

$$\pi(g_\lambda(\boldsymbol{x})\boldsymbol{\theta}(\lambda)|\mathbf{z}, \lambda, \phi) \sim \mathcal{N}_n(g_\lambda(\boldsymbol{x})\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda),\ g_\lambda(\boldsymbol{x})\Sigma_{\text{pred}}(\lambda, \lambda')g_\lambda(\boldsymbol{x})^t). \tag{2.37}$$

By integrating over the whole uncertainty of λ, we have:

$$\pi(g_\lambda(\boldsymbol{x})\boldsymbol{\theta}(\lambda)|\mathbf{z}, \phi) = \int \pi(g_{\lambda'}(\boldsymbol{x})\boldsymbol{\theta}(\lambda')|\mathbf{z}, \phi, \lambda')\pi(\lambda'|w)d\lambda'. \tag{2.38}$$

Using the cross-validation technique [Hastie et al., 2009; Bachoc, 2013], we can derive for any configuration $\boldsymbol{x}_i$, the predictive distribution of the associated numerical model for any λ, for $1 \leq i \leq n_{\text{exp}}$:

$$\pi(g_\lambda(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda)|\mathbf{z}_{-i}, \lambda, \phi) \sim \mathcal{N}\left(g_\lambda(\boldsymbol{x}_i)^t\bar{\boldsymbol{\theta}}_{\text{pred},-i}(\lambda), g_\lambda(\boldsymbol{x}_i)^t\Sigma_{\text{pred},-i}(\lambda, \lambda)g_\lambda(\boldsymbol{x}_i)\right), \tag{2.39}$$

where $\mathbf{z}_{-i}$ is equal to $\mathbf{z}$ without the column i, $\bar{\boldsymbol{\theta}}_{\text{pred},-i}$ and $\Sigma_{\text{pred},-i}$ are given by Theorem 2 where $\mathbf{z}$ is replaced by $\mathbf{z}_{-i}$. If the compensation hypothesis is achieved by the numerical model, then for any $\lambda_1 \neq \lambda_2$, the distributions $\pi(g_{\lambda_1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda_1)|\mathbf{z}_{-i}, \lambda_1)$, $\pi(g_{\lambda_2}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda_2)|\mathbf{z}_{-i}, \lambda_2)$ and $\pi(g_\lambda(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda)|\mathbf{z}_{-i}, \phi)$ are expected to be similar to each other. Thus the predictive credibility interval of the distribution of the random variable

$$g_{\lambda_1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda_1) - g_\lambda(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda_2)|\mathbf{z}_{-i}, \lambda_1, \lambda_2, \phi \tag{2.40}$$

is likely to cover 0. This random variable follows a normal distribution of mean and variance given respectively by:

$$\begin{aligned}
\mu_i(\lambda_1, \lambda_2) &:= g_{\lambda_1}(\boldsymbol{x}_i)^t\bar{\boldsymbol{\theta}}_{\text{pred},-i}(\lambda_1) - g_{\lambda_2}(\boldsymbol{x}_i)^t\bar{\boldsymbol{\theta}}_{\text{pred},-i}(\lambda_2), \\
\sigma_i^2(\lambda_1, \lambda_2) &:= g_{\lambda_1}(\boldsymbol{x}_i)^t\Sigma_{\text{pred},-i}(\lambda_1, \lambda_1)g_{\lambda_1}(\boldsymbol{x}_i) - 2g_{\lambda_1}(\boldsymbol{x}_i)^t\Sigma_{\text{pred},-i}(\lambda_1, \lambda_2)g_{\lambda_2}(\boldsymbol{x}_i) \\
&\quad + g_{\lambda_2}(\boldsymbol{x}_i)^t\Sigma_{\text{pred},-i}(\lambda_2, \lambda_2)g_{\lambda_2}(\boldsymbol{x}_i).
\end{aligned}$$

From this distribution, we can compute an empirical coverage probability for 0 for $1-\alpha$ level, from N different i.i.d. sample pairs $(\lambda_1, \lambda_2) \sim \pi(\lambda|w) \times \pi(\lambda|w)$:

$$\hat{\Delta}(\alpha, \boldsymbol{x}_i) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{\left\{0 \in \left[\mu_i(\lambda_1^j, \lambda_2^j) \pm q_{1-\frac{\alpha}{2}} \sqrt{\sigma_i^2(\lambda_1^j, \lambda_2^j)}\right]\right\}} \tag{2.41}$$

where $q_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$ quantile of the standard Gaussian distribution. The objective is to obtain $\hat{\Delta}(\alpha, \boldsymbol{x}_i)$ as close as possible to the theoretical probability of interval $1-\alpha$. For instance, considering a standard level $\alpha = 5\%$, if the coverage probability $\hat{\Delta}(5\%, \boldsymbol{x}_i)$ is lower than 95%, then the compensation hypothesis must be questioned.

2.5 Numerical examples

2.5.1 Example 1 in 1D ($\boldsymbol{\theta} \in \mathbb{R}$)

We consider the following probabilistic model:

$$\boldsymbol{z}_i = r + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}, \tag{2.42}$$

where $\epsilon_i \sim \mathcal{N}(0, 2)$ and the chosen r system is assumed to be constant and equal to 5. We model it by

$$y_{\boldsymbol{\theta},\lambda} = \lambda \boldsymbol{\theta}, \tag{2.43}$$

where

$$\pi(\lambda \mid \boldsymbol{w}) \sim \mathscr{U}[1,\ 10] \tag{2.44}$$

Equation (2.42) can thus be rewritten as,

$$\boldsymbol{z}_i = \lambda \boldsymbol{\theta} + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}. \tag{2.45}$$

The prior distribution on $\boldsymbol{\Theta}_m$ is chosen with:

- a Matérn 5/2 covariance function given in Equation (2.18),
- a constant mean function: $m_\beta(\lambda) = \beta$.

A Latin hypercuble sampling (LHS) [McKay et al., 1979] is used to sample $\lambda \sim \pi(\lambda|w)$ and generate the design D_m. The GP hyperparameters $\phi = (\beta, \psi, \sigma^2)$ are estimated by marginal likelihood maximization (Equation (2.31)). Appendix 2.7.2.3 details this estimation. Once ϕ is estimated and denoting $\hat{\phi}$ its estimator, the predictive distribution $\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{z}\mathbf{1}^t, \hat{\phi})$ is computed for a vector $\lambda^\star \in [1, 10]^k$ with k new realizations, and then compared to the target distribution $\pi_{\text{target}}(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{z}\mathbf{1}_k^t)$. Figure 2.2 presents the results obtained with $n_{\text{exp}} = 50$ i.i.d. samples generated by Equation (2.42) , $m = 10$ and $k = 500$. On this figure, we have printed the true function $\boldsymbol{\theta}(\lambda) = \frac{r}{\lambda}$ as well as the means of the predictive and target distributions denoted respectively by $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ and $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda)$. The 95% credibility intervals associated with these two distributions are also represented. It can be seen that the predicted mean $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ is relatively close to both the true and target functions and that the predicted credibility intervals of 95% cover them relatively well. We have not plotted $\bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda)$ (respectively its credibility interval of 95%). In fact, we can easily notice that $\bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda)$ and $\Sigma_{\text{targetGP}}(\lambda, \lambda)$ are very close to $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda)$ and $\Sigma_{\text{target}}(\lambda, \lambda)$ respectively because the estimated variance $\hat{\sigma}^2$ of the GP is greater than $\Sigma_{\text{target}}(\lambda, \lambda)$. These quantities are related by

$$\Sigma_{\text{targetGP}}(\lambda, \lambda) = \frac{\Sigma_{\text{target}}(\lambda, \lambda)}{1 + \frac{\Sigma_{\text{target}}(\lambda, \lambda)}{\hat{\sigma}^2}} \approx \frac{\sigma_\epsilon^2}{n_{\text{exp}} \lambda^2} \tag{2.46}$$

$$\bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda) = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + n_{\text{exp}} \lambda^2 \hat{\sigma}^2} \hat{\beta} + \frac{\bar{\boldsymbol{z}}}{\lambda} \frac{1}{1 + \frac{\sigma_\epsilon^2}{n_{\text{exp}} \lambda^2 \hat{\sigma}^2}} \approx \bar{\boldsymbol{\theta}}_{\text{target}}(\lambda) \tag{2.47}$$

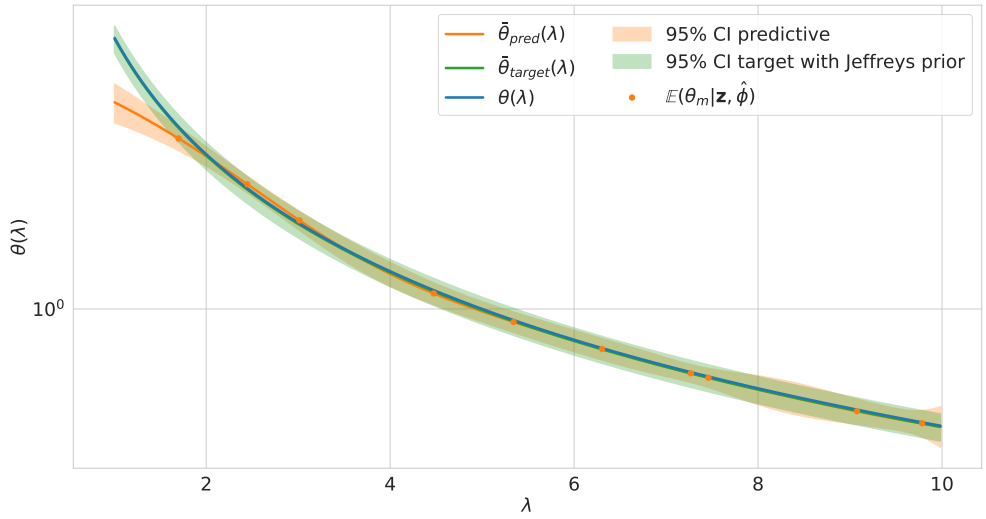


Figure 2.2: True function $\boldsymbol{\theta}(\lambda)$, predicted mean denoted by $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$, target mean denoted by $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda)$ (with Jeffreys prior) and associated 95% credibility intervals, obtained from a sample of $n_{\text{exp}} = 50$ observations and a design D_m with $m = 10$. The y-axis is in log-scale.

where $\Sigma_{\text{target}}(\lambda, \lambda) = \frac{\sigma_\epsilon^2}{n_{\text{exp}}\lambda^2} = \frac{2}{n_{\text{exp}}\lambda^2}$, $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda) = \frac{\bar{z}}{\lambda}$, $\hat{\sigma}^2 = 8.72$, $\hat{\beta} = 2.98$.

For a given vector $\lambda^\star \in [1, 10]^{N_\lambda}$ with $N_\lambda = 1000$ new i.i.d. realizations of λ, we have computed the empirical MSE of the following estimators:

- $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ the mean of the predictive distribution given by Theorem 2 with $p = 1$,
- $\bar{\boldsymbol{\theta}}_{\text{target}}(\theta(\lambda^\star)|\boldsymbol{z})$ the mean of the target distribution with Jeffreys prior given in Equation (2.32) with $p = 1$,
- $\bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda^\star)$ the mean of the target distribution with a Gaussian prior on $\boldsymbol{\theta}(\lambda^\star)$ whose hyperparameters are estimated from D_m and $\boldsymbol{z}$. It is given bin Equation (2.33) with $p = 1$.

This procedure is randomly replicated 100 times for two sample sizes $n \in \{50, 100\}$ of observations $\boldsymbol{z}$ and three sample sizes $m \in \{10, 15, 20\}$ of D_m. Each replication is performed from independent samples of $\boldsymbol{z}$ and independent LHS designs D_m. Figure 2.3 shows the boxplots of the empirical MSE thus obtained. For both sample sizes n, two phenomena are observed as m increases:

- the boxplot of the empirical MSE of $\bar{\theta}_{pred}(\lambda^\star)$ comes close to that of $\bar{\theta}_{targetGP}(\lambda^\star)$,
- the results obtained for the two target distributions but calculated with two different priors differ little. The impact of the prior choice is therefore limited in this case, mostly due to the use of empirical Bayes.

2.5.2 Example 2 in 2D ($\boldsymbol{\theta} \in \mathbb{R}^2$)

We consider the following analytical example:

- an additive physical system of interest: $r(\boldsymbol{x}) = r_1(\boldsymbol{x}) + r_2(\boldsymbol{x})$ with $r_1(\boldsymbol{x}) \neq r_2(\boldsymbol{x})$ $\forall\, \boldsymbol{x} \in [-4, 4]$,

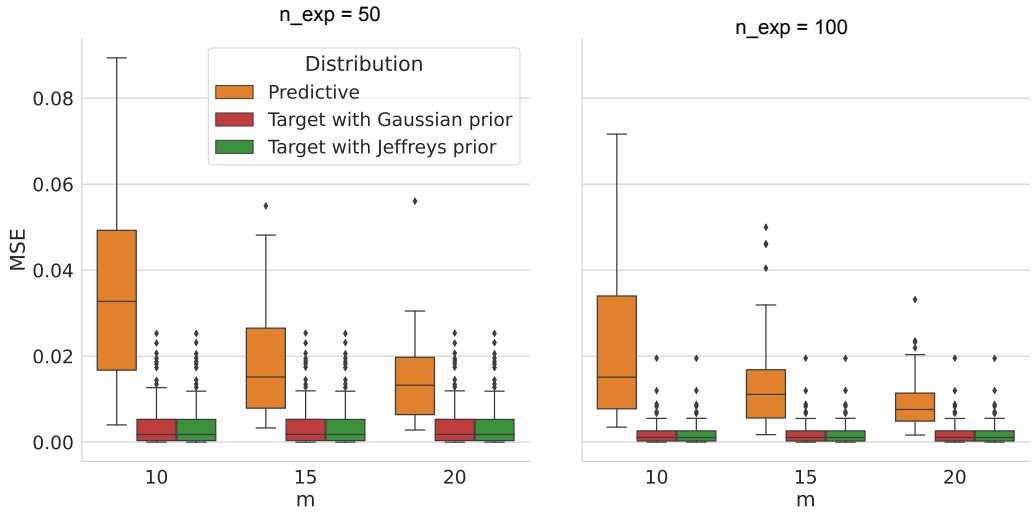


Figure 2.3: Boxplots of the MSE of the means of predictive and target distributions for observations samples of size $n_{\text{exp}} \in \{50, 100\}$ and from a design D_m with $m \in \{10, 15, 20\}$ replicated 100 times.

- the experimental data are linked to the system via:

$$\boldsymbol{z}_i = r(\boldsymbol{x}_i) + \epsilon_i,\ -4 \leq \boldsymbol{x}_i \leq 4,\ i = 1, \cdots, n_{\text{exp}}, \tag{2.48}$$

$$\epsilon_i \sim \mathscr{N}(0, \sigma^2_{\epsilon_i}),\ \sigma^2_{\epsilon_i} := (0.06|r(\boldsymbol{x}_i)|)^2. \tag{2.49}$$

We postulate the following linear numerical model, supposed to represent $r(\boldsymbol{x})$ and verifying the compensation hypothesis:

$$r(\boldsymbol{x}_i) = g_\lambda(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda),\ \boldsymbol{\theta}(\lambda) \in \mathbb{R}^2, \tag{2.50}$$

where the components of $\boldsymbol{\theta}(\lambda)$ are given by:

$$\boldsymbol{\theta}_1(\lambda) = \lambda \sin(10\lambda) + 1,\ \boldsymbol{\theta}_2(\lambda) = \sin(2\pi\lambda/10) + 0.2\sin(20\pi\lambda/2.5) + 1.75, \tag{2.51}$$

and $\pi(\lambda|w) \sim \mathscr{U}[0, 1]$. The two functions $\boldsymbol{\theta}_l(\lambda)$ $(1 \leq l \leq 2)$ are chosen from [Surjanovic and Bingham, 2013]. We need to construct $g_\lambda(\boldsymbol{x}_i)$ in such a way that the numerical model $g_\lambda(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda)$ remains unchanged in λ. To do this, we propose to set:

$$g_1(\boldsymbol{x}_i, \lambda) = \frac{r_1(\boldsymbol{x}_i)}{\boldsymbol{\theta}_1(\lambda)} = \frac{\boldsymbol{x}_i^2 + \boldsymbol{x}_i + 1}{\boldsymbol{\theta}_1(\lambda)} \text{ and } g_2(\boldsymbol{x}_i, \lambda) = \frac{r_2(\boldsymbol{x}_i)}{\boldsymbol{\theta}_2(\lambda)} = \frac{\boldsymbol{x}_i^2 + \boldsymbol{x}_i + 4}{\boldsymbol{\theta}_2(\lambda)}. \tag{2.52}$$

With this choice, the compensation hypothesis is verified and moreover, the matrix $\boldsymbol{\Delta}$ of Theorem 1 is invertible. The prior distribution on $\boldsymbol{\Theta}_m$ is chosen from two GP prior for $\boldsymbol{\theta}_1(\lambda)$ and $\boldsymbol{\theta}_2(\lambda)$ with for each:

- a Matérn 5/2 covariance function given in Equation (2.18),
- a constant mean function: $m_{\beta_l}(\lambda) = \beta_l$, $l \in \{1, 2\}$.

Once $\phi = (\beta_l, \sigma_l^2, \psi_l)_{1 \leq l \leq 2}$ is estimated, the predictive distribution $\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z}, \hat{\phi})$ is computed for a vector $\lambda^\star \in [1, 10]^k$ with $k = 500$ new realizations, then compared to the target distribution $\pi_{target}(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{z}\mathbf{1}_k^t)$. Figures 2.4 (a) and (b) present respectively the two components $\boldsymbol{\theta}_1(\lambda)$ et $\boldsymbol{\theta}_2(\lambda)$ of the true function $\boldsymbol{\theta}(\lambda)$ as well as the two means of both the predictive and target distributions calculated with a sample of $n_{\text{exp}} = 50$ i.i.d. observations generated by Equation (2.48) and a design D_m of size $m = 10$. The 95% credibility intervals

cover relatively well the two components of the true function ((a) for $\boldsymbol{\theta}_1(\lambda)$ and (b) for $\boldsymbol{\theta}_2(\lambda)$). We also observe that the marginal predictors are able to approximate well the components of the true function and those of the target function. Figure 2.5 presents the comparison between the marginal predictive and target densities for a specific $\lambda^\star$, equal to 0.45. For this λ realization, the three densities are in good agreeement.

For a given vector $\lambda^\star \in [1, 10]^{N_\lambda}$ with $N_\lambda = 1000$ new i.i.d. realizations of λ, we have computed the empirical MSE of the following estimators:

- $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ the mean of the predictive distribution given by Theorem 2 with $p = 2$,
- $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda^\star)$ the mean of the target distribution with Jeffreys prior given in Equation (2.32) with $p = 2$,
- $\bar{\boldsymbol{\theta}}_{\text{targetGP}}(\lambda^\star)$ the mean of the target distribution with a Gaussian prior on $\boldsymbol{\theta}(\lambda^\star)$ whose hyperparameters are estimated from D_m and $\boldsymbol{z}$. It is given in Equation (2.33) with $p = 2$.

This procedure is randomly replicated 100 times for two sample sizes $n_{\text{exp}} \in \{50, 100\}$ of observations $\boldsymbol{z}$ and three sample sizes $m \in \{10, 15, 20\}$ of D_m. Results are given by Figures 2.6 and 2.7, for $\boldsymbol{\theta}_1(\lambda)$ and $\boldsymbol{\theta}_2(\lambda)$ respectively. We can see that for $m = 10$, the GP-LinCC method predicts better the first component than the second one. This could be due to the shape of these two components. However, when m increases, it predicts well the components of $\boldsymbol{\theta}(\lambda)$.

2.5.3 Example 3 in 1D ($\theta \in \mathbb{R}$) with falsification of the compensation hypothesis

In this example, we will test the GP-LinCC method in the case where the compensation hypothesis is violated. The sampling of experimental data now depends on λ. We have chosen a specific sample λ_0 such that:

$$\boldsymbol{z}_i = r_{\lambda_0}(\boldsymbol{x}_i) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}, \tag{2.53}$$
$$\boldsymbol{z}_i = g_{\lambda_0}(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda_0) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}, \tag{2.54}$$

with r_λ not being equal to r_{λ_0} if and only if λ is not equal to λ_0. For any λ, the observations $\boldsymbol{z}$ can actually be related to the numerical model outputs via:

$$\boldsymbol{z}_i = g_\lambda(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda) + b(\boldsymbol{x}_i, \lambda) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}, \tag{2.55}$$

where the discrepancy term $b(\boldsymbol{x}_i, \lambda)$ is equal to:

$$b(\boldsymbol{x}_i, \lambda) = g_{\lambda_0}(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda_0) - g_\lambda(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda) = g_\lambda(\boldsymbol{x}_i)^t \left(\frac{g_{\lambda_0}(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda_0)}{g_\lambda(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda)} - 1\right) \boldsymbol{\theta}(\lambda) \tag{2.56}$$
$$= (\alpha(\boldsymbol{x}_i, \lambda) - 1) g_\lambda(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda), \tag{2.57}$$

where

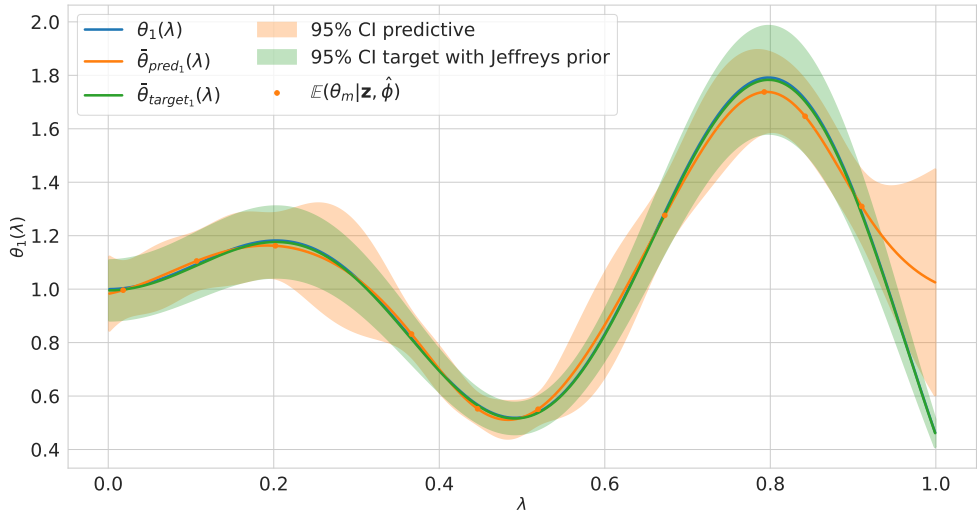
$$\alpha(\boldsymbol{x}_i, \lambda) = \frac{g_{\lambda_0}(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda_0)}{g_\lambda(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda)} = \frac{r_{\lambda_0}(\boldsymbol{x}_i)}{r_\lambda(\boldsymbol{x}_i)}. \tag{2.58}$$

Equation (2.55) becomes:

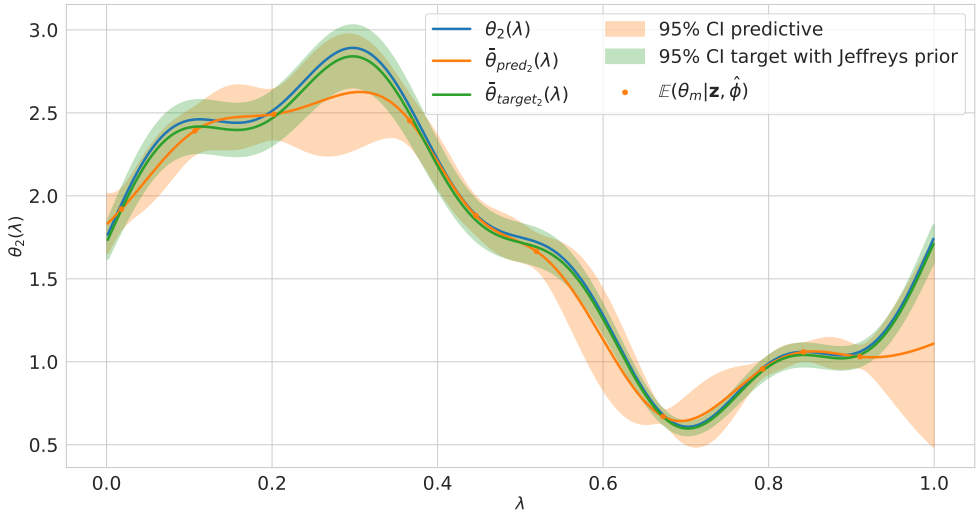
$$\boldsymbol{z}_i = g_\lambda(\boldsymbol{x}_i)^t \alpha(\boldsymbol{x}_i, \lambda) \boldsymbol{\theta}(\lambda) + \epsilon_i,\ 1 \leq i \leq n_{\text{exp}}. \tag{2.59}$$

As $\alpha(\boldsymbol{x}_i, \lambda)$ is neglected by the GP-LinCC method, the mean estimators $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ are equal to

$$\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda) = \left(g_\lambda(\boldsymbol{x})^t \Sigma_\epsilon^{-1} g_\lambda(\boldsymbol{x})\right)^{-1} g_\lambda(\boldsymbol{x})^t \Sigma_\epsilon^{-1} \boldsymbol{z}.$$

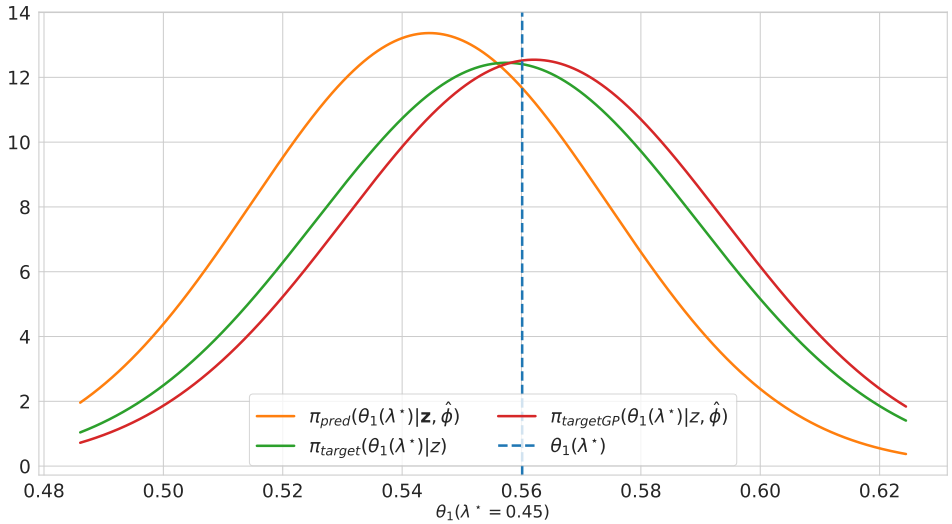


(a) True function $\boldsymbol{\theta}_1(\lambda)$, predicted mean $\bar{\boldsymbol{\theta}}_{\text{pred}_1}(\lambda)$, target mean $\bar{\boldsymbol{\theta}}_{\text{target}_1}(\lambda)$ and associated 95% credibility intervals, obtained with a sample of $n_{\text{exp}} = 50$ observations and with $m = 10$.

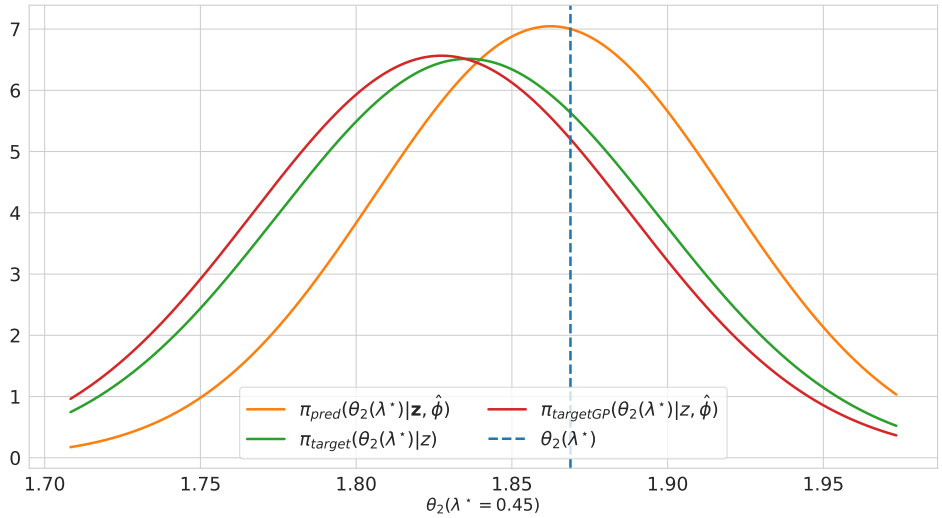


(b) True function $\boldsymbol{\theta}_2(\lambda)$, predicted mean $\bar{\boldsymbol{\theta}}_{\text{pred}_2}(\lambda)$, target mean $\bar{\boldsymbol{\theta}}_{\text{target}_2}(\lambda)$ and associated 95% credibility intervals, obtained with a sample of $n_{\text{exp}} = 50$ observations and with $m = 10$.

Figure 2.4: Comparison of the three functions $\boldsymbol{\theta}(\lambda)$, $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ and $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda)$, for the two components of $\boldsymbol{\theta}(\lambda)$.



(a) First marginal probability densities of the predictive and the both targets distributions at $\lambda^\star = 0.45$.



(b) Second marginal probability densities of the predictive and the both targets distributions at $\lambda^\star = 0.45$.

Figure 2.5: Comparison of the three marginal probability densities obtained from a sample of $n_{\text{exp}} = 50$ observations and D_m with $m = 10$.

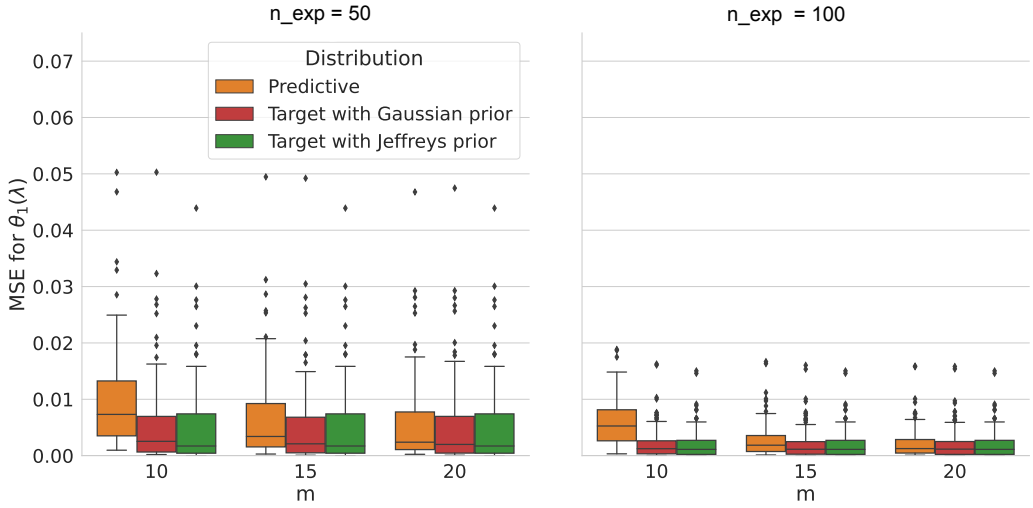


Figure 2.6: Boxplots of the MSE of the first marginal predictive and both target distributions for a design D_m of size $m \in \{10, 15, 20\}$ and a sample of size $n_{\text{exp}} \in \{50, 100\}$ of the observations replicated 100 times.

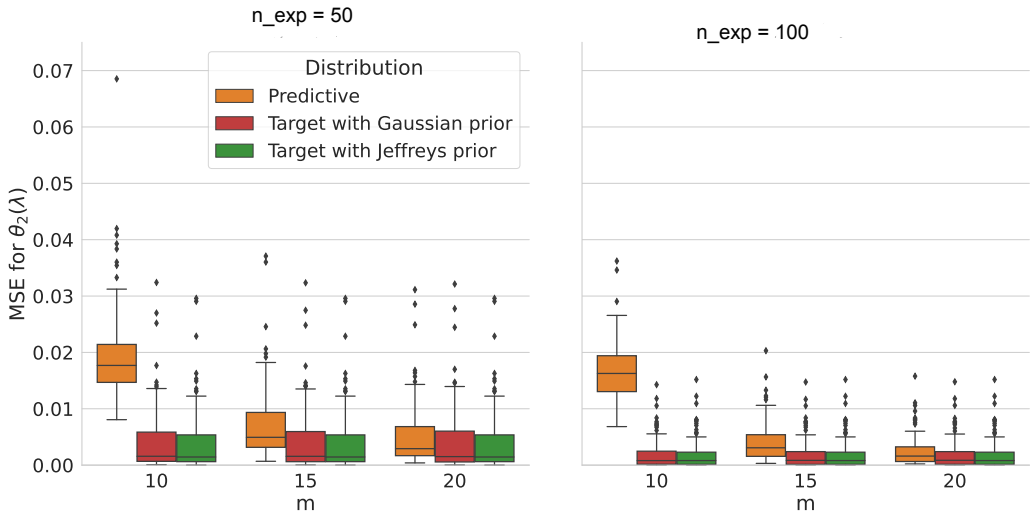


Figure 2.7: Boxplots of the MSE of the second marginal predictive and both target distributions for a design D_m of size $m \in \{10, 15, 20\}$ and a sample of size $n_{\text{exp}} \in \{50, 100\}$ of the observations repeated 100 times.

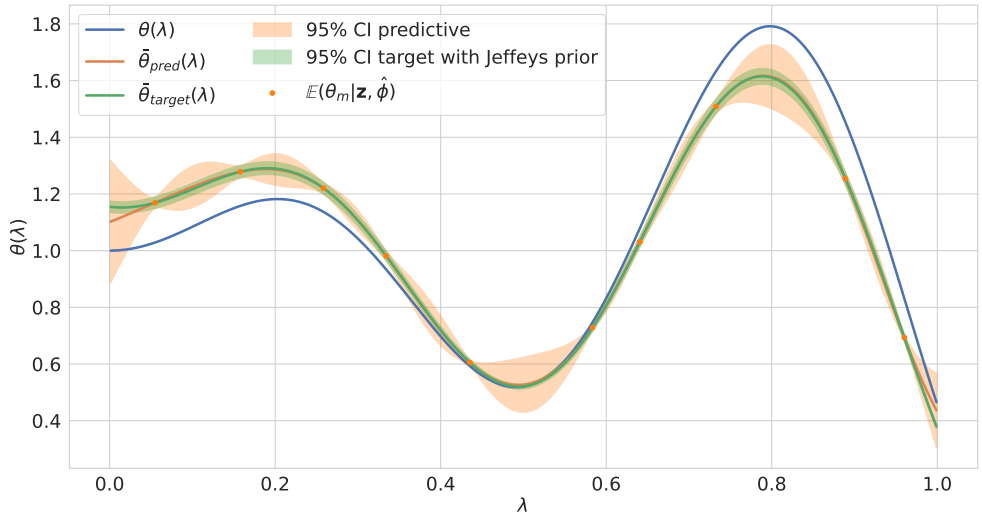


Figure 2.8: True function $\boldsymbol{\theta}(\lambda)$, predicted mean noted $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$, target mean noted $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda)$ and associated 95% credibility intervals, obtained with a sample of $n_{\text{exp}} = 50$ observations and a design D_m of size $m = 10$.

By propagating these estimators through the model, we obtain the following calibrated predictions that depend on λ:

$$\widehat{r}_\lambda(\boldsymbol{x}_i) := g_\lambda(\boldsymbol{x}_i)^t \bar{\boldsymbol{\theta}}_{pred}(\lambda),\ 1 \leq i \leq n_{\text{exp}} \Longrightarrow \widehat{r}_\lambda(\boldsymbol{x}) = g_\lambda(\boldsymbol{x})\bar{\boldsymbol{\theta}}_{pred}(\lambda). \tag{2.60}$$

For our test case, we choose the following physical system

$$r_\lambda(\boldsymbol{x}_i) = 3\boldsymbol{x}_i^2 + 2\lambda^2\boldsymbol{x}_i + 1 + \lambda,\ -2 \leq \boldsymbol{x}_i \leq 2, \tag{2.61}$$

$\pi(\lambda \mid \boldsymbol{w}) \sim \mathscr{U}[0, 1]$ and the numerical model $g_\lambda(\boldsymbol{x})\boldsymbol{\theta}(\lambda)$ such that

$$g_\lambda(\boldsymbol{x}_i) = \frac{r_\lambda(\boldsymbol{x}_i)}{\boldsymbol{\theta}(\lambda)}, \boldsymbol{\theta}(\lambda) = 1 + \lambda \sin(10\lambda). \tag{2.62}$$

The i.i.d. observations $\boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n)^t$ are generated with $\lambda_0 = 0.5$ and the variance $\sigma_i^2 = 0.06(r_{\lambda_0}(\boldsymbol{x}_i))^2$ for $1 \leq i \leq n$ by Equation (2.53). Once $\phi = (\beta_l, \sigma_l^2, \psi_l)_{1 \leq l \leq 2}$ is estimated, the predictive distribution is computed for a vector $\lambda^\star \in [1, 10]^k$ with $k = 500$ new realizations, then compared to the target distribution $\pi_{\text{target}}(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{z}\mathbf{1}_k^t)$. Figure 2.8 presents the true function $\boldsymbol{\theta}(\lambda)$, the predicted function $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ and the target function $\bar{\boldsymbol{\theta}}_{\text{target}}(\lambda)$ computed from a sample of $\boldsymbol{z}$ and a design D_m of size $m = 10$, as well as the associated 95% credibility intervals. We can see that the predicted function does not match the true function whereas, as expected, it approximates well the target function. The predictors $\widehat{r}_\lambda(\boldsymbol{x})$ for different λ and the physical system $r_{\lambda_0}(\boldsymbol{x})$ are plotted in Figure 2.9 where we can observe that these predictors vary with the chosen value of λ. This reveals that the calibrated model does not compensate in λ. This absence of compensation is also confirmed by the predictive distributions $\pi(g_{\lambda_j}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda)|\mathbf{z}_{-i}, \lambda_j, \hat{\phi})$, $\boldsymbol{x}_i \in \{\boldsymbol{x}_1, \boldsymbol{x}_4\}$ for $1 \leq j \leq 4$ as presented in Figure 2.10. The empirical coverage probabilities $\hat{\Delta}(5\%, \boldsymbol{x}_i)$, computed with $N = 5000$ pairs of samples (λ_1, λ_2), presented in Figure 2.11 do not exceed the threshold of 95%, revealing once again that the compensation hypothesis is not satisfied. These results are coherent with the simulation procedure of the observations $\boldsymbol{z}$. Indeed, since $\boldsymbol{z}$ has been simulated with a nominal value $\lambda_0 = 0.5$, the predicted function $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ is not able to predict well the true function for new realizations of λ not close to λ_0. This can be observed in Figure 2.8.

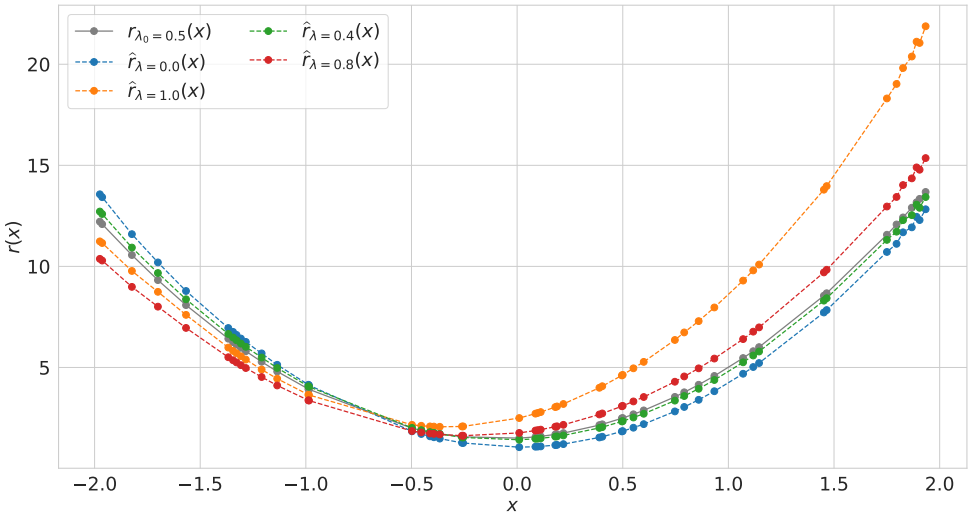
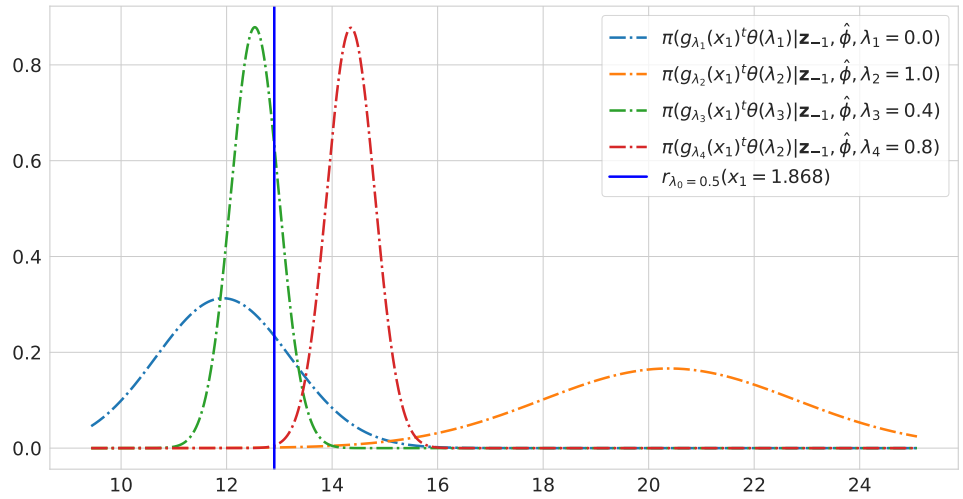
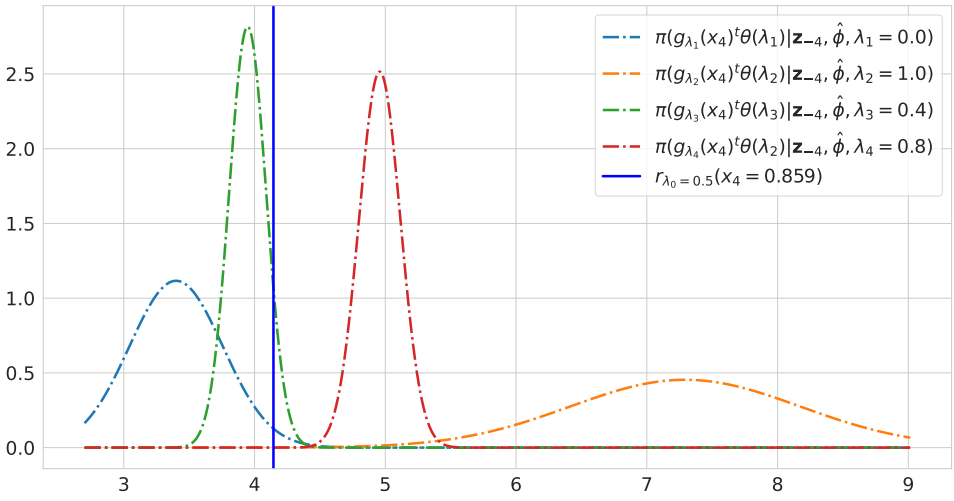


Figure 2.9: Predictors $\widehat{r}_\lambda(\boldsymbol{x})$ for different values of λ compared to $r_{\lambda_0}(\boldsymbol{x})$.

The GP-LinCC method based on both linearization and Gaussian processes has allowed us to perform a joint calibration of the calibration parameters $\boldsymbol{\theta}(\lambda_j)$ for a set of training samples $\lambda_j \in D_m$. From this calibration, we have derived a predictive distribution for some new realizations $\lambda^\star$ whose mean is the predictor of $\boldsymbol{\theta}(\lambda^\star)$ and covariance matrix is the predictive covariance matrix between the components of $\boldsymbol{\theta}(\lambda^\star)$. The accuracy of the predictive mean function $\bar{\boldsymbol{\theta}}_{\text{pred}}(\lambda)$ directly depends on both the size of the design D_m and n_{exp}.



(a) Predictive densities $\pi(g_{\lambda_j}(\boldsymbol{x}_1)^t\theta(\lambda)|\mathbf{z}_{-i},\lambda_j,\hat{\phi})$ for $1 \leq j \leq 4$.



(b) Predictive densities $\pi(g_{\lambda_j}(\boldsymbol{x}_4)^t\theta(\lambda)|\mathbf{z}_{-i},\lambda_j,\hat{\phi})$ for $1 \leq j \leq 4$.

Figure 2.10: Predictive densities $\pi(g_{\lambda_j}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda)|\mathbf{z}_{-i},\lambda_j,\hat{\phi})$, $\boldsymbol{x}_i \in \{\boldsymbol{x}_1,\boldsymbol{x}_4\}$ for $1 \leq j \leq 4$.

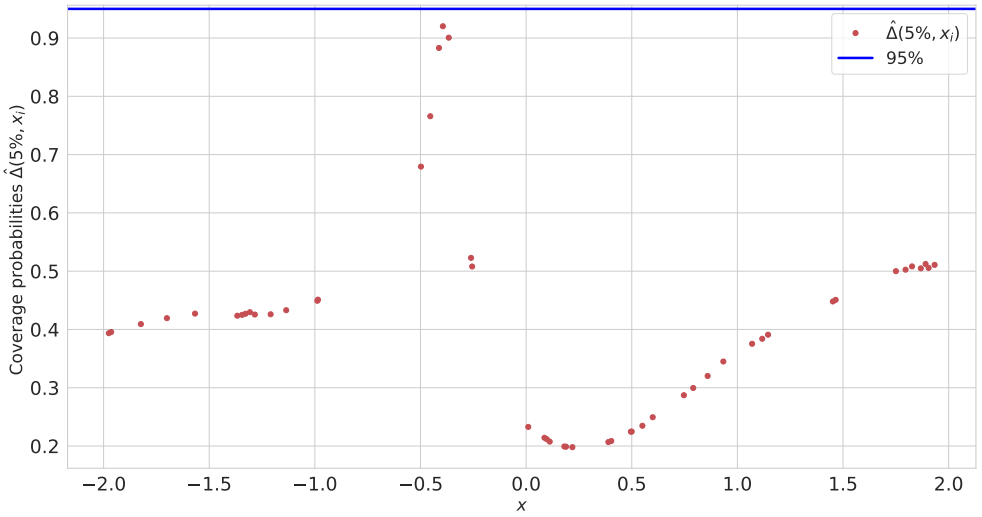


Figure 2.11: Coverage probabilities $\left\{\widehat{\Delta}(5\%, \boldsymbol{x}_i)\right\}_{1\leq i\leq n_{\text{exp}}}$, computed with $N = 5000$ pairs of samples (λ_1, λ_2) (see Equation (2.41)).

2.6 Conclusion and perspectives

In this paper, we have proposed a new method to address the problem of conditional calibration in the framework of two chained numerical models. The parameters $\boldsymbol{\theta}$ of the second model of the chain are calibrated conditionally on the parameters λ of the first model whose probability distribution is known. By leveraging some modeling assumptions consistent with the targeted application framework, the new method GP-LinCC, offers an analytical resolution, avoiding the use of a MCMC algorithm.

The GP-LinCC method can learn the relationship between $\boldsymbol{\theta}$ and λ via the calibration function $\boldsymbol{\theta}(\lambda)$ from available experimental data and from a numerical design of the model for a set of λ. To do so, it relies on three hypotheses. First, the prior distribution for $\boldsymbol{\theta}(\lambda)$ is assumed to be a Gaussian process of given mean and covariance functions (but whose parameters are to be estimated). Then, a compensation hypothesis supported by expert opinion is made to ensure the consistency of the calibration problem. This hypothesis states that the experimental data provide negligible information on the uncertainty of λ and, as a result, on the corresponding simulated quantities. Finally, the last assumption states that the output of the second model is linear in $\boldsymbol{\theta}(\lambda)$. Under these three hypotheses, we have demonstrated that the GP-LinCC method provides a Gaussian predictive probability distribution of $\boldsymbol{\theta}(\lambda^\star)$ for any new set of realizations $\lambda^\star$.

We have implemented this method on analytical examples for small dimensions of the parameters $\boldsymbol{\theta}$ and the results obtained are convincing. Moreover, we propose a way to assess the accuracy of the predictive distribution by comparing it with a target distribution (unknown in practice but computable for analytical test cases). The latter corresponds to the posterior distribution that we would obtain if we knew the linearization coefficients of the model in $\lambda^\star$ while assuming a Jeffreys prior on $\boldsymbol{\theta}(\lambda^\star)$. The predictive distribution has also been compared to another target distribution obtained with a Gaussian prior whose hyperparameters are estimated on the numerical design for λ.

Further work is planned to apply the GP-LinCC method to the fuel application in pressurized water reactors which has motivated this methodological work: namely the calibration of the parameters of the gas behavior model conditionally on the thermal conductivity λ. However, a preliminary sensitivity analysis must be done before deploying the GP-LinCC

method to this physical problem. Indeed, the large dimension of $\boldsymbol{\theta}$ (more than ten parameters) requires a pre-selection of the most important parameters. To achieve this, we propose to carry out a Global sensitivity analysis can be carried out in order to identify the most influencing parameters to be calibrated. More precisely, we propose to use the multivariate version of the sensitivity indices based on the Hilbert-Schmidt independence criterion (HSIC). First introduced by Gretton et al. [2005b] and recently, HSIC extensions for multivariate and functional outputs have then been proposed by El Amri and Marrel [2024].

Finally, the parameters $\boldsymbol{\theta}$ may have bounded variation ranges related to their physical meaning. It would be necessary to incorporate such bound constraints in the GP-LinCC method to guarantee that the results make physical sense. The predictive distribution provided by the GP-LinCC method will thereby transform into a truncated multivariate normal distribution [Da Veiga and Marrel, 2020; López-Lopera et al., 2018].

Acknowledgments

We would like to acknowledge Merlin Keller, research engineer at EDF R&D, who provided us with information on cut-off models. This was a great help in the mathematical formalization of the conditional calibration problem.

2.7 Supplementary material

2.7.1 Some useful mathematical results

2.7.1.1 Sherman-Morrisson formula

Let A and $B \in \mathbb{R}^{n\times n}$ be two invertible matrices. Then we have:

$$(A^{-1} + B^{-1})^{-1} = A - A(A + B)^{-1}A = B - B(A + B)^{-1}B. \tag{2.63}$$

2.7.1.2 Woodbury-Sherman-Morrisson identity

Let $Z \in \mathbb{R}^{n\times n}$, $W \in \mathbb{R}^{m\times m}$, U and $V \in \mathbb{R}^{n\times m}$ be matrices. Suppose that Z and W are invertible. Then, the Woodbury identity and its associated determinant are:

$$(Z + UWV^t)^{-1} = Z^{-1} - Z^{-1}U(W^{-1} + V^tZ^{-1}U)^{-1}V^tZ^{-1}, \tag{2.64}$$

$$|Z + UWV^t| = |Z||W||W^{-1} + V^tZ^{-1}U|. \tag{2.65}$$

2.7.1.3 Vectorization

Vectorization is an operator that transforms any matrix $A \in \mathbb{R}^{m\times p}$ into a column vector denoted $\vec{A} \in \mathbb{R}^{mp}$. This operation consists in stacking the components of A successively, from the first to the last column of A. For example,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2} \Rightarrow \vec{A} = \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix} \in \mathbb{R}^4.$$

2.7.1.4 Gaussian process

Definition 1

A Gaussian process (GP) is a collection of random variables, any finite number of which is a realization of a multivariate normal distribution. The basic idea of GP regression is to consider that the available observations (or realizations) of a variable of interest can be modeled by a GP prior. Suppose that we have m observations $\boldsymbol{\Theta}_m = (\boldsymbol{\theta}(\lambda_j))_{1\leq j\leq m}$ of the variable of interest which are realizations of a GP prior where $\boldsymbol{\theta}(\lambda_j) \in \mathbb{R}$, $\lambda_j \in \mathbb{R}^s (s \geq 1)$ and $D_m = (\lambda_1, \cdots, \lambda_m)^t$ is a numerical design. A GP is fully defined by its mean function $m_\beta(\lambda)$ and its covariance function $\sigma^2 K_\psi(\lambda, \lambda')$. The predictive GP distribution is therefore given by the GP conditioned by the known observations $\boldsymbol{\Theta}_m$. More precisely, this conditional distribution for any new set of $\lambda^\star = (\lambda_1^\star, \cdots, \lambda_k^\star)^t$ can be obtained analytically from the following joint distribution:

$$\begin{pmatrix} \boldsymbol{\theta}(\lambda^\star) \\ \boldsymbol{\Theta}_m \end{pmatrix} \sim \mathscr{N}_{m+k}\left(\begin{pmatrix} m_\beta(\lambda^\star) \\ m_\beta(D_m) \end{pmatrix}, \sigma^2 \begin{pmatrix} K(\lambda^\star, \lambda^\star) & K(\lambda^\star, D_m)^t \\ K(\lambda^\star, D_m) & K(D_m, D_m) \end{pmatrix} \right),$$

where

- $m_\beta(D_m) = (m_\beta(\lambda_1), \cdots, m_\beta(\lambda_m))^t \in \mathbb{R}^m$ is the mean vector of GP evaluated in each location of D_m,
- $K(D_m, D_m) = (K_\psi(\lambda_i, \lambda_j))_{1\leq i,j\leq m} \in \mathbb{R}^{m\times m}$ is the correlation matrix at sample of D_m,
- $K(\lambda^\star, D_m) \in \mathbb{R}^{k\times m}$ is the correlation matrix between $\lambda^\star$ and D_m.

By applying the conditioning formula of Gaussian vectors to the above joint distribution, we obtain that the conditional vector $\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{\Theta}_m$ is still a GP characterized by its mean given by:

$$\bar{\boldsymbol{\theta}}(\lambda^\star) := \mathbb{E}(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{\Theta}_m, \beta, \sigma^2, \psi) = m_\beta(\lambda^\star) + K(\lambda^\star, D_m)K(D_m, D_m)^{-1}(\boldsymbol{\Theta}_m - m_\beta(D_m)),$$

and its covariance function:

$$\Sigma_{\text{pred}}(\lambda^\star, \lambda^{\star\prime}) := \sigma^2(K(\lambda^\star, \lambda^\star) - K(\lambda^\star, D_m)K(D_m, D_m)^{-1}K(D_m, \lambda^{\star\prime})),$$

where the GP hyperparameters $\phi := (\beta, \sigma^2, \psi)$ are not known in practice and have to be estimated. This estimation can be done by marginal likelihood maximization.

2.7.2 Proof of the results of Section 2.4

2.7.2.1 Proof of Theorem 1

Proof: From Bayes rule, we have:

$$\pi(\boldsymbol{\Theta}_m|\mathbf{z}, \phi) = \frac{\mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi)}{\int \mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi)d\boldsymbol{\Theta}_m}.$$

We rewrite the likelihood function as a function of $\vec{\boldsymbol{\Theta}}_m$ and we obtain:

$$\mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m) = \frac{1}{\sqrt{2\pi}^{n_{\text{exp}}m}|\Sigma_\epsilon|^{m/2}}\prod_{j=1}^{m}\exp\big(-\frac{1}{2}\big(\boldsymbol{z} - g_{\lambda_j}(\boldsymbol{x})\boldsymbol{\theta}(\lambda_j)\big)^t\Sigma_\epsilon^{-1}\big(\boldsymbol{z} - g_{\lambda_j}(\boldsymbol{x})\boldsymbol{\theta}(\lambda_j)\big)\big)$$

$$\propto \exp -\frac{1}{2}\big(m\boldsymbol{z}^t\Sigma_\epsilon^{-1}\boldsymbol{z} - 2\sum_{j=1}^{m}\boldsymbol{z}^t\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x})\boldsymbol{\theta}(\lambda_j) + \sum_{j=1}^{m}\boldsymbol{\theta}(\lambda_j)^t g_{\lambda_j}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x})\boldsymbol{\theta}(\lambda_j)\big)$$

$$= \frac{\sqrt{2\pi}^{-n_{\text{exp}}m}}{|\Sigma_\epsilon|^{m/2}}\exp -\frac{1}{2}\big(m\boldsymbol{z}^t\Sigma_\epsilon^{-1}\boldsymbol{z} - 2\boldsymbol{z}^t\Sigma_\epsilon^{-1}G\vec{\boldsymbol{\Theta}}_m + \vec{\boldsymbol{\Theta}}_m^t\boldsymbol{\Delta}^{-1}\vec{\boldsymbol{\Theta}}_m\big),$$

where

$$G = (g_{\lambda_1}(\boldsymbol{x}), \cdots, g_{\lambda_m}(\boldsymbol{x})) \in \mathbb{R}^{n_{\text{exp}}\times pm},$$
$$\boldsymbol{\Delta}^{-1} = \text{diag}\big(g_{\lambda_1}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_1}(\boldsymbol{x}), \cdots g_{\lambda_m}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_m}(\boldsymbol{x})\big) \in \mathbb{R}^{pm\times pm}.$$

Then, we have:

$$\mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi) = \frac{\sqrt{2\pi}^{-n_{\text{exp}}m}}{|\Sigma_\epsilon|^{m/2}}\exp -\frac{1}{2}\big(m\boldsymbol{z}^t\Sigma_\epsilon^{-1}\boldsymbol{z} - 2\boldsymbol{z}^t\Sigma_\epsilon^{-1}G\vec{\boldsymbol{\Theta}}_m + \vec{\boldsymbol{\Theta}}_m^t\boldsymbol{\Delta}^{-1}\vec{\boldsymbol{\Theta}}_m\big)\times$$
$$\frac{1}{\sqrt{2\pi}^{pm}|\mathbf{K}_\phi|^{1/2}}\exp -\frac{1}{2}(\vec{\boldsymbol{\Theta}}_m - \vec{M}_\beta)^t\mathbf{K}_\phi^{-1}(\vec{\boldsymbol{\Theta}}_m - \vec{M}_\beta).$$

By grouping the terms in $\vec{\boldsymbol{\Theta}}_m$, we obtain:

$$\mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi) = \frac{\sqrt{2\pi}^{-m(n_{\text{exp}}+p)}|\mathbf{K}_\phi|^{-1/2}}{|\Sigma_\epsilon|^{m/2}}\exp -\frac{1}{2}\big(m\boldsymbol{z}^t\Sigma_\epsilon^{-1}\boldsymbol{z} + \vec{M}_\beta^t\mathbf{K}_\phi^{-1}\vec{M}_\beta\big)\times$$
$$\exp -\frac{1}{2}\Big[-2\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t\vec{\boldsymbol{\Theta}}_m + \vec{\boldsymbol{\Theta}}_m^t\big(\boldsymbol{\Delta}^{-1} + \mathbf{K}_\phi^{-1}\big)\vec{\boldsymbol{\Theta}}_m\Big]. \quad (2.66)$$

We conclude:

$$\pi(\boldsymbol{\Theta}_m|\mathbf{z}, \phi) \propto \exp -\frac{1}{2}\Big[-2\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t\vec{\boldsymbol{\Theta}}_m + \vec{\boldsymbol{\Theta}}_m^t\big(\boldsymbol{\Delta}^{-1} + \mathbf{K}_\phi^{-1}\big)\vec{\boldsymbol{\Theta}}_m\Big]$$
$$\Rightarrow \pi(\boldsymbol{\Theta}_m|\mathbf{z}, \phi) \sim \mathscr{N}_{pm}(\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z}, \phi], \boldsymbol{\Sigma}_\phi),$$

where

$$\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi] = \boldsymbol{\Sigma}_\phi\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big) \in \mathbb{R}^{pm},$$

$$\boldsymbol{\Sigma}_\phi = \big(\boldsymbol{\Delta}^{-1} + \mathbf{K}_\phi^{-1}\big)^{-1} \in \mathbb{R}^{pm\times pm}.$$

■

2.7.2.2 Proof of Theorem 2

Proof: We consider any new set of realizations $\lambda^\star = (\lambda_1^\star, \cdots, \lambda_k^\star)^t$ and the predictive distribution associated to $\boldsymbol{\theta}(\lambda^\star)$ is obtained by integrating the Gaussian conditional distribution $\pi(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{\Theta}_m,\phi)$ over the posterior probability measure $\pi(\boldsymbol{\Theta}_m|\mathbf{z},\phi)d\boldsymbol{\Theta}_m$:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \int \pi(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{\Theta}_m,\phi)\pi(\boldsymbol{\Theta}_m|\mathbf{z},\phi)d\boldsymbol{\Theta}_m,$$

with $\pi(\boldsymbol{\theta}(\lambda^\star)|\boldsymbol{\Theta}_m,\phi) \sim \mathscr{N}_{pk}(\vec{\mu}_{\text{cond}}(\lambda^\star), \Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}))$ where

$$\begin{aligned}
\vec{\mu}_{\text{cond}}(\lambda^\star) &= \vec{m}_\beta(\lambda^\star) + \mathbf{C}(\lambda^\star, D_m)\mathbf{K}_\phi^{-1}(\vec{\boldsymbol{\Theta}}_m - \vec{M}_\beta) =: \vec{\mu}^\star + X^t\vec{\boldsymbol{\Theta}}_m,\\
\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) &= \mathbf{C}(\lambda^\star,\lambda^{\star\prime}) - \mathbf{C}(\lambda^\star, D_m)\mathbf{K}_\phi^{-1}\mathbf{C}(D_m,\lambda^\star),
\end{aligned}$$

and

$$\begin{aligned}
X^t &:= \mathbf{C}(\lambda^\star, D_m)\mathbf{K}_\phi^{-1},\\
\vec{\mu}^\star &:= \vec{m}_\beta(\lambda^\star) + X^t\vec{M}_\beta.
\end{aligned}$$

Then, we have:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \int \Big\{ \frac{\sqrt{2\pi}^{-pk}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})|^{1/2}} \exp -\frac{1}{2}\Big[\big(\vec{\boldsymbol{\Theta}}(\lambda^\star) - (\vec{\mu}^\star - X^t\vec{\boldsymbol{\Theta}}_m)\big)^t\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}$$
$$\times\big(\vec{\boldsymbol{\Theta}}(\lambda^\star) - (\vec{\mu}^\star + X^t\vec{\boldsymbol{\Theta}}_m)\big)\Big] \frac{\sqrt{2\pi}^{-pm}}{|\boldsymbol{\Sigma}_\phi|^{1/2}} \exp -\frac{1}{2}\Big[\big(\vec{\boldsymbol{\Theta}}_m - \mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)^t\boldsymbol{\Sigma}_\phi^{-1}\big(\vec{\boldsymbol{\Theta}}_m - \mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)\Big] d\boldsymbol{\Theta}_m \Big\}.$$

By setting $\vec{\boldsymbol{\Theta}}^\star := \vec{\boldsymbol{\Theta}}(\lambda^\star) - \vec{\mu}^\star$ and by expanding the expression in the integral while grouping the terms in $\vec{\boldsymbol{\Theta}}_m$, we obtain:

$$\begin{aligned}
\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \int \Big\{ &\frac{\sqrt{2\pi}^{-p(k+m)}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})|^{1/2}|\boldsymbol{\Sigma}_\phi|^{1/2}} \exp -\frac{1}{2}\Big(\vec{\boldsymbol{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\boldsymbol{\Theta}}^\star +\\
&\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]^t\boldsymbol{\Sigma}_\phi^{-1}\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\Big)\\
\times \exp -\frac{1}{2}\Big[&-2\big(\vec{\boldsymbol{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t + \mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]^t\boldsymbol{\Sigma}_\phi^{-1}\big)\vec{\boldsymbol{\Theta}}_m +\\
&\vec{\boldsymbol{\Theta}}_m^t(\boldsymbol{\Sigma}_\phi^{-1} + X\Sigma^{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t)\Big] d\boldsymbol{\Theta}_m \Big\}.
\end{aligned}$$

The expression of $\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi)$ becomes:

$$\begin{aligned}
\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = &\frac{\sqrt{2\pi}^{-pk}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})|^{1/2}|\boldsymbol{\Sigma}_\phi|^{1/2}} \exp -\frac{1}{2}\Big(\vec{\boldsymbol{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\boldsymbol{\Theta}}^\star +\\
&\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]^t\boldsymbol{\Sigma}_\phi^{-1}\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\Big)\\
\times \Big\{ \int \frac{1}{\sqrt{2\pi}^{pm}} \exp -\frac{1}{2}\Big[&-2\big(\vec{\boldsymbol{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t + \mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]^t\boldsymbol{\Sigma}_\phi^{-1}\big)\vec{\boldsymbol{\Theta}}_m +\\
&\vec{\boldsymbol{\Theta}}_m^t(\boldsymbol{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t)\Big] d\boldsymbol{\Theta}_m \Big\}.
\end{aligned}$$

Using the integration formula for multivariate Gaussian densities, the expression $\Big\{\cdots\Big\}$ becomes:

$$|\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}|^{1/2}\exp\frac{1}{2}\Big[\big(X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star + \mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big)^t\times$$
$$\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}\big(X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star + \mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big)\Big].$$

Therefore, we can write:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \frac{\sqrt{2\pi}^{-pk}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})|^{1/2}|\mathbf{\Sigma}_\phi|^{1/2}}\exp-\frac{1}{2}\Big(\vec{\mathbf{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star +$$
$$\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]^t\mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\Big)\times\Big\{|\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}|^{1/2}$$
$$\exp\frac{1}{2}\Big[\big(X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star + \mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big)^t\times$$
$$\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}\big(X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star + \mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big)\Big]\Big\}. \quad (2.67)$$

By (2.65), we have:

$$|\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t| = |\mathbf{\Sigma}_\phi^{-1}||\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}||\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\mathbf{\Sigma}_\phi X|.$$

By replacing it in (2.67), we obtain:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \frac{\sqrt{2\pi}^{-pk}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\mathbf{\Sigma}_\phi X|^{1/2}}\exp-\frac{1}{2}\Big(\vec{\mathbf{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star +$$
$$\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]^t\mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\Big)\exp\frac{1}{2}\Big\{\big(X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star + \mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big)^t\times$$
$$\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}\big(X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star + \mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big)\Big\}. \quad (2.68)$$

We expand the expression $\Big\{\cdots\Big\}$ in (2.68):

$$\vec{\mathbf{\Theta}}^{\star t}\Big[\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\Big]\vec{\mathbf{\Theta}}^\star +$$
$$2\vec{\mathbf{\Theta}}^{\star t}\Big[\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}\mathbf{\Sigma}_\phi^{-1}\Big]\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi] +$$
$$\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]^t\{\mathbf{\Sigma}_\phi^{-1}\big(\mathbf{\Sigma}_\phi^{-1} + X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}\mathbf{\Sigma}_\phi^{-1}\}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]. \quad (2.69)$$

By the Woodbury identity (2.64) (with $Z = \Sigma^{\text{cond}}(\lambda^\star,\lambda^{\star\prime})$, $W = \mathbf{\Sigma}_\phi$, $U = V = X^t$), the expression $\Big[\cdots\Big]$ is equal to:

$$\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1} - \big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\mathbf{\Sigma}_\phi X\big)^{-1}.$$

The expression $\big[\cdots\big]$ becomes:

$$\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big(I_{pm}+\mathbf{\Sigma}_\phi X\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\big)^{-1}\underset{(*)}{=}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\Big[I_{pm}-\mathbf{\Sigma}_\phi X$$

$$\times\big(I_{pk}+\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\mathbf{\Sigma}_\phi X\big)^{-1}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\Big]=\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t$$

$$\times\Big[I_{pm}-\mathbf{\Sigma}_\phi X\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}\Big]=\Big[\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}-$$

$$\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\mathbf{\Sigma}_\phi X\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}\Big]X^t=$$

$$\Big[\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}-$$

$$\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\mathbf{\Sigma}_\phi X\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}\Big]X^t. \quad (2.70)$$

Therefore, the above equality becomes:

$$\Big[\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)-\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\mathbf{\Sigma}_\phi X\Big]\times$$
$$\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}X^t=\Big[I_{pk}+\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\mathbf{\Sigma}_\phi X-$$
$$\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t\mathbf{\Sigma}_\phi X\Big]\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}X^t=\Big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})$$
$$+X^t\mathbf{\Sigma}_\phi X\Big)^{-1}X^t.$$

Note that in step $\underset{(*)}{=}$ of (2.70), we used (2.64) with $Z=I_{pm}$, $W=I_{pk}, U=\mathbf{\Sigma}_\phi X$ and $V^t=\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}X^t$.

By applying the Woodbury identity (2.64) with $Z=\mathbf{\Sigma}_\phi^{-1}$, $W=\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}$, $U=V=X$, the expression $\{\cdots\}$ of (2.69) becomes:

$$\mathbf{\Sigma}_\phi^{-1}\Big[\mathbf{\Sigma}_\phi-\mathbf{\Sigma}_\phi X\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}X^t\mathbf{\Sigma}_\phi\Big]\mathbf{\Sigma}_\phi^{-1}=\mathbf{\Sigma}_\phi^{-1}-X\Big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+$$
$$X^t\mathbf{\Sigma}_\phi X\Big)^{-1}X^t.$$

Therefore, the expression $\Big\{\cdots\Big\}$ of (2.68) is equal to:

$$\vec{\mathbf{\Theta}}^{\star t}\big[\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}-\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}\big]+2\vec{\mathbf{\Theta}}^{\star t}\big[\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}X^t\big]$$
$$\times\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]+\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]^t\{\mathbf{\Sigma}_\phi^{-1}-X\big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})+X^t\mathbf{\Sigma}_\phi X\big)^{-1}X^t\}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi].$$

The above expression is equal to:

$$\vec{\mathbf{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\mathbf{\Theta}}^\star+\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]^t\mathbf{\Sigma}_\phi^{-1}\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]-\big(\vec{\mathbf{\Theta}}^\star-X^t\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big)^t\Big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})$$
$$+X^t\mathbf{\Sigma}_\phi X\Big)^{-1}\big(\vec{\mathbf{\Theta}}^\star-X^t\mathbb{E}[\mathbf{\Theta}_m|\mathbf{z},\phi]\big).$$

As a result, the predictive distribution is given by:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \frac{\sqrt{2\pi}^{-pk}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\boldsymbol{\Sigma}_\phi X|^{1/2}} \exp -\frac{1}{2}\Big(\vec{\boldsymbol{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\boldsymbol{\Theta}}^\star + \mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]^t\boldsymbol{\Sigma}_\phi^{-1}\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\Big) \times \exp\frac{1}{2}\Big\{\vec{\boldsymbol{\Theta}}^{\star t}\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime})^{-1}\vec{\boldsymbol{\Theta}}^\star + \mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]^t\boldsymbol{\Sigma}_\phi^{-1}\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi] - \big(\vec{\boldsymbol{\Theta}}^\star - X^t\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)^t \times \big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\boldsymbol{\Sigma}_\phi X\big)^{-1}\big(\vec{\boldsymbol{\Theta}}^\star - X^t\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)\Big\}.$$

Finally, the predictive distribution becomes:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \frac{\sqrt{2\pi}^{-pk}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\boldsymbol{\Sigma}_\phi X|^{1/2}} \exp\frac{1}{2}\Big\{ - \big(\vec{\boldsymbol{\Theta}}^\star - X^t\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)^t \times \big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\boldsymbol{\Sigma}_\phi X\big)^{-1}\big(\vec{\boldsymbol{\Theta}}^\star - X^t\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)\Big\},$$

which can be rewritten as:

$$\pi_{\text{pred}}(\boldsymbol{\theta}(\lambda^\star)|\mathbf{z},\phi) = \frac{\sqrt{2\pi}^{-pk}}{|\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\boldsymbol{\Sigma}_\phi X|^{1/2}} \exp -\frac{1}{2}\Big\{\big(\vec{\boldsymbol{\Theta}}(\lambda^\star) - \vec{\mu}^\star - X^t\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)^t \times \big(\Sigma_{\text{cond}}(\lambda^\star,\lambda^{\star\prime}) + X^t\boldsymbol{\Sigma}_\phi X\big)^{-1}\big(\vec{\boldsymbol{\Theta}}(\lambda^\star) - \vec{\mu}^\star - X^t\mathbb{E}[\boldsymbol{\Theta}_m|\mathbf{z},\phi]\big)\Big\}.$$

■

2.7.2.3 Marginal likelihood expression and hyperparameters tuning

The optimal hyperparameters $\hat{\phi}$ are obtained by maximizing the marginal likelihood:

$$\widehat{\phi} := \underset{\phi=(\beta_l,\sigma_l^2,\psi_k)_{l=1}^p}{\text{argmax}} \int \mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi)d\boldsymbol{\Theta}_m. \tag{2.71}$$

2.7.2.3.1 Marginal likelihood expression

Proof: Starting from Equation (2.66), the above integral is equal to:

$$\int \mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi)d\boldsymbol{\Theta}_m = \frac{\sqrt{2\pi}^{-mn_{\text{exp}}}|\mathbf{K}_\phi|^{-1/2}}{|\Sigma_\epsilon|^{m/2}} \exp -\frac{1}{2}\big(m\boldsymbol{z}^t\Sigma_\epsilon^{-1}\boldsymbol{z} + \vec{M}_\beta^t\mathbf{K}_\phi^{-1}\vec{M}_\beta\big)\times \Big(\int \frac{1}{\sqrt{2\pi}^{pm}} \exp -\frac{1}{2}\Big[- 2\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t\vec{\boldsymbol{\Theta}}_m + \vec{\boldsymbol{\Theta}}_m^t\boldsymbol{\Sigma}_\phi^{-1}\vec{\boldsymbol{\Theta}}_m\Big]d\boldsymbol{\Theta}_m\Big),$$

where

$$(\cdots) = |\boldsymbol{\Sigma}_\phi|^{1/2}\exp -\frac{1}{2}\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t\boldsymbol{\Sigma}_\phi\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big).$$

Therefore, we write:

$$\int \mathscr{L}(\mathbf{z}|\boldsymbol{\Theta}_m)\pi(\boldsymbol{\Theta}_m|\phi)d\boldsymbol{\Theta}_m = \frac{\sqrt{2\pi}^{-mn_{\text{exp}}}|\mathbf{K}_\phi|^{1/2}}{|\Sigma_\epsilon|^{m/2}}|\boldsymbol{\Sigma}_\phi|^{1/2}\exp -\frac{1}{2}\Big[m\boldsymbol{z}^t\Sigma_\epsilon^{-1}\boldsymbol{z} + \vec{M}_\beta^t\mathbf{K}_\phi^{-1}\vec{M}_\beta - \big\{\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t\boldsymbol{\Sigma}_\phi\big(\mathbf{K}_\phi^{-1}\vec{M}_\beta + G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)\big\}\Big],$$

where

$$\{\cdots\} = \vec{M}_\beta^t\mathbf{K}_\phi^{-1}\boldsymbol{\Sigma}_\phi\mathbf{K}_\phi\vec{M}_\beta + 2\vec{M}_\beta^t\mathbf{K}_\phi\boldsymbol{\Sigma}_\phi G^t\Sigma_\epsilon^{-1}\boldsymbol{z} + \boldsymbol{z}^t\Sigma_\epsilon^{-1}G\boldsymbol{\Sigma}_\phi G^t\Sigma_\epsilon^{-1}\boldsymbol{z}.$$

By using the Sherman-Morrisson formula given by (2.63) (with $A = \mathbf{K}_\phi^{-1}$ and $B = \boldsymbol{\Delta}^{-1}$), we obtain:

$$\vec{M}_\beta^t\mathbf{K}_\phi^{-1}\boldsymbol{\Sigma}_\phi\mathbf{K}_\phi^{-1}\vec{M}_\beta = \vec{M}_\beta^t\Big(\mathbf{K}_\phi^{-1} - (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1}\Big)\vec{M}_\beta,$$

$$2\vec{M}_\beta^t\mathbf{K}_\phi^{-1}\mathbf{\Sigma}_\phi G^t\Sigma_\epsilon^{-1}\boldsymbol{z} = 2\vec{M}_\beta^t\mathbf{K}_\phi^{-1}\mathbf{\Sigma}_\phi\mathbf{\Delta}^{-1}\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} = 2\vec{M}_\beta^t\mathbf{K}_\phi^{-1}\big(\mathbf{\Delta}^{-1}+\mathbf{K}_\phi^{-1}\big)^{-1}\mathbf{\Delta}^{-1} \times \mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} = 2\vec{M}_\beta^t(\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z},$$

and

$$\boldsymbol{z}^t\Sigma_\epsilon^{-1}G\mathbf{\Sigma}_\phi^{-1}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} = \boldsymbol{z}^t\Sigma_\epsilon^{-1}G\mathbf{\Delta}\mathbf{\Delta}^{-1}\big(\mathbf{\Delta}^{-1}+\mathbf{K}_\phi^{-1}\big)^{-1}\mathbf{\Delta}^{-1}\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} \underset{(\star)}{=} \boldsymbol{z}^t\Sigma_\epsilon^{-1}G\mathbf{\Delta} \times \big(\mathbf{\Delta}^{-1}-(\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\big)\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}.$$

Note that at step $\underset{(\star)}{=}$, we use again the Sherman-Morrisson formula given by (2.63) (with $A=\mathbf{\Delta}^{-1}$ and $B=\mathbf{K}_\phi^{-1}$). Therefore, we have:

$$\{\cdots\} = \vec{M}_\beta^t\mathbf{K}_\phi^{-1}\vec{M}_\beta - \vec{M}_\beta^t(\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\vec{M}_\beta + 2\vec{M}_\beta^t(\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} + \boldsymbol{z}^t\Sigma_\epsilon^{-1}G\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} -$$

$$\boldsymbol{z}^t\Sigma_\epsilon^{-1}G\mathbf{\Delta}(\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} = \vec{M}_\beta^t\mathbf{K}_\phi^{-1}\vec{M}_\beta + \boldsymbol{z}^t\Sigma_\epsilon^{-1}G\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z} - \big(\vec{M}_\beta - \mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t(\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\big(\vec{M}_\beta - \mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big).$$

As a result, the marginal likelihood is equal to:

$$\int \mathscr{L}(\mathbf{z}|\mathbf{\Theta}_m)\pi(\mathbf{\Theta}_m|\phi)d\mathbf{\Theta}_m = \frac{\sqrt{2\pi}^{-mn_{\exp}}|\mathbf{K}_\phi|^{-1/2}}{|\Sigma_\epsilon|^{m/2}}|\mathbf{\Sigma}_\phi|^{1/2}\exp -\frac{1}{2}\Big[\big(\vec{M}_\beta - \mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t\times (\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\big(\vec{M}_\beta-\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)\Big] = \exp -\Big[\frac{1}{2}\big(\vec{M}_\beta-\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)^t\times(\mathbf{\Delta}+\mathbf{K}_\phi)^{-1}\big(\vec{M}_\beta-\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}\big)\Big]\times \frac{|\mathbf{\Delta}|^{1/2}}{\sqrt{2\pi}^{n_{\exp}m}|\Sigma_\epsilon|^{m/2}|\mathbf{\Delta}+\mathbf{K}_\phi|^{1/2}},$$

because

$$m\boldsymbol{z}^t\Sigma_\epsilon^{-1}\boldsymbol{z} = \boldsymbol{z}^t\Sigma_\epsilon^{-1}G\mathbf{\Delta}G^t\Sigma_\epsilon^{-1}\boldsymbol{z}.$$

In fact, we have:

$$\begin{aligned}\Sigma_\epsilon^{-1}G\mathbf{\Delta}G^t &= \Sigma_\epsilon^{-1}\sum_{j=1}^m g_{\lambda_j}(\boldsymbol{x})(g_{\lambda_j}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x}))^{-1}g_{\lambda_j}(\boldsymbol{x})^t\\ &= \sum_{j=1}^m \Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x})(g_{\lambda_j}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x}))^{-1}g_{\lambda_j}(\boldsymbol{x})^t\\ &= mI_{n_{\exp}}.\end{aligned}$$

The last equality can be proved by contradiction. Indeed, if

$$\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x})(g_{\lambda_j}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x}))^{-1}g_{\lambda_j}(\boldsymbol{x})^t \neq I_{n_{\exp}},$$

then, we have:

$$g_{\lambda_j}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x})(g_{\lambda_j}(\boldsymbol{x})^t\Sigma_\epsilon^{-1}g_{\lambda_j}(\boldsymbol{x}))^{-1}g_{\lambda_j}(\boldsymbol{x})^t \neq g_{\lambda_j}(\boldsymbol{x})^t,$$

which yields the absurd result : $g_{\lambda_j}(\boldsymbol{x})^t \neq g_{\lambda_j}(\boldsymbol{x})^t, \forall\, j$. ■

2.7.2.3.2 Hyperparameters tuning Equation (2.71) gives the optimal hyperparameters:

$$\widehat{\phi} := \underset{\phi=(\beta_l,\sigma_l^2,\psi_k)_{l=1}^p}{\operatorname{argmax}} \mathbf{L}(\phi),$$

where $\mathbf{L}(\phi)$ is the marginal likelihood. Taking the opposite of the marginal log likelihood, we obtain:

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}}\ \boldsymbol{\ell}(\phi), \tag{2.72}$$

where

$$\boldsymbol{\ell}(\phi) = (\vec{M}_\beta - \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z})^t (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1} (\vec{M}_\beta - \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z}) - \log|\boldsymbol{\Delta}| + n_{\text{exp}} m \log(2\pi) + m \log|\Sigma_\epsilon| + \log|\boldsymbol{\Delta} + \mathbf{K}_\phi|.$$

Suppose that $\vec{M}_\beta \in \mathbb{R}^{pm}$ is of the following form:

$$\vec{M}_\beta = \mathbf{H}\beta,$$

where $\beta = (\beta_1, \cdots, \beta_p)^t \in \mathbb{R}^{q_1}$ with $q_1 \geq p$ et $\mathbf{H} \in \mathbb{R}^{pm \times q_1}$. Then, the optimum in β is given by the following proposition:

Proposition 1

The optimum in β is equal to:

$$\hat{\beta}(\sigma^2, \psi) = (\mathbf{H}^t (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1} \mathbf{H})^{-1} \mathbf{H}^t (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1} \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z},$$

where $\sigma^2 = (\sigma_1^2, \cdots, \sigma_p^2)^t \in \mathbb{R}_+^p$ and $\psi = (\psi_1, \cdots, \psi_p)^t \in \mathbb{R}^{q_2}$ with $q_2 \geq p$.

Therefore, the optimization problem given in Equation (2.72) becomes:

$$\hat{\sigma}^2, \hat{\psi} = \underset{\sigma^2, \psi}{\text{argmin}}\ \boldsymbol{\ell}(\hat{\beta}(\sigma^2, \psi), \sigma^2, \psi). \tag{2.73}$$

This optimization problem will be solved by using numerical optimization algorithms.

Proof of proposition 1

Proof: Let us derive $\boldsymbol{\ell}$ with respect to β using chain rule derivative:

$$\begin{aligned}
\frac{\partial \ell(\beta, \sigma^2, \psi)}{\partial \beta} &= \frac{\partial \ell(\beta, \sigma^2, \psi)}{\partial (\mathbf{H}\beta - \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z})} \times \frac{\partial (\mathbf{H}\beta - \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z})}{\partial \beta} \\
&= \frac{\partial \ell(\beta, \sigma^2, \psi)}{\partial (\mathbf{H}\beta - \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z})} \times \mathbf{H} \\
&= 2(\mathbf{H}\beta - \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z})^t (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1} \mathbf{H}.
\end{aligned}$$

Then, we have:

$$\begin{aligned}
&\frac{\partial \ell(\beta, \sigma^2, \psi)}{\partial \beta} = 0 \\
&\iff (\mathbf{H}\beta - \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z})^t (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1} \mathbf{H} = 0 \\
&\iff \mathbf{H}^t (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1} \mathbf{H}\beta = \mathbf{H}^t (\boldsymbol{\Delta} + \mathbf{K}_\phi)^{-1} \boldsymbol{\Delta} G^t \Sigma_\epsilon^{-1} \boldsymbol{z}.
\end{aligned}$$

■

Chapter 3

Kernel-based parameter screening for Bayesian calibration

Contents

Abstract

Numerical simulation is widely used in many fields of engineering to study complex physical systems. The numerical models, designed to represent faithfully physical phenomena under study, are affected by uncertainties of different natures (numerical, stochastic or epistemic) that affect the accuracy of numerical models. Some of these uncertainties can be reduced and quantified by a model calibration process such as deterministic or Bayesian calibration. The letter quantifies and reduces such uncertainties by means of probability distributions. Additionally, Global sensitivity analysis allow us to quantify how input model uncertainties impact the variability of the quantity of interest. In this work, we propose a new global sensitivity analysis method based on the Hilbert-Schmidt Independence Criterion (HSIC) in support of Bayesian calibration in the context of two chained physical models. Our specific goal is to identify the calibration parameters $\boldsymbol{\theta}$ of the second model, taking into account all uncertainties associated with the parameters λ of the first model. To achieve this, our method incorporates the uncertainty associated with λ into the computation of the HSIC indices. Our results show that the independence tests associated with this method can identify the calibration parameters $\boldsymbol{\theta}$ to be considered in the Bayesian calibration process. Furthermore, we demonstrate that the corresponding estimators are consistent and have convergence rates comparable to those of traditional Monte Carlo estimators. Finally, we have applied this new approach in nuclear fuel simulation, to find the most influential calibration parameters $\boldsymbol{\theta}$ of a fission gas behavior model while accounting for the uncertainty of the thermal conductivity λ of an upstream thermal model.

Keywords. Bayesian calibration, global sensitivity analysis, surrogate modeling, dependence measures, Hilbert-Schmidt independence criterion, independence testing.

3.1 Introduction

Like the power generation industry, many high-impact industrial sectors rely on numerical simulations to investigate complex physical phenomena. The numerical models used to mimic the behavior of systems or facilities often involve a large number of control variables, such as experimental conditions or geometric design specifications [Santner et al., 2003]. In addition, another sort of inputs, called calibration parameters, are also required to run the simulations. Being generally not measurable with an experimental setup, such parameters are tainted by epistemic uncertainty [Campbell, 2006]. In a formal mathematical framework, a numerical model is represented by a function $y : (\boldsymbol{x}, \boldsymbol{\theta}) \in \mathscr{D}_{\boldsymbol{X}} \times \mathscr{D}_{\boldsymbol{\Theta}} \subset \mathbb{R}^d \times \mathbb{R}^p \longrightarrow \mathbb{R}^q$ $(q \geq 1)$ denoting the set of control variables and calibration parameters respectively. When there is a need to insist that the calibration parameters are fixed, the notation $y_{\boldsymbol{\theta}}(\cdot)$ will be used instead of $y(\cdot, \boldsymbol{\theta})$. This model serves as an abstract representation of the studied physical system. The experimental observations, when available, are denoted by $\{(\boldsymbol{x}_i, \boldsymbol{z}_i)\}_{1 \leq i \leq n_{\text{exp}}}$, with n_{exp} being the number of observations.

In the context of uncertainty quantification [De Rocquigny et al., 2008], model calibration is a crucial step to enhance the accuracy and robustness of numerical simulations [Bachoc et al., 2014; Santner et al., 2018]. The calibration step aims to set the optimal parameter value(s) for $\boldsymbol{\theta}$ so that the simulator y sticks as closely as possible to the underlying physical reality. To do so, some experimental data of the system must be available. There are basically two types of calibration: deterministic calibration and Bayesian calibration [Kennedy and O'Hagan, 2001; Campbell, 2006; Trucano et al., 2006; Wong et al., 2017; Gu and Wang, 2018]. The first one tries to find the optimal value(s) of $\boldsymbol{\theta}$ by minimizing a loss criterion (most often based on the least-square loss function) which quantifies the discrepancy between the available experimental data and the corresponding simulation data [Wong et al., 2017; Gu and Wang, 2018]. On the contrary, if using the second approach, the uncertainty related to calibration parameters is studied in a purely probabilistic framework, and the theory of Bayesian statistics allow us to extract maximum information from the experimental data [Kennedy and O'Hagan, 2001; Damblin, 2015; Wu et al., 2018; Gu and Wang, 2018; Willmann et al., 2022].

More precisely, Bayesian calibration is a framework that combines prior knowledge, experimental data and a probabilistic equation linking these experimental data to the outputs

of the numerical model in order to derive a more accurate estimate of the unknown model parameters. This approach is particularly useful when data are limited or subject to uncertainty, as it allows using all the existing sources of information (expert opinions and experimental data) to ultimately provide a robust and informative estimate of calibration parameters. As already said, Bayesian calibration is based on three ingredients. First, the prior distribution, $\pi(\boldsymbol{\theta})$ quantifies the prior knowledge on $\boldsymbol{\theta}$. This prior information often results from elicitation of expert judgment and/or statistician choices [Kass and Wasserman, 1996]. Second, the implementation of Bayesian calibration is conditioned upon the availability of a set of experimental data $\boldsymbol{z}$. These data contain valuable information, but they also carry a certain amount of uncertainty (mainly due to measurement errors). The last ingredient is to set a probabilistic equation that enables us to assess the plausibility of the experimental data conditionally to any value of the calibration parameters. The Bayes' theorem can then be applied to derive the posterior probability distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$ which is interpreted as an update to the prior distribution $\pi(\boldsymbol{\theta})$ in view of the experimental data $\boldsymbol{z}$ [Kennedy and O'Hagan, 2001]. In the end, the maximum a posteriori (MAP), i.e. the mode of $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$, is most commonly used to run the calibrated numerical model.

In the general case of a complex non-linear numerical model, the posterior probability distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$ cannot be calculated analytically. However, as this distribution is known up to a constant, it can be simulated with Markov chain Monte Carlo (MCMC) methods [Robert et al., 1999]. Among them, the Metropolis-Hastings algorithm (MH) is widely used in practice [Chib and Greenberg, 1995]. The MH algorithm implies an intensive use of the numerical model y (thousands of model evaluations), which can be an issue if the simulation code is CPU-time expensive. A standard approach to limit model calls, in practice, is to replace the model in the MCMC sampler with an inexpensive mathematical function $\widehat{y}$, called metamodel (or surrogate model) [Currin et al., 1991][1]. When the number p of parameters to be calibrated becomes larger, additional difficulties arise. In particular, the mixing time (i.e., the waiting time before the Markov chain is sufficiently close to the steady state distribution) becomes excessively long, which dramatically increases the computational burden. To remedy this problem, adaptive variants of the MH algorithm [Andrieu and Thoms, 2008; Liu et al., 2009] make it possible to progressively update the proposal distribution, and thus redress the mixing properties. However, such adaptive MH algorithms are complex and difficult to tune efficiently. For this reason, in many recent works dedicated to high-dimensional Bayesian calibration, a preliminary step of dimension reduction is completed before using the (classic) MH algorithm. Bayesian calibration is thus limited to a subset of parameters which are selected by means of a sensitivity analysis (SA) procedure [Van Oijen et al., 2011; Zambrano-Bigiarini et al., 2013; Xu et al., 2016; Wu et al., 2017, 2019; Nagel et al., 2020; Gou et al., 2020; Dighe et al., 2022; Perret et al., 2022].

The goal pursued by a SA is to understand how the uncertainty affecting an output of interest can be apportioned between several sources of uncertainty. In the specific context of an input-output computer code, performing a SA means estimating a collection of sensitivity indices/measures (at least one index per input) from the available simulation data. The results of a SA can be used to rank the inputs (i.e., to establish a clear hierarchy of the influential inputs on the output) [Hamby, 1993, 1994, 1995] and/or to screen the inputs (i.e. to discard or at least put less focus on the non-influential inputs) [Morris, 1991; Morris and Mitchell, 1995; Campolongo et al., 2007, 2011]. Unlike local methods that only measure sensitivity in the neighborhood of a nominal point [Morris, 1991], global sensitivity analysis (GSA) endeavors to account for the entire (joint) distribution of all sources of uncertainty [Saltelli et al., 2008]. In the 40 years that SA has been around, very different approaches have emerged, such as those based on the decomposition of the output variance [Sobol, 2001;

[1] This metamodel is built from a training set (obtained by spending the computational budget) and must be sufficiently representative of y (over the whole variation domain) [Sacks et al., 1989; Currin et al., 1991; Santner et al., 2003].

Saltelli, 2002], the output gradient [Kucherenko et al., 2009], regression strategies [Clouvel et al., 2023], cooperative game strategies [Owen and Prieur, 2017; Iooss and Prieur, 2019] or kernels methods [Da Veiga, 2015; De Lozzo and Marrel, 2016]. For more details, the reader can refer to the following monographs Saltelli et al. [2008] and Da Veiga et al. [2021]. For a given problem, the most suitable SA method depends on many factors, including the number of inputs, the simulation budget (i.e., the maximum number of simulations and therefore the number of input-output samples), the presence of correlations between the inputs and the mathematical nature of the output object, the given-data context [2], and able to adapt to both scalar and vector cases.

Let us now present the application context to better highlight its specificities and better understand what motivates the second part of this work. The ALCYONE fuel application, is composed of interdependent physical models that imitate the mechanical, thermal and chemical behaviors of fuel rods in the core of pressurized water reactors [Michel et al., 2021]. In this work, we are only focused on the chaining of two of them: the thermal model and the fission gas behavior model. The thermal model simulates the evolution of the temperature within the fuel rod during the fission reaction in the core and provides as output, the associated temperature field. This temperature field is subject to uncertainties because it depends on an uncertain quantity called thermal conductivity whose probability distribution has been estimated from experimental data of thermal component [Bouloré et al., 2023]. Then, the role of the gas behavior model is to simulate the behavior of the fission products (fuel swelling and release of fission gas atoms). It provides several (scalar) outputs, including the one of interest here, namely the released gas fraction. The fission gas behavior model also depends on another uncertain parameter called λ, which is the thermal conductivity of the temperature model, whose epistemic uncertainty is provided in [Bouloré et al., 2023]. In contrast, the uncertainty of $\boldsymbol{\theta}$ remains to be estimated and should be impacted by that of λ. Estimating the posterior distribution of $\boldsymbol{\theta}$ conditional on λ can be difficult when the dimension of $\boldsymbol{\theta}$ is high (a dozen). Therefore, our goal in this paper is to reduce this dimension of $\boldsymbol{\theta}$ by a GSA while taking into account the uncertainty brought by λ. To do this, we first need a sensitive measure that can take account of this double uncertainty: that of $\boldsymbol{\theta}$ and that of λ. Next, the estimation of the associated indices should be as rapidly as feasible. Then, the possibility of setting up a decision rule procedure that will determine whether each component of $\boldsymbol{\theta}$ is influential or not. Therefore, we focus on indices based on the Hilbert-Schmidt Independence Criterion (HSIC) [Da Veiga, 2015].

This paper is intended to contribute to:

- The proposal of a sensitivity measure that takes into account the uncertainty of the thermal conductivity.
- The construction of consistent estimators for this new sensitivity measure, which incorporates the uncertainty of the thermal conductivity λ. Both associated independence tests and p-values will also be presented to decide whether or not each calibration parameter is influential.

The outline of the paper is as follows. Section 3.2 recalls the main principles of model calibration, and highlights why parameter screening is an important step in high-dimensional Bayesian calibration. The difficulties inherent to this specific GSA problem are laid out and we emphasize that most classic GSA methods fail to cope with them. Section 3.3 provides some mathematical reminders on the HSIC, before showing that the associated sensitivity indices are well suited to screening-oriented GSA, and perfectly meets the additional specificities of Bayesian calibration. In Section 3.4, in order to illustrate the lessons from Section 3.3 in the simplest possible context, the thermal conductivity is set to its nominal value, and

[2] The given-data context is the context in which the data for a given physical system are imposed in order to perform a statistical study, such as a sensitivity analysis study. These data are often very limited in number, about a few hundred.

the study is confined to the fission gas model only. GSA is then performed in three different ways, but always using HSIC indices. Section 3.5, where it is shown how to adapt the HSIC methodology to the particular context of chained codes, is the heart of this work. In Section 3.6, the methodology proposed in Section 3.5 is applied to ALCYONE, and more particularly to the chaining between the thermal and fission gas behavior models. An integrated version of the HSIC is notably used to incorporate the uncertainty coming from the Bayesian calibration of the first numerical model. In particular, it is demonstrated that the resulting sensitivity indices properly characterize independence in this bi-level uncertainty framework, which underlines their usefulness from a screening perspective. Then, four different estimators are proposed and their statistical properties are deeply investigated. Since the values taken by such kernel-based indices are always difficult to interpret, the final decision (regarding which parameters to select) is only trustworthy if it is supported by tests of independence. For this reason, two test procedures, which are much inspired from what is done for classic HSIC indices, are developed in the last part of Section 3.5. Section 3.7 concludes the paper.

3.2 Parameter screening for Bayesian calibration

3.2.1 Bayesian calibration for black-box numerical models

A physical system of interest, represented as $r(\boldsymbol{x}) \in \mathscr{D}_{\boldsymbol{Y}} \subset \mathbb{R}^q$ ($q \geq 1$), can be viewed as a mathematical function characterizing the possible relationship between the input vector $\boldsymbol{x}$ and the output $r(\boldsymbol{x})$. The input vector, $\boldsymbol{x} \in \mathscr{D}_{\boldsymbol{X}} \subset \mathbb{R}^d$, mainly contains variables known as control variables [Santner et al., 2003]. Most often, these control variables represent some experimental conditions and some geometric specifications of the physical system. Let $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ be a deterministic numerical model intended to faithfully simulate $r(\boldsymbol{x})$. The general probabilistic equation that links the experimental data of the physical system $\boldsymbol{z} = \{\boldsymbol{z}_i\}_{i=1}^{n_{\text{exp}}}$ to the model outputs $\{y_{\boldsymbol{\theta}}(\boldsymbol{x}_i)\}_{i=1}^{n_{\text{exp}}}$ for input experimental configurations $\{\boldsymbol{x}_i\}_{i=1}^{n_{\text{exp}}}$ is given by:

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta}}(\boldsymbol{x}_i) + b(\boldsymbol{x}_i) + \epsilon_i, \quad 1 \leq i \leq n_{\text{exp}}. \tag{3.1}$$

Here, ϵ_i represents the experimental uncertainty, often assumed to follow a zero mean multivariate Gaussian distribution with a known covariance matrix $\Sigma_{\epsilon_i} = \text{diag}(\sigma^2_{\epsilon_{i,1}}, \cdots, \sigma^2_{\epsilon_{i,q}})$. The term $b(\boldsymbol{x})$, called the model discrepancy in [Kennedy and O'Hagan, 2001], is the gap between $y_{\boldsymbol{\theta}}(\boldsymbol{x})$ and $r(\boldsymbol{x})$ when the numerical model is executed at the optimal but unknown calibration parameter value $\boldsymbol{\theta} \in \mathscr{D}_{\boldsymbol{\Theta}} \subset \mathbb{R}^p$ [Kennedy and O'Hagan, 2001; Wu et al., 2018]. In general, the model discrepancy $b(\boldsymbol{x})$ is modeled by a Gaussian process [Kennedy and O'Hagan, 2001; Bachoc et al., 2014; Wu et al., 2018; Gu and Wang, 2018], but, in many cases, it is neglected and assumed to be part of the experimental uncertainty [Damblin et al., 2018; Willmann et al., 2022]. Then, we consider the following simplified probabilistic equation:

$$\boldsymbol{z}_i = y_{\boldsymbol{\theta}}(\boldsymbol{x}_i) + \epsilon_i, \quad 1 \leq i \leq n_{\text{exp}}. \tag{3.2}$$

The posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$ is computed from Bayes' formula:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{z}) \propto \mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{3.3}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution quantifying the uncertainty of the calibration parameters before collecting $\boldsymbol{z}$. The likelihood $\mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta})$ is a measure of the correspondence between the observed data $\boldsymbol{z}$ and the outputs of the model $y(\boldsymbol{x}, \boldsymbol{\theta})$ for a given value of the parameters $\boldsymbol{\theta}$. Therefore, it is a crucial indicator of the level of plausibility of $\boldsymbol{\theta}$ in view of $\boldsymbol{z}$. In this context, the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$ is an update of the prior distribution $\pi(\boldsymbol{\theta})$ which is obtained after leveraging the additional information brought by $\boldsymbol{z}$.

In general case, the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$ is known up to a constant. As a consequence, its estimation most often requires Markov Chain Monte Carlo (MCMC) algorithms

[Robert et al., 1999] such as Metropolis-Hastings (MH) algorithm [Chib and Greenberg, 1995; Andrieu and Thoms, 2008]. The aim of the MH algorithm is to simulate a sequence of samples $\{\boldsymbol{\theta}^{(l)}\}_{l=1}^{N_{\text{MH}}}$[3] whose distribution converges towards the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$. MH algorithm then starts with an initial state $\boldsymbol{\theta}^{(0)} \sim \pi(\boldsymbol{\theta})$. It generates a new state proposal $\boldsymbol{\theta}^{\text{new}}$ at each iteration $1 \leq l \leq N_{\text{MH}}$, using a given proposal distribution. The acceptance of this proposal depends on a probability determined by comparing the ratio of the likelihoods computed at $\boldsymbol{\theta}^{\text{new}}$ and the current state $\boldsymbol{\theta}^{(l)}$ weighted by the density ratio of the proposal distributions at $\boldsymbol{\theta}^{\text{new}}$ and $\boldsymbol{\theta}^{(l)}$. If this ratio is greater than a random number generated uniformly on $[0, 1]$, the proposition is accepted as (the) new state (i.e., $\boldsymbol{\theta}^{(l+1)} = \boldsymbol{\theta}^{\text{new}}$), otherwise the current state is unchanged (i.e., $\boldsymbol{\theta}^{(l+1)} = \boldsymbol{\theta}^{(l)}$). By proceeding iteratively, this mechanism produces a Markov chain that converges to the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{z})$. It is important to note that this method efficiently explores the space $\mathscr{D}_\theta$ of the calibration parameters $\boldsymbol{\theta}$. At each iteration, the computation of the likelihood requires the model y to be evaluated at all the experimental configurations $\boldsymbol{x} = \{x_i\}_{i=1}^{n_{\text{exp}}}$ for each new value $\boldsymbol{\theta}^{\text{new}}$. A significant challenge thus arises from the computational expense associated with the numerical model y. To overcome this issue, a commonly employed approach is to replace $y(\cdot, \cdot)$ with a surrogate model, also referred to as a metamodel (or emulator) and denoted by $\widehat{y}$ [Sacks et al., 1989; Currin et al., 1991; Santner et al., 2003]. This substitution allows the computational burden to be greatly reduced while maintaining essential features of the original model.

Once a surrogate model $\widehat{y}$ is fitted to the available training data set (composed of simulated data), the output can be predicted for any new value of $\boldsymbol{x}$ and $\boldsymbol{\theta}$. This approach is particularly useful when a probabilistic model is available for the control variables and the best type of surrogate model depends on the mathematical nature of the control variables [Wu et al., 2018; Damblin et al., 2018]. Sometimes, a surrogate model $\widehat{y}_i : \mathscr{D}_{\boldsymbol{\Theta}} \to \mathscr{D}_{\boldsymbol{Y}}$ is built for each function $y_i(\boldsymbol{\theta}) := y_{\boldsymbol{\theta}}(\boldsymbol{x}_i)$, which is obtained by restricting y to a specific experimental configuration. This metamodeling strategy becomes particularly relevant when the experimental configurations are significantly different from each other, such as for example in nuclear engineering [Damblin and Gaillard, 2020]. However, a drawback of this approach is the necessity to train as many surrogates as there are experimental configurations (here n_{exp}). In some situations, it is even better to build a surrogate model for the likelihood $\mathscr{L}(\boldsymbol{z} \mid \boldsymbol{\theta})$ than for the numerical model y itself [Lebel et al., 2019]. Gaussian process (GP) regression appears to be the most popular approach [Rasmussen et al., 2006; Gramacy, 2020]. Polynomial chaos expansion (PCE) metamodels are also promoted in the literature [Nagel et al., 2020; Dighe et al., 2022].

Unfortunately, everything becomes much harder when the number of calibration parameters exceeds several dozens. This difficulty is very common in many application fields, as illustrated in [Wu et al., 2017; Van Oijen et al., 2011; Dighe et al., 2022] where $p \in \{36, 48, 11\}$ respectively. In particular, the increase in the number of calibration parameters raises two majors issues:

- Building a metamodel with good capabilities turns out to be very difficult because the number of hyperparameters to be optimized grows too fast. For GP models, various strategies have been explored in recent years to overcome the limitations of standard GP regression (advanced screening techniques in [Iooss and Marrel, 2019; Marrel et al., 2022], estimation algorithm based on multi-objective constrained optimization in [Marrel and Iooss, 2023], construction of an active subspace in [Zhou et al., 2022]). A review of existing solutions is proposed in [Binois and Wycoff, 2022].

- Inadequate mixing within the MCMC chain, also known as poor mixing, which occurs when the Markov chain generated by the MH algorithm fails to explore the entire parameter space $\mathscr{D}_{\boldsymbol{\Theta}}$. This is often due to the challenge of tuning the proposal distribution of the MH algorithm itself [Haario et al., 2001; Andrieu and Thoms, 2008; Liu

[3] N_{MH} is typically large (i.e. $N_{\text{MH}} \geq 10^6$).

et al., 2009] especially in high dimension. This poor mixing can lead to potential identifiability problems of the calibration parameters, resulting in uninformative posterior distributions characterized by highly correlated marginals [Liu et al., 2009]. Adaptive algorithms have been proposed in the literature to overcome this problem [Haario et al., 2001; Andrieu and Thoms, 2008].

For both problems, the simplest solution to implement is to select the most influential parameters and set all others to their nominal values. This allows us to recover a low-dimensional problem, and two difficulties mentioned above are considerably mitigated. In order to be rigorous in the choice of relevant parameters, it is often advisable to make a decision based on the results of a GSA.

3.2.2 Description of the GSA problem arising from Bayesian calibration

In the Bayesian calibration literature, GSA techniques are often used to identify the relevant model parameters to be calibrated in order to reduce the complexity of MH algorithm and thus to avoid the potential identifiability problems [Van Oijen et al., 2011; Zambrano-Bigiarini et al., 2013; Xu et al., 2016; Wu et al., 2017, 2019; Nagel et al., 2020; Gou et al., 2020; Dighe et al., 2022; Perret et al., 2022]. For example Wu et al. [2019] performed a dimension reduction of the model parameters by GSA technique before the calibration process. In their case, the numerical model is not CPU intensive (so there is no need to use a metamodel), so they use GSA only to facilitate the use of the MH algorithm. This GSA task is performed by computing Sobol' indices [Sobol, 1993]. Using practical examples in nuclear engineering, Wu et al. [2019] show that this dimension reduction leads to a better identifiability of the calibration parameters.

When it arises in the context of Bayesian calibration, the GSA problem presents specificities which must be kept in mind:

- The declared objective is only to screen calibration parameters. This implies that a binary decision (selection vs. eviction) has to be made for each parameter, nothing more. In particular, it is not asked to rank the parameters by order of influence. Of course, being able to perform both screening and ranking is even better but screening is sufficient to complete the objective.

- One may wish to study the sensitivity of several random objects, including:

  - **Approach A:** the (scalar) output for one specific experimental configuration:

$$\forall\ 1 \leq i \leq n_{\text{exp}}, \boldsymbol{Y}_i(\boldsymbol{\Theta}) := y\left(\boldsymbol{x}_i, \boldsymbol{\Theta}\right). \tag{3.4}$$

    If proceeding like that, the same GSA must be performed n_{exp} times and the GSA results consist of n_{exp} collections of p sensitivity indices (for each component of $\boldsymbol{\Theta}$).

  - **Approach B:** the vector containing the outputs associated with all the experimental configurations:

$$\boldsymbol{Y}(\boldsymbol{\Theta}) := \left[\boldsymbol{Y}_i(\boldsymbol{\Theta})\right]_{1 \leq i \leq n_{\text{exp}}}. \tag{3.5}$$

  - **Approach C:** the (generalized) least-squares function, since this is a usual quantity we want to minimize in the calibration process:

$$L(\boldsymbol{\Theta}) = \left(\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{\Theta})\right)^t \Sigma_\epsilon^{-1} \left(\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{\Theta})\right), \tag{3.6}$$

    where $\Sigma_\epsilon = \text{diag}\left(\Sigma_{\epsilon_1}, \cdots, \Sigma_{\epsilon_{n_{\text{exp}}}}\right)$.

The dimension reduction enabled by Approach B is expected to improve the predictive performance of the metamodel. Note that Approach B is not equivalent to Approach C where GSA is designed to meet the needs of Bayesian calibration. Even though they often give fairly similar answers, these approaches provide complementary insights. Thus, it would not be surprising to find a parameter $\boldsymbol{\Theta}_j$ having influence on $\boldsymbol{Y}(\boldsymbol{\Theta})$ but not on $L(\boldsymbol{\Theta})$, and vice versa. Moreover, if observations of $\boldsymbol{Y}(\boldsymbol{\Theta})$ are available, they can immediately be converted into observations of $L(\boldsymbol{\Theta})$. Therefore, all the three approaches can be carried out simultaneously, provided that the GSA method is able to handle scalar and vector output objects.

- The design of experiments is not necessarily chosen in the interest of GSA, except for inexpensive numerical models [Wu et al., 2019]. Indeed, as soon as it is essential to build a metamodel, this latter task becomes the nub of the matter, because the quality of the Bayesian calibration of parameters heavily depends on the quality of the fitted metamodel. The design is build to favor the estimation and predictivity of the metamodel as much as possible, and the preliminary GSA step must therefore adapt to this design. This situation is called the *given-data* context. This often goes hand in hand with the *small-data* context, that is, the small size of the dataset. To support the metamodeling step, the input samples must be scattered to maximize the coverage of the variation domain. For this, it is often recommended to adopt a space-filing sampling strategy, such as Latin Hypercube Sampling (LHS) [McKay et al., 1979; Santner et al., 2003]. Among other things, LHS designs are appreciated for their ability to maintain satisfactory space coverage when uncertain parameters are eliminated from the study. This ensures efficient surrogate modeling in the reduced parameter space, even before knowing which parameters about to be selected. In this work, we will not use LHS designs because this would make the theoretical developments much harder, if not impossible. To circumvent this difficulty while remaining in a framework where the data are scarce and imposed, small-size Monte Carlo designs will be used throughout the following.

3.2.3 Review of screening techniques used for model calibration

In the papers addressing the issue of screening before Bayesian calibration, the literature highlights the classical Sobol' indices [Wu et al., 2019; Nagel et al., 2020; Dighe et al., 2022]. These indices [Sobol, 1993, 2001] are based on the decomposition of the output variance. In this decomposition, each fraction of output variance corresponds to the influence of a single parameter $\boldsymbol{\Theta}_j$ (main effect of $\boldsymbol{\Theta}_j$) or a group of parameters $\boldsymbol{\Theta}_u$ with $u \subseteq \{1, \cdots, p\}$ (interaction effect of $\boldsymbol{\Theta}_u$). The variance decomposition is called ANOVA (ANalysis Of VariAnce) and directly stems from the Sobol'-Hoeffding decomposition of the numerical model [Hoeffding, 1992; Sobol, 1993]. Assuming that $F := f(\boldsymbol{\Theta})$[4] is scalar, the first-order Sobol' index S_j and total-order Sobol' index T_j associated to $\boldsymbol{\Theta}_j$ are respectively given by:

$$S_j = \frac{\mathbb{V}\left(\mathbb{E}\left[F \mid \boldsymbol{\Theta}_j\right]\right)}{\mathbb{V}\left(F\right)} \quad \text{and} \quad T_j = 1 - \frac{\mathbb{V}\left(\mathbb{E}\left[F \mid \boldsymbol{\Theta}_{-j}\right]\right)}{\mathbb{V}\left(F\right)}, \tag{3.7}$$

where $\boldsymbol{\Theta}_{-j}$ represents the vector $\boldsymbol{\Theta}$ when $\boldsymbol{\Theta}_j$ is removed. S_j quantifies the main effect of $\boldsymbol{\Theta}_j$ whereas T_j also accounts for all the interaction effects with the other parameters. When $\boldsymbol{F} := [F_l]_l$ is a random vector, a simple way to perform variance-based GSA is to compute the first-order (resp. total-order) Sobol' index S_{jl} (resp. T_{jl}) with respect to each scalar output F_k [De Lozzo and Marrel, 2017]. For each input parameter Θ_j, a generalized Sobol' index S_j (resp. T_j) can then be defined as the weighted sum of all the indices S_{jk} (resp. T_{jl}). In particular, the weight associated to S_{jl} is equal to the ratio between the variance of F_l and the sum of all variances [Gamboa et al., 2014]. Although the interpretation of Equation (3.7)

[4] The function f is assumed to be square integrable.

is straightforward, the accurate estimation of Sobol' indices for real-life numerical models requires a significant number of evaluation of the numerical model. To be more precise, the first-order and total-order Sobol' indices are estimated with traditional methods such as the *Pick-Freeze* where $N = n(p+2)$ model evaluations are required [Saltelli, 2002] and n is at least equal to a few hundreds (since the convergence rate is $\frac{1}{\sqrt{n}}$ [Janon et al., 2014]). This result in an unaffordable simulation budget which cannot be considered as early as for $p \geq 10$ [Iooss and Lemaître, 2015]. An alternative approach is to estimate these indices using a metamodel (e.g., GP, PCE, etc.) [Sudret, 2008; Marrel et al., 2009; Crestaux et al., 2009], but this option is tractable only for a small number of input parameters. New advanced methods for cheaper estimation of the first-order Sobol' indices have emerged over the past few years, such as the technique based on replicated LHS designs (allowing for $N = 2n$) simulations [Tissot and Prieur, 2015; Damblin and Ghione, 2021] and the method based on the notion of Chatterjee's correlation [Chatterjee, 2021] (allowing for $N = n$) simulations [Gamboa et al., 2022]. However, the independence between $\mathbf{\Theta}_j$ and $f(\mathbf{\Theta})$ is not characterized by the nullity of S_j, but only by that of T_j. Note that there are several initiatives to reduce the cost of estimating higher-order indices, such as the method based on nearest neighbors [Broto et al., 2020] and the method based on smoothing kernels [Da Veiga et al., 2023]. These two approaches have a nice theoretical properties but they prove not to be sufficiently effective in practice. This is particularly true for high-order Sobol' indices.

In contrast to the Sobol' indices, we also have sensitivity indices computed by the Morris method [Morris, 1991]. The Morris method is a more rudimentary screening method which is sometimes used to identify the most impacting calibration parameters [Zambrano-Bigiarini et al., 2013; Xu et al., 2016], most often for lack of anything better. As opposed to Sobol' indices, this derivative-based approach seeks to average a collection of local sensitivity measures. More precisely, the model gradient (with respect to the calibration parameters) is estimated at a few points of the parameter space. At each point, and in each direction of the parameter space, the partial derivative is approximated (with a finite difference scheme) after applying a slight perturbation (at this point and in this direction). Then, the mean values and standard deviations of all gradient estimates are computed, and these two collections of sensitivity measures allow one to classify the parameters in three groups:

1. those with negligible influence,
2. those with large linear effects without interactions,
3. those with large non-linear or interaction effects.

This method has the huge advantage of relying on very few model evaluations (a few dozens at most) but it is not truly adapted to the specific constraints of the calibration problem. First, there is no possible characterization of the probabilistic independence between $\mathbf{\Theta}_j$ and F. Second, it is not tailored to handle a vector output $F = [F_l]_{1 \leq l \leq q}$, unless you repeat the standard method for each output F_l. Finally, the estimation of partial derivatives requires a specific type of designs, which is contrary to the *given-data* context.

For the two screening methods presented so far, inference cannot be reasonably considered from a given Monte Carlo design. However, there exists a class of sensitivity measures for which this problem does not arise. For each input-output pair, the idea is to measure the distance between the joint distribution $\mathbb{P}_{\mathbf{\Theta}_j F}$ and the product of marginal distributions $\mathbb{P}_{\mathbf{\Theta}_j} \otimes \mathbb{P}_F$:

$$\forall\ 1 \leq j \leq p, \qquad \mathrm{S}_j^{\delta} := \delta\left(\mathbb{P}_{\mathbf{\Theta}_j F}, \mathbb{P}_{\mathbf{\Theta}_j} \otimes \mathbb{P}_F\right). \tag{3.8}$$

Let $\mathscr{M}_+^1(\mathscr{D}_j)$ be the space of all probability measures on $\mathscr{D}_j := \mathscr{D}_{\mathbf{\Theta}_j} \times \mathscr{D}_F$. It notably contains the two distributions of interest. In the most general setting, the function

$$\delta : \mathscr{D}_j \to \mathbb{R}_+$$

used to measure the discrepancy between $\mathbb{P}_{\boldsymbol{\Theta}_j F}$ and $\mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F$ is only assumed to be a dissimilarity measure. In other words, it verifies at least $\delta(\mathbb{P}, \mathbb{P}) = 0$ for any $\mathbb{P} \in \mathscr{M}_+^1(\mathscr{D}_j)$, and sometimes stronger properties. $\mathbb{P}_{\boldsymbol{\Theta}_j F}$ is the true distribution of the pair $(\boldsymbol{\Theta}_j, F)$, that is to say the distribution which is generated by the propagation of uncertainties through the map $f : \boldsymbol{\theta} \in \mathscr{D}_{\boldsymbol{\Theta}} \to f(\boldsymbol{\theta}) \in \mathscr{D}_F$. On the contrary, $\mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F$ corresponds to the distribution that would have been obtained in a hypothetical situation of probabilistic independence between $\boldsymbol{\Theta}_j$ and F, and as such can be seen as a fictitious distribution serving as reference. In short, $\mathbb{P}_{\boldsymbol{\Theta}_j F}$ represents the true influence of $\boldsymbol{\Theta}_j$ (on F) whereas $\mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F$ represents what happens in case of no influence. The closer $\mathbb{P}_{\boldsymbol{\Theta}_j F}$ and $\mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F$ are, the less influence $\boldsymbol{\Theta}_j$ has on F. Many different sensitivity measures obey this formalism:

- Even if there are seldom introduced as specific cases of Equation (3.8), the well-known Spearman's ρ [Spearman, 1904] and Kendall's τ [Kendall, 1949] are two examples of dependence measures. Both of them can be expressed as a kind of integrated difference between the true copula of the pair (copula associated to $\mathbb{P}_{\boldsymbol{\Theta}_j F}$) and the independence copula (copula associated to $\mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F$). Additional details on this very point can be found in [Plischke and Borgonovo, 2015]. By reverting to the original mathematical definition of Spearman's ρ (Pearson's linear correlation coefficient between the rank variables associated to $\boldsymbol{\Theta}_j$ and F) and Kendall's τ (difference between the probabilities of concordance and discordance), they can be easily estimated from a given Monte Carlo design. For Spearman's ρ, an estimator is obtained with the empirical correlation between rank statistics. For Kendall's τ, a U-statistic (see Chapter 5 in [Serfling, 1980]) allows to estimate the probabilities of concordance and discordance. Due to this ease of estimation, Kendall's rank correlations were used for parameter screening (before model calibration) in [Minunno et al., 2013]. Unfortunately, they still suffer from several drawbacks. First, the nullity of Kendall's τ is not equivalent to independence. It only indicates that each of the four quadrants of the distribution has the same probability. Second, there is no possible generalization to vector outputs. Almost the same remarks holds for Spearman's ρ.

- If the dissimilarity measure δ is also a divergence, then one has:

$$\forall\, \mathbb{P}_1, \mathbb{P}_2 \in \mathscr{M}_+^1(\mathscr{D}_j), \quad \delta(\mathbb{P}_1, \mathbb{P}_2) = 0 \iff \mathbb{P}_1 = \mathbb{P}_2, \tag{3.9}$$

  and this makes S_j^δ a dependence measure adapted to screening since:

$$\mathrm{S}_j^\delta = \delta\left(\mathbb{P}_{\boldsymbol{\Theta}_j F}, \mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F\right) = 0 \iff \mathbb{P}_{\boldsymbol{\Theta}_j F} = \mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F \iff \boldsymbol{\Theta}_j \perp\!\!\!\perp F. \tag{3.10}$$

  Famous examples include the total-variation distance [Borgonovo, 2007], the mutual information [Auder and Iooss, 2008] or any other f-divergence [Rahman, 2016]. Unfortunately, in most cases, the value $\delta\left(\mathbb{P}_{\boldsymbol{\Theta}_j F}, \mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F\right)$ cannot be estimated without a preliminary step of inference where the density functions (or the cumulative distribution functions) of $\mathbb{P}_{\boldsymbol{\Theta}_j F}$ and $\mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F$ have to be estimated. The problem does not lies in the data because both marginal and joint observations can be found in the dataset. The problem is rather that most density estimation techniques lack robustness when little data is available. Many sampling tricks may be found in the literature [Plischke et al., 2013; Wei et al., 2014; Derennes et al., 2018] but no theoretical result secures their convergence. In addition, these sensitivity measures are not intended for non-scalar outputs.

It is interesting to note that the two previous types of sensitivity measures have complementary properties. On the one hand, Kendall's τ can be directly and easily estimated from a provided Monte Carlo design (but fails to characterize independence). On the other hand, the dependence measures based on a divergence are able to detect any type of dependence

(but their estimation requires a preliminary step of density estimation). In fact, there is a way to combine all the advantages, while remaining within the framework delimited by Equation (3.8). To achieve this, a prerequisite is to equip each random object with a (covariance) kernel. Then, for each pair of random objects, the two selected kernels allow us to define a kernel-based dissimilarity measure called the maximum mean discrepancy (MMD) [Gretton et al., 2006]. When quantified with the MMD, the distance $\delta\left(\mathbb{P}_{\boldsymbol{\Theta}_j F}, \mathbb{P}_{\boldsymbol{\Theta}_j} \otimes \mathbb{P}_F\right)$ is named the Hilbert-Schmidt independence criterion (HSIC) [Gretton et al., 2005a]. The original idea to use the HSIC in GSA is to the credit of Da Veiga [2015] who first highlighted the advantages of this approach compared to the existing ones. Since then, the method has been enriched with new extensions [De Lozzo and Marrel, 2016; El Amri and Marrel, 2022, 2024], while gradually disseminating as a reference practice [Delipei et al., 2019; Marie et al., 2021; Rollón de Pinedo et al., 2021]. Whether HSIC indices have become pretty popular for feature selection, they have not yet been applied to parameter screening in the particular context of Bayesian calibration, at least to the best of our knowledge. This is rather surprising since they cope perfectly with all the constraints already mentioned.

- First, the HSIC can be accurately estimated from a small amount of input-output samples.
- Second, the HSIC framework is completely generic, thus can deal with scalar and vector outputs in exactly the same way.
- Third, under some conditions on the kernels, an HSIC index is equal to zero if and only if the two objects under study are independent of each other.

To better understand what allows HSIC indices to meet all the expected conditions, the next section tackles the difficult task of succinctly presenting the mathematical foundations of the HSIC.

3.3 Benefits brought by HSIC indices

Defining the HSIC requires a dive into the universe of reproducing kernels (better known as covariance kernels) [Berlinet and Thomas-Agnan, 2011]. The concept behind this criterion is to equip each random variable with a kernel function which is (virtually) linked to a reproducing kernel Hilbert space (RKHS). This function space is generated from the only knowledge of the kernel, and inherits most of the kernel smoothness properties [Steinwart and Christmann, 2008]. In addition, the Moore-Aronszajn theorem [Aronszajn, 1950] states that there is a one-to-one mapping between kernels and RKHSs. An RKHS is a valuable ally for manipulating (or even comparing) probability distributions. Indeed, for a random object equipped with a given kernel, all possible probability distributions can be embedded into the RKHS associated with this kernel. It is then possible to compare two distributions by only measuring the squared distance of their respective embeddings in the RKHS. This distance, called the HSIC, is therefore a kernel-based dependence measure.

In the following two sections, we will use the generic notations X and Y for the input and output respectively to introduce the mathematical details of the HSIC criterion and the independence tests based on HSIC. Note that the random objects X and Y can be, for example, variables, vectors, excursion sets, or any other random quantity taking its values in a separable metric space [Gretton et al., 2005a].

3.3.1 Definition and estimation

HSIC, based on kernel approaches, is used to measure the dependence between a given input variable X of variation domain $\mathscr{D}_{\boldsymbol{X}}$ and the random object Y of variation domain $\mathscr{D}_{\boldsymbol{Y}}$. It is

defined as the square of maximum mean discrepancy (MMD) between $\mathbb{P}_{XY}$ and $\mathbb{P}_X \otimes \mathbb{P}_Y$:

$$\text{HSIC(X, Y)} := \text{MMD}^2(\mathbb{P}_{XY},\ \mathbb{P}_X \otimes \mathbb{P}_Y). \tag{3.11}$$

As this MMD is a kernel-based dissimilarity measure, the HSIC belongs to the class of sensitivity measures described by Equation (3.8). More precisely, one has:

$$\text{HSIC(X,Y)} = \|\mu_{\mathbb{P}_{XY}} - \mu_{\mathbb{P}_X \otimes \mathbb{P}_Y}\|^2_{\mathscr{H}\otimes\mathscr{G}}, \tag{3.12}$$

where

- $\mathscr{H}$ is the RKHS induced by K_x,
- $\mathscr{G}$ is the RKHS induced by K_y,
- $\mathscr{H} \otimes \mathscr{G}$ is tensor product RKHS built from $\mathscr{H}$ and $\mathscr{G}$,
- $\mu_{\mathbb{P}_{XY}}$ is the kernel mean embedding of $\mathbb{P}_{XY}$ in $\mathscr{H} \otimes \mathscr{G}$,
- $\mu_{\mathbb{P}_X \otimes \mathbb{P}_Y}\|^2_{\mathscr{H}\otimes\mathscr{G}}$ is the kernel mean embedding of $\mathbb{P}_X \otimes \mathbb{P}_Y$ in $\mathscr{H} \otimes \mathscr{G}$.

For a given kernel $K : \mathscr{Z} \times \mathscr{Z} \longrightarrow \mathbb{R}$ verifying $\mathbb{E}_Z[K(Z,Z)] < +\infty$, the *kernel mean embedding* of the probability distribution ν is the function defined by:

$$\begin{aligned} \mu_\nu : \mathscr{Z} &\longrightarrow \mathbb{R} \\ z &\longmapsto \mathbb{E}_Z[K(z,Z)] = \int_{\mathscr{Z}} K(z,z')d\nu(z'). \end{aligned} \tag{3.13}$$

Not only μ_ν belongs to $\mathbb{R}^{\mathscr{Z}}$, but above all it belongs to the RKHS $\mathscr{H}$ induced by K. This can be demonstrated by applying the Riesz-Fréchet theorem to the continuous linear form

$$h \in \mathscr{H} \mapsto \mathbb{E}_\nu[h(Z)] \in \mathbb{R}.$$

Morever, the kernel K is referred to as characteristic when the operator $\nu \mapsto \mu_\nu$ is injective. This means that each probability distribution ν has a unique representative (or a unique image) in the RKHS induced by the kernel K. Consequently, two different distributions will have different two images in the RKHS [Fukumizu et al., 2008; Muandet et al., 2017; Simon-Gabriel and Schölkopf, 2018]. Besides, if K_x and K_y are two characteristic kernels, the HSIC is equal to zero if and only if X and Y are independent [Gretton, 2015]:

$$\text{HSIC}(X,Y) = 0 \iff X \perp\!\!\!\perp Y, \tag{3.14}$$

where the symbol $\perp\!\!\!\perp$ denotes an independence situation between the random objects X and Y. As we will be see later in Section 3.3.2, this property is crucial for testing independence between X and Y.

As regards estimation, it is demonstrated in [Gretton et al., 2005a, 2007] that Equation (3.11) can be rewritten as:

$$\begin{aligned} \text{HSIC}(X,Y) = \ & \mathbb{E}_{\substack{XY \\ X'Y'}}[K_x(X,X')K_y(Y,Y')] + \mathbb{E}_{X'}[K_x(X,X')]\,\mathbb{E}_{Y'}[K_y(Y,Y')] \\ & -2\,\mathbb{E}_{\substack{XY \\ X'Y''}}[K_x(X,X')K_y(Y,Y'')], \end{aligned} \tag{3.15}$$

where (X,Y), (X',Y') and (X'',Y'') are three independent pairs of variables following the same (joint) distribution $\mathbb{P}_{XY}$. In light of Equation (3.15), the HSIC can be easily estimated from a Monte Carlo design $D_n := \left\{\left(X^{(l)}, Y^{(l)}\right)\right\}_{1\leq l\leq n}$. It suffices to estimate (with either a U-statistic or a V-statistic) each expectation found in Equation (3.15). In the end, as highlighted in [Gretton et al., 2007], two estimators can be considered:

$$\widehat{H}^{\mathrm{U}}(D_n) := \frac{1}{(n)_2} \sum_{(l,l')\in\mathscr{I}_n^2} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(l)}, Y^{(l')}\right) + \frac{1}{(n)_4} \sum_{(l,l',t,t')\in\mathscr{I}_n^4} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(t)}, Y^{(t')}\right) - 2\frac{1}{(n)_3} \sum_{(l,l',t)\in\mathscr{I}_n^3} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(l)}, Y^{(t)}\right), \tag{3.16}$$

$$\begin{aligned}\widehat{H}^{\mathrm{V}}(D_n) &:= \frac{1}{n^2} \sum_{(l,\, l')\in\mathscr{I}_n^2} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(l)}, Y^{(l')}\right) + \frac{1}{n^4} \sum_{(l,l',t,t')\in\mathscr{I}_n^4} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(t)}, Y^{(t')}\right) - 2\frac{1}{n^3} \sum_{(l,l',t)\in\mathscr{I}_n^3} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(l)}, Y^{(t)}\right) \\ &= \frac{1}{n^2}\mathrm{Tr}\left(\boldsymbol{H}\boldsymbol{L}_x\boldsymbol{H}\boldsymbol{L}_y\right) = \frac{1}{n^2}\mathrm{Tr}\left((\boldsymbol{H}\boldsymbol{L}_x\boldsymbol{H})(\boldsymbol{H}\boldsymbol{L}_y\boldsymbol{H})\right),\end{aligned} \tag{3.17}$$

where

- $\mathrm{Tr}(\cdot)$ is the trace operator for square matrices and $\boldsymbol{I}_n$ is identity matrix.
- $(n)_r := \frac{n!}{(n-r)!}$, $\mathscr{I}_n^r := \{(i_1, \cdots, i_r) :\ 1 \leq i_1 \neq i_2 \neq \cdots \neq i_r \leq n\}$ with $\#\mathscr{I}_n^r = (n)_r$.
- $\mathscr{J}_n^r := \{(i_1, \cdots, i_r)\ 1 \leq i_1, \cdots, i_r \leq n\}$ with $\#\mathscr{J}_n^r = n^r$.
- $\boldsymbol{L}_x := \left[K_x\left(X^{(l)}, X^{(l')}\right)\right]_{l,l'=1}^n$ and $\boldsymbol{L}_y := \left[K_y\left(Y^{(l)}, Y^{(l')}\right)\right]_{l,l'=1}^n$ are the Gram matrices.
- $\boldsymbol{H} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^t$ is the centering matrix with $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ the identity matrix and $\mathbf{1}_n \in \mathbb{R}^n$ the vector filled by taking 1 everywhere.

As any U-statistic estimator, $\widehat{H}^{\mathrm{U}}(D_n)$ is unbiased (see Chapter 5 in [Serfling, 1980]). This estimator of the HSIC may take negative values, which is pretty disturbing (knowing that the HSIC is a non-negative quantity). The V-statistic $\widehat{H}^{\mathrm{V}}(D_n)$ has complementary properties. It is biased (see Chapter 6 in [Serfling, 1980]), asymptotically unbiased and its non-negativity is guaranteed by Equation (3.17). Even if estimators $\widehat{H}^{\mathrm{U}}(D_n)$ and $\widehat{H}^{\mathrm{V}}(D_n)$ are expressed as double, triple and quadruple sums, there exist calculation tricks to get their values at the cost of only $\mathscr{O}(n^2)$ operations [Song et al., 2012]. For $\widehat{H}^{\mathrm{U}}(D_n)$, the summation sets are $\mathscr{I}_n^2$, $\mathscr{I}_n^3$ and $\mathscr{I}_n^4$. Note that $\mathscr{I}_n^2$ (resp. $\mathscr{I}_n^3$, $\mathscr{I}_n^4$) is composed of all pairs (resp. triplets, quadruplets) picked without replacement among $\{1, \cdots, n\}$. On the contrary, for the summation sets $\mathscr{J}_n^2$, $\mathscr{J}_n^3$ and $\mathscr{J}_n^4$ involved in Equation (3.17), replacement must be considered when enumerating tuples.

3.3.2 Independence testing based on HSIC

We are interested in determining whether the two random objects X and Y are independent of each other. The idea is to make a decision between the two following statements:

$$(\mathrm{H}_0) :\ X \perp\!\!\!\perp Y \text{ vs. } (\mathrm{H}_1) :\ X \not\perp\!\!\!\perp Y, \tag{3.18}$$

on the only basis of the observed data. Regarding that very point, let D_n^{obs} denote the available dataset (resulting from $N = n$ calls to the numerical model), as opposed to D_n which denotes the underlying random quantity. Remember that the nullity of the HSIC (defined by means

of two characteristic kernels) is equivalent to the probabilistic independence between X and Y. Consequently, the two statements can be reformulated as:

$$(\mathrm{H}_0): \ \mathrm{HSIC}(X,Y)=0 \text{ vs. } (\mathrm{H}_1): \ \mathrm{HSIC}(X,Y)>0. \tag{3.19}$$

Here, the test statistic is very naturally $\widehat{H}^{\mathrm{stat}}(D_n)$ with $\mathrm{stat}\in\{\mathrm{U},\mathrm{V}\}$. The decision rule is to reject (H_0) if the observed value $\widehat{H}^{\mathrm{stat}}(D_n^{\mathrm{obs}})$ of $\widehat{H}^{\mathrm{stat}}(D_n)$ exceeds the quantile of order $1-\alpha$ of the distribution of $\widehat{H}^{\mathrm{stat}}(D_n)$ under (H_0). The quantity α is called the *type-I error* of the *level* of the test and is often set to 5% (or 10%). In practice, the concept of p-value is commonly used, especially to return an answer richer than the simple binary decision of accepting or rejecting (H_0). The p-value is defined as the probability that the test statistic takes an even more extreme value:

$$p_{\mathrm{val}}=\mathbb{P}_{\mathscr{H}_0}\left(\widehat{H}^{\mathrm{stat}}(D_n)\geq\widehat{H}^{\mathrm{stat}}(D_n^{\mathrm{obs}})\right). \tag{3.20}$$

Then (H_0) is accepted if $p_{\mathrm{val}}\geq\alpha$, and rejected otherwise. When (H_0) is accepted (resp. rejected), it is decided that X and Y are independent (resp. dependent), which indicates that X has negligible (resp. significant) influence on Y. In order to estimate the p-value, the distribution of $\widehat{H}^{\mathrm{stat}}(D_n^{\mathrm{obs}})$ under (H_0) is essential. This distribution is unknown but three different methods can be used to compute a satisfactory approximation:

- **The asymptotic test procedure.** It was proved in [Gretton et al., 2007] that the rescaled test statistic $n\widehat{H}^{\mathrm{stat}}(D_n)$ converges in law to a (shifted) spectral distribution, that is a series of chi-squared distributions with positive weights. More precisely, the weights are determined by the eigenvalues of the Mercer expansions (with respect to $\mathbb{P}_X$ and $\mathbb{P}_Y$) of the double-centered kernels built from K_x and K_y [Zhang et al., 2018]. This distribution can be approximated by a Gamma distribution for $\widehat{H}^{\mathrm{V}}(D_n)$ and by a Pearson type-III distribution for $\widehat{H}^{\mathrm{U}}(D_n)$. In both cases, the parameters of the approximated distribution are estimated with the method of moments (based on asymptotic formulas for the two moments).

- **The permutation-based test procedure.** When n is not sufficiently large for the asymptotic approach to be used, the distribution of $\widehat{H}^{\mathrm{stat}}(D_n)$ under (H_0) can still be simulated by randomly permuting the samples of Y (while those of X remain in the same order). Remember that a permutation is a one-to-one mapping from $\{1,\cdots,n\}$ to $\{1,\cdots,n\}$. For a given permutation $\tau\in\mathbb{S}_n$, the idea is to recompute the HSIC estimate on the following permuted design:

$$D_n^{\tau}=\left\{\left(X^{(l)},Y^{(\tau(l))}\right)\right\}_{l=1}^{n}.$$

  The resulting value $\widehat{H}^{\mathrm{stat}}(D_n^{\tau})$ can be seen as a new realization of the test statistic under H_0. After performing B_{perm} permutations, the p-value can be estimated from the collection of simulated realizations:

$$\widehat{p}_{\mathrm{val}}=\frac{1}{B_{\mathrm{perm}}}\sum_{b=1}^{B_{\mathrm{perm}}}\mathbf{1}_{\left\{\widehat{H}^{\mathrm{stat}}(D_n^{\tau_b})>\widehat{H}^{\mathrm{stat}}(D_n^{\mathrm{obs}})\right\}}. \tag{3.21}$$

  For more comprehensive insights into this methodology, the interested reader is invited to consult De Lozzo and Marrel [2016] and El Amri and Marrel [2022].

- **The non-asymptotic Gamma test procedure.** When the sample size n is around a few hundreds, the asymptotic regime has probably not yet been reached. For a sample size of this type, the permutation-based test procedure can be computationally intensive, especially if the number of parameters is large ($p\geq 10$) or the number of random permutations to be completed is set very high ($B_{\mathrm{perm}}\geq 10^3$). In response

to this, a test procedure imagined in [El Amri and Marrel, 2024] makes it possible to drastically reduce [5] the complexity of the p-value estimation step. This procedure requires the test statistic to be the V-statistical estimator of the HSIC index, because the method relies entirely on the algebraic formula (Equation (3.17)) specific to this HSIC estimator. The starting point is to notice that the histogram of the HSIC values $\left\{\widehat{H}^{\text{stat}}(D_n^{\tau_b})\right\}_{1\leq b\leq B_{\text{perm}}}$ obtained after a sequence of B_{perm} random permutations looks like that of a Gamma distribution (unimodal law with positive values only and slight positive skewness). This is not surprising as the asymptotic distribution of the test statistic under (H_0) is precisely approximated by a Gamma distribution. The idea is then to determine the parameters of a Gamma distribution which fits correctly to this histogram. Of course, these parameters must be estimated without performing any permutation because, if the $\left\{\widehat{H}^{\text{stat}}(D_n^{\tau_b})\right\}_{1\leq b\leq B_{\text{perm}}}$ values are available, it is better to use the non-parametric estimate (as no arguable parametric assumption is made). Recall that the V-statistic estimator of the HSIC is of the form $n^2\widehat{H}^{\text{V}}(D_n) = \text{Tr}(\boldsymbol{AW})$ with $\boldsymbol{A} = \boldsymbol{HL}_x\boldsymbol{H}$ and $\boldsymbol{W} = \boldsymbol{HL}_y\boldsymbol{H}$. As pointed out in [El Amri and Marrel, 2024], when a permutation τ is applied to the output samples, the matrix $\boldsymbol{W}$ is transformed into a matrix $\boldsymbol{W}^{\tau\tau} := \left[\boldsymbol{W}_{\tau(i)\tau(j)}\right]_{1\leq i,j\leq n}$ which is the image of $\boldsymbol{W}$ after permuting its rows and columns according to the instructions given by τ. For a random permutation τ (following the uniform distribution on $\mathbb{S}_n$ [6]), the exact values of the first three moments of the random variable $\text{Tr}\left(\boldsymbol{AW}^{\tau\tau}\right)$ can be calculated via elementary algebraic operations on $\boldsymbol{A}$ and $\boldsymbol{W}$. In particular, the mean and variance are given by:

$$\mathbb{E}_{\text{perm}}(\text{Tr}(\boldsymbol{AW})) = \frac{1}{n!}\sum_{\tau\in\mathbb{S}_n}\text{Tr}\left(\boldsymbol{AW}^{\tau\tau}\right) = \frac{T}{n-1}, \tag{3.22}$$

$$\begin{aligned}\mathbb{V}_{\text{perm}}(\text{Tr}(\boldsymbol{AW})) = \frac{T_1}{n} &+ \frac{(T^2-T_1)+2(T_2-T_1)+4T_1}{n(n-1)} \\ &+ \frac{4(2T_1-T_2)+2(2T_1-T^2)}{n(n-1)(n-2)} \\ &+ \frac{2T_2-6T_1+T^2}{n(n-1)(n-2)(n-3)},\end{aligned} \tag{3.23}$$

with

$$T = \text{Tr}(\boldsymbol{A})\text{Tr}(\boldsymbol{W}), \tag{3.24}$$

$$T_1 = \sum_{ll}(\boldsymbol{A}_{ll})^2\sum_{ll}(\boldsymbol{W}_{ll})^2, \tag{3.25}$$

$$T_2 = \sum_{lh}(\boldsymbol{A}_{lh})^2\sum_{lh}(\boldsymbol{W}_{lh})^2. \tag{3.26}$$

These analytical formulas were established in [Kazi-Aoual et al., 1995], where it is notably precised that the matrices $\boldsymbol{A}$ and $\boldsymbol{W}$ must be square, symmetric and double centered (which is the case here). Let $\boldsymbol{Z}$ be a random variable following a Gamma distribution $\Gamma(a,b)$ with a and b denoting the shape and scale parameters respectively. Recall that $\mathbb{E}[\boldsymbol{Z}] = ab$ and $\mathbb{V}(\boldsymbol{Z}) = ab^2$, or equivalently $a = \mathbb{E}[\boldsymbol{Z}]^2/Var(\boldsymbol{Z})$ and $b = Var(\boldsymbol{Z})/\mathbb{E}[\boldsymbol{Z}]$. Therefore, in order to provide a parametric approximation of the permutation distribution, a reasonable strategy is to take a Gamma distribution with parameters:

$$\widehat{a} = \frac{\mathbb{E}_{\text{perm}}(\text{Tr}(\boldsymbol{AW}))^2}{\mathbb{V}_{\text{perm}}(\text{Tr}(\boldsymbol{AW}))}, \quad \text{and} \quad \widehat{b} = \frac{\mathbb{V}_{\text{perm}}(\text{Tr}(\boldsymbol{AW}))}{\mathbb{E}_{\text{perm}}(\text{Tr}(\boldsymbol{AW}))}. \tag{3.27}$$

[5] When the p-values are estimated with the permutation approach, $\mathcal{O}(B_{\text{perm}}pn^2)$ operations are required. In practice, this may take some time, but this will always be negligible compared to the time needed to evaluate the numerical model.

[6] $\mathbb{S}_n$ is the set of permutations of $\{1,\cdots,n\}$.

Such a distribution is then used to estimate the p-value:

$$\widehat{p}_{\text{val}} = 1 - F_{\Gamma(\widehat{a},\widehat{b})}\left(\widehat{H}^{\text{V}}(D_n^{\text{obs}})\right), \tag{3.28}$$

where $F_{\Gamma(\widehat{a},\widehat{b})}(.)$ is the cumulative distribution function of Gamma distribution.

3.3.3 HSIC-based parameter screening

Let us now return to the problem of parameter screening before model calibration. It simply comes down to estimating the HSIC (and its p-value) between each parameter $\boldsymbol{\Theta}_j$ and the output of interest $f(\boldsymbol{\Theta})$. Recall that f can take three different forms, namely $\boldsymbol{Y}_i(\boldsymbol{\Theta})$, $\boldsymbol{Y}(\boldsymbol{\Theta})$ or $L(\boldsymbol{\Theta})$. They correspond to the three approaches (A, B, C) proposed in Section 3.2.2. In fact, only two situations need to be distinguished: the case of scalar outputs and that of vector outputs (Approach B). As everything said in Section 3.3 holds for any pair of objects living in (possibly different) separable spaces, there is no difficulty in applying the HSIC methodology to parameter screening, provided that appropriate kernel functions are used. Indeed, the choice of kernels is an important step, which deserves particular attention. The most commonly used (characteristic) kernel is the Gaussian kernel:

$$\forall\ (u,u') \in \mathbb{R}^2, \qquad K_\gamma(u,u') = \exp\left[-\tfrac{1}{2}\left(\tfrac{u-u'}{\gamma}\right)^2\right], \tag{3.29}$$

where γ is a bandwidth parameter. It can be set with a simple rule of thumb [Da Veiga, 2015; De Lozzo and Marrel, 2016] (empirical standard deviation or median) or tuned to maximize the HSIC value [El Amri and Marrel, 2024]. Even if it is defined on $\mathbb{R}$, the Gaussian kernel (once its bandwidth is correctly adjusted) is perfectly suited to a random variable taking its inputs in a compact domain (like a bounded interval). Hence, the Gaussian kernel will be assigned to all input parameters, as well as for f (when it is simply scalar). The only delicate point is the choice of an output kernel when $f = [f_i]_{1\leq i\leq n_{\text{exp}}}$ is a vector output (Approach B). It is rather not recommended to use a multivariate Gaussian kernel since this kernel is parametrized by a single (scalar) bandwidth parameter, which will be a problem if the component of f do not share the same physical scale. It is much preferable to assign a Gaussian kernel to each output f_i, and to adjust the parameter γ_i accordingly. A simple tensorization then allows one to build a kernel for f. It can be easily proved that such a multivariate kernel is characteristic. Two different proofs are provided in Appendix 3.8.1.1. The first one is a simple illustration of a very general result recently established in [Ziegel et al., 2022]. The second one (which is also true for any characteristic translation-invariant kernel defined on $\mathbb{R}$) relies on much older arguments found in [Sriperumbudur et al., 2010a].

Now that it is clear how to estimate HSIC indices in the specific context of parameter screening for model calibration, all that remains to do is to perform GSA on the industrial test case, which is here the fission gas behavior model (with nominal value of the thermal conductivity) of the ALCYONE application.

3.4 Application to the gas behavior model of the ALCYONE code

3.4.1 Presentation of the ALCYONE application

The ALCYONE application is composed of interdependent physical models that represent the mechanical, thermal and chemical behaviors of fuel rods in the core of pressurized water reactors (multi-physics application). Among these different models, we will only focus on the thermal and fission gas behavior models [Michel et al., 2021]. The thermal model simulates the evolution of the temperature within the fuel rod during the fission reaction and provides as output, the associated temperature field. The fission gas behavior model, as a function of

the thermal model, continuously represents the behavior of the fission products (fuel swelling and release of fission gas atoms) during the fission reaction.

Figure 1.5 shows the chaining of the two models (represented by the blue boxes) and their integration into the overall simulation process. The (uncertain) input variables of these two models are respectively the thermal conductivity $\lambda \in \mathscr{D}_{\mathbf{\Lambda}} \subseteq \mathbb{R}$ and the model parameters $\boldsymbol{\theta} \in \mathscr{D}_{\mathbf{\Theta}} \subset \mathbb{R}^p$. The quantity $\mathbf{\Lambda}$ (resp. $\mathbf{\Theta}$) denotes the random variable (resp. vector) representing the uncertainty on λ (resp. $\boldsymbol{\theta}$). It is assumed that: *a*) the components of $\mathbf{\Theta}$ are mutually independent and *b*) $\mathbf{\Lambda}$ and $\mathbf{\Theta}$ are independent of each other. The fission gases produced by irradiation are distributed as follows:

- gas dissolved in the grain,
- precipitated gas in intragranular bubbles,
- precipitated gas in intergranular bubbles.

When the interganular bubbles at the Uranium dioxide grain junction are sufficiently covered to be interconnected, we obtain the phenomenon known as Release Gas Fraction (RGF) in the free volume of the fuel rod modeled by the behavior model. Fuel restructuring (small-grain structure, highly porous medium) occurs at high burnup rates. In the event of an accident, this restructuring has a major impact on the behavior of fission gases. All these physical mechanisms are represented in the fission gas behavior model, which has the following uncertain model parameters:

- $\mathbf{\Theta}_1$ ($\mathrm{k_{eff}}$): nucleation efficiency coefficient,
- $\mathbf{\Theta}_2$ ($\mathrm{b_{gbb}}$): parameter involved in the ballistic gas re-solution term of intergranular bubbles in the grain,
- $\mathbf{\Theta}_3$ ($\mathrm{k_v}$): parameter involved in the shape of bubbles on the grain boundary (direct effect on the rate of coverage of the grain boundary by bubbles and therefore on the threshold above which the gas release is triggered),
- $\mathbf{\Theta}_4$ ($\mathrm{Fc_{lim\ coalesc}}$): threshold of grain boundary coverage by intergranular bubbles to initiate intergranular bubble coalescence,
- $\mathbf{\Theta}_5$ ($\mathrm{A_2}$): pre-exponential coefficient of the mixed term of the diffusion coefficient (irradiation-accelerated thermal diffusion),
- $\mathbf{\Theta}_6$ ($\mathrm{A_3}$): coefficient of the athermal term of the diffusion coefficient (diffusion only dependent on fission density),
- $\mathbf{\Theta}_7$ ($\mathrm{k_\alpha}$): parameter involved in the rate of irradiation defect production and therefore affecting restructuring kinetics,
- $\mathbf{\Theta}_8$ ($\mathrm{t_{kthres}}$): threshold temperature above which defect annealing is activated during the restructuring process,
- $\mathbf{\Theta}_9$ (α_{bhsd}): coefficient of proportionality between the gas diffusion coefficient in the grain in the healthy zone and that used in the restructured fragment of the restructured zone,
- $\mathbf{\Theta}_{10}$ ($\mathrm{b_{hbs}}$): frequency at which gas trapped in peripheral bubbles returns to solution in grain,
- $\mathbf{\Theta}_{11}$ ($\mathrm{R_{fhbs}}$): fragment-size into which the gas diffuses in restructured areas with high burnup rates.

For given values of $\boldsymbol{\theta}$ and λ the fission gas behavior model computes the following quantities:

- the swelling deformations (bubble volume, solid swelling, etc.),
- the RGF in each fuel rod in free volume,
- the distribution of retained gas in different populations (dissolved, precipitated bubbles, etc.).

3.4.2 Strategies considered to conduct GSA

In this paper, we focus on the RGF of the fission gas behavior model (here $q = 1$). Among the model parameters to be calibrated, we wish to eliminate those which have a negligible influence on either the simulated values of RGF (to help build a surrogate model for each fuel rod) or the distance between the simulated and observed values of RGF (to help perform Bayesian calibration). The epistemic uncertainty affecting λ is not taking into at this step, and we fixed it at its nominal value $\lambda_{\text{nom}} = 1$. Such uncertainty is incorporate into the GSA framework in Section 3.5. First, a Monte Carlo numerical design is generated for the calibration parameters:

$$D_{\boldsymbol{\Theta},n} = \left\{\boldsymbol{\Theta}^{(l)}\right\}_{1\leq l\leq n} \in \mathbb{R}^{n\times p} \quad \text{with} \quad \boldsymbol{\Theta}^{(l)} := \left[\boldsymbol{\Theta}_1^{(l)}, \cdots, \boldsymbol{\Theta}_p^{(l)}\right]^t \in \mathbb{R}^p.$$

Finally, running the numerical model for $(\boldsymbol{x},\ D_{\boldsymbol{\Theta},n}, \lambda_{\text{nom}})$ provides the outputs of the RGF simulations for the entire set of experimental configurations $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n_{\text{exp}}})^t$. These output are denoted by:

$$D_{\boldsymbol{Y},n} = \cup_{i=1}^{n_{\text{exp}}} D_{\boldsymbol{Y}_i,n} \quad \text{with} \quad D_{\boldsymbol{Y}_i,n} = \left\{Y_i^{(l)}\right\}_{1\leq l\leq n} \quad \text{and} \quad Y_i^{(l)} := y\left(\boldsymbol{x}_i, \boldsymbol{\Theta}^{(l)}, \lambda_{\text{nom}}\right).$$

Below, the deterministic function $y(\cdot,\cdot,\cdot)$ is the result of the chaining illustrated in Figure 1.5:

$$y : \mathscr{D}_{\boldsymbol{X}} \times \mathscr{D}_{\boldsymbol{\Theta}} \times \mathscr{D}_{\boldsymbol{\Lambda}} \subset \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R} \longrightarrow \mathscr{D}_{\boldsymbol{Y}} \subset \mathbb{R}^q \ (q = 1) \tag{3.30}$$

$$(\boldsymbol{x}, \boldsymbol{\theta}, \lambda_{\text{nom}}) \longmapsto y(\boldsymbol{x},\ \boldsymbol{\theta},\ \lambda_{\text{nom}}) := \left[y_{\boldsymbol{\theta}}^2 \circ y_{\lambda_{\text{nom}}}^1\right](\boldsymbol{x}). \tag{3.31}$$

In order to determine the most influential parameters on the RGF (for a nominal conductivity λ_{nom}), GSA is performed in three different ways:

- **Approach A**: sensitivity for the the RGF produced by each fuel rod. The HSIC is computed (with either U-statistic or V-statistic) between each calibration parameter $\boldsymbol{\Theta}_j$ and the RGF associated to the i-th experimental configuration $\boldsymbol{x}_i$, namely $\boldsymbol{Y}_i := y(x_i, \boldsymbol{\Theta}, \lambda_{\text{nom}}) =: f_i^{\text{A}}(\boldsymbol{\Theta}) \in \mathbb{R}$. This results in n_{exp} collections of $p = 11$ HSIC indices:
  - Sensitivity measures: $\forall\ 1 \leq j \leq p,\ \forall\ 1 \leq i \leq n_{\text{exp}},\ \widehat{H}_j^{\text{stat}}\left(D_{n,i}^{\text{A}}\right)$.
  - Samples: $D_{n,i}^{\text{A}} := \{D_{\boldsymbol{\Theta},n}; D_{\boldsymbol{Y}_i,n}\},\ 1 \leq i \leq n_{\text{exp}}$.
- **Approach B**: sensitivity of the vector gathering all RGF values. The HSIC is computed between each calibration parameter $\boldsymbol{\Theta}_j$ $(1 \leq j \leq p)$ and the vector

$$\boldsymbol{Y}(\boldsymbol{\Theta}) := y(\boldsymbol{x}, \boldsymbol{\Theta}, \lambda_{\text{nom}}) = [\boldsymbol{Y}_i(\boldsymbol{\Theta})]_{1\leq i\leq n_{\text{exp}}} =: f^{\text{B}}(\boldsymbol{\Theta}) \in \mathbb{R}^{n_{\text{exp}}} :$$

  - Sensitivity measures: $\forall\ 1 \leq j \leq p,\ \widehat{H}_j^{\text{stat}}\left(D_n^{\text{B}}\right)$.
  - Samples: $D_n^{\text{B}} := \{D_{\boldsymbol{\Theta},n}; D_{\boldsymbol{Y},n}\}$.

- **Approach C**: sensitivity of the least-square function. The HSIC is computed between each calibration parameter $\boldsymbol{\Theta}_j$ $(1 \leq j \leq p)$ and the least-squares function

$$L(\boldsymbol{\Theta}) = (\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{\Theta}))^t \, \Sigma_\epsilon^{-1} \, (\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{\Theta})) =: f^{\mathrm{C}}(\boldsymbol{\Theta}) \in \mathbb{R}_+ :$$

  - Sensitivity measures: $\forall\ 1 \leq j \leq p,\ \widehat{H}_j^{\mathrm{stat}}\left(D_n^{\mathrm{C}}\right)$.
  - Samples: $D_n^{\mathrm{C}} := \{D_{\boldsymbol{\Theta},n}; D_{\boldsymbol{L},n}\}$ with

$$D_{\boldsymbol{L},n} := \cup_{l=1}^{n} L^{(l)} \quad \text{with} \quad L^{(l)} = \left(\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{\Theta}^{(l)})\right)^t \Sigma_\epsilon^{-1} \left(\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{\Theta}^{(l)})\right) =: f^{\mathrm{C}}(\boldsymbol{\Theta}^{(l)}) \in \mathbb{R}_+,$$

  where $\boldsymbol{z}$ and $\Sigma_\epsilon = \mathrm{diag}(\sigma_{\epsilon_1}^2, \cdots, \sigma_{\epsilon_{n_{\mathrm{exp}}}}^2)$ represent respectively the experimental data and the covariance matrix of the experimental uncertainty.

It is worth noting that the approaches (A, B and C) are applied exclusively as post-treatment procedures, carried out after the computation of the outputs $\boldsymbol{Y}$ through simulation, and can be executed simultaneously. Specifically, the datasets $D_{n,i}^{\mathrm{A}}$, D_n^{B}, and D_n^{C} are derived from $D_{\boldsymbol{\Theta},n}$ and $D_{\boldsymbol{Y},n}$. Regardless of the chosen approach (A, B or C), we simply apply the methodologies outlined in Section 3.3.3 to estimate the HSIC indices and the corresponding p-values for the independence tests. Consequently, the notations A, B and C (accompanying the data D_n and the function f) are omitted for the remainder of this work.

Approaches A and B are particularly well suited for parameter reduction before constructing metamodels for the functions $\boldsymbol{\theta} \mapsto y_i(\boldsymbol{\theta})$. On the other hand, Approach C is specifically tailored for calibration purposes and involves dimension reduction prior to running the MH algorithm. In scenarios where a metamodel is unnecessary, Approach C should be preferred. It will therefore be interesting to look at the results of A/B vs. C.

3.4.3 Results

The three approaches described above are applied on the ALCYONE test case in order to identify the less influential parameters within $\boldsymbol{\Theta}$ (for a nominal value of the thermal conductivity). The decision to accept (or reject) (H_0): $\boldsymbol{\Theta}_j \perp\!\!\!\perp f(\boldsymbol{\Theta})$ is made on the basis of a *type-I error* α equal to 5%.

For all approaches, since the number of samples is relatively small, the p-values are estimated with the permutation-based test procedure, and in particular with the formula given in Equation (3.21). The p-values of Approach A are shown in Figure 3.1, where the x-axis represents the RGF values $(1 \leq i \leq n_{\mathrm{exp}} = 40)$ ordered by decreasing burnup rate. For all the n_{exp} outputs, the influential calibration parameters are $\mathrm{R_{fhbs}}$, k_α, $\mathrm{A_3}$ and $\mathrm{A_2}$. The parameter $\mathrm{b_{hbs}}$ only appears to be influential for a small number of RGF outputs (those corresponding to a high burnup rate). The remaining calibration parameters have almost no influence on the RGF, and this is true for all the fuel rods. Table 3.1 summarizes the results of Figure 3.1 in terms of influence detection rates on all $n_{\mathrm{exp}} = 40$ RGFs. The detection rate on all n_{exp} RGFs for each parameter $\boldsymbol{\Theta}_j$ is defined as the ratio between the number of times H_0 is rejected and n_{exp}. More specifically, we have:

$$\frac{\#\left\{(\mathrm{H}_0) : \boldsymbol{\Theta}_j \perp\!\!\!\perp Y_i(\boldsymbol{\Theta}) \text{ is rejected for } \ 1 \leq i \leq n_{\mathrm{exp}}\right\}}{n_{\mathrm{exp}}}.$$

The p-values related to Approach B are presented in Table 3.2. We can see that the parameters selected at the threshold $\alpha = 5\%$ are $\mathrm{R_{fhbs}}$, $\mathrm{b_{hbs}}$, k_α, $\mathrm{A_3}$ and $\mathrm{A_2}$. As regards Approach C, which is based on the lest-squares function, the same parameters are identified as influential (except $\mathrm{b_{hbs}}$). However, the related p-value is quite close to the limit threshold $\alpha = 5\%$ (see Table 3.3). Note that the non-asymptotic Gamma approximation (described in Section 3.3.2) gives the same results as the permutation approximation. It is rather reassuring to see that the three GSA strategies lead to the same conclusions. Furthermore, the results obtained for

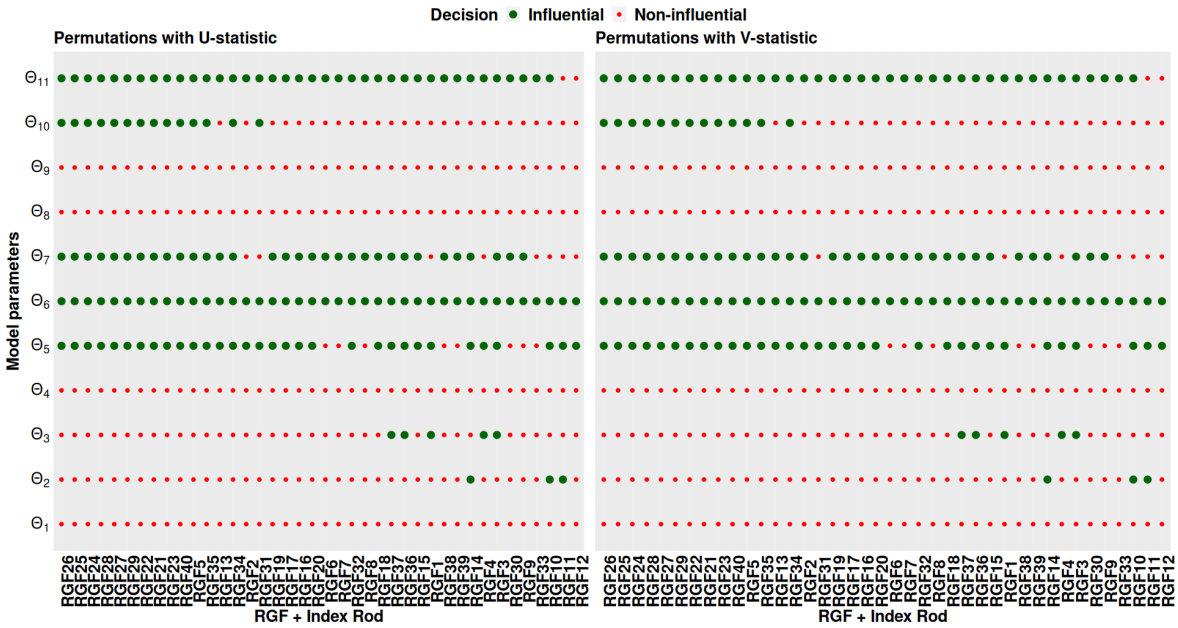


Figure 3.1: Summary of the results obtained when independence is tested with Approach A. On the left (resp. right) hand side, HSIC indices are computed with U-statistic (resp. V-statistics). The associated detection rates are given in Table 3.1.

Model Parameters \ Detection rates	with U-statistic	with V-statistic
$\mathbf{\Theta}_1 = k_{eff}$	0%	0%
$\mathbf{\Theta}_2 = b_{gbb}$	7.5%	7.5%
$\mathbf{\Theta}_3 = k_v$	12.5%	12.5%
$\mathbf{\Theta}_4 = Fc_{lim\ coalesc}$	0%	0%
$\mathbf{\Theta}_5 = A_2$	80%	80%
$\mathbf{\Theta}_6 = A_3$	100%	100%
$\mathbf{\Theta}_7 = k_\alpha$	80%	82.5%
$\mathbf{\Theta}_8 =$ tkthres	0%	0%
$\mathbf{\Theta}_9 = \alpha_{bhsd}$	0%	0%
$\mathbf{\Theta}_{10} = b_{hbs}$	35%	32.5%
$\mathbf{\Theta}_{11} = R_{fhbs}$	95%	95%

Table 3.1: Detection rates obtained for Approach A.

Model parameters \ p-values	$\widehat{p}^B_{U,\,l}$	$\widehat{p}^B_{V,\,l},\ 1 \leq l \leq p$
$\mathbf{\Theta}_1 = k_{eff}$	0.972	0.973
$\mathbf{\Theta}_2 = b_{gbb}$	0.25	0.241
$\mathbf{\Theta}_3 = k_v$	0.372	0.406
$\mathbf{\Theta}_4 = Fc_{limcoalesc}$	0.909	0.891
$\mathbf{\Theta}_5 = A_2$	0	0
$\mathbf{\Theta}_6 = A_3$	0	0
$\mathbf{\Theta}_7 = k_\alpha$	0	0
$\mathbf{\Theta}_8 =$ tkthres	0.869	0.861
$\mathbf{\Theta}_9 = \alpha_{bhsd}$	0.856	0.864
$\mathbf{\Theta}_{10} = b_{hbs}$	0.022	0.011
$\mathbf{\Theta}_{11} = R_{fhbs}$	0	0

Table 3.2: p-value estimates for Approach B (with the permutation-based test procedure).

p-values / Model parameters	with U-statistic	with V-statistic
$\mathbf{\Theta}_1 = \mathrm{k_{eff}}$	0.538	0.554
$\mathbf{\Theta}_2 = \mathrm{b_{gbb}}$	0.741	0.761
$\mathbf{\Theta}_3 = \mathrm{k_v}$	0.127	0.11
$\mathbf{\Theta}_4 = \mathrm{Fc_{limcoalesc}}$	0.801	0.799
$\mathbf{\Theta}_5 = \mathrm{A_2}$	0	0
$\mathbf{\Theta}_6 = \mathrm{A_3}$	0.006	0.006
$\mathbf{\Theta}_7 = \mathrm{k_\alpha}$	0	0
$\mathbf{\Theta}_8 = \mathrm{tkthres}$	0.757	0.745
$\mathbf{\Theta}_9 = \alpha_{\mathrm{bhsd}}$	0.729	0.799
$\mathbf{\Theta}_{10} = \mathrm{b_{hbs}}$	0.081	0.092
$\mathbf{\Theta}_{11} = \mathrm{R_{fhbs}}$	0	0

Table 3.3: p-value estimates for Approach C (with the permutation-based test procedure).

Approaches A, B, and C are not affected by the choice of the HSIC estimator (U-statistic vs. V-statistic).

In practice, the thermal conductivity cannot be regarded as a constant parameter. On the contrary, it is subject to an epistemic uncertainty quantified in the Bayesian setting by a posterior distribution $\mathbb{P}_{\mathbf{\Lambda}}$, which corresponds $\pi(\lambda \mid \boldsymbol{w})$ in Chapter 2, from a preliminary calibration of the thermal model. More precisely, this distribution has been estimated thanks to measurements w (of the output of the thermal conductivity) in a previous work by [Bouloré et al., 2023]. Of course, the uncertainty affecting this conductivity parameter must be incorporated into the study by adapting both the Bayesian calibration and GSA methods. This situation is illustrated in Figure 3.2 where the thermal and fission gas behavior models are referred to as Models 1 and 2 respectively. A conditional Bayesian calibration method has been recently developed in Chapter 2 in order to derive the posterior distribution of $\mathbf{\Theta}$ using a) the posterior distribution of $\mathbf{\Lambda}$, b) the prior distribution of $\mathbf{\Theta}$ and c) the experimental observations $\boldsymbol{z} = \{\boldsymbol{z}_i\}_{i=1}^{n_{\mathrm{exp}}}$. This method works pretty well as long as the number of p of calibration parameters remains small. Otherwise, a preliminary step of GSA becomes essential and the way to proceed this *conditional GSA* (i.e. GSA in presence of the extra uncertainty brought by $\mathbf{\Lambda}$ is not addressed in Chapter 2. Our paper provides an innovative solution on this point.

A first solution to achieve the *conditional GSA* would be to repeat the GSA as many times as there are thermal conductivity of interest. In this way, average indicators such as average HSIC indices and detection rates can be calculated. However, these averaged p-values do cannot decide whether a model parameter is influential or not, in the sense of a conventional statistical test. This has led us to define new indices suitable for incorporating the uncertainty of $\mathbf{\Lambda}$. These new sensitivity measures, which are the main topic of the next section, have a theoretical basis and the corresponding estimators are consistent with convergence rates similar to those of classical Monte Carlo estimators.

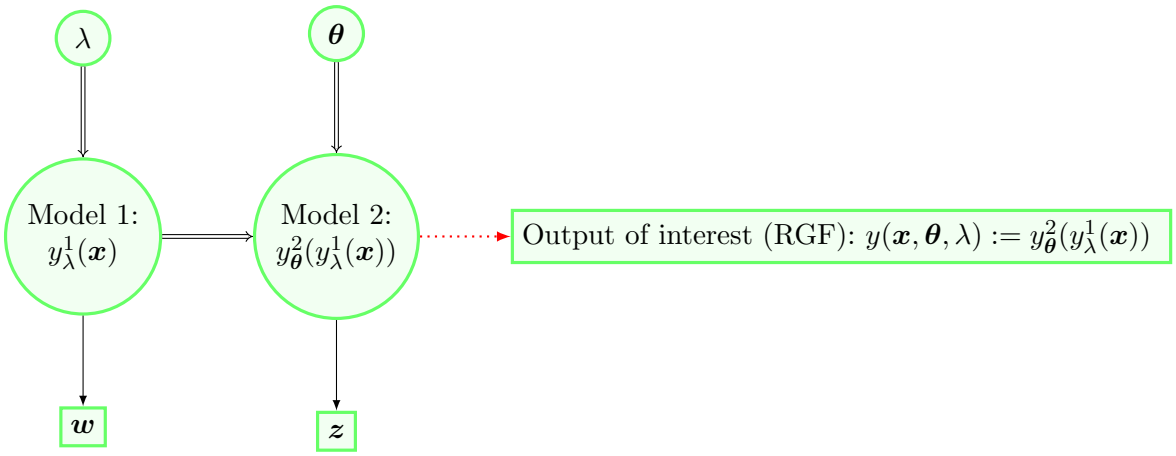


Figure 3.2: Two chained numerical models. Single and double arrows indicate stochastic and logical dependencies respectively. Square nodes are used to represent the observed data used for calibration tasks.

3.5 Development of a screening-oriented GSA methodology

Our contribution to conditional GSA can be summarized in four successive steps. First, we define sensitivity measures adapted to the problem of conditional GSA using the expectation of the HSIC indices with respect to $\boldsymbol{\Lambda}$. We show that these sensitivity measures characterize independence (in the presence of the additional uncertainty provided by $\boldsymbol{\Lambda}$). We then propose four possible estimators of these sensitivity measures and study their theoretical properties (bias, consistency, convergence rate). Finally, we propose two test procedures (one parametric and one non-parametric) to rigorously test (H_0) in the presence of the uncertainty associated with $\boldsymbol{\Lambda}$.

3.5.1 Construction and estimation of a sensitivity measure able to cope with the bi-level uncertainty framework

In the previous section, the HSIC was used to measure the level of independence between $\boldsymbol{\Theta}_j$ and $f(\boldsymbol{\Theta}) = \widetilde{f}(\boldsymbol{\Theta}, \lambda_{\mathrm{nom}})$. The function $\widetilde{f}(\cdot, \cdot)$ is nothing but a different way to rewrite the function $f(\cdot, \cdot)$ introduced in Section 3.3.3. The main advantage of $\widetilde{f}$ over f is to insist on the rule played by the thermal conductivity. More precisely, the focus was on the indices $H_j(\lambda_{\mathrm{nom}}) := \mathrm{HSIC}\left(\boldsymbol{\Theta}_j, \widetilde{f}(\boldsymbol{\Theta}, \lambda_{\mathrm{nom}})\right)$ defined as follows:

$$\begin{aligned} H_j(\lambda_{\mathrm{nom}}) = \mathbb{E}_{\boldsymbol{\Theta}\boldsymbol{\Theta}'}\left[K_j\left(\boldsymbol{\Theta}_j,\ \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta},\ \lambda_{\mathrm{nom}}),\ \widetilde{f}\left(\boldsymbol{\Theta}',\ \lambda_{\mathrm{nom}}\right)\right)\right] \\ + \mathbb{E}_{\boldsymbol{\Theta},\ \boldsymbol{\Theta}'}\left[K_j\left(\boldsymbol{\Theta}_j,\ \boldsymbol{\Theta}'_j\right)\right] \mathbb{E}_{\boldsymbol{\Theta}\boldsymbol{\Theta}'}\left[K_y\left(\widetilde{f}(\boldsymbol{\Theta},\ \lambda_{\mathrm{nom}}),\ \widetilde{f}\left(\boldsymbol{\Theta}',\ \lambda_{\mathrm{nom}}\right)\right)\right] \\ - 2\mathbb{E}_{\boldsymbol{\Theta}\boldsymbol{\Theta}'\boldsymbol{\Theta}''}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta},\ \lambda_{\mathrm{nom}}),\ \widetilde{f}\left(\boldsymbol{\Theta}'',\ \lambda_{\mathrm{nom}}\right)\right)\right]. \end{aligned} \tag{3.32}$$

where $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}'$ and $\boldsymbol{\Theta}''$ are independent and identically distributed according to $\mathbb{P}_{\boldsymbol{\Theta}}$ (i.e., $\pi(\boldsymbol{\theta})$ in Chapter 2). In this section, the thermal conductivity is no longer fixed at the nominal value λ_{nom} and is now random and modeled by a random variable with probability distribution $\mathbb{P}_{\boldsymbol{\Lambda}}$ [Bouloré et al., 2023]. Because of the randomness introduced by $\boldsymbol{\Lambda}$, the quantity $H_j(\boldsymbol{\Lambda})$ becomes random as well:

$$\begin{aligned} H_j(\boldsymbol{\Lambda}) = \mathbb{E}_{\boldsymbol{\Theta}\boldsymbol{\Theta}'}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right)\right)\right] \\ + \mathbb{E}_{\boldsymbol{\Theta}\boldsymbol{\Theta}'}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right)\right] \mathbb{E}_{\boldsymbol{\Theta}\boldsymbol{\Theta}'}\left[K_y\left(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right)\right)\right] \\ - 2\mathbb{E}_{\boldsymbol{\Theta}\boldsymbol{\Theta}'\boldsymbol{\Theta}''}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}'', \boldsymbol{\Lambda}\right)\right)\right]. \end{aligned} \tag{3.33}$$

Indeed, as the expectations in Equation (3.33) are only taken over $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}'$ and $\boldsymbol{\Theta}''$, we do not integrate with respect to $\boldsymbol{\Lambda}$ and therefore $H_j(\boldsymbol{\Lambda})$ is a random variable. Note that the random variable $\boldsymbol{\Lambda}$ is assumed to be independent of $\boldsymbol{\Theta}$. As a consequence, we can rewrite $H_j(\boldsymbol{\Lambda})$ using conditional expectations:

$$\begin{aligned} H_j(\boldsymbol{\Lambda}) = \mathbb{E}\left[K_j\left(\boldsymbol{\Theta}_j,\ \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta},\boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}',\ \boldsymbol{\Lambda}\right)\right) \mid \boldsymbol{\Lambda}\right] \\ + \mathbb{E}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right)\right] \mathbb{E}\left[K_y\left(\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}),\ \widetilde{f}\left(\boldsymbol{\Theta}',\ \boldsymbol{\Lambda}\right)\right) \mid \boldsymbol{\Lambda}\right] \\ - 2\mathbb{E}\left[K_j\left(\boldsymbol{\Theta}_j,\ \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}),\ \widetilde{f}\left(\boldsymbol{\Theta}'',\ \boldsymbol{\Lambda}\right)\right) \mid \boldsymbol{\Lambda}\right]. \end{aligned} \tag{3.34}$$

We can switch from Equation (3.33) to Equation (3.34) only because $\boldsymbol{\Lambda}$ is independent of $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}'$ and $\boldsymbol{\Theta}''$.

3.5.1.1 Definition of a specific sensitivity measure $\mathscr{H}_j$

We aim to identify the non-influential model parameter $\boldsymbol{\Theta}_j$ of the fission gas model while accounting for the uncertainty of the thermal conductivity $\boldsymbol{\Lambda}$ in the HSIC criterion. This can be done by computing the mean value of each random variable $H_j(\boldsymbol{\Lambda})$ [Hart et al., 2017]:

$$\forall\ 1 \leq j \leq p,\ \ \mathscr{H}_j = \mathbb{E}_{\boldsymbol{\Lambda}}\left(H_j\left(\boldsymbol{\Lambda}\right)\right). \tag{3.35}$$

To determine whether the indices $\mathscr{H}_j$ are well adapted to parameter screening, it must be checked whether they characterize independence. In other words, it must be checked whether $\mathscr{H}_j = 0$ is equivalent to $\boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})$. The proposition below indicates that it is actually what happens.

Proposition 2

For all $j \in \{1, \cdots, p\}$, the nullity of $\mathscr{H}_j$ is directly related to the independence between the random variables $\boldsymbol{\Theta}_j$ and $\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda})$. More precisely, we can express it as:

$$\mathscr{H}_j = 0 \iff \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}). \tag{3.36}$$

A proof of Proposition 2 can be found in Appendix 3.8.4.

Remark 1

Another property of the random variable $H_j(\boldsymbol{\lambda})$ could also have been used, for instance its upper bound:

$$\sup_{\boldsymbol{\lambda} \in \mathscr{D}_{\boldsymbol{\Lambda}}} H_j(\boldsymbol{\lambda}). \tag{3.37}$$

It is pretty simple to show that this quantity characterizes well the independence between $\boldsymbol{\Theta}_j$ and $\widetilde{f}\left(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}\right)$. However, it seems much harder to estimate (accurately) from a reasonable amount of simulated data.

3.5.1.2 Estimation of $\mathscr{H}_j$

The simplest way to estimate $\mathscr{H}_j$ is to adopt a two-step sampling strategy. First, a m-sample $D_{\boldsymbol{\Lambda},m} := \left\{\Lambda^{(k)}\right\}_{1 \leq k \leq m}$ of $\mathbb{P}_{\boldsymbol{\Lambda}}$ is generated and the empirical mean of $H_j(\boldsymbol{\lambda})$ is used to approximate the expectation of $H_j(\lambda)$ over $\boldsymbol{\Lambda}$:

$$\widehat{\mathscr{H}_j} = \tfrac{1}{m}\textstyle\sum_{k=1}^{m} H_j\left(\boldsymbol{\Lambda}^{(k)}\right) \quad \text{with} \quad H_j\left(\boldsymbol{\Lambda}^{(\boldsymbol{k})}\right) = \mathrm{HSIC}\left(\boldsymbol{\Theta}_j, \widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}^{(\boldsymbol{k})})\right). \tag{3.38}$$

Then, since $H_j\left(\boldsymbol{\Lambda}^{(\boldsymbol{k})}\right)$ cannot be calculated analytically, it must also be estimated. By adding this second level of approximation in Equation (3.38), $\widehat{\mathscr{H}_j}$ becomes $\widehat{\widehat{\mathscr{H}_j}}$:

$$\widehat{\widehat{\mathscr{H}_j}}^{\mathrm{stat}} = \frac{1}{m}\sum_{k=1}^{m} \widehat{H}_j^{\mathrm{stat}}\left(\boldsymbol{\Lambda}^{(k)}\right), \tag{3.39}$$

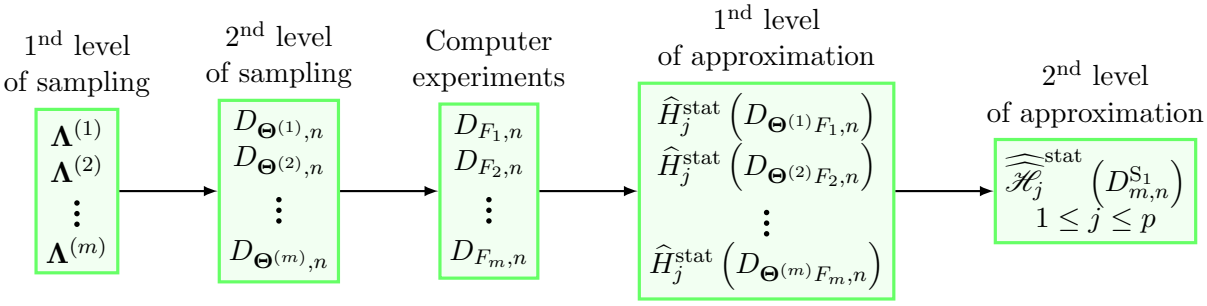


Figure 3.3: The first sampling strategy (S1) based on the use of a specific $D_{\boldsymbol{\Theta}^{(k)},n}$ for each conductivity value $\boldsymbol{\Lambda}^{(k)}$.

with stat $\in \{\mathrm{U},\ \mathrm{V}\}$ denotes either the U-statistic or V-statistic estimator of $H_j\left(\boldsymbol{\Lambda}^{(\boldsymbol{k})}\right)$. The question is how to obtain sufficiently many samples of the random variables $\boldsymbol{\Theta}_j$ and $F_k := \widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}^{(\boldsymbol{k})})$ to be able to estimate (accurately) all indices $H_j\left(\boldsymbol{\Lambda}^{(\boldsymbol{k})}\right) = \mathrm{HSIC}\left(\boldsymbol{\Theta}_j, F_k\right)$ $(1 \leq k \leq m)$. For this, two sampling strategies can be considered:

(S1) The first sampling strategy is to generate a (new) Monte Carlo design of $\boldsymbol{\Theta} \sim \mathbb{P}_{\boldsymbol{\Theta}}$ for each (new) conductivity value $\boldsymbol{\Lambda}^{(k)}$ found in $D_{\boldsymbol{\Lambda},m}$. In terms of probabilistic modeling, that means the random vector $\boldsymbol{\Theta}^{(k)}$ (which a copy of $\boldsymbol{\Theta} \sim \mathbb{P}_{\boldsymbol{\Theta}}$) is associated to $\boldsymbol{\Lambda}^{(k)}$ and then used to draw a n-sample of $\mathbb{P}_{\boldsymbol{\Theta}}$ denoted by

$$D_{\boldsymbol{\Theta}^{(k)},n} := \left\{\boldsymbol{\Theta}^{(k,l)}\right\}_{1\leq l\leq n}.$$

The corresponding values of F_k are then computed with $D_{\boldsymbol{\Theta}^{(k)},n}$, $\boldsymbol{\Lambda}^{(k)}$ and $\widetilde{f}$, and gathered in a dataset

$$D_{F_k,n} = \left\{F_k^{(l)}\right\} \quad \text{with} \quad F_k^{(l)} := \widetilde{f}(\boldsymbol{\Theta}^{(k,l)}, \boldsymbol{\Lambda}^{(\boldsymbol{k})}).$$

Furthermore, in order to simplify future developments, some other notations need to be introduced:

$$D_{m,n}^{\mathrm{S1}} = \left\{D_{\boldsymbol{\Theta}^{(k)}F_k,n}\right\}_{1\leq k\leq m} \quad \text{with} \quad D_{\boldsymbol{\Theta}^{(k)}F_k,n} = \left\{D_{\boldsymbol{\Theta}^{(k)},n}; D_{F_k,n}\right\}.$$

In particular, the dataset $D_{\boldsymbol{\Theta}^{(k)}F_k,n}$ is composed of n realizations of the pair $(\boldsymbol{\Theta}, F_k)$ and can be used to estimate all indices $H_j\left(\boldsymbol{\Lambda}^{(k)}\right)$ for $1 \leq j \leq p$. Likewise, the dataset $D_{m,n}^{\mathrm{S1}}$ contains all the useful data to estimate the indices $\mathscr{H}_j$. This sampling strategy is illustrated in Figure 3.3.

(S2) The second sampling strategy is to use the same Monte Carlo design $D_{\boldsymbol{\Theta},n}$ for all the conductivity values. In this context, the notations from (S1) become:

- $D_{\boldsymbol{\Theta},n} := \left\{\boldsymbol{\Theta}^{(l)}\right\}_{1\leq l\leq n}$,
- $\forall\, 1 \leq k \leq m$, $D_{F_k,n} = \left\{F_k^{(l)}\right\}$ with $\widetilde{f}(\boldsymbol{\Theta}^{(l)}, \boldsymbol{\Lambda}^{(\boldsymbol{k})})$, $D_{\boldsymbol{\Theta}F_k,n} = \{D_{\boldsymbol{\Theta},n}; D_{F_k,n}\}$,
- $D_{m,n}^{\mathrm{S2}} = \{D_{\boldsymbol{\Theta}F_k,n}\}_{1\leq k\leq m}$.

This sampling strategy is illustrated in Figure 3.4.

Note that the Monte Carlo designs $D_{\boldsymbol{\Theta}^{(k)},n}$ are mutually independent designs. They are also independent from the design $D_{\boldsymbol{\Lambda},n}$. Since Strategies (S1) and (S2) and two HSIC estimators (U-statistic and V-statistic) are available, four different estimators of $\mathscr{H}_j$ can be built:

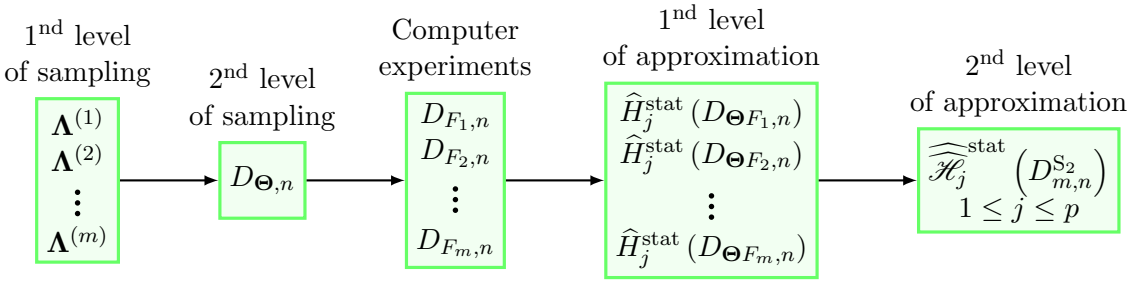


Figure 3.4: The second sampling strategy (S2) based on the use of a specific $D_{\mathbf{\Theta},n}$ for all the conductivity values $\mathbf{\Lambda}^{(k)}$ $(1 \leq k \leq m)$.

- **Estimator 1:** $\widehat{\widehat{\mathscr{H}_j}}^{\text{U}}\left(D_{m,n}^{\text{S}_1}\right) = \frac{1}{m}\sum_{k=1}^{m} \widehat{H}_j^{\text{U}}\left(D_{\mathbf{\Theta}^{(k)}F_k,n}\right) =: \widehat{\widehat{\mathscr{H}_j}}^{1}$,
- **Estimator 2:** $\widehat{\widehat{\mathscr{H}_j}}^{\text{V}}\left(D_{m,n}^{\text{S}_1}\right) = \frac{1}{m}\sum_{k=1}^{m} \widehat{H}_j^{\text{U}}\left(D_{\mathbf{\Theta}^{(k)}F_k,n}\right) =: \widehat{\widehat{\mathscr{H}_j}}^{2}$,
- **Estimator 3:** $\widehat{\widehat{\mathscr{H}_j}}^{\text{U}}\left(D_{m,n}^{\text{S}_2}\right) = \frac{1}{m}\sum_{k=1}^{m} \widehat{H}_j^{\text{U}}\left(D_{\mathbf{\Theta}F_k,n}\right) =: \widehat{\widehat{\mathscr{H}_j}}^{3}$,
- **Estimator 4:** $\widehat{\widehat{\mathscr{H}_j}}^{\text{V}}\left(D_{m,n}^{\text{S}_2}\right) = \frac{1}{m}\sum_{k=1}^{m} \widehat{H}_j^{\text{V}}\left(D_{\mathbf{\Theta}F_k,n}\right) =: \widehat{\widehat{\mathscr{H}_j}}^{4}$.

To denote any of these four estimators, one can write $\widehat{\widehat{\mathscr{H}_j}}$. In order to find out which estimator best meets the needs of conditional GSA, the statistical properties (bias, consistency, convergence rate) of the estimators $\widehat{\widehat{\mathscr{H}_j}}^{1}$, $\widehat{\widehat{\mathscr{H}_j}}^{2}$, $\widehat{\widehat{\mathscr{H}_j}}^{3}$ and $\widehat{\widehat{\mathscr{H}_j}}^{4}$ need to be investigated:

- Knowing if $\widehat{\widehat{\mathscr{H}_j}}$ is biased allows us to know whether the estimator $\widehat{\widehat{\mathscr{H}_j}}$ returns the target value $\widehat{\widehat{\mathscr{H}_j}}$ on average.
- Regardless of an estimator is biased or not, studying its consistency allows to know if the estimation error tends to 0 when the number of samples involved in the dataset ($D_{m,n}^{\text{S}_1}$ and $D_{m,n}^{\text{S}_2}$) becomes infinitely large.
- For two consistent estimators, it is preferable to choose the one which converges at the fastest rate to the target value. Indeed, for a prescribed simulation budget (i.e., for prescribed number N of input-output evaluations of the computer code), this estimator guarantees better accuracy.

Let us recall that $\widehat{\widehat{\mathscr{H}_j}}$ is consistent if $\widehat{\widehat{\mathscr{H}_j}}$ converges in probability towards $\mathscr{H}_j$:

$$\forall\, \widetilde{\epsilon} > 0, \quad \lim_{m,n\to\infty} \mathbb{P}\left(\left|\widehat{\widehat{\mathscr{H}_j}} - \mathscr{H}_j\right| > \widetilde{\epsilon}\right) = 0. \tag{3.40}$$

A sufficient condition for this is to show that $\widehat{\widehat{\mathscr{H}_j}}$ converges towards $\mathscr{H}_j$ in in the 2-th mean (or in L^2-norm):

$$\lim_{m,n\to\infty} \mathbb{E}\left[\left|\widehat{\widehat{\mathscr{H}_j}} - \mathscr{H}_j\right|^2\right] = 0. \tag{3.41}$$

This type of convergence is stronger than convergence in probability. However, it has the advantage of often being easier to demonstrate, and of revealing convergence rates more readily.

In Proposition 3, it is shown that for the four estimators are consistent but only two of them are unbiased. Beyond that, upper bounds for the convergence rates are also provided. Such insights are truly valuable because they allow us to better grasp how each construction

option (sampling strategy and HSIC estimator) impacts the statistical properties of $\widehat{\widehat{\mathscr{H}_j}}$. Furthermore, Proposition 3 also helps to distribute the simulation budget ($N = mn$) between the two levels of sampling. For fixed N, the question is indeed how to adjust m and n.

Proposition 3

It is assumed that the kernels $\{K_j\}_{1\leq j\leq p}$ and K_y are bounded and under Strategies $(S1)$ and $(S2)$, we obtain the following results:

1. **Properties of Estimator 1:** $\widehat{\widehat{\mathscr{H}_j}}^1$ is unbiased, consistent and
$$\mathbb{E}\left[\left|\widehat{\widehat{\mathscr{H}_j}}^1 - \mathscr{H}_j\right|^2\right] = \mathscr{O}\left(\frac{1}{m}\right).$$
2. **Properties of Estimator 2:** $\widehat{\widehat{\mathscr{H}_j}}^2$ is biased, consistent and
$$\mathbb{E}\left[\left|\widehat{\widehat{\mathscr{H}_j}}^2 - \mathscr{H}_j\right|^2\right] = \mathscr{O}\left(\frac{1}{m} + \frac{1}{n^2}\right).$$
3. **Properties of Estimator 3:** $\widehat{\widehat{\mathscr{H}_j}}^3$ is unbiased, consistent and
$$\mathbb{E}\left[\left|\widehat{\widehat{\mathscr{H}_j}}^3 - \mathscr{H}_j\right|^2\right] = \mathscr{O}\left(\frac{1}{m} + \frac{1}{n}\right).$$
4. **Properties of Estimator 4:** $\widehat{\widehat{\mathscr{H}_j}}^4$ is biased, consistent and
$$\mathbb{E}\left[\left|\widehat{\widehat{\mathscr{H}_j}}^4 - \mathscr{H}_j\right|^2\right] = \mathscr{O}\left(\frac{1}{m} + \frac{1}{n}\right).$$

The detailed proofs can be found in Appendix 3.8.5. It should be mentioned that these developments are inspired from ideas found in a recent work by Fellmann et al. [2023].

Remark 2

It is worth noting that the assumption of bounded kernels is pretty standard in the literature related to HSIC indices especially to obtain concentration inequalities (see Theorem 4 in [Song et al., 2012]). This assumption is not very stringent as it is verified by most of the translation-kernel kernels used in statistical applications (Gaussian kernels, exponential kernels, Matérn kernels).

Remark 3

It should be clearly understood that a mean square (MSE) $\mathbb{E}\left[\left|\widehat{\widehat{\mathscr{H}_j}} - \mathscr{H}_j\right|^2\right]$ equal to $\mathscr{O}\left(\frac{1}{m}\right)$ does not mean that the MSE is asymptotically equivalent to $\frac{1}{m}$ as $m \to \infty$. It is only means that the MSE can be upper bounded by $\frac{C}{m}$ where C is a positive constant (which may be very large). If we had a use case for which the exact values of the indices $\{\mathscr{H}_j\}_{1\leq j\leq p}$ were known, we could have compared the convergence rates of Proposition 3 with the true convergence rates, and the latters could have been much faster. Thus, the rates from Proposition 3 must be seen as worst-case rates.

Proposition 3 is very instructive as regards the statistical properties of the four estimators:

- **Estimator 1:** It is unbiased and this is basically due to the exclusive presence of U-statistic. The MSE decays at a rate of $\frac{1}{m}$. In that light, it seems much smarter to take large values for m (number of points used to estimate the expectation over $\mathbf{\Lambda}$) rather than n (number of points used to estimate the HSIC). However, as will be seen latter, n cannot be taken too small, at the risk of deteriorating too much the statistical power associated with HSIC-based tests of independence. For this reason, it is often advised to preserve $n \geq 100$.

- **Estimator 2:** It is biased and this bias naturally comes from the use of V-statistics (instead of U-statistics in Estimator 1), the approximation error cannot be completely eliminated by simply increasing m. Indeed, this bias introduced by V-statistics is responsible for the term $\frac{1}{n^2}$ and it can only be reduced by playing on n.

- **Estimators 3 and 4:** Same remarks as for Estimators 1 and 2 as regards the lack/presence of bias. Their convergence rate, namely $\mathscr{O}\left(\frac{1}{m}+\frac{1}{n}\right)$, proves to be slower than that of Estimators 1 and 2. This pretty discredits (S2) in favor of (S1) which therefore emerges as the best sampling strategy to reach maximum accuracy for a limited simulation budget. That is why *a*) Estimators 3 and 4 are not implemented in Section 3.6 and *b*) test procedures are only developed for Estimators 1 and 2 in the next section. Moreover, if you have a look at the proof (see Appendix 3.8.5) related to $\widehat{\widehat{\mathscr{H}_j}}^4$ it can even be proved that the MSE is equal to $\mathscr{O}\left(\frac{1}{m}+\frac{1}{n}+\frac{1}{n^2}\right)$ where:

  - the term $\frac{1}{m}$ is due to the empirical mean over the m values of the thermal conductivity,
  - the term $\frac{1}{n}$ is caused by the use of (S2),
  - the term $\frac{1}{n^2}$ comes from the use of V-statistic (and the resulting bias).

In this section, a sensitivity measure adapted to the conditional GSA problem has been put forth. The resulting sensitivity indices $\{\mathscr{H}_j\}_{1\leq j\leq p}$ seem to be well-adapted to the screening of the calibration parameters since each index $\mathscr{H}_j$ is able to characterize the input-output independence between $\mathbf{\Theta}_j$ and $\widetilde{f}(\mathbf{\Theta},\mathbf{\Lambda})$. Besides, two consistent estimators have been built and their convergence rates match classic Monte Carlo standards, which suggests that the indices can be estimated accurately even with a small amount of simulation data. However, just as for basic HSIC indices, the values of $\{\mathscr{H}_j\}_{1\leq j\leq p}$ cannot be compared to each other (because they may correspond to different MMD scales, especially if the calibration parameters are distinct variation ranges and/or are equipped with different kernels). In addition, the decision to accept/reject $\mathbf{\Theta}_j$ cannot be made on the only basis of the estimate returned by $\widehat{\widehat{\mathscr{H}_j}}$. To go further, a test procedure must be developed.

3.5.2 Two independence tests for parameter screening in the particular context of conditional calibration

The objective is to develop a data-driven procedure (only based on the simulation data $D_{m,n} := D^{\mathrm{S}_1}_{m,n}$ previously used for GSA) to make a decision between the following statements:

$$(\mathrm{H}_0) : \mathbf{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\mathbf{\Theta},\mathbf{\Lambda}) \quad \text{vs.} \quad (\mathrm{H}_1) : \mathbf{\Theta}_j \not\perp\!\!\!\perp \widetilde{f}(\mathbf{\Theta},\mathbf{\Lambda}).$$

In virtue of Proposition 2, these statements may be rewritten as:

$$(\mathrm{H}_0) : \mathscr{H}_j = 0 \quad \text{vs.} \quad (\mathrm{H}_1) : \mathscr{H}_j > 0.$$

Naturally, the test statistic will be either $\widehat{\widehat{\mathscr{H}_j}}^1$ or $\widehat{\widehat{\mathscr{H}_j}}^2$ (see Section 3.5.1) and (H_0) will be rejected when the test statistic takes on too large a value. By *too large*, it is meant by greater

than the quantile of order $1-\alpha$ (where $\alpha = 5\%$ is the *type-I error*) of the distribution $\widehat{\widehat{\mathscr{H}_j}}$ under (H_0). Let $D_{m,n}^{\mathrm{obs}}$ denote the available dataset (obtained after N runs of the numerical model), as opposed to $D_{m,n}$ which is the associated underlying random design. The p-value of the test is defined by:

$$p_j = \mathbb{P}_{\mathrm{H}_0}\left(\widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}\right) > \widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}^{\mathrm{obs}}\right)\right). \tag{3.42}$$

Of course, its estimation cannot be considered without knowledge (even approximate) of the distribution of $\widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}\right)$ under (H_0). Therefore, all that remains is to determine the distribution of $\widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}\right)$ under (H_0).

In Section 3.3.2, it was said that the distribution of an HSIC index (under the null hypothesis) could be approximated in three different ways:

(a) When n is very large, the HSIC (once multiplied by n) follows asymptotically a spectral distribution [Gretton et al., 2007; Zhang et al., 2018] that can be approximated by a Gamma distribution. Its parameters are then computed from the method of moments (based on asymptotic formulas of the first two moments).

(b) When n is very small, the distribution of the HSIC can only be simulated with a random permutation mechanism and the p-value is estimated with a non-parametric approach [De Lozzo and Marrel, 2016; El Amri and Marrel, 2022].

(c) In any intermediate situation (for example when n is equal to a few hundreds), the (discrete) distribution of the HSIC (for a fixed dataset, and after completing all possible permutations of the output data) can again be approximated by a (continuous) Gamma distribution. In that case, the parameters are computed from the method of moments (based this time on non-asymptotic but still exact formulas of the first two moments [Kazi-Aoual et al., 1995; El Amri and Marrel, 2024].

Here, the idea is trying to do the same in the particular context of $\widehat{\widehat{\mathscr{H}_j}}$. The asymptotic framework (a), which would be to find the asymptotic distribution of $\widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}\right)$ under (H_0) when both m and n tend to infinity, seems too difficult to leverage. It is left for future work. On the opposite, the test procedures (b) and (c) can be easily adapted to our problem, as shown in the next two sections.

Remark 4

Testing the probabilistic independence between $\boldsymbol{\Theta}_j$ and $\widetilde{f}\left(\boldsymbol{\Theta}, \boldsymbol{\Lambda}\right)$ is not equivalent to testing the following statement:

$$(\widetilde{\mathrm{H}}_0): \ \forall\, \lambda \in \mathscr{D}_{\boldsymbol{\Lambda}}, \ \boldsymbol{\Theta} \perp\!\!\!\perp \widetilde{f}\left(\boldsymbol{\Theta}, \lambda\right).$$

Indeed, $(\widetilde{\mathrm{H}}_0)$ implies H_0 but the reverse is false because, one has:

$$\begin{aligned}
(\mathrm{H}_0) &\iff \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}\left(\boldsymbol{\Theta}, \boldsymbol{\Lambda}\right) \iff \mathscr{H}_j = 0 \\
&\iff \mathrm{HSIC}\left(\boldsymbol{\Theta}_j, \widetilde{f}\left(\boldsymbol{\Theta}, \boldsymbol{\Lambda}\right)\right) = 0 \text{ a.s.} \\
&\iff \exists\, \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}} \subseteq \mathscr{D}_{\boldsymbol{\Lambda}} \text{ such that } \mathbb{P}_{\boldsymbol{\Lambda}}\left(\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}\right) = 1 \text{ and } \forall\, \boldsymbol{\lambda} \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}, \mathrm{HSIC}\left(\boldsymbol{\Theta}_j, \widetilde{f}\left(\boldsymbol{\Theta}, \lambda\right)\right) = 0 \\
&\iff \exists\, \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}} \subseteq \mathscr{D}_{\boldsymbol{\Lambda}} \text{ such that } \mathbb{P}_{\boldsymbol{\Lambda}}\left(\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}\right) = 1 \text{ and } \forall\, \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}, \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}\left(\boldsymbol{\Theta}, \lambda\right) \\
&\not\Longleftarrow (\widetilde{H}_0).
\end{aligned}$$

3.5.2.1 Permutation-based test procedure

For an HSIC-based test of independence, remember that, when the sample size n becomes too small for the asymptotic procedure to be used, the distribution of the test statistic under the null hypothesis is simulated by means of a permutation-based mechanism (operating on the available dataset). Here, the idea is to draw inspiration from this approach to test

$$(\mathrm{H}_0): \ \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}\left(\boldsymbol{\Theta}, \boldsymbol{\Lambda}\right).$$

As described in Section 3.3.2, the idea behind the permutation method is to recompute the HSIC estimate from a design that contains permutations of the output vector (without permuting the inputs). In this section, we adopt this idea to simulate the test statistic $\mathscr{H}_j$ by applying permutations adapted to the $D_{m,n}$ design (concatenation of m mutually independent designs $\{D_{F_k,n}\}_{1\leq k\leq m}$). In this context, we apply a set of permutations $\tau_{b\bullet} : (\mathbb{S}_n)^m \longrightarrow (\mathbb{S}_n)^m$ to $D_{m,n}$ to obtain the following permuted design:

$$D_{m,n}^{\tau_{b\bullet}} = \left\{D_{\boldsymbol{\Theta}^{(k)},n}; D_{F_k,n}^{\tau_{bk}}\right\}_{1\leq k\leq m} \quad \text{with} \quad D_{F_k,n}^{\tau_{bk}} = \left\{F_k^{\tau_{bk}(l)}\right\}_{1\leq l\leq n},$$

where the random permutations $\{\tau_{bk}\}_{1\leq k\leq m}$ are one-to-one mappings from $\{1\cdots,n\}$ to $\{1\cdots,n\}$ and are applied to the designs $\{D_{F_k,n}\}_{1\leq k\leq m}$. In this way, we obtain a new realization of the test statistic given by:

$$\widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}^{\tau_{b\bullet}}\right).$$

Therefore, by repeating this procedure B_{perm} times, we can estimate the p-value non-parametrically as follows:

$$\widehat{p}_j = \frac{1}{B_{\text{perm}}} \sum_{b=1}^{B_{\text{perm}}} \mathbf{1}_{\left\{\widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}^{\tau_{b\bullet}}\right) > \widehat{\widehat{\mathscr{H}_j}}\left(D_{m,n}^{\text{obs}}\right)\right\}}. \tag{3.43}$$

Figure 3.5 summarizes this procedure built upon $B_{\text{perm}} \times m$ independent random permutations. For more details on approximation by random permutation, the interested reader may refer to De Lozzo and Marrel [2016] and El Amri and Marrel [2022].

3.5.2.2 Non-asymptotic Gamma test procedure

The non-asympotic Gamma approximation dedicated to V-statistic estimators, described in [El Amri and Marrel, 2024] and recalled in Section 3.3.2, can be used to approximate the probability distribution of $n^2\widehat{\widehat{\mathscr{H}_j}}^2$ with a Gamma distribution $\Gamma(\widetilde{a}_j, \widetilde{b}_j)$ without performing any permutation. Instead, it solely relies on the first two moments of the exact permutation distribution under the null hypothesis. To do this, the test statistic is rewritten as follows:

$$\widehat{\widehat{\mathscr{H}_j}}^2 = \frac{1}{mn^2} \sum_{k=1}^{m} Tr\left(\boldsymbol{A}_{jk}\boldsymbol{W}_k\right), \tag{3.44}$$

with

- $\boldsymbol{A}_{jk} = \boldsymbol{H}\boldsymbol{L}_{jk}\boldsymbol{H}$ being a double-centered, square symmetric matrix, where $\boldsymbol{L}_{jk}$ is the Gram matrix of $\boldsymbol{\Theta}_j$ constructed from the data $D_{\boldsymbol{\Theta}^{(k)},n}$,
- $\boldsymbol{W}_k = \boldsymbol{H}\boldsymbol{L}_{y_k}\boldsymbol{H}$ being a double-centered, square symmetric matrix, where $\boldsymbol{L}_{y_k}$ is the Gram matrix of $\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})$ constructed from the data $D_{F_k,n}$.

Using the results of Kazi-Aoual et al. [1995] and El Amri and Marrel [2024], for any $1 \leq k \leq m$, we can deduce the shape and scale parameters of Gamma distribution $\Gamma\left(a_j^{[k]}, b_j^{[k]}\right)$ for each

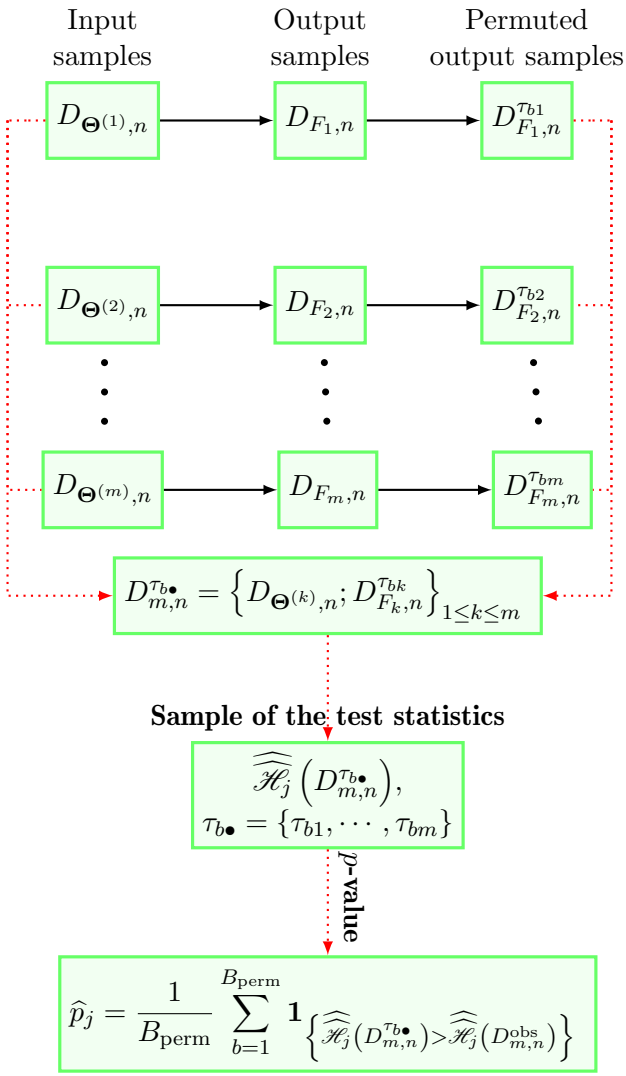


Figure 3.5: Summary diagram of the different stages of the permutation-based test procedure used to test (H_0). The input material only consists of simulation data $D_{m,n} = \left\{D_{\boldsymbol{\Theta}^{(k)},n}; D_{F_k,n}\right\}_{1\leq k\leq m}$ to estimate the sensitivity measures $\{\mathscr{H}_j\}_{1\leq j\leq p}$. The output results are the estimated p-values.

term $Tr\left(\boldsymbol{A}_{jk}\boldsymbol{W}_k\right)$ by the method of moments:

$$a_j^{[k]} = \frac{\mathbb{E}_{\text{perm}}\left[\text{Tr}(\boldsymbol{A}_{jk}\boldsymbol{W}_k)\right]^2}{\mathbb{V}_{\text{perm}}(\text{Tr}(\boldsymbol{A}_{jk}\boldsymbol{W}_k))}, \tag{3.45}$$

$$b_j^{[k]} = \frac{\mathbb{V}_{\text{perm}}(\text{Tr}(\boldsymbol{A}_{jk}\boldsymbol{W}_k))}{\mathbb{E}_{\text{perm}}\left[\text{Tr}(\boldsymbol{A}_{jk}\boldsymbol{W}_k)\right]}. \tag{3.46}$$

Then, by calculating the mean and variance of $\widehat{\widehat{\mathscr{H}_j}}^2$ under the permutation distribution, we derive the shape and scale parameters of Gamma distribution $\Gamma(\widetilde{a}_j,\ \widetilde{b}_j)$ by:

$$\begin{cases} \widetilde{a}_j = \dfrac{\mathbb{E}_{\substack{\text{perm} \\ \tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}_j}}^2\right)^2}{\mathbb{V}_{\substack{\text{perm} \\ \tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}_j}}^2\right)} = \dfrac{\left(\sum\limits_{k=1}^{m} a_j^{[k]} b_j^{[k]}\right)^2}{\sum\limits_{k=1}^{m} a_j^{[k]}\left(b_j^{[k]}\right)^2}, \\ \widetilde{b}_j = \dfrac{\mathbb{V}_{\substack{\text{perm} \\ \tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}_j}}^2\right)}{\mathbb{E}_{\substack{\text{perm} \\ \tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}_j}}^2\right)} = \dfrac{\sum\limits_{k=1}^{m} a_j^{[k]}\left(b_j^{[k]}\right)^2}{m\sum\limits_{k=1}^{m} a_j^{[k]} b_j^{[k]}}. \end{cases} \tag{3.47}$$

As a result, the p-value can be approximated by:

$$\widehat{p}_j = 1 - F_{\Gamma(\widetilde{a}_j,\,\widetilde{b}_j)}\left(\widehat{\widehat{\mathscr{H}_j}}^2\left(D_{m,n}^{\text{obs}}\right)\right), \tag{3.48}$$

where $F_{\Gamma(\widetilde{a}_j,\,\widetilde{b}_j)}(\cdot)$ is the cumulative distribution function of Gamma distribution $\Gamma(\widetilde{a}_j,\ \widetilde{b}_j)$. Figure 3.6 summarizes the scheme for estimating the shape and scale parameters of the final Gamma distribution $\Gamma(\widetilde{a}_j,\ \widetilde{b}_j)$. See Appendix 3.8.6.1 for details of the calculations.

We have defined a new sensitivity measure that takes into account the dual-level parametric uncertainty due to the chaining between Models 1 and 2 (Figure 3.2). This (proposed) kernel-based sensitivity measure includes the parametric uncertainty of Model 1 into the screening process for targeting the uncertain parameters of Model 2. As already said, this new sensitivity measure whose estimators are consistent characterizes the independence between $\boldsymbol{\Theta}_j$ and $\widetilde{f}\left(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}\right)$. In our application case, the ALCYONE application (see Section 3.4.1), we are interested in screening the uncertain parameters of the fission gas behavior model, while taking into account the uncertainty of the thermal conductivity coming from the thermal model. Therefore, we can use this proposed measure to detect the non-influential parameters of the fission gas behavior model conditionally on the uncertainty of the thermal conductivity.

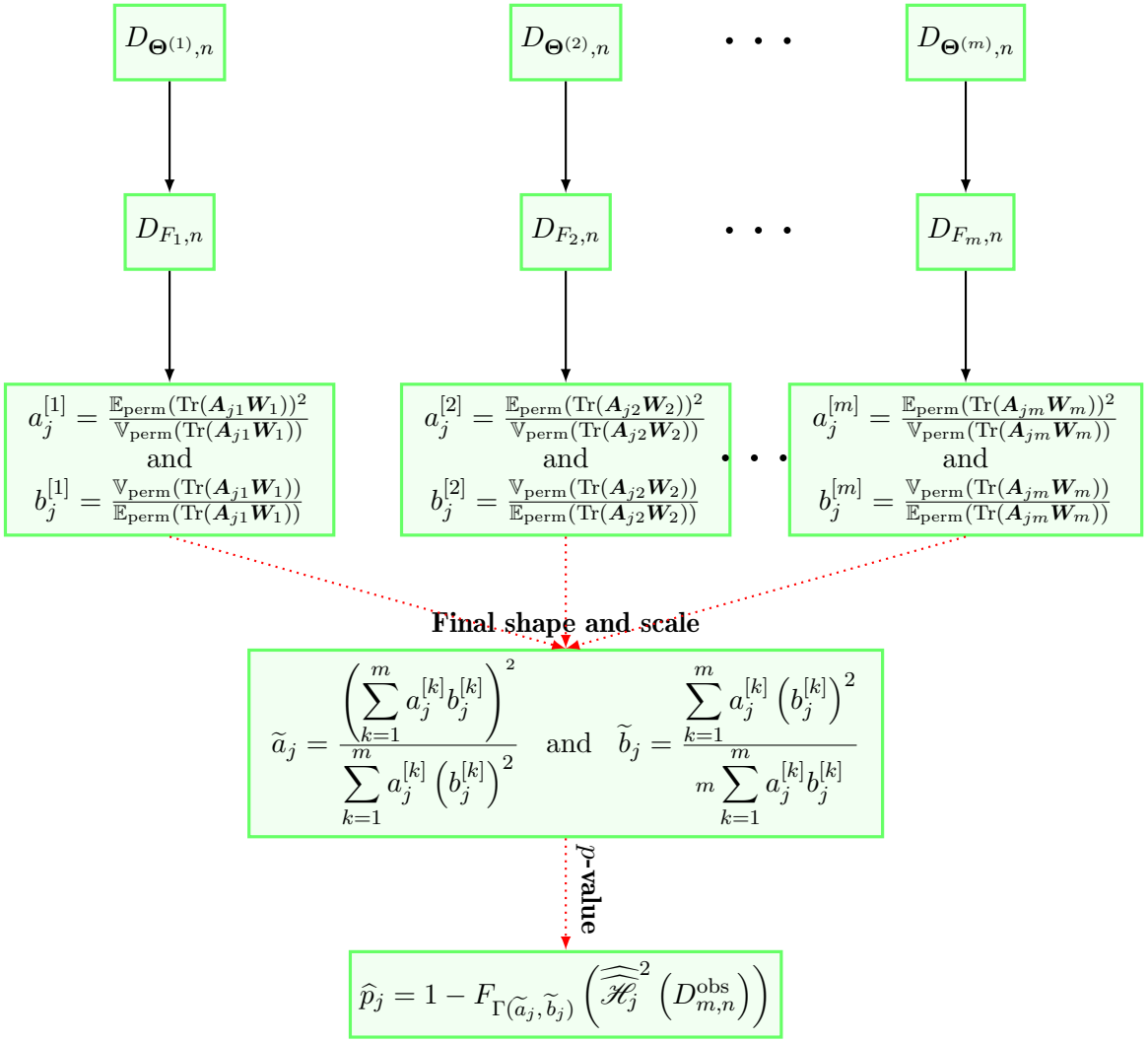


Figure 3.6: Summary diagram of the different stages of the non-asymptotic Gamma test procedure used to test (H_0). The input material only consists of simulation data $D_{m,n} = \left\{D_{\boldsymbol{\Theta}^{(k)},n}; D_{F_k,n}\right\}_{1 \leq k \leq m}$ to estimate the sensitivity measures $\{\mathscr{H}_j\}_{1 \leq j \leq p}$. The output results are the estimated p-values.

3.6 Results obtained on the ALCYONE test case

The new HSIC-based measures $\mathscr{H}_j$ and associated independence tests are now applied to the ALCYONE test case to identify the non-influential calibration parameters of the fission gas behavior model taking account the uncertainty of the conductivity parameter of the thermal model. To achieve this, the GSA methodology developed in Section 3.5 is implemented. In particular, the following steps are carried out:

(1) A Monte Carlo design $D_{\boldsymbol{\Lambda},m} = \left\{\boldsymbol{\Lambda}^{(k)}\right\}_{1\leq k\leq m}$ $(m = 20)$ is generated according to the posterior distribution inferred in [Bouloré et al., 2023].

(2) The input Monte Carlo designs $D_{\boldsymbol{\Theta}^{(1)},n}, \cdots, D_{\boldsymbol{\Theta}^{(m)},n}$ $(n = 200)$ are generated according to the sampling strategy (S1).

(3) The output designs $D_{F_1,n}, \cdots, D_{F_m,n}$ are computed for the three different types of output objects detailed in Section 3.4.2:

- Approach A: $\widetilde{f}_i(\boldsymbol{\theta}, \lambda) = y(\boldsymbol{x}_i, \boldsymbol{\theta}, \lambda)$ for $1 \leq i \leq n_{\text{exp}}$ $(n_{\text{exp}} = 40)$,
- Approach B: $\widetilde{f}(\boldsymbol{\theta}, \lambda) = [y(\boldsymbol{x}_i, \boldsymbol{\theta}, \lambda)]_{1\leq i\leq n_{\text{exp}}}$,
- Approach C: $\widetilde{f}(\boldsymbol{\theta}, \lambda) = \displaystyle\sum_{i=1}^{n_{\text{exp}}} \left(\frac{\boldsymbol{z}_i - y(\boldsymbol{x}_i, \boldsymbol{\theta}, \lambda)}{\sigma_{\epsilon_i}}\right)^2$.

This allows us to build the input-output designs $D_{\boldsymbol{\Theta}^{(1)},n}, \cdots, D_{\boldsymbol{\Theta}^{(m)},n}$ and the complete dataset $D^{\text{S1}}_{m,n} = \left\{D_{\boldsymbol{\Theta}^{(k)}F_k,n}\right\}_{1\leq k\leq m}$ with $D_{\boldsymbol{\Theta}^{(k)}F_k,n} = \left\{D_{\boldsymbol{\Theta}^{(k)},n}; D_{F_k,n}\right\}$.

(4) Each index $\mathscr{H}_j$ is estimated by means of $\widehat{\widehat{\mathscr{H}}}^1_j$ (mean of U-statistics) or $\widehat{\widehat{\mathscr{H}}}^2_j$ (mean of V-statistics).

(5) The two test procedures (see Sections 3.5.2.1 and 3.5.2.2) are performed to obtain the p-values $\widehat{p}_j$ $(1 \leq j \leq p)$. The permutation-based test procedure is run for both types of estimators, while the non-asymptotic Gamma test procedure is only run for $\widehat{\widehat{\mathscr{H}}}^2_j$.

(6) Decisions to accept/reject calibration parameters are made on the basis of a *type-I error* $\alpha = 5\%$.

First, Approach A is performed for each experimental configuration $\{\boldsymbol{x}_i\}_{1\leq i\leq n_{\text{exp}}}$. The obtained p-values are shown in Figure 3.7, where the x-axis represents the RGF values ordered by decreasing burning rate. We can see that for all n_{exp} fuel rods, the influential calibration parameters are R_{fhbs}, k_α, A_3, A_2 and b_{hbs}. It is worth mentioning that the parameters k_v and b_{gbb} appear to be influential for less than a half of the RGF values. Tables 3.4 summarizes the results from Figure 3.7 in terms of detection rates of each model parameter $\boldsymbol{\Theta}_j$ $(1 \leq j \leq p)$ on the set of n_{exp} RGF values. This table displays the detection rates obtained with $\widehat{\widehat{\mathscr{H}}}^1_j$ and $\widehat{\widehat{\mathscr{H}}}^2_j$. The results of Approach B are then represented in Table 3.5 and we can see that the parameters selected at the threshold $\alpha = 5\%$ are R_{fhbs}, k_α, A_3, A_2 and b_{gbb}. By way of comparison, Approach C which focuses on the least-squares function, selects R_{fhbs}, k_α, A_3, A_2 and b_{hbs}, whereas b_{gbb} is not selected (Table 3.6). This is reasonably consistent with the results of Approach A, where this parameter is kept for 9 fuel rods having low burnup rates. Approach B uses a tensorised kernel that scales all outputs to the same level. This could explain why b_{hbs} is not selected. Finally, the model parameters retained for calibration will be R_{fhbs}, k_α, A_3, A_2 and b_{hbs}. Note that the non-asymptotic Gamma test procedure (Figure 3.6) gives almost the same results than the permutation-based test procedure. Appendix 3.8

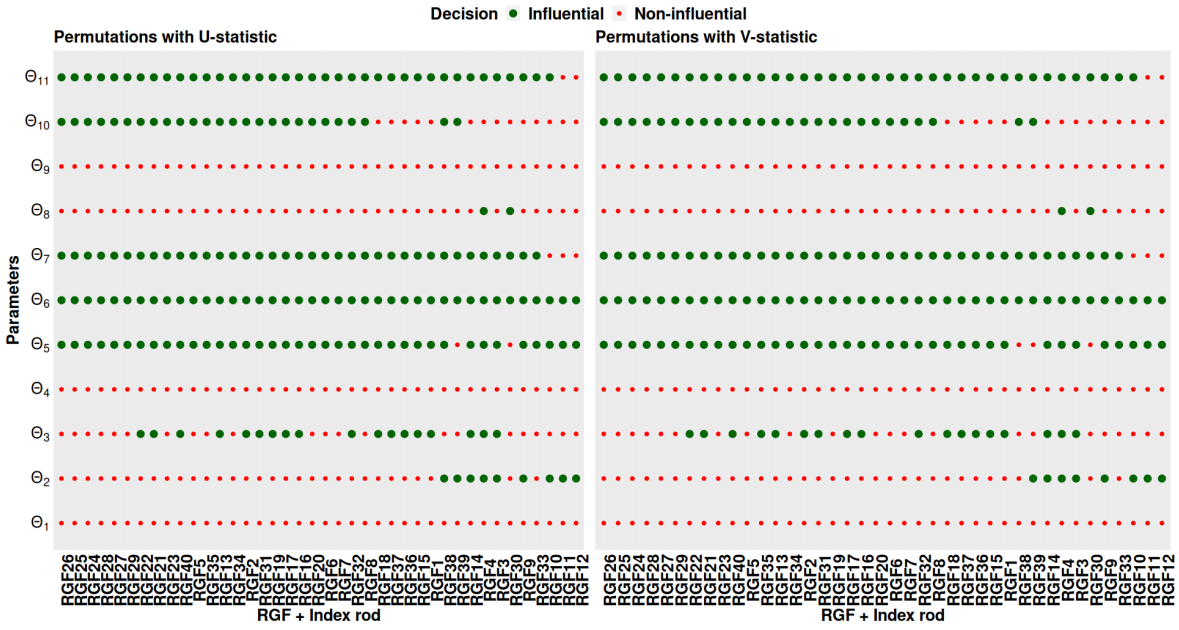


Figure 3.7: Summary of the results obtained when independence is tested with Approach A. On the left (resp. right) hand side, HSIC indices are computed with U-statistic (resp. V-statistics). The associated detection rates are given in Table 3.4.

Detection rates / Model parameters	with U-statistic	with V-statistic
$\mathbf{\Theta}_1 = \mathrm{k_{eff}}$	0%	0%
$\mathbf{\Theta}_2 = \mathrm{b_{gbb}}$	22.5%	20%
$\mathbf{\Theta}_3 = \mathrm{k_v}$	47.5%	47.5%
$\mathbf{\Theta}_4 = \mathrm{Fc_{limcoalesc}}$	0%	0%
$\mathbf{\Theta}_5 = \mathrm{A_2}$	95%	95%
$\mathbf{\Theta}_6 = \mathrm{A_3}$	100%	100%
$\mathbf{\Theta}_7 = \mathrm{k_\alpha}$	92.5%	92.5%
$\mathbf{\Theta}_8 =$tkthres	5%	5%
$\mathbf{\Theta}_9 = \alpha_{\mathrm{bhsd}}$	0%	0%
$\mathbf{\Theta}_{10} = \mathrm{b_{hbs}}$	65%	65%
$\mathbf{\Theta}_{11} = \mathrm{R_{fhbs}}$	95%	95%

Table 3.4: Detection rates obtained for Approach A.

shows the superimposition of histograms $\left\{\widehat{\widehat{\mathscr{H}_j}}^2\left(D_{m,n}^{\tau_b \bullet}\right)\right\}_{1\leq b\leq B_{\mathrm{perm}}}$ with $B_{\mathrm{perm}} = 10^3$ (which are obtained from the permutation-based test procedure presented in Section 3.5.2.1) and the Gamma distribution densities $\Gamma(\widetilde{a}_j, \widetilde{b}_j)$ obtained by the procedure described in Section 3.5.2.2.

p-values / Model parameters	with U-statistic	with V-statistic
k_{eff}	1	1
b_{gbb}	0	0
k_v	0.01	0.015
$Fc_{limcoalesc}$	1	1
A_2	0	0
A_3	0	0
k_α	0	0
tkthres	0.968	0.974
α_{bhsd}	1	1
b_{hbs}	0.661	0.672
R_{fhbs}	0	0

Table 3.5: *p*-value estimates for Approach B (with the permutation-based test procedure).

p-values / Model parameters	with U-statistic	with V-statistic
k_{eff}	0.999	0.999
b_{gbb}	0.999	0.998
k_v	0	0
$Fc_{limcoalesc}$	1	1
A_2	0	0
A_3	0	0
k_α	0	0
tkthres	0.881	0.912
α_{bhsd}	1	1
b_{hbs}	0.001	0.002
R_{fhbs}	0	0

Table 3.6: *p*-value estimates for Approach C (with the permutation-based test procedure).

3.7 Conclusions and perspectives

In this paper, we carried out a global sensitivity analysis based on HSIC indices to identify the most influential calibration parameters of a fission gas behavior model conditionally on the uncertainty conveyed by a parameter coming from a thermal model. We proposed a sensitivity measure specifically adapted to the needs of the problem, and which is notably capable of absorbing the additional uncertainty (related to the thermal conductivity). We have proposed four estimators based on Monte Carlo sampling schemes. Despite the presence of two sampling levels, all the estimators are consistent, and their convergence speeds are comparable to those of classic Monte Carlo estimators, which offers the possibility of accurately estimating the proposed sensitivity indices, even when the simulation budget is limited.

Furthermore, we managed to construct independence test procedure based on two of these estimators. This allows us to perform the screening of calibration parameters through a data-driven decision-making process. Since the proposed sensitivity measure has proved capable of characterizing independence, the associated test procedure rests on a sound theoretical basis. Thanks to the ability of HSIC to measure the dependence between two random objects having very different mathematical natures, the sensitivity of several quantities of interest was examined (sensitivity of the RGF for a particular experimental configuration, sensitivity of all RGF values, sensitivity of the least-squares function). In the end, 6 of the 11 initial parameters, namely R_{fhbs}, k_{α}, A_3, A_2, b_{hbs} and k_v, were identified as significantly influential and will be kept in the calibration process.

It would be interesting to extend the theoretical developments to Latin Hypercube sampling (LHS) designs McKay et al. [1979] and their variants for future theoretical extensions. We will investigate the consistency of the four estimators associated with LHS designs. Finding a possible asymptotic procedure for these four estimators would also be interesting.

3.8 Supplementary material

3.8.1 Some basic reminders on kernels

Definition 2

Let $\mathscr{X}$ be a separable metric space. A function $K : \mathscr{X} \times \mathscr{X} \longrightarrow \mathbb{R}$ and Let us put

$$\boldsymbol{K}_n = [K(x_i, x_j)]_{1 \leq i,j \leq n}$$

the associated Gram matrix. Suppose that K is symmetric (i.e, $K(x, y) = K(y, x)$). Then, K is said to be:

- **A positive definite kernel** if $\boldsymbol{K}_n$ is positive definite. This means that:

$$\forall\, n \geq 1,\ x_1, \cdots, x_n \in \mathscr{X},\quad \forall\, c = (c_1, \cdots, c_n)^t \in \mathbb{R}^n,\quad c^t \boldsymbol{K}_n c = \sum_{i,j=1}^n c_i c_j K(x_i, x_j) \geq 0.$$

- **A strictly positive definite** if $\boldsymbol{K}_n$ is strictly positive definite. This means that:

$$\forall\, n \geq 1,\ x_1 \neq \cdots \neq x_n \in \mathscr{X},\quad \forall\, c = (c_1, \cdots, c_n)^t \in \mathbb{R}^n \setminus \{\mathbf{0}\},\quad c^t \boldsymbol{K}_n c > 0.$$

For any $\mathscr{X}$ be a separable metric space, the space of all Borel probability measures on $\mathscr{X}$ is denoted by $\mathscr{M}_1^+(\mathscr{X})$.

Definition 3

Let $\mathscr{X}$ be a separable metric space. A measurable and bounded kernel K is said to be **integrally strictly positive definite** if

$$\forall\, \nu \in \mathscr{M}_1^+(\mathscr{X}),\quad \int_{\mathscr{X}} \int_{\mathscr{X}} K(x, y) d\nu(x) d\nu(y) > 0.$$

Definition 4

For a given kernel $K : \mathscr{Z} \times \mathscr{Z} \longrightarrow \mathbb{R}$ verifying $\mathbb{E}_\nu[K(Z, Z)] < +\infty$, the *kernel mean embedding* of the probability distribution ν is the function defined by:

$$\begin{aligned} \mu_\nu : \mathscr{Z} &\longrightarrow \mathbb{R} \\ z &\longmapsto \mathbb{E}_\nu[K(z, Z)] = \int_{\mathscr{Z}} K(z, z') d\nu(z'). \end{aligned}$$

K is said to be a *characteristic kernel* if the operator $\nu \in \mathscr{M}_1^+(\mathscr{Z}) \longrightarrow \mu_\nu \in \mathscr{H}$ is injective. This means that:

$$\forall\ (\nu_1, \nu_2) \in \mathscr{M}_1^+(\mathscr{Z})^2,\ \|\mu_{\nu_1} - \mu_{\nu_2}\|_{\mathscr{H}} = 0 \quad \Rightarrow \quad \nu_1 = \nu_2,$$

In particular, we have

$$\forall\, \nu \in \mathscr{M}_1^+(\mathscr{Z}),\quad \|\mu_\nu\|_{\mathscr{H}}^2 = \int_{\mathscr{Z}} \int_{\mathscr{Z}} K(z, z') d\nu(z) d\nu(z') > 0.$$

With this result from the above equation, it is not surprising that there is a connection between characteristic kernels and strictly positive defined kernels. It is shown in theorem 7 in [Sriperumbudur et al., 2010b] that there is indeed a relationship between these two types of kernels.

Theorem 3

(**Theorem** 7 in [Sriperumbudur et al., 2010b]) Let K be an integrally strictly positive

definite on a separable metric space $\mathscr{X}$. Then K is characteristic.

This result suggests that for a kernel to be characteristic (a fundamental property for HSIC-based independence testing), it is sufficient for it to be integrally strictly positive definite. In Proposition 5.2 in [Ziegel et al., 2022], they show that a family of specific kernels is integrally strictly positive definite (then characteristic according to Theorem 7 in [Sriperumbudur et al., 2010b]).

Proposition 4

(**Proposition** 5.2 in [Ziegel et al., 2022]) Let $\mathscr{X}$ be a separable metric space, $\mathscr{F}$ a separable Hilbert space and $T : \mathscr{X} \longrightarrow \mathscr{F}$ an injective measurable map. In addition, let $\varphi : [0, \infty) \longrightarrow \mathbb{R}$ be a function such that the map $k_\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}$ defined by:

$$\forall\ \boldsymbol{z}, \boldsymbol{z}' \in \mathbb{R}^d, \quad k_\varphi\left(\boldsymbol{z}, \boldsymbol{z}'\right) := \varphi\left(\|\boldsymbol{z} - \boldsymbol{z}'\|^2_{\mathbb{R}^d}\right)$$

is a characteristic kernel. Then, the map $K : \mathscr{X} \longrightarrow \mathbb{R}$ defined by:

$$\forall\ x, x' \in \mathscr{X}, \quad K(x, x') := \varphi\left(\|T(x) - T(x')\|^2_{\mathscr{F}}\right)$$

is an integrally strictly positive definite kernel.

Assumption 1

Let $K(x, x') = \psi(x - x')$ where ψ is continuous and integrable function:

$$\forall\ d \geq 1, \quad \psi \in C^0(\mathbb{R}^d, \mathbb{R}) \cap L^1(\mathbb{R}^d).$$

In practice, the kernels used in the statistical learning and uncertainty quantification community (Gaussian kernels, exponential kernels, Matérn kernels, etc.) often verify Assumption 1 [Sriperumbudur et al., 2010b,a; Kanagawa et al., 2018]. In our case, we are only interested in translation-invariant kernels that verify Assumption 1.

Definition 5

Let $K : \mathbb{R}^q \times \mathbb{R}^q \longrightarrow \mathbb{R}$ $(q \geq 1)$ a kernel which satisfies Assumption 1. Then K is said to be translation-invariant and there exists ψ such that:

$$\forall\ x, x' \in \mathbb{R}^q, \quad K(x, x') = \psi(x - x').$$

The function ψ is called signature.

The following proposition gives a necessary and sufficient condition for the tensorization of several translation-invariant kernels (satisfy Assumption 1) to be characteristic.

Proposition 5

(**Proposition** 8 **in** [Sriperumbudur et al., 2010a]) Let K be a translation-invariant kernel on $\mathbb{R}^d$ and satisfies Assumption 1. Let ψ be a associated signature. Then K is characteristic on $\mathscr{M}_1^+(\mathbb{R}^d)$ if and if only the support of the Fourier transformation of ψ is equal to $\mathbb{R}^d$:

$$\operatorname{supp}\left(\widehat{\psi}\right) = \mathbb{R}^d.$$

The Fourier transform of ψ is given by:

$$\widehat{\psi}(w) = \frac{1}{\sqrt{2\pi}^d} \int_{\mathbb{R}^d} e^{-iw^t x} \psi(x) dx.$$

3.8.1.1 Characteristic kernels for some random objects

When Y is a random vector taking values in $\mathbb{R}^q$, additional work is required to obtain a characteristic kernel associated with Y. In [Szabó and Sriperumbudur, 2017], it is demonstrated that when $q \geq 3$, tensorizing q characteristic kernels $\{k_j\}_{j=1}^q$ generally does not guarantee that the resulting kernel $K_q = \otimes_{j=1}^q k_j$ is characteristic. To show that K_q remains characteristic, we will use two propositions (Propositions 6 and 7). The first one is based is based on the works of [Ziegel et al., 2022; Sriperumbudur et al., 2010b] to prove that the tensorization of Gaussian kernels is still characteristic. The second one is more general result (compared to the first one) and based on the works of [Sriperumbudur et al., 2010a] to prove that the tensorization of translation-invariant kernels is characteristic whether the support the Fourier Transform of the signature ψ (associated to K_q) is $\mathbb{R}^q$.

Proposition 6

Let $k_1^{\mathscr{N}}, \cdots k_q^{\mathscr{N}}$ be Gaussian kernels whose each $k_l^{\mathscr{N}}$ (i.e., characteristic) is given by:

$$\forall\, x, x' \in \mathbb{R}, \quad k_l^{\mathscr{N}}(x, x') = \exp\left[-\tfrac{1}{2}\left(\tfrac{x-x'}{\gamma_l}\right)^2\right].$$

Then, the kernel $K_q = \otimes_{l=1}^q k_l^{\mathscr{N}}$ is characteristic.

Proof: We tensorize $q \geq 1$ characteristic Gaussian kernels $\left\{k_1^{\mathscr{N}}\right\}_{1\leq l\leq q}$ and obtain:

$$\begin{aligned} K_q(x,x') = & \prod_{l=1}^q k_l^{\mathscr{N}}(x_l, x_l') = & \prod_{l=1}^q \exp\left(-\frac{1}{2}\left(\frac{x_l - x_l'}{\gamma_l}\right)^2\right) \\ K_q(x,x') = & \exp\left(-\sum_{l=1}^q \frac{1}{2}\left(\frac{x_l - x_l'}{\gamma_l}\right)^2\right) = & \varphi\left(\|T(x) - T(x')\|^2\right), \end{aligned}$$

where $\varphi : x \mapsto \exp(-x)$ and $T : x \mapsto \frac{1}{\sqrt{2}}\left(\frac{x_1}{\gamma_1}, \cdots, \frac{x_q}{\gamma_q}\right)$ is a homothety from $\mathbb{R}^q$ into $\mathbb{R}^q$. This is an injective operator. Therefore, according to Proposition 4 and Theorem 3, K_q is characteristic for all $q \geq 1$. ■

Proposition 7

Let $k_1, \cdots k_q$ be q translation-invariant kernels. In addition, assume that each k_l verifies Assumption 1(with $d = 1$) and characteristic. Then the kernel $K_q = \otimes_{l=1}^q k_l$ is characteristic.

Proof: We start again with the tensorization of the q translation-invariant kernels:

$$K_q = \otimes_{j=1}^q k_j.$$

Furthermore, we have:

$$\forall\, 1 \leq l \leq q, \quad \exists\, \psi_l \in C^0(\mathbb{R}, \mathbb{R}) \cap L^1(\mathbb{R}) : \quad k_l(x_l, x_l') = \psi_l(x_l - x_l').$$

Thus, the tensorized kernel K_q is equal to the tensorization of the signatures $\{\psi_l\}_{1\leq l\leq q}$:

$$\forall\, x, x' \in \mathbb{R}^q, \quad K_q(x, x') = \prod_{l=1}^q \psi_l(x_l - x_l') =: \psi(x, x'),$$

where

$$\begin{aligned} \psi : \quad \mathbb{R}^q \times \mathbb{R}^q & \longrightarrow \mathbb{R} \\ (x, x') & \longmapsto \psi(x, x') = \prod_{l=1}^q \psi_l(x_j - x_l'). \end{aligned}$$

It remains to show that a) $\widehat{\psi} = \prod_{l=1}^{q} \widehat{\psi}_l$, b) each $\widehat{\psi}_l$ has support $\mathbb{R}$ and c) $\widehat{\psi}$ has support $\mathbb{R}^q$. This proves that the tensorized kernel is characteristic. Let us prove a) and b). Let $w \in \mathbb{R}^q$:

$$\widehat{\psi}(w) = \frac{1}{\sqrt{2\pi}^q} \int_{\mathbb{R}^q} e^{-iw^t x} \psi(x) dx = \int_{\mathbb{R}^q} \prod_{l=1}^{q} \left(\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-iw_l x_l} \psi_l(x_l) dx_l \right).$$

Now, for all $1 \leq l \leq q$ the function $x_l \longmapsto \frac{1}{\sqrt{2\pi}} e^{-iw_l x_l} \psi_l(x_l)$ is integrable with respect to the Lebesgue measure on $\mathbb{R}$. In fact, one has:

$$\forall\, w_l \in \mathbb{R}, \quad \int_{\mathbb{R}} \left| \frac{1}{\sqrt{2\pi}} e^{-iw_l x_l} \psi_l(x_l) \right| dx_l \leq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |\psi_l(x_l)|\, dx_l = \frac{1}{\sqrt{2\pi}} \|\psi_l\|_{L^1(\mathbb{R})} < \infty$$

according Assumption 1. The Fourier transform of ψ_l:

$$\forall\, w_l \in \mathbb{R}, \quad \widehat{\psi}_l(w_l) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-iw_l x_l} \psi_l(x_l) dx_l$$

exists. Then by Proposition 5 (with $d = 1$), one has:

$$\mathrm{supp}\left(\widehat{\psi}_l\right) = \mathbb{R}.$$

In addition, by Fubini's theorem, one has:

$$\widehat{\psi}(w) = \prod_{l=1}^{q} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-iw_l x_l} \psi_l(x_l) dx_l = \prod_{l=1}^{q} \widehat{\psi}_l(w_l).$$

Then we have:

$$\widehat{\psi} = \prod_{l=1}^{q} \widehat{\psi}_l \quad \text{and} \quad \forall\, w \in \mathbb{R}^q, \quad \widehat{\psi}(w) = \prod_{l=1}^{q} \widehat{\psi}_l(w) \Rightarrow \mathrm{supp}\left(\widehat{\psi}\right) = \prod_{l=1}^{q} \mathrm{supp}\left(\widehat{\psi}_l\right) = \mathbb{R}^q.$$

According to Proposition 5 (with $d = q$), K_q is characteristic. ■

3.8.2 U-statistics and V-statistics estimators

Definition 6

(**U-statistics and V-statistics**). Let $Z_1, \cdots, Z_{n'}$ be $n' \geq 2$ independent variables following the same distribution $\mathbb{P}_Z$ defined on $\mathscr{Z}$. Consider the following quantity:

$$\zeta := \mathbb{E}_{\mathbb{P}_Z^{\otimes n'}} \left[\eta\left(Z_1, \cdots, Z_{n'}\right) \right] = \int_{\mathscr{Z}} \cdots \int_{\mathscr{Z}} \eta\left(z_1, \cdots, z_{n'}\right) d\mathbb{P}_Z(z_1) \cdots d\mathbb{P}_Z(z_{n'}),$$

where η is any function, called kernel, assumed to be symmetric. If η is not symmetric, it can be replaced by a symmetric version given by:

$$\forall\, (z_1, \cdots, z_{n'}), \quad \widetilde{\eta}(z_1, \cdots, z_{n'}) = \frac{1}{n'!} \sum_{\sigma \in \mathbb{S}_{n'}} \eta(z_{\sigma(1)}, \cdots, z_{\sigma(n')}). \tag{3.49}$$

The expectation obtained with $\widetilde{\eta}$ is indeed equal to that obtained with η:

$$\mathbb{E}_{\mathbb{P}_Z^{\otimes n'}} \left[\widetilde{\eta}(Z_1, \cdots, Z_{n'}) \right] = \frac{1}{n'!} \sum_{\sigma \in \mathbb{S}_{n'}} \mathbb{E}_{\mathbb{P}_Z^{\otimes n'}} \left[\eta(Z_{\sigma(1)}, \cdots, Z_{\sigma(n')}) \right] = \mathbb{E}_{\mathbb{P}_Z^{\otimes n'}} \left[\eta(Z_1, \cdots, Z_{n'}) \right].$$

The vectors $(Z_1, \cdots, Z_{n'})$ and $(Z_{\sigma(1)}, \cdots, Z_{\sigma(n')})$ follow that the same distribution (any $\sigma \in \mathbb{S}_{n'}$). In all that follow, we will assume that η is not necessarily symmetric and we will use $\widetilde{\eta}$ to recall the formulas for the U-statistics and V-statistics, as well as the formulas

for the HSIC indices. U-statistic estimator of ζ obtained on a sample $(Z^{(1)}, \cdots, Z^{(n)})$ of size $n \geq n'$ is defined by:

$$\widehat{\zeta}^{\mathrm{U}} = \binom{n}{n'}^{-1} \sum_{\boldsymbol{i} \in \mathscr{I}_n^{n'}} \widetilde{\eta}\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right), \tag{3.50}$$

where

$$\mathscr{I}_n^{n'} = \{\boldsymbol{i} := (i_1, \cdots, i_{n'}) : \ 1 \leq i_1 \neq i_2 \cdots \neq i_{n'} \leq n\} \quad \text{and} \quad \binom{n}{n'} = \tfrac{n!}{n'!(n-n')!}.$$

The von Mises statistic associated to the above U-statistic, also commonly called V-statistic, for estimation of ζ is given by:

$$\widehat{\zeta}^{\mathrm{V}} = \frac{1}{n^{n'}} \sum_{\boldsymbol{i} \in \mathscr{I}_n^{n'}} \widetilde{\eta}\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right), \tag{3.51}$$

where

$$\mathscr{I}_n^{n'} := \{(i_1, \cdots, i_{n'}) : \ 1 \leq i_1, \cdots, i_{n'} \leq n\}.$$

$\widehat{\zeta}^{\mathrm{U}}$ is unbiased whereas $\widehat{\zeta}^{\mathrm{V}}$ is biased. This can be understood very easily. Indeed, in Equation (3.50), because of the constraint $\boldsymbol{i} \in \mathscr{I}_n^{n'}$ (which ensures $i_1 \neq i_2 \neq \ldots \neq i_{n'}$), the expectation of each term $\widetilde{\eta}\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right)$ is exactly equal to ζ. On the contrary, the additional terms involved in the V-statistic violate the constraint $\boldsymbol{i} \in \mathscr{I}_n^{n'}$ and their expectations are not equal to ζ (in general). This is the reason why $\widehat{\zeta}^{\mathrm{V}}$ is biased while $\widehat{\zeta}^{\mathrm{U}}$ is not. For more details on these two estimators (see Chapters 5 and 6 in [Serfling, 1980]).

Lemma 1

(Lemma A of Section 5.2.1, [Serfling, 1980]). The variance of $\widehat{\eta}^{\mathrm{U}}$ is bounded:

$$0 \leq \mathbb{V}\left(\widehat{\zeta}^{\mathrm{U}}\right) \leq \frac{n'}{n} \mathbb{V}\left(\widetilde{\eta}\left(Z_1, \cdots, Z_{n'}\right)\right). \tag{3.52}$$

Moreover, when n tends to infinity, the variance $\mathbb{V}\left(\widehat{\zeta}^{\mathrm{U}}\right)$ tends to 0.

Lemma 2

(Lemma of Section 5.7.3, [Serfling, 1980]). Let $r \in \mathbb{N}^\star$ and suppose that

$$\forall\ i \leq i_1, \cdots, i_{n'} \leq n', \quad \mathbb{E}_{\mathbb{P}_Z^{\otimes n'}}\left(\left|\eta\left(Z_{i_1}, \cdots, Z_{i_{n'}}\right)\right|^r\right) < \infty.$$

Then, we have:

$$\mathbb{E}\left[|\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}|^r\right] = \mathscr{O}(n^{-r}). \tag{3.53}$$

We would like to know the expression of the constant (as a function of $\widetilde{\eta}$) behind $\mathscr{O}(n^{-r})$ in Equation (3.53). To achieve this, we first propose a bound for the difference of $\widehat{\eta}^{\mathrm{U}} - \widehat{\eta}^{\mathrm{V}}$ in L^r norm in the following proposition:

Proposition 8

Under same assumptions than Lemma 2, one has:

$$\forall\ r \in \mathbb{N}^\star, \quad \mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}\right|^r\right] \leq \left(\tfrac{n'(n'-1)}{n}\right)^r C(r, \widetilde{\eta}), \tag{3.54}$$

with

$$C(r, \widetilde{\eta}) := \max_{1 \leq i_1, \cdots, i_{n'} \leq n'} \mathbb{E}\left[|\widetilde{\eta}(Z_{i_1}, \cdots, Z_{i_{n'}})|^r\right]. \tag{3.55}$$

Proof: To prove Proposition 8, we will follow the different steps proposed by [Serfling, 1980] to establish Equation (3.53) and we give all details. First, we will show the following equality:

$$n^{n'}(\widehat{\eta}^{\mathrm{U}} - \widehat{\eta}^{\mathrm{V}}) = (n^{n'} - (n)_{n'})(\widehat{\eta}^{\mathrm{U}} - W_n), \tag{3.56}$$

where

$$(n)_{n'} := m! \binom{n}{n'} = \frac{n!}{n'!(n-n')!} = \frac{n!}{(n-n')!} = n(n-1)\cdots(n-n'+1),$$

and W_n is the average of all terms $\eta(Z_{i_1}, \cdots, Z_{i_{n'}})$ with at least one equality $i_a = i_b,\ a \neq b$. Then we will show that:

$$(n^{n'} - (n)_{n'}) = \mathscr{O}(n^{n'-1}), \tag{3.57}$$

to finally obtain (3.54). To establish (3.56), we begin to define the following set:

$$\mathscr{E}_n^{n'} := \left\{ \boldsymbol{i} = (i_1, \cdots, i_{n'}) \in \{1, \cdots, n\}^{n'} : \exists\ (a,b) \in \{1, \cdots, m\}^2 \text{ such that } i_a = i_b \text{ and } a \neq b \right\},$$

and its cardinal is equal to:

$$\begin{aligned}
\mathrm{Card}\left(\mathscr{E}_n^{n'}\right) &= \mathrm{Card}\left(\{1, \cdots, n\}^{n'}\right) - \mathrm{Card}\left(\{1 \leq i_1 \neq i_2 \cdots \neq i_{n'} \leq n\}\right) \\
&= n^{n'} - n'! \binom{n}{n'} = n^{n'} - (n)_{n'},
\end{aligned}$$

Furthermore, we have the following equality:

$$\begin{aligned}
n^{n'}\widehat{\zeta}^{\mathrm{V}} &= \sum_{\boldsymbol{i} \in \mathscr{J}_n^{n'}} \widetilde{\eta}\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right) \\
&= \sum_{1 \leq i_1 \neq i_2 \cdots \neq i_{n'} \leq n} \eta\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right) + \sum_{(i_1, \cdots, i_{n'}) \in \mathscr{E}_n^{n'}} \eta\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right) \\
&= n'! \sum_{1 \leq i_1 < i_2 \cdots < i_{n'} \leq n} \eta\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right) + \sum_{(i_1, \cdots, i_{n'}) \in \mathscr{E}_n^{n'}} \eta\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right) \\
&= (n)_{n'}\widehat{\zeta}^{\mathrm{U}} + \left(n^{n'} - (n)_{n'}\right) \frac{1}{n^{n'} - (n)_{n'}} \sum_{(i_1, \cdots, i_m) \in \mathscr{E}_n^{n'}} \eta\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right) \\
&= (n)_{n'}\widehat{\zeta}^{\mathrm{U}} + \left(n^{n'} - (n)_{n'}\right) W_n,
\end{aligned}$$

with

$$W_n := \frac{1}{n^{n'} - (n)_{n'}} \sum_{(i_1, \cdots, i_{n'}) \in \mathscr{E}_n^{n'}} \eta\left(Z^{(i_1)}, \cdots, Z^{(i_{n'})}\right).$$

Then, we have:

$$\begin{aligned}
n^{n'}\widehat{\zeta}^{\mathrm{V}} &= (n)_{n'}\widehat{\zeta}^{\mathrm{U}} + \left(n^{n'} - (n)_{n'}\right) W_n \\
n^{n'}\left(\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}\right) &= n^{n'}\widehat{\zeta}^{\mathrm{U}} - (n)_{n'}\widehat{\zeta}^{\mathrm{U}} - \left(n^{n'} - (n)_{n'}\right) W_n \\
n^{n'}\left(\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}\right) &= \left(n^{n'} - (n)_m\right)\left(\widehat{\zeta}^{\mathrm{U}} - W_n\right),
\end{aligned} \tag{3.58}$$

Then, Equation (3.56) is satisfied. Let us now show Equality (3.57). This can be done by recurrence. For $n' = 1$, we have:

$$n^{n'} - (n)_{n'} = n - n = 0 = \mathscr{O}(1). \tag{3.59}$$

Let us assume that the result is true for $n' \geq 1$ and let us show that the result remains at for order $n' + 1$.

$$n^{n'} - (n)_{n'} = \mathscr{O}(n^{n'-1}) \iff \exists\ C_{n'} > 0 : \forall\ n \geq 1,\ n^{n'} - (n)_{n'} \leq C_{n'} n^{n'-1}.$$

At order $n'+1$, one has:

$$\begin{aligned} n^{n'+1} - n_{(n'+1)} &= n^{(n'+1)} - n(n-1)(n-2)\cdots(n-n') \\ &= n\left[n^{n'} - (n-1)(n-1)(n-2)\cdots(n-n')\right] \\ &= n\left[n^{n'} - (n)_{n'} + (n)_{n'} - (n-1)_{(n')}\right]. \end{aligned} \tag{3.60}$$

On the one hand, since the result is assumed to hold at order n', one has:

$$n^{n'} - (n)_{n'} \leq C_{n'} n^{n'-1}. \tag{3.61}$$

And the other,one has:

$$\begin{aligned} (n)_{n'} - (n-1)_{n'} &= n(n-1)(n-2)\cdots(n-n'+1) - (n-1)(n-2)(n-3)\cdots(n-1+n'+2)(n-1-n'+1) \\ &= (n-2)(n-3)\cdots(n-n'+1)\left[n-(n-n')\right] \\ &= (n-2)(n-3)\cdots(n-n'+1)n' \\ &\leq n^{n'-1}n', \end{aligned} \tag{3.62}$$

because all $n'-1$ factors of $(n-2)(n-3)\cdots(n-n'+1)$ are less than n. By injecting (3.61) and (3.62) into (3.60), we obtain:

$$n^{n'+1} - (n)_{n'+1} \leq (n' + C_{n'})n^{n'-1} =: C_{n'+1}n^{n'-1}$$

By recursion, we have the following relationship between the constants $C_{n'}$:

$$C_{n'+1} = n' + C_{n'} = n' + (n'-1) + C_{n'-1} = n' + (n'-1) + (n'-2) + \cdots + 1 + C_1,$$

and $C_1 = 0$ according to Equation (3.59). Then, we have:

$$C_{n'+1} = \sum_{k=1}^{n'} k = \frac{(n'+1)n'}{2}. \tag{3.63}$$

Therefore, we obtain:

$$n^{n'} - (n)_{n'} \leq \frac{(n'-1)n'}{2} n^{n'-1}, \ \forall\ n' \geq 1. \tag{3.64}$$

Note that, Equations (3.56) and (3.57) established, let us now show how to obtain Equation (3.54). For $r \in \mathbb{N}^\star$, it was assumed that:

$$\forall\ 1 \leq i_1, \cdots, i_{n'} \leq n',\ \mathbb{E}_{\mathbb{P}_Z^{\otimes n'}}\left[\left|\eta\left(Z_{i_1}, \cdots, Z_{i_{n'}}\right)\right|^r\right] < \infty.$$

Therefore, the following quantity is well-defined:

$$C(r, \widetilde{\eta}) = \max_{1 \leq i_1, \cdots, i_{n'} \leq n'} \mathbb{E}\left[\left|\widetilde{\eta}(Z_{i_1}, \cdots, Z_{i_{n'}})\right|^r\right] < \infty.$$

In addition, by combining (3.58) and (3.61), we obtain

$$\begin{aligned} \left(\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}\right) &\leq C_{n'} n^{-1}(\widehat{\zeta}^{\mathrm{U}} - W_n) \\ \left|\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}\right|^r &\leq \left(\frac{C_{n'}}{n}\right)^r \left|\widehat{\zeta}^{\mathrm{U}} - W_n\right|^r \\ \mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}\right|^r\right] &\leq \left(\frac{C_{n'}}{n}\right)^r \mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}} - W_n\right|^r\right]. \end{aligned} \tag{3.65}$$

Using the triangle inequality for $\|\bullet\|_{L^r}$ ($t \mapsto t^r$ an increasing function on $[0,\ +\infty)$) gives:

$$\begin{aligned} \mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}} - W_n\right|^r\right] &= \left\|\widehat{\zeta}^{\mathrm{U}} - W_n\right\|_{L^r}^r \\ &\leq \left(\|\widehat{\zeta}^{\mathrm{U}}\|_{L^r} + \|W_n\|_{L^r}\right)^r. \end{aligned}$$

The convexity of $t \in [0,\ +\infty) \mapsto t^r$ allows us to write:

$$\left(\|\widehat{\zeta}^{\mathrm{U}}\|_{L^r} + \|W_n\|_{L^r}\right)^r \leq 2^{r-1}\left(\|\widehat{\zeta}^{\mathrm{U}}\|_{L^r}^r + \|W_n\|_{L^r}^r\right).$$

Therefore, one can write:

$$\mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}} - W_n\right|^r\right] \leq 2^{r-1}\left(\mathbb{E}\left[|\widehat{\zeta}^{\mathrm{U}}|^r\right] + \mathbb{E}\left[|W_n|^r\right]\right). \tag{3.66}$$

Moreover, one has:

$$\begin{aligned}\left|\widehat{\zeta}^{\mathrm{U}}\right|^r &\leq \left(\binom{n}{n'}^{-1} \sum_{\boldsymbol{i}\in\mathscr{I}_n^{n'}} \left|\eta\left(Z^{(i_1)},\cdots,Z^{(i_{n'})}\right)\right|\right)^r \\ &\leq \binom{n}{n'}^{-1} \sum_{\boldsymbol{i}\in\mathscr{I}_n^{n'}} \left|\eta\left(Z^{(i_1)},\cdots,Z^{(i_{n'})}\right)\right|^r.\end{aligned}$$

This last inequality is obtained by applying Jensen's inequality in the discrete case to the convex function $t \in [0,\ +\infty[\mapsto t^r$ at the points

$$t_{\boldsymbol{i}} := \left|\eta\left(Z^{(i_1)},\cdots,Z^{(i_m)}\right)\right|$$

with the weights:

$$a_{\boldsymbol{i}} := \binom{n}{n'}^{-1} \quad \text{knowing that} \quad \sum_{\boldsymbol{i}\in\mathscr{I}_n^{n'}} a_{\boldsymbol{i}} = 1.$$

Thus, we have:

$$\begin{aligned}\mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}}\right|^r\right] &\leq \binom{n}{n'}^{-1} \sum_{\boldsymbol{i}\in\mathscr{I}_n^{n'}} \mathbb{E}\left[\left|\eta\left(Z^{(i_1)},\cdots,Z^{(i_{n'})}\right)\right|^r\right] \\ &\leq \binom{n}{n'}^{-1} \sum_{\boldsymbol{i}\in\mathscr{I}_n^{n'}} \mathbb{E}_{\mathbb{P}_Z^{\otimes n'}}\left[\left|\eta\left(Z_{i_1},\cdots,Z_{i_{n'}}\right)\right|^r\right] \\ &\leq \mathbb{E}_{\mathbb{P}_Z^{\otimes n'}}\left[\left|\eta\left(Z_{i_1},\cdots,Z_{i_{n'}}\right)\right|^r\right] \leq C(r,\widetilde{\eta}).\end{aligned} \tag{3.67}$$

With the same arguments as for (3.67), it can be show that:

$$\mathbb{E}\left[|W_n|^r\right] \leq C(r,\widetilde{\eta}). \tag{3.68}$$

Injecting Equations (3.67) and (3.68) in (3.66), leads to:

$$\mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}} - W_n\right|^r\right] \leq 2^r C\left(r,\widetilde{\eta}\right). \tag{3.69}$$

Finally, bringing together Equations (3.65) and (3.69) give:

$$\begin{aligned}\mathbb{E}\left[\left|\widehat{\zeta}^{\mathrm{U}} - \widehat{\zeta}^{\mathrm{V}}\right|^r\right] &\leq \left(\frac{C_{n'}}{n}\right)^r 2^r C(r,\widetilde{\eta}) \\ &\leq \left(\frac{n'(n'-1)}{n}\right)^r C(r,\widetilde{\eta}).\end{aligned}$$

■

3.8.2.1 Case of HSIC index estimation between X and Y

Let K_x and K_y be two kernels associated with X and Y and defined on $\mathscr{X}\times\mathscr{X}$ and $\mathscr{Y}\times\mathscr{Y}$ respectively. The HSIC index between the random variables X and Y is defined by:

$$\mathrm{HSIC(X,Y)} = \mathbb{E}\left[K_x(X_1,\ X_2)K_y(Y_1,\ Y_2)\right] + \left[K_x(X_1,\ X_2)\right]\mathbb{E}\left[K_y(Y_1,\ Y_2)\right] - 2\mathbb{E}\left[K_x(X_1,\ X_2)K_y(Y_1,\ Y_3)\right], \tag{3.70}$$

with $Z_1 = (X_1, Y_1)$, $Z_2 = (X_2, Y_2)$ and $Z_3 = (X_3, Y_3)$ are three independent copies of the vector

$$Z = (X, Y) \sim \mathbb{P}_Z = \mathbb{P}_{XY}.$$

Note that the HSIC can be rewritten as a single expectation:

$$\begin{aligned}\mathrm{HSIC}(X, Y) = \mathbb{E}[K_x(X_1,\ X_2)K_y(Y_1,\ Y_2) + K_x(X_1, X_2)K_y(Y_3, Y_4) \\ - 2K_x(X_1,\ X_2)K_y(Y_1, Y_3)],\end{aligned}$$

where $Z_1 = (X_1, Y_1)$, $Z_2 = (X_2, Y_2)$, $Z_3 = (X_3, Y_3)$ and $Z_4 = (X_4, Y_4)$ are four independent copies of Z. Finally, $\mathrm{HSIC}(X, Y)$ is the expectation of function η under the distribution $\mathbb{P}_Z^{\otimes 4}$:

$$\mathrm{HSIC}(X, Y) = \mathbb{E}_{\mathbb{P}_Z^{\otimes 4}}\left[\eta(Z_1, Z_2, Z_3, Z_4)\right], \tag{3.71}$$

with η being defined by:

$$\begin{array}{rccc}\eta: & \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} & \longrightarrow & \mathbb{R} \\ & (z_1, z_2, z_3, z_4) & \longmapsto & K_x(x_1, x_2)K_y(y_1, y_2) + K_x(x_1, x_2)K_y(y_3, y_4) \\ & & & -2\ K_x(x_1, x_2)K_y(y_1, y_3),\end{array} \tag{3.72}$$

with $\mathbb{Z} = \mathscr{X} \times \mathscr{Y}$ and $z_i = (x_i, y_i)$, $1 \leq i \leq 4$. The functions K_x and K_y are symmetric, but the function η is not, so we need to symmetrize η before applying formulas (3.50) and (3.51). So we define the symmetrized version of η according to Equation (3.49) (with $n' = 4$):

$$\widetilde{\eta}(z_1, z_2, z_3, z_4) = \frac{1}{4!}\sum_{\sigma \in \mathbb{S}_4} \eta(z_{\sigma(1)}, z_{\sigma(2)}, z_{\sigma(3)}, z_{\sigma(4)}). \tag{3.73}$$

Then, the expressions of $\widehat{\zeta}^{\mathrm{U}}$ and $\widehat{\zeta}^{\mathrm{V}}$ are given respectively by:

$$\widehat{\zeta}^{\mathrm{U}} = \binom{n}{n'}^{-1} \sum_{1 \leq i_1 < \cdots < i_4 \leq n} \widetilde{\eta}\left(Z^{(i_1)}, \cdots, Z^{(i_4)}\right),$$

$$\widehat{\zeta}^{\mathrm{V}} = \frac{1}{n^4} \sum_{1 \leq i_1, \cdots, i_4 \leq n} \widetilde{\eta}\left(Z^{(i_1)}, \cdots, Z^{(i_4)}\right).$$

The expressions of these estimators ($\widehat{\zeta}^{\mathrm{U}}$ and $\widehat{\zeta}^{\mathrm{V}}$) with kernels K_x and K_y (see Appendix C.2 in [Sarazin et al., 2023]) are given respectively by:

$$\begin{aligned}\widehat{\zeta}^{\mathrm{U}} := \frac{1}{(n)_2} \sum_{(l,l') \in \mathscr{I}_n^2} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(l)}, Y^{(l')}\right) \\ + \frac{1}{(n)_4} \sum_{(l,l',t,t') \in \mathscr{I}_n^4} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(t)}, Y^{(t')}\right) \\ - 2\frac{1}{(n)_3} \sum_{(l,l',t) \in \mathscr{I}_n^3} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(l)}, Y^{(t)}\right),\end{aligned} \tag{3.74}$$

$$\begin{aligned}\widehat{\zeta}^{\mathrm{V}} := \frac{1}{n^2} \sum_{(l,l') \in \mathscr{I}_n^2} K_x\left(X^{(l)}, X^{(l')}\right) K_{y_a}\left(Y^{(l)}, Y^{(l')}\right) \\ + \frac{1}{n^4} \sum_{(l,l',t,t') \in \mathscr{I}_n^4} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(t)}, Y^{(t')}\right) \\ - 2\frac{1}{n^3} \sum_{(l,\ l',\ t) \in \mathscr{I}_n^3} K_x\left(X^{(l)}, X^{(l')}\right) K_y\left(Y^{(l)}, Y^{(t)}\right).\end{aligned} \tag{3.75}$$

Proposition 9

If variables $Z_1, \cdots, Z_m$ are independent and have the same distribution, then we have:

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{k=1}^{m} Z_k\right)^2\right] = \frac{1}{m}\mathbb{E}\left[Z_1^2\right] + \frac{m-1}{m}\left(\mathbb{E}\left[Z_1\right]\right)^2. \tag{3.76}$$

The proof is straightforward.

3.8.3 Supplementary material for Section 3.5

Lemma 3

(**Lemma** 1 in [Herin et al., 2022]) Let $X = (X_1, \cdots, X_p)^t$ be random input vector mutually independent and $f(X) \in L^2(\mathbb{P}_X)$ be a model. One has, $\forall\ 1 \leq j \leq p$,

$$T_j = 1 - \frac{\mathbb{V}(\mathbb{E}[f(X)|X_{-j}])}{\mathbb{V}(f(X))} = 0 \iff f(X) = h(X_{-j}) \text{ almost surely (a.s.)},$$

where the measurable function $h(X_{-j}) = \mathbb{E}\left[f(X) \mid X_{-j}\right]$.

The proof of Lemma 3 can be found in Appendix B in [Herin et al., 2022]. We recall that the nullity of T_j is equivalent to independence between X_j and $f(X)$. Thus, the result of Lemma 3 can be rewritten as follows:

$$X_j \perp\!\!\!\perp f(X) \iff f(X) = h(X_{-j}) \text{ a.s.} \tag{3.77}$$

Consequently, when we use Lemma 3 in what follows, we refer to the equivalence of Equation (3.77).

3.8.4 Proof of Proposition 2

Proof: First, let us examine the case where the support of $\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda})$ is included in $\mathbb{R}$. The objective is to prove that:

$$\mathscr{H}_j = 0 \iff \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}). \tag{3.78}$$

We start with $\mathscr{H}_j = \mathbb{E}_{\boldsymbol{\Lambda}}(H_j(\boldsymbol{\Lambda})) = 0$ and we aim to show that $\boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda})$. As the HSIC is non-negative, one has:

$$\mathbb{E}_{\boldsymbol{\Lambda}}(\underbrace{H_j(\boldsymbol{\Lambda})}_{\geq 0}) = 0 \iff H_j(\boldsymbol{\Lambda}) = 0 \text{ a.s.} \tag{3.79}$$

$$\iff \exists\ \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}} \subseteq \mathscr{D}_{\boldsymbol{\Lambda}} \text{ such that } \mathbb{P}_{\boldsymbol{\Lambda}}(\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}) = 1 \text{ and } \forall\ \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}, H_j(\lambda) = 0. \tag{3.80}$$

On the other hand, the HSIC nullity is equivalent to independence. Hence, one has:

$$\forall\ \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}, \quad \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta},\ \lambda). \tag{3.81}$$

Using Lemma 3 for each $\forall\ \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}$, one has:

$$\exists\ h:\ \widetilde{f}(\boldsymbol{\Theta}, \lambda) = h(\boldsymbol{\Theta}_{-j},\ \lambda) \text{ a.s.}, \tag{3.82}$$

where $h(\boldsymbol{\Theta}_{-j}, \lambda) = \mathbb{E}\left(\widetilde{f}(\boldsymbol{\Theta}, \lambda) \mid \boldsymbol{\Theta}_{-j}\right)$. Furthermore, one has the following equivalence:

$$\begin{aligned}\widetilde{f}(\boldsymbol{\Theta}, \lambda) = h(\boldsymbol{\Theta}_{-j}, \lambda) \text{ a.s.} \iff &\exists\ \widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}} \subseteq \mathscr{D}_{\boldsymbol{\Theta}} \text{ such that } \mathbb{P}_{\boldsymbol{\Theta}}(\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}) = 1 \\ &\text{and } \forall\ \theta \in \widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}, \widetilde{f}(\theta,\ \lambda) = h(\theta_{-j}, \lambda).\end{aligned} \tag{3.83}$$

In short, Equations (3.82) and (3.81) may be summarized as follows:

$$\forall\ \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}},\ H_j(\lambda) = 0 \iff \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta},\ \lambda)\ \forall\ \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}} \iff \widetilde{f}(\boldsymbol{\Theta},\ \lambda) = h(\boldsymbol{\Theta}_{-j},\ \lambda) \text{ a.s.} \tag{3.84}$$

To prove independence between $\boldsymbol{\Theta}_j$ and $\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})$, it is sufficient to check that:

$$\mathbb{E}\left[\varphi(\boldsymbol{\Theta}_j)\psi(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})\right)] = \mathbb{E}\left[\varphi(\boldsymbol{\Theta}_j)\right]\mathbb{E}\left[\psi(\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda})\right], \tag{3.85}$$

for all bounded Borelian functions φ and ψ. Let φ and ψ be two bounded Borelians functions. The equality between the joint and marginal expectations can be proved by the following way:

$$\begin{aligned}
\mathbb{E}\left[\varphi(\boldsymbol{\Theta}_j)\psi(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})\right] &= \int_{\mathscr{D}_{\boldsymbol{\Lambda}} \times \mathscr{D}_{\boldsymbol{\Theta}}} \varphi\left(\theta_j\right) \psi\left(\widetilde{f}(\theta, \lambda)\right) d\mathbb{P}_{\boldsymbol{\Theta}}(\theta) d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \\
&= \int_{\mathscr{D}_{\boldsymbol{\Lambda}}} \left(\int_{\mathscr{D}_{\boldsymbol{\Theta}}} \varphi(\theta_j)\psi\left(\widetilde{f}(\theta, \lambda)\right) d\mathbb{P}_{\boldsymbol{\Theta}}(\theta) \right) d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \text{ with Fubini's theorem} \\
&= \int_{\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}} \left(\int_{\mathscr{D}_{\boldsymbol{\Theta}}} \varphi(\theta_j)\psi(\widetilde{f}(\theta, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta}}(\theta) \right) d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \text{ since } \mathbb{P}\left(\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}\right) = 1 \\
&= \int_{\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}} \left(\int_{\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}} \varphi(\theta_j)\psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta}}(\theta) \right) d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \text{ since } \mathbb{P}\left(\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}\right) = 1 \\
&= \int_{\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}} \left(\int_{\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}} \varphi(\theta_j)\psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta}_{-j}}(\theta_{-j}) \right) d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \text{ since } \boldsymbol{\Theta}_j \perp\!\!\!\perp \boldsymbol{\Theta}_{-j}.
\end{aligned}$$

In addition, for each $\lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}$, one has:

$$\begin{aligned}
\int_{\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}} \varphi(\theta_j)\psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta}_{-j}}(\theta_{-j}) &= \int_{\mathscr{D}_{\boldsymbol{\Theta}}} \varphi(\theta_j)\psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta}_{-j}}(\theta_{-j}) \\
&= \int_{\mathscr{D}_{\boldsymbol{\Theta}_j} \times \mathscr{D}_{\boldsymbol{\Theta}_{-j}}} \varphi(\theta_j)\psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta_j}}(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta}_{-j}}(\theta_{-j}) \\
&= \left(\int_{\mathscr{D}_{\boldsymbol{\Theta}_j}} \varphi(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta_j}}(\theta_j) \right) \left(\int_{\mathscr{D}_{\boldsymbol{\Theta}_{-j}}} \psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta_{-j}}}(\theta_{-j}) \right).
\end{aligned}$$

Furthermore, one has:

$$\begin{aligned}
\int_{\mathscr{D}_{\boldsymbol{\Theta}_{-j}}} \psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta_{-j}}}(\theta_{-j}) &= \left(\underbrace{\int_{\mathscr{D}_{\boldsymbol{\Theta}_j}} d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j)}_{=1} \right) \left(\int_{\mathscr{D}_{\boldsymbol{\Theta}_{-j}}} \psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta_{-j}}}(\theta_{-j}) \right) \\
&= \int_{\mathscr{D}_{\boldsymbol{\Theta}_j} \times \mathscr{D}_{\boldsymbol{\Theta}_{-j}}} \psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta_{-j}}}(\theta_{-j}) d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j) \text{ by Fubini's theorem.}
\end{aligned}$$

Hence, one has:

$$\begin{aligned}
\mathbb{E}\left[\varphi(\boldsymbol{\Theta}_j)\psi(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})\right] &= \int_{\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}} \left(\int_{\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}} \varphi(\theta_j)\psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta}_{-j}}(\theta_{-j}) \right) d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \\
&= \int_{\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}} \left[\left(\int_{\mathscr{D}_{\boldsymbol{\Theta}_j}} \varphi(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta_j}}(\theta_j) \right) \left(\int_{\mathscr{D}_{\boldsymbol{\Theta}_{-j}}} \psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta_{-j}}}(\theta_{-j}) \right) \right] d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \\
&= \left(\int_{\mathscr{D}_{\boldsymbol{\Theta}_j}} \varphi(\theta_j) d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j) \right) \times \int_{\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}} \left[\int_{\mathscr{D}_{\boldsymbol{\Theta}_j} \times \mathscr{D}_{\boldsymbol{\Theta}_{-j}}} \psi(h(\theta_{-j}, \lambda)) d\mathbb{P}_{\boldsymbol{\Theta_{-j}}}(\theta_{-j}) d\mathbb{P}_{\boldsymbol{\Theta}_j}(\theta_j) \right] d\mathbb{P}_{\boldsymbol{\Lambda}}(\lambda) \\
&= \mathbb{E}\left[\varphi(\boldsymbol{\Theta}_j)\right]\mathbb{E}\left[h(\boldsymbol{\Theta}, \boldsymbol{\Lambda})\right] = \mathbb{E}\left[\varphi(\boldsymbol{\Theta}_j)\right]\mathbb{E}\left[\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})\right].
\end{aligned}$$

Then $\boldsymbol{\Theta}_j$ and $\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda})$ are independent.

Let us now the reverse statement. It is assumed that $\boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})$ and the objective is to prove that $\mathbb{E}_{\boldsymbol{\Lambda}}\left[H_j(\boldsymbol{\Lambda})\right] = 0$.

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\Lambda}}\left[H_j(\boldsymbol{\Lambda})\right] = \mathbb{E}\Big[& K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right)\right)\Big] \\
& + \mathbb{E}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right)\right]\mathbb{E}\Big[K_y\left(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right)\right)\Big] \\
& - 2\mathbb{E}\Big[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}'', \boldsymbol{\Lambda}\right)\right)\Big].
\end{aligned}$$

We aim to prove that:

$$\mathbb{E}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right)\right)\right] = \mathbb{E}\left[K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right)\right] \mathbb{E}\left[K_y\left(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right)\right)\right].$$

By hypothesis, one has:

$$\boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}\left(\boldsymbol{\Theta}, \boldsymbol{\Lambda}\right) \quad \text{and} \quad \boldsymbol{\Theta}'_j \perp\!\!\!\perp \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right).$$

In addition, one has:

$$\boldsymbol{\Theta} \perp\!\!\!\perp \boldsymbol{\Theta}' \perp\!\!\!\perp \boldsymbol{\Lambda}.$$

Hence, one has:

$$\begin{bmatrix} \boldsymbol{\Theta} \\ \boldsymbol{\Theta}' \end{bmatrix} \perp\!\!\!\perp \begin{bmatrix} \widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}) \\ \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right) \end{bmatrix} \Rightarrow K_j\left(\boldsymbol{\Theta}_j, \boldsymbol{\Theta}'_j\right) \perp\!\!\!\perp K_y\left(\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}), \widetilde{f}\left(\boldsymbol{\Theta}', \boldsymbol{\Lambda}\right)\right).$$

Hence, the desired equality. Proceed in the same way for the other terms. As a result, we obtain:

$$\mathbb{E}_{\boldsymbol{\Lambda}}\left[H_j(\boldsymbol{\Lambda})\right] = 0.$$

■

Remark 5

When $\widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}) = \left\{\widetilde{f}_l(\boldsymbol{\Theta}, \boldsymbol{\Lambda})\right\}_{1 \leq l \leq q}$ and each $g_l(\boldsymbol{\Theta}, \boldsymbol{\Lambda}) \in \mathbb{R}$, we can show the following equivalence

$$\mathscr{H}_j = 0 \iff \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}).$$

by component-by-component argument. More precisely, starting with $\mathscr{H}_j = 0$:

$$\mathscr{H}_j = 0 \Rightarrow \exists\, \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}} \subseteq \mathscr{D}_{\boldsymbol{\Lambda}} \text{ such that } \mathbb{P}_{\boldsymbol{\Lambda}}(\widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}) = 1 \text{ and } \forall\, \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}, \quad \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}(\boldsymbol{\Theta}, \lambda) \Rightarrow \boldsymbol{\Theta}_j \perp\!\!\!\perp \widetilde{f}_l(\boldsymbol{\Theta}, \lambda). \quad \forall\, 1 \leq l \leq q.$$

Thus, the following steps can be taken:

1. for each $1 \leq l \leq q$, use Lemma 3 to obtain the measurable function h_l:

$$h_l\left(\boldsymbol{\Theta}_{-j}, \lambda\right) = \widetilde{f}_l(\boldsymbol{\Theta}, \lambda) \text{ a.s.} \iff \exists\, \widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta},l} \subseteq \mathscr{D}_{\boldsymbol{\Theta}} \text{ such that } \mathbb{P}_{\boldsymbol{\Theta}}(\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta},l}) = 1 \text{ and } \forall\, \theta \in \widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta},l},\ \widetilde{f}_l(\theta, \lambda) = h_l(\theta_{-j}, \lambda).$$

2. Take $\forall\, \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}$, $\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}} := \cap_{l=1}^{q} \widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta},l} \subseteq \mathscr{D}_{\boldsymbol{\Theta}}$ with $\mathbb{P}_{\boldsymbol{\Theta}}\left(\widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}\right) = 1$ (because the intersection of almost surely events remains almost surely.)
3. Let us put $h(\cdot, \lambda) := \{h_l(\cdot, \lambda)\}_{1 \leq l \leq q}$, $\forall\, \lambda \in \widetilde{\mathscr{D}}_{\boldsymbol{\Lambda}}$. Then, one has:

$$\forall\, \theta \in \widetilde{\mathscr{D}}^{\lambda}_{\boldsymbol{\Theta}}, \widetilde{f}\left(\theta, \lambda\right) = h\left(\theta, \lambda\right).$$

4. We continue the proof from Equation (3.85).

The reverse statement is obtained in the same way for $q = 1$.

3.8.5 Proof of Proposition 3

For a fixed value of λ, the HSIC index is given by:

$$\begin{aligned} H_j(\lambda) = \mathbb{E}_{\boldsymbol{\Theta}^1\boldsymbol{\Theta}^2}\left[K_j\left(\boldsymbol{\Theta}^1_j, \boldsymbol{\Theta}^2_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta}^1, \lambda), \widetilde{f}\left(\boldsymbol{\Theta}^2, \lambda\right)\right)\right] \\ + \mathbb{E}_{\boldsymbol{\Theta}^1\boldsymbol{\Theta}^2}\left[K_j\left(\boldsymbol{\Theta}^1_j, \boldsymbol{\Theta}^2_j\right)\right] \mathbb{E}_{\boldsymbol{\Theta}^3\boldsymbol{\Theta}^4}\left[K_y\left(\widetilde{f}(\boldsymbol{\Theta}^3, \lambda), \widetilde{f}\left(\boldsymbol{\Theta}^4, \lambda\right)\right)\right] \\ - 2\mathbb{E}_{\boldsymbol{\Theta}^1\boldsymbol{\Theta}^2\boldsymbol{\Theta}^3}\left[K_j\left(\boldsymbol{\Theta}^1_j, \boldsymbol{\Theta}^2_j\right) K_y\left(\widetilde{f}(\boldsymbol{\Theta}^1, \lambda), \widetilde{f}\left(\boldsymbol{\Theta}^3, \lambda\right)\right)\right], \end{aligned} \tag{3.86}$$

where $\mathbf{\Theta}^1, \mathbf{\Theta}^2, \mathbf{\Theta}^3$ and $\mathbf{\Theta}^4$ are four independent copies of $\mathbf{\Theta}$. In the light of what was done in Section 3.8.2.1, $H_j(\lambda)$ can written as a single expectation:

$$H_j(\lambda) = \mathbb{E}_{\mathbf{\Theta}^1\mathbf{\Theta}^2\mathbf{\Theta}^3\mathbf{\Theta}^4}\left[\widetilde{\eta}_\lambda\left(\mathbf{\Theta}^1, \mathbf{\Theta}^2, \mathbf{\Theta}^3, \mathbf{\Theta}^4\right)\right], \tag{3.87}$$

where

$$\widetilde{\eta}_\lambda\left(\theta^1, \theta^2, \theta^3, \theta^4\right) := \frac{1}{4!}\sum_{\sigma\in\mathbb{S}_4}\eta_\lambda\left(\theta^{\sigma(1)}, \theta^{\sigma(2)}, \theta^{\sigma(3)}, \theta^{\sigma(4)}\right). \tag{3.88}$$

and

$$\begin{aligned}\eta_\lambda:\ &\mathscr{D}_{\mathbf{\Theta}}\times\mathscr{D}_{\mathbf{\Theta}}\times\mathscr{D}_{\mathbf{\Theta}}\times\mathscr{D}_{\mathbf{\Theta}}\longrightarrow\mathbb{R}\\ &(\theta^1,\theta^2,\theta^3,\theta^4)\longmapsto K_j\left(\theta_j^1,\theta_j^2\right)K_y\left(\widetilde{f}\left(\theta^1,\lambda\right),\widetilde{f}\left(\theta^2,\lambda\right)\right)+K_j\left(\theta_j^1,\theta_j^2\right)K_yK_y\left(\widetilde{f}\left(\theta^3,\lambda\right),\widetilde{f}\left(\theta^4,\lambda\right)\right)\\ &\qquad-2K_j\left(\theta_j^1,\theta_j^2\right)K_y\left(\widetilde{f}\left(\theta^1,\lambda\right),\widetilde{f}\left(\theta^3,\lambda\right)\right).\end{aligned} \tag{3.89}$$

The assumption of bounded kernels implies that:

$$\exists\ M_j>0:\ \forall\ \left(\theta_j,\theta_j'\right)\in\mathscr{D}^2_{\mathbf{\Theta}_j},\left|K_j\left(\theta_j,\theta_j'\right)\right|\leq M_j<\infty, \tag{3.90}$$

$$\exists\ M_y>0:\ \forall\ (y,y')\in\mathscr{D}^2_{\boldsymbol{Y}},\ \left|K_y\left(y,y'\right)\right|\leq M_y<\infty. \tag{3.91}$$

Consequently, the functions η_λ and $\widetilde{\eta}_\lambda$ are also bounded:

$$\forall\ \lambda\in\mathscr{D}_{\mathbf{\Lambda}},\ \forall\ \left(\theta^1,\theta^2,\theta^3,\theta^4\right)\in\mathscr{D}^4_{\mathbf{\Theta}},\ \left|\eta_\lambda\left(\theta^1,\theta^2,\theta^3,\theta^4\right)\right|\leq 4M_jM_y, \tag{3.92}$$

$$\forall\ \lambda\in\mathscr{D}_{\mathbf{\Lambda}},\ \forall\ \left(\theta^1,\theta^2,\theta^3,\theta^4\right)\in\mathscr{D}^4_{\mathbf{\Theta}},\ \left|\widetilde{\eta}_\lambda\left(\theta^1,\theta^2,\theta^3,\theta^4\right)\right|\leq\left|\eta_\lambda\left(\theta^1,\theta^2,\theta^3,\theta^4\right)\right|\leq 4M_jM_y. \tag{3.93}$$

According to (3.93), the variance of $\widetilde{\eta}_\lambda\left(\mathbf{\Theta}^1,\mathbf{\Theta}^2,\mathbf{\Theta}^3,\mathbf{\Theta}^4\right)$ is bounded:

$$\mathbb{V}\left(\widetilde{\eta}_\lambda\left(\mathbf{\Theta}^1,\mathbf{\Theta}^2,\mathbf{\Theta}^3,\mathbf{\Theta}^4\right)\right)\leq 16M_j^2M_y^2. \tag{3.94}$$

Still according to (3.93), the constant $C(r,\widetilde{\eta})$ of Proposition 8 is finite:

$$\forall\ r\in\mathbb{N}^\star\ , C(r,\widetilde{\eta})=\max_{1\leq i_1,i_2,i_3,i_4\leq 4}\mathbb{E}\left[\left|\widetilde{\eta}_\lambda\left(\mathbf{\Theta}^{i_1},\mathbf{\Theta}^{i_2},\mathbf{\Theta}^{i_3},\mathbf{\Theta}^{i_4}\right)\right|^r\right]\leq\left(4M_jM_y\right)^r. \tag{3.95}$$

3.8.5.1 Consistency of Estimator 1

We begin by showing that Estimator 1 is unbiased.

$$\begin{aligned}\mathbb{E}\left[\widehat{\widehat{\mathscr{H}}}_j^1\right]&=\frac{1}{m}\sum_{k=1}^m\widehat{H}_j^{\mathrm{U}}\left(D_{\mathbf{\Theta}^{(k)}F_k,n}\right)\quad\text{with}\quad D_{\mathbf{\Theta}^{(k)}F_k,n}=\{D^{(k)}_{\mathbf{\Theta},n};F_k\}\\ &=\mathbb{E}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\mathbf{\Theta}F,n}\right)\right]\text{ as the designs }\left\{D_{\mathbf{\Theta}^{(k)}F_k,n}\right\}_{\leq k\leq m}\text{ are i.i.d.}\\ &=\mathbb{E}_{\mathbf{\Lambda}}\left[\mathbb{E}_{D_{\mathbf{\Theta},n}}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\mathbf{\Theta}F,n}\right)\mid\mathbf{\Lambda}\right]\right]\\ &=\mathbb{E}_{\mathbf{\Lambda}}\left[\mathbb{E}_{D_{\mathbf{\Theta},n}}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\mathbf{\Theta}F,n}\right)\right]\right]\text{ since }\mathbf{\Lambda}\perp\!\!\!\perp D_{\mathbf{\Theta},n}.\end{aligned}$$

For a fixed value $\lambda\in\mathscr{D}_{\mathbf{\Lambda}}$, $\widehat{H}_j^{\mathrm{U}}\left(D_{\mathbf{\Theta}F,n}\right)$ is simply a U-statistic estimator of $\mathrm{HSIC}\left(\mathbf{\Theta}_j,F\right)$ and is therefore unbiased:

$$\mathbb{E}_{D_{\mathbf{\Theta},n}}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\mathbf{\Theta}F,n}\right)\right]=H_j\left(\lambda\right).$$

Finally, we have:

$$\mathbb{E}\left[\widehat{\widehat{\mathscr{H}}}_j^1\right]=\mathbb{E}_{\mathbf{\Lambda}}\left[H_j\left(\mathbf{\Lambda}\right)\right]=\mathscr{H}_j.$$

Now, let us address the question of convergence of square mean. It is only a matter of founding an appropriate upper bound for the mean square error.

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}}^1 - \mathscr{H}_j\right)^2\right] = \mathbb{V}\left(\widehat{\widehat{\mathscr{H}_j}}^1\right) \text{ as } \widehat{\widehat{\mathscr{H}_j}}^1 \text{ is unbiased}$$
$$= \frac{1}{m}\mathbb{V}\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right) \text{ as the designs } \left\{D_{\boldsymbol{\Theta}^{(k)}F_k,n}\right\}_{1\leq k\leq m} \text{ are i.i.d.}$$

The total variance formula yields:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}}^1 - \mathscr{H}_j\right)^2\right] = \frac{1}{m}\left(\mathbb{V}_{\boldsymbol{\Lambda}}\left(\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right) \mid \boldsymbol{\Lambda}\right]\right) + \mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{V}_{D_{\boldsymbol{\Theta},n}}\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right) \mid \boldsymbol{\Lambda}\right)\right]\right) \tag{3.96}$$

$$= \frac{1}{m}\left(\mathbb{V}_{\boldsymbol{\Lambda}}\left(\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right]\right) + \mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{V}_{D_{\boldsymbol{\Theta},n}}\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right)\right]\right), \tag{3.97}$$

because $\boldsymbol{\Lambda} \perp\!\!\!\perp D_{\boldsymbol{\Theta},n}$. In addition, one has:

$$\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right] = H_j(\boldsymbol{\Lambda}). \tag{3.98}$$

Applying the result of Equation (3.52) to $\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)$ and then applying the result of Equation (3.94), we finally obtain:

$$\mathbb{V}\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right) \leq \frac{4}{n} \times \left(16M_j^2M_y^2\right). \tag{3.99}$$

Therefore, we have:

$$\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{V}_{D_{\boldsymbol{\Theta},n}}\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right)\right] \leq \frac{64M_j^2M_y^2}{n} \tag{3.100}$$

$$\mathbb{V}_{\boldsymbol{\Lambda}}\left(\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right]\right) = \mathbb{V}\left(H_j(\boldsymbol{\Lambda})\right) \leq 16M_j^2M_y^2. \tag{3.101}$$

Injecting (3.100) and (3.101) into (3.97), leads to:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}}^1 - \mathscr{H}_j\right)^2\right] \leq \frac{16M_j^2M_y^2 + \frac{64M_j^2M_y^2}{n}}{m} = \mathscr{O}\left(\frac{1}{m}\right). \tag{3.102}$$

3.8.5.2 Consistency of Estimator 2

We begin by showing that Estimator 2 is biased.

$$\mathbb{E}\left[\widehat{\widehat{\mathscr{H}_j}}^2\right] = \frac{1}{m}\sum_{k=1}^{m}\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}^{(k)}F_k,n}\right)$$
$$= \mathbb{E}\left[\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right] \text{ as the designs } \left\{D_{\boldsymbol{\Theta}^{(k)}F_k,n}\right\}_{\leq k\leq m} \text{ are i.i.d}$$
$$= \mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) \mid \boldsymbol{\Lambda}\right]\right]$$
$$= \mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right]\right] \text{ since } \boldsymbol{\Lambda} \perp\!\!\!\perp D_{\boldsymbol{\Theta},n}.$$

For a fixed value $\lambda \in \mathscr{D}_{\boldsymbol{\Lambda}}$, $\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right)$ is simply a V-statistic estimator of $\mathrm{HSIC}\left(\boldsymbol{\Theta}_j, F\right)$ and is therefore biased:

$$\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right] \neq H_j\left(\lambda\right).$$

Finally, we have:

$$\mathbb{E}\left[\widehat{\widehat{\mathscr{H}_j}}^2\right] \neq \mathscr{H}_j.$$

Let us now show the convergence in square mean:

Proof:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2} - \mathscr{H}_j\right)^2\right] = \mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2} - \widehat{\widehat{\mathscr{H}}}_j^{\,1} + \widehat{\widehat{\mathscr{H}}}_j^{\,1} - \mathscr{H}_j\right)^2\right] \tag{3.103}$$

$$\leq 2\,\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2} - \widehat{\widehat{\mathscr{H}}}_j^{\,1}\right)^2\right] + 2\,\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^{\,1} - \mathscr{H}_j\right)^2\right]. \tag{3.104}$$

According to Equation (3.102), the first term is equal to:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^{\,1} - \mathscr{H}_j\right)^2\right] = \mathscr{O}\left(\frac{1}{m}\right). \tag{3.105}$$

As regards the second term, it can be rewritten as:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2} - \widehat{\widehat{\mathscr{H}}}_j^{\,1}\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{m}\sum_{k=1}^{m} A_{jk}\right)^2\right], \tag{3.106}$$

where $A_{jk} = \widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - \widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)$. The variables $A_{j1}, \cdots, A_{jm}$ are independent and have the same distribution because they are computed from the designs $\left\{D_{\boldsymbol{\Theta}^{(k)}F_k,n}\right\}_{\leq k \leq m}$ that have the same properties. Then applying (3.76) to $A_{j1}, \cdots, A_{jm}$, we obtain:

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{k=1}^{m} A_{jk}\right)^2\right] = \frac{1}{m}\mathbb{E}\left[\left(\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - \widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right)^2\right] + \frac{m-1}{m}\left(\mathbb{E}\left[\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - \widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right]\right)^2. \tag{3.107}$$

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{k=1}^{m} A_{jk}\right)^2\right] = \frac{1}{m}\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},\boldsymbol{n}}}\left[\left(\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - \widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right)^2\right]\right] + \frac{m-1}{m}\left(\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},\boldsymbol{n}}}\left[\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - \widehat{H}_{\mathrm{U},\,j}\left(D_{\boldsymbol{\Theta}F,n}\right)\right]\right]\right)^2, \tag{3.108}$$

According to Lemma 2 (applied with $r = 2$), we have:

$$\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},\boldsymbol{n}}}\left[\left(\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - \widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right)^2\right]\right] = \mathscr{O}\left(\frac{1}{n^2}\right). \tag{3.109}$$

■

Likewise, after applying Jensen's inequality and Lemma 2 (applied with $r = 2$), it can easily seen that:

$$\left(\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},\boldsymbol{n}}}\left[\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - \widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right)\right]\right]\right)^2 = \mathscr{O}\left(\frac{1}{n^2}\right). \tag{3.110}$$

Consequently, Equation (3.108) becomes:

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{k=1}^{m} A_{jk}\right)^2\right] = \frac{1}{m}\mathscr{O}\left(\frac{1}{n^2}\right) + \frac{m-1}{m}\mathscr{O}\left(\frac{1}{n^2}\right) = \mathscr{O}\left(\frac{1}{mn^2}\right) + \mathscr{O}\left(\frac{1}{n^2}\right) = \mathscr{O}\left(\frac{1}{n^2}\right) \tag{3.111}$$

Combining Equations (3.105) and (3.111) provides the final convergence rate:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2} - \mathscr{H}_j\right)^2\right] = \mathscr{O}\left(\frac{1}{m}\right) + \mathscr{O}\left(\frac{1}{n^2}\right). \tag{3.112}$$

3.8.5.3 Consistency of Estimator 3

We can show that Estimator 3 is unbiased by using the fact that the designs $\{D_{\boldsymbol{\Theta}F_k,n}\}_{\leq k\leq m}$ follow the same distribution. It remains to show consistency. Let us put $\widehat{H}_j := \frac{1}{m}\sum_{k=1}^{m} H_j(\boldsymbol{\Lambda}^{(k)})$.

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}^3} - \mathscr{H}_j\right)^2\right] = \mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}^3} - \widehat{H}_j + \widehat{H}_j - \mathscr{H}_j\right)^2\right] \tag{3.113}$$

$$\leq 2\,\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}} - \widehat{H}_j\right)^2\right] + 2\,\mathbb{E}\left[\left(\widehat{H}_j - \mathscr{H}_j\right)^2\right]. \tag{3.114}$$

In the above, the first term can be bounded by:

$$2\,\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}^3} - \widehat{H}_j\right)^2\right] = 2\,\mathbb{E}\left[\frac{1}{m}\sum_{k=1}^{m}\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F_k,n}\right) - H_j(\boldsymbol{\Lambda}^{(k)})\right)^2\right] \tag{3.115}$$

$$\leq \frac{2}{m}\sum_{k=1}^{m}\mathbb{E}\left[\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F_k,n}\right) - H_j(\boldsymbol{\Lambda}^{(k)})\right)^2\right], \tag{3.116}$$

where we have used Cauchy-Schwarz formula in $\mathbb{R}^m$ on the vectors $\frac{1}{\sqrt{m}}\left\{\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F_k,n}\right) - H_j(\boldsymbol{\Lambda}^{(k)})\right\}_{k=1}^{m}$ and $(1,\cdots,1)^t$. Then, one has:

$$2\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}^3} - \widehat{H}_j\right)^2\right] \leq 2\mathbb{E}\left[\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right) - H_j(\boldsymbol{\Lambda})\right)^2\right], \tag{3.117}$$

because the designs $\left\{D_{\boldsymbol{\Theta},n}, \boldsymbol{\Lambda}^{(k)}\right\}_{1\leq k\leq m}$ follow the same distribution. By the independence between $\boldsymbol{\Lambda}$ and $D_{\boldsymbol{\Theta},n}$:

$$2\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}^3} - \widehat{H}_j\right)^2\right] \leq 2\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right) - H_j(\boldsymbol{\Lambda})\right)^2\right]\right]. \tag{3.118}$$

Now, according to what was done in the demonstration of the consistency of Estimator 1, we have

$$\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{U}}\left(D_{\boldsymbol{\Theta}F,n}\right) - H_j(\boldsymbol{\Lambda})\right)^2\right]\right] = \mathscr{O}\left(\frac{1}{n}\right) \text{ (see Equation (3.100)).} \tag{3.119}$$

Then, Equation (3.117) becomes:

$$2\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}_j}^3} - \widehat{H}_j\right)^2\right] = \mathscr{O}\left(\frac{1}{n}\right). \tag{3.120}$$

The second term in the right hand side of Equation (3.114) is equal to:

$$2\mathbb{E}\left[\left(\widehat{H}_j - \mathscr{H}_j\right)^2\right] = 2\mathbb{E}\left[\left(\widehat{H}_j - \mathbb{E}\left[\widehat{H}_j\right]\right)^2\right] \tag{3.121}$$

$$= \mathbb{V}\left(\widehat{H}_j\right) = \mathbb{V}\left(\frac{1}{m}\sum_{k=1}^{m} H_j(\boldsymbol{\Lambda}^{(k)})\right) = \frac{\mathbb{V}\left(H_j(\boldsymbol{\Lambda})\right)}{m} \text{ because the variable } \{\boldsymbol{\Lambda}^{(k)}\}_{k=1}^{m} \text{ are i.i.d.} \tag{3.122}$$

$$= \mathscr{O}\left(\frac{1}{m}\right). \tag{3.123}$$

Finally, (3.114) becomes:

$$\mathbb{E}\left[\left(\widehat{\mathscr{H}}_{\mathrm{U},\ D_{\boldsymbol{\Theta},n},\ j} - \mathscr{H}_j\right)^2\right] \leq \mathscr{O}\left(\frac{1}{n}\right) + \mathscr{O}\left(\frac{1}{m}\right) = \mathscr{O}\left(\frac{1}{m} + \frac{1}{n}\right). \tag{3.124}$$

3.8.5.4 Consistency of estimator 4

Proof: Like Estimator 2, Estimator 4 is biased due to the use of V-statistics. Let us now show consistency with the same technique as for Estimator 3. In particular, Equation (3.114) is not affected by the choice of U-statistics or V-statistics. This gives us:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^4 - \mathscr{H}_j\right)^2\right] \leq 2\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^4 - \widehat{H}_j\right)^2\right] + 2\mathbb{E}\left[\left(\widehat{H}_j - \mathscr{H}_j\right)^2\right]. \tag{3.125}$$

The second term does not change:

$$2\mathbb{E}\left[\left(\widehat{H}_j - \mathscr{H}_j\right)^2\right] = \mathscr{O}\left(\frac{1}{m}\right). \tag{3.126}$$

For the same reasons as for obtaining (3.118), the first term of (3.125) is bounded by:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^4 - \widehat{H}_j\right)^2\right] \leq \mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - H_j(\boldsymbol{\Lambda})\right)^2\right]\right] \tag{3.127}$$

Then, we obtain:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{V}}\left(D_{\boldsymbol{\Theta}F,n}\right) - H_j(\boldsymbol{\Lambda})\right)^2\right]\right] &\leq 2\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{V}} - \widehat{H}_j^{\mathrm{U}}\right)^2\right]\right] \\ &\quad + 2\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{U}} - H_j(\boldsymbol{\Lambda})\right)^2\right]\right]. \end{aligned} \tag{3.128}$$

Now, from the consistency proofs of Estimators 1 and 2, we have the following results:

$$\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{V}} - \widehat{H}_j^{\mathrm{U}}\right)^2\right]\right] = \mathscr{O}\left(\frac{1}{n^2}\right) \tag{3.129}$$

$$\mathbb{E}_{\boldsymbol{\Lambda}}\left[\mathbb{E}_{D_{\boldsymbol{\Theta},n}}\left[\left(\widehat{H}_j^{\mathrm{U}} - H_j(\boldsymbol{\Lambda})\right)^2\right]\right] = \mathscr{O}\left(\frac{1}{n}\right). \tag{3.130}$$

Finally, Equation (3.125) becomes:

$$\mathbb{E}\left[\left(\widehat{\widehat{\mathscr{H}}}_j^4 - \mathscr{H}_j\right)^2\right] \leq 2\left(2\left(\mathscr{O}\left(\frac{1}{n^2}\right) + \mathscr{O}\left(\frac{1}{n}\right)\right) + \mathscr{O}\left(\frac{1}{m}\right)\right) \tag{3.131}$$

$$\leq \mathscr{O}\left(\frac{1}{n} + \frac{1}{m}\right). \tag{3.132}$$

■

3.8.6 Non-asymptotic Gamma test procedure of Section 3.5.2.2

3.8.6.1 Computation of the shape and scale parameters of the Gamma distribution

Proof: We start with Equation (3.44):

$$\widehat{\widehat{\mathscr{H}}}_j^2 = \frac{1}{mn^2}\sum_{k=1}^{m} Tr\left(\boldsymbol{A}_{jk}\boldsymbol{W}_k\right), \tag{3.133}$$

with

- $\boldsymbol{A}_{jk} = \boldsymbol{H}\boldsymbol{L}_{jk}\boldsymbol{H}$ being a double-centered, square symmetric matrix, where $\boldsymbol{L}_{jk}$ is the Gram matrix of $\boldsymbol{\Theta}_j$ constructed from the data $D_{\boldsymbol{\Theta}^{(k)},n}$,
- $\boldsymbol{W}_k = \boldsymbol{H}\boldsymbol{L}_{y_k}\boldsymbol{H}$ being a double-centered, square symmetric matrix, where $\boldsymbol{L}_{y_k}$ is the Gram matrix of $\widetilde{f}(\boldsymbol{\Theta},\ \boldsymbol{\Lambda})$ constructed from the data $D_{F_k,n}$.

In this method, we fit a Gamma distribution to the histogram of permutations

$$\left\{\widehat{\widehat{\mathscr{H}}}_j^2\left(D_{m,n}^{\tau_b\bullet}\right)\right\}_{1\leq b\leq B_{\mathrm{perm}}}$$

using the method of moments [El Amri and Marrel, 2024; Kazi-Aoual et al., 1995]. To do this, we first rewrite the test statistic as a trace of a product of two square, symmetrical and double-centered matrices, then we use the results of the first two moments provided by [Kazi-Aoual et al., 1995], and finally deduce the shape and scale parameters of the Gamma distribution using the method of moments. The expectation and variance of $\widehat{\widehat{\mathscr{H}}}_j^{\,2}$ under the permutation distribution are given by:

$$\mathbb{E}_{\substack{\text{perm}\\(\tau_1,\cdots,\tau_m)}}\left[\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right] = \frac{1}{mn^2}\sum_{k=1}^{m}\mathbb{E}_{\substack{\text{perm}\\\tau_k}}\left[Tr(\boldsymbol{A}_{jk}\boldsymbol{W}_k^{\tau_k\tau_k})\right] = \frac{1}{m}\sum_{k=1}^{m}\mathbb{E}_{\substack{\text{perm}\\\tau_k}}\left[\widehat{T}_j^{[k]}\right], \tag{3.134}$$

$$\mathbb{V}_{\substack{\text{perm}\\\tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right) = \frac{1}{m^2n^4}\sum_{k=1}^{m}\mathbb{V}_{\substack{\text{perm}\\\tau_k}}\left(Tr(\boldsymbol{A}_{jk}\boldsymbol{W}_k^{\tau_k\tau_k})\right) = \frac{1}{m^2}\sum_{k=1}^{m}\mathbb{V}_{\substack{\text{perm}\\\tau_k}}\left(\widehat{T}_j^{[k]}\right). \tag{3.135}$$

where

$$\widehat{T}_j^{[k]} := \frac{1}{n^2}Tr\left(\boldsymbol{A}_{jk}\boldsymbol{W}_k^{\tau_k\tau_k}\right). \tag{3.136}$$

These two equalities are valid because $\tau_1,\cdots,\tau_m$ are drawn randomly and independently. We fit the histogram of permutations $\left\{\widehat{\widehat{\mathscr{H}}}_j^{\,2}\left(D_{m,n}^{\tau_b\bullet}\right)\right\}_{1\leq b\leq B_{\text{perm}}}$ to the Gamma distribution $\Gamma(\widetilde{a}_j,\widetilde{b}_j)$. Then, we deduce the shape $\widehat{a}_j$ and scale $\widehat{b}_j$ parameters of the Gamma distribution:

$$\begin{cases}\widetilde{a}_j = \dfrac{\mathbb{E}_{\substack{\text{perm}\\\tau_1,\cdots,\tau_m}}\left[\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right]^2}{\mathbb{V}_{\substack{\text{perm}\\\tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right)},\\[2ex] \widetilde{b}_j = \dfrac{\mathbb{V}_{\substack{\text{perm}\\\tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right)}{\mathbb{E}_{\substack{\text{perm}\\\tau_1,\cdots,\tau_m}}\left[\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right]}.\end{cases} \tag{3.137}$$

However, for each $1\leq k\leq m$, we can fit the Gamma distribution $\Gamma(a_j^{[k]},b_j^{[k]})$ to the sample

$$\left\{\widehat{T}_j^{[k]}\left(D_{\boldsymbol{\Theta}^{(k)}F_k,n}^{\tau_b k}\right)\right\}_{1\leq b\leq B_{\text{perm}}} \quad \text{with} \quad D_{\boldsymbol{\Theta}^{(k)}F_k,n}^{\tau_b k} = \left\{D_{\boldsymbol{\Theta}^{(k)},n};D_{F_k,n}^{\tau_b k}\right\}$$

by solving the following system of equations (Figure 3.6):

$$a_j^{[k]}b_j^{[k]} = \mathbb{E}_{\substack{\text{perm}\\\tau_k}}\left(\widehat{T}_j^{[k]}\right) \tag{3.138}$$

$$a_j^{[k]}\left(b_j^{[k]}\right)^2 = \mathbb{V}_{\substack{\text{perm}\\\tau_k}}\left(\widehat{T}_j^{[k]}\right) \tag{3.139}$$

Then Equations (3.134) and (3.135) become:

$$\mathbb{E}_{\substack{\text{perm}\\\tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right) = \frac{1}{m}\sum_{k=1}^{m}a_j^{[k]}b_j^{[k]}, \tag{3.140}$$

$$\mathbb{V}_{\substack{\text{perm}\\\tau_1,\cdots,\tau_m}}\left(\widehat{\widehat{\mathscr{H}}}_j^{\,2}\right) = \frac{1}{m^2}\sum_{k=1}^{m}a_j^{[k]}\left(b_j^{[k]}\right)^2. \tag{3.141}$$

As a result, System (3.137) also becomes:

$$\begin{cases} \widetilde{a}_j = \dfrac{\left(\dfrac{1}{m}\displaystyle\sum_{k=1}^{m} a_j^{[k]} b_j^{[k]}\right)^2}{\frac{1}{m^2}\displaystyle\sum_{k=1}^{m} a_j^{[k]} \left(b_j^{[k]}\right)^2} = \dfrac{\left(\displaystyle\sum_{k=1}^{m} a_j^{[k]} b_j^{[k]}\right)^2}{\displaystyle\sum_{k=1}^{m} a_j^{[k]} \left(b_j^{[k]}\right)^2}, \\ \widetilde{b}_j = \dfrac{\frac{1}{m^2}\displaystyle\sum_{k=1}^{m} a_j^{[k]} \left(b_j^{[k]}\right)^2}{\frac{1}{m}\displaystyle\sum_{k=1}^{m} a_j^{[k]} b_j^{[k]}} = \dfrac{\displaystyle\sum_{k=1}^{m} a_j^{[k]} \left(b_j^{[k]}\right)^2}{m\displaystyle\sum_{k=1}^{m} a_j^{[k]} b_j^{[k]}}. \end{cases}$$

Therefore, the p-value is computed as follows:

$$\widehat{p}_j = 1 - F_{\Gamma(\widetilde{a}_j, \widetilde{b}_j)}\left(D_{m,n}^{\text{obs}}\right) \text{ with } 1 \leq j \leq p, \tag{3.142}$$

where $F_{\Gamma(\widetilde{a}_j, \widetilde{b}_j)}(\cdot)$ is the cumulative function of Gamma distribution. ■

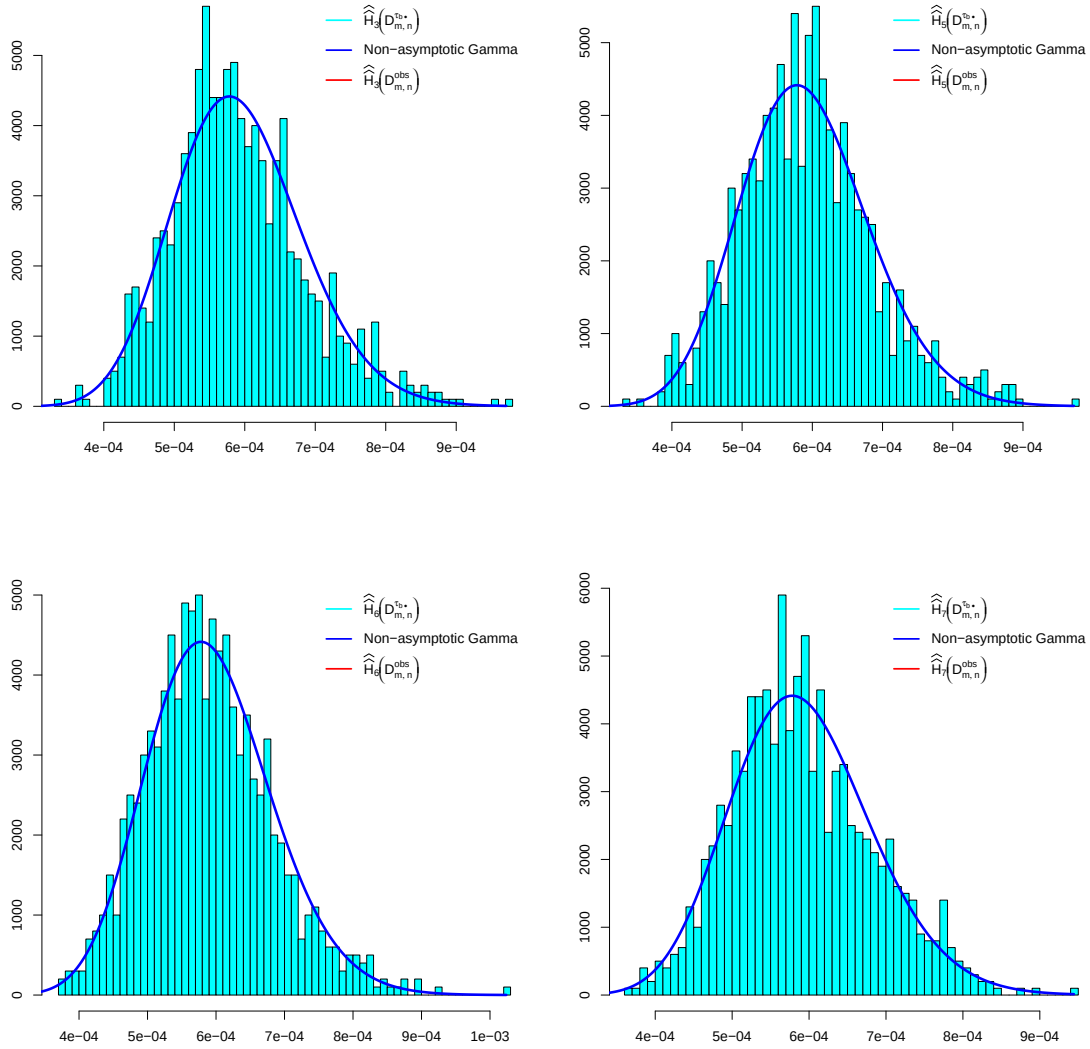


Figure 3.8: Superposition of histograms of $\left\{\widehat{\widehat{\mathscr{H}}}_j^{\,2}\left(D_{m,n}^{\tau_b\bullet}\right)\right\}_{1\le b\le B_{\mathrm{perm}}}$ ($B_{\mathrm{perm}} = 10^3$) and Gamma distribution densities $\Gamma(\widetilde{a}_j, \widetilde{b}_j)$ ($1 \le j \le p$) relative to the calibration parameters $\mathrm{k_v}$, A_2 and A_3 (Approach C).

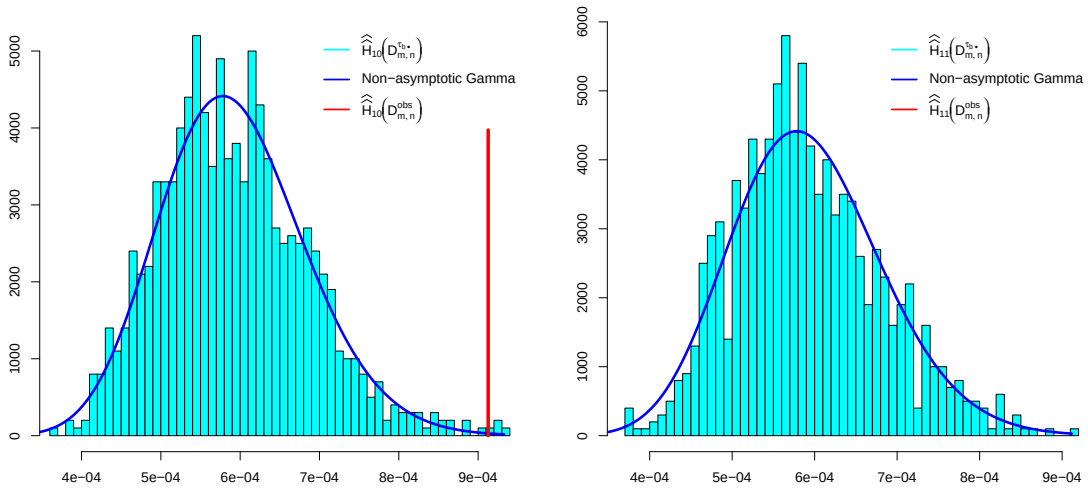


Figure 3.9: Superposition of histograms of $\left\{\widehat{\widehat{\mathscr{H}}}_j^{\,2}\left(D_{m,n}^{\tau_b\bullet}\right)\right\}_{1\leq b\leq B_{\mathrm{perm}}}$ ($B_{\mathrm{perm}} = 10^3$) and Gamma distribution densities $\Gamma(\widetilde{a}_j, \widetilde{b}_j)$ ($1 \leq j \leq p$) to the calibration parameters $\mathrm{b}_{\mathrm{hbs}}$ and $\mathrm{R}_{\mathrm{fhbs}}$ (Approach C).

Chapter 4

GP-LinCC approach: application in fuel simulation

Contents

4.1 Introduction

In this chapter, the GP-LinCC method developed in Chapter 2 for linear models is used to calibrate the uncertain parameters $\boldsymbol{\theta}$ of the fission gas behavior model conditional on the uncertainty of the thermal conductivity λ of the thermal model. To accomplish this, the chapter is divided into two main parts.

The first part is devoted to the construction and validation of linear models of the fission gas behavior model (also known as the CARACAS code) for different values of the thermal conductivity. This part is divided into three sections:

- Section 4.2 focuses on an exploratory phase of the relationship between the RGF (Released Gas Fraction) simulated by the CARACAS code and the $p = 6$ parameters identified as influential in Chapter 3 (see Section 3.6). The aim of this exploratory analysis is to identify potential transformations to be applied to the RGF and/or the input parameters in order to obtain more accurate linear models.
- Section 4.3 briefly reviews Bayesian linear regression, ANOVA decomposition, Sobol' indices and the use of the latter two to assess the potential impact of epistemic uncertainties on linear regression coefficients into the calibration process.
- Section 4.4 is devoted to the validation of the linear models thus constructed.

The second part starts from Section 4.5 and is devoted to the implementation of the GP-LinCC approach with the log-linear model built in the first part. It is divided into six subsections:

- The first two subsections focus on GP-LinCC calibration for two cases ($\boldsymbol{\theta} \in \mathbb{R}^2$ and $\boldsymbol{\theta} \in \mathbb{R}^4$). Specifically, we will perform a calibration for two uncertain parameters (i.e., $\boldsymbol{\theta} \in \mathbb{R}^2$) and increase two other uncertain parameters (i.e., $\boldsymbol{\theta} \in \mathbb{R}^4$) in the calibration process.
- The next three subsections are dedicated to the comparison of the predictions of the calibrated CARACAS code with the predictions of the exponential model (application of the exponential transformation to the log-linear model), the comparison of the experimental RGF with the predictions of the exponential model and with the predictions of the calibrated CARACAS code.
- The last subsection focuses on the inspection of the compensation hypothesis.

Finally, Section 4.6 concludes the chapter. Before starting the first part, we recall the definition of the $p = 6$ calibration parameters identified in Chapter 3, which are the explanatory variables of the linear models to be built:

- $\boldsymbol{\theta}_1$ (k_v): parameter involved in the shape of bubbles on the grain boundary (direct effect on the rate of coverage of the grain boundary by bubbles and therefore on the threshold above which the gas release is triggered),
- $\boldsymbol{\theta}_2$ (A_2): pre-exponential coefficient of the mixed term of the diffusion coefficient (irradiation-accelerated thermal diffusion),
- $\boldsymbol{\theta}_3$ (A_3): coefficient of the athermal term of the diffusion coefficient (diffusion only dependent on fission density),
- $\boldsymbol{\theta}_4$ (k_α): parameter involved in the rate of irradiation defect production and therefore affecting restructuring kinetics,
- $\boldsymbol{\theta}_5$ (b_{hbs}): frequency at which gas trapped in peripheral bubbles returns to solution in grain,

- $\boldsymbol{\theta}_6$ ($\mathrm{R_{fhbs}}$): fragment-size into which the gas diffuses in restructured areas with high burnup rates.

We also introduce the following notations. These will be used throughout the rest of this study:

- Transformed parameters:

$$\widetilde{\boldsymbol{\theta}} = \left(\exp(\boldsymbol{\theta}_1), \exp(\boldsymbol{\theta}_2), \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \exp(\boldsymbol{\theta}_5), \boldsymbol{\theta}_6^{-1}\right)^t. \tag{4.1}$$

- Inverse of the paramter $\mathrm{R_{fhbs}}$: $\mathrm{InvR_{fhbs}} := (\mathrm{R_{fhbs}})^{-1}$.

In addition, **we remind that when we aim to estimate the RGF of a fuel rod, we need to run the full ALCYONE code, which in its convergence loop calls the CARACAS model whose parameters are $\boldsymbol{\theta}$ (see Figure 1.5). This is why, in the remainder of this document, when we refer to the result of the CARACAS code, we actually mean the result of the ALCYONE-CARACAS computation**.

4.2 Preliminary data analysis

The purpose of Bayesian conditional calibration is to provide a posterior distribution of $\boldsymbol{\theta}$ conditional on the thermal conductivity λ. For this, we have measured RGF data for $n_{\mathrm{exp}} = 40$ fuel rods. The measurement uncertainty associated with each fuel rod, of the order of 6% $\boldsymbol{z}_i$, is modeled by a zero mean Gaussian distribution of variance $\sigma^2_{\epsilon_i}$, given by the following equation:

$$\mathbb{P}([\boldsymbol{z}_i \pm 6\%\ \boldsymbol{z}_i]) = 1-\alpha \iff \sigma_{\epsilon_i} = \frac{6\%\ \boldsymbol{z}_i}{q_{1-\frac{\alpha}{2}}},\ 1 \leq i \leq n_{\mathrm{exp}}, \tag{4.2}$$

where $\boldsymbol{z}_i$ is the measured experimental RGF fuel for the i-th rod and $q_{1-\frac{\alpha}{2}}$ is the quantile of order $1-\frac{\alpha}{2}$ of $\mathcal{N}(0,1)$.

We also carried out $n = 200$ simulations of the fission gas behavior model (i.e, the simulations of the CARACAS code) for each fuel rod i ($1 \leq i \leq n_{\mathrm{exp}}$). A prior sensitivity analysis performed on these simulations allowed us to reduce the number of parameters to be considered for the conditional calibration. Thus, of the 11 initial parameters, only $p = 6$ parameters $\boldsymbol{\theta}$ are retained for calibration (see Section 3.6 in Chapter 3). This dimension reduction simplifies the calibration problem by focusing only on those parameters that are essential for the fission gas behavior model to reproduce as best as possible the experimental data. In addition, we found that the output of the fission behavior model (namely RGF output) is globally linear with respect to the parameters $\boldsymbol{\theta}$ for a fixed λ. Consequently, we suppose that for any λ_k ($1 \leq k \leq m$), the simulated RGF of the rod i ($1 \leq i \leq n_{\mathrm{exp}}$) can be modeled by the following linear model:

$$\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) = \mathbf{1} g_{\lambda_k,0}(\boldsymbol{x}_i) + \boldsymbol{\Theta}_k g_{\lambda_k}(\boldsymbol{x}_i) + \varepsilon_{i,\lambda_k} = \begin{pmatrix} \mathbf{1} & \boldsymbol{\Theta}_k \end{pmatrix} \begin{pmatrix} g_{\lambda_k,0}(\boldsymbol{x}_i) \\ g_{\lambda_k,1}(\boldsymbol{x}_i) \end{pmatrix} + \epsilon_{i,\lambda_k} = \boldsymbol{\Theta}_{\mathrm{reg},k} g_{\lambda_k}(\boldsymbol{x}_i) + \varepsilon_{i,\lambda_k}, \tag{4.3}$$

where

- the vector $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n_{\mathrm{exp}}})^t$ represents the experimental configurations including, for example, the linear heat rate, the geometry and the burnup of each fuel rod i ($1 \leq i \leq n_{\mathrm{exp}}$).
- The numerical experimental design of size n containing $p = 6$ input parameters is given by:

$$\boldsymbol{\Theta}_k = \left[\boldsymbol{\theta}_1^{(k,l)}, \cdots, \boldsymbol{\theta}_p^{(k,l)}\right]_{1 \leq l \leq n} \in \mathbb{R}^{n \times p}. \tag{4.4}$$

The index k is used to differentiate the numerical designs $\{\boldsymbol{\Theta}_k\}_{k=1}^m$ where each design is associated with a single λ_k. Then the vector

$$\boldsymbol{\theta}^{(k,l)} = \left(\boldsymbol{\theta}_1^{(k,l)}, \cdots, \boldsymbol{\theta}_p^{(k,l)}\right)^t$$

means the l-th sample of $\boldsymbol{\Theta}_k$. Let us denote now:

$$\boldsymbol{\Theta}_{\text{reg},k} = \begin{pmatrix} \mathbf{1} & \boldsymbol{\Theta}_k \end{pmatrix}.$$

- The vector of RGF simulations of size n each rod $1 \leq i \leq n_{\text{exp}}$ is given by:

$$\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) = \left[Y_{\lambda_k}^{(l)}(\boldsymbol{x}_i)\right]_{1\leq l\leq n} \quad \text{with} \quad \boldsymbol{Y}_{\lambda_k}^{(l)}(\boldsymbol{x}_i) := y_{\boldsymbol{\theta}^{(k,l)},\lambda_k}(\boldsymbol{x}_i)$$

with

$$y_{\boldsymbol{\theta}^{(k,l)},\lambda_k}(\boldsymbol{x}_i) := \left[y^2_{\boldsymbol{\theta}^{(k,l)}} \circ y^1_{\lambda_k}\right](\boldsymbol{x}_i) \in \mathbb{R}. \tag{4.5}$$

- The quantity $g_{\lambda_k}(\boldsymbol{x}_i) \in \mathbb{R}^{p+1}$ represents the coefficients to be estimated associated with the matrix $\boldsymbol{\Theta}_{\text{reg},k}$ of explanatory variables.

- The vector $\varepsilon_{i,\lambda_k} = \left(\epsilon_{i,\lambda_k}^{(1)}, \cdots, \epsilon_{i,\lambda_k}^{(n)}\right)^t | \sigma^2_{i,\lambda_k} \sim \mathcal{N}(0, \sigma^2_{i,\lambda_k} I_n)$ is the vector of the realization of linear model uncertainty, where each component is an independent and identically distributed (i.i.d.) normal distribution $\mathcal{N}(0, \sigma^2_{i,\lambda_k})$ with σ^2_{i,λ_k} potentially different for each $\boldsymbol{x}_i$ and each λ_k.

We can graphically display the relationships between the outputs $\{\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\}_{1\leq i\leq n}$ and the explanatory variables $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_p$ in order to identify some possible transformations that could improve the predictive capability of the linear models. To do this, we generate the $p \times n_{\text{exp}}$ scatter plots of $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)$ as a function of $\boldsymbol{\theta}_j$ $(1 \leq j \leq p)$ associated with the thermal conductivity λ_k.

In this analysis, we focus on a specific example, namely the scatter plot of the fuel rod $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ (see Figure 4.1). This rod has the highest burnup of the n_{exp} rods.To make it easier to visualize and deduce possible input transformations, we overlay a locally weighted three-cubed polynomial approximation (red curves) [Cleveland, 1979]. An interesting observation is that we might consider applying an inverse transformation to the parameter $\boldsymbol{\theta}_6$ as the RGF of this rod appears to vary inversely with $\boldsymbol{\theta}_6$. We will refer to this new parameter as $\boldsymbol{\theta}_6^{-1}$ (i.e., $\text{InvR}_{\text{fhbs}}$). However, for the other parameters, it is unclear how to find a suitable transformation. Additionally, the histogram of the simulations $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ (see Figure 4.2) seems to be well-fitted by a log-normal distribution. Based on this observation, we applied a logarithmic transformation to $\boldsymbol{\theta}_2, \boldsymbol{\theta}_4$ and $\boldsymbol{\theta}_6$ as well as the outputs and plotted the corresponding scatter plots (see Figure 4.3). The resulting graphs suggest a linear relationship between $\log \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ and

$$\log(\widetilde{\boldsymbol{\theta}}) = \left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \log(\boldsymbol{\theta}_3), \log(\boldsymbol{\theta}_4), \boldsymbol{\theta}_5, \log(\boldsymbol{\theta}_6^{-1})\right)^t, \tag{4.6}$$

where $\widetilde{\boldsymbol{\theta}}$ is given in Equation (4.1).

We proceed with two different linear regressions for a better selection of linear regressions for different values of the thermal conductivity λ:

- **A linear regression** between the inputs and the output $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ without any logarithmic transformation of the variables (Equation (4.3) with $i = 26$):

$$\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26}) = \mathbf{1} g_{\lambda_k,0}(\boldsymbol{x}_{26}) + \boldsymbol{\Theta}_k g_{\lambda_k}(\boldsymbol{x}_{26}) + \varepsilon_{26,\lambda_k} = \boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_{26}) + \varepsilon_{26,\,\lambda_k}, \tag{4.7}$$

with

$$\boldsymbol{\Theta}_k = \left[\boldsymbol{\theta}_1^{(k,l)}, \cdots, (\boldsymbol{\theta}_p^{-1})^{(k,l)}\right]_{1\leq l\leq n} \in \mathbb{R}^{n\times p} \quad \text{and} \quad \boldsymbol{\Theta}_{\text{reg},k} = \begin{pmatrix} \mathbf{1} & \boldsymbol{\Theta}_k \end{pmatrix}.$$

We use the *lm* function in R software to estimate the coefficients $g_{\lambda_k}(\boldsymbol{x}_{26})$ with the thermal conductivity $\lambda_k = 1.099$. We recall that the index k is used only to differentiate the numerical designs $\{\boldsymbol{\Theta}_k\}_{k=1}^m$ where each design is associated with a single λ_k.

- **A log-linear regression** by applying the logarithmic transformation to both the output $y_{\boldsymbol{\theta}^{(k,l)},\lambda_k}(\boldsymbol{x}_i)$ and inputs $\widetilde{\boldsymbol{\theta}}^{(k,l)}$ for each $1 \leq l \leq n$:

$$\log \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26}) = \mathbf{1} g_{\lambda_k,0}(\boldsymbol{x}_{26}) + \log\left(\widetilde{\boldsymbol{\Theta}}_k\right) g_{\lambda_k}(\boldsymbol{x}_{26}) + \varepsilon_{26,\lambda_k},$$

with

$$\log\left(\widetilde{\boldsymbol{\Theta}}_k\right) = \left\{\log\left(\widetilde{\boldsymbol{\Theta}}_k^{(l)}\right)\right\}_{1\leq l\leq n} \quad \text{with} \quad \log\left(\widetilde{\boldsymbol{\Theta}}_k^{(l)}\right) = \log(\widetilde{\boldsymbol{\theta}}^{(k,l)}).$$

We again use *lm* to estimate the coefficients $g_{\lambda_k}(\boldsymbol{x}_{26})$ with $\lambda_k = 1.099$. Let us denote:

$$\log \widetilde{\boldsymbol{\Theta}}_{\text{reg},k} = \left(\mathbf{1} \quad \log\left(\widetilde{\boldsymbol{\Theta}}_k\right)\right).$$

Remark 6

1. The exploratory study carried out on $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ can be generalized to the other fuel rods. The conclusions would be practically the same.
2. Both models (linear and log-linear) are used for all fuel rods ($1 \leq i \leq n_{\text{exp}}$) and all λ_k ($1 \leq k \leq m$).
3. It should be noted that in both cases (i.e. linear and log-linear regressions) the coefficients $g_{\lambda_k}(\boldsymbol{x}_{26})$ are not the same quantities and are estimated on different training data. In fact, in the first case they are estimated on simulated input-output data without logarithmic transformation, and in the second case they are estimated on transformed simulated input-output data (with logarithmic transformation). Consequently, the coefficients $g_{\lambda_k}(\boldsymbol{x}_{26})$ are not the same in the two cases. Note that this observation remains valid throughout the rest of this chapter.

Figure 4.4 presents the scatter plot of the true RGF values against those predicted by each of the two regression models for $\lambda_k = 1.099$. For simplicity, we present the regression methodology below using a simple linear regression model. However, the results can easily be extended to the log-linear model without loss of generality by replacing $\{\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\}_{1\leq i\leq n}$ and $\{\boldsymbol{\Theta}_{\text{reg},k}\}_{1\leq k\leq m}$ with $\{\log \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\}_{1\leq i\leq n}$ and $\{\log \widetilde{\boldsymbol{\Theta}}_{\text{reg},k}\}_{1\leq k\leq m}$, respectively. Then, regardless of the regression model, we propose to estimate the regression coefficients in a Bayesian manner.

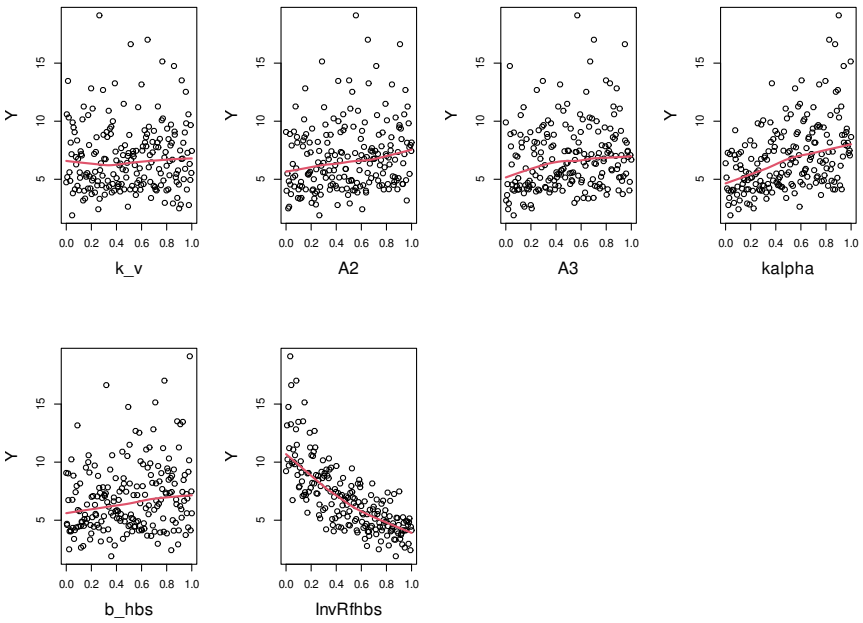


Figure 4.1: Scatter plots between the p inputs $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5, \boldsymbol{\theta}_6^{-1}\}$ and the output $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ associated with $\lambda_k = 1.099$. The red curve represents the local polynomial approximation.

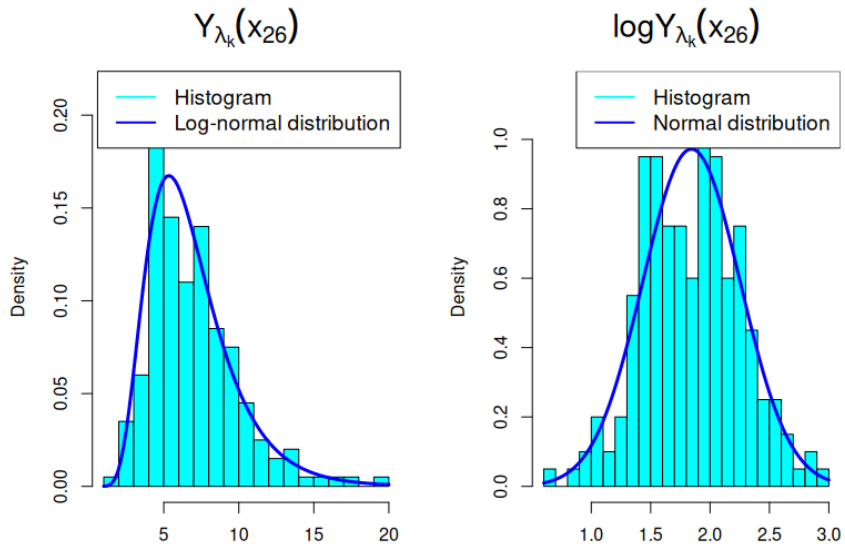


Figure 4.2: Histograms of the data $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ (left) and $\log \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ (right) with the respective overlay of the densities of the estimated log-normal and normal distribution on the data, for $\lambda_k = 1.099$.

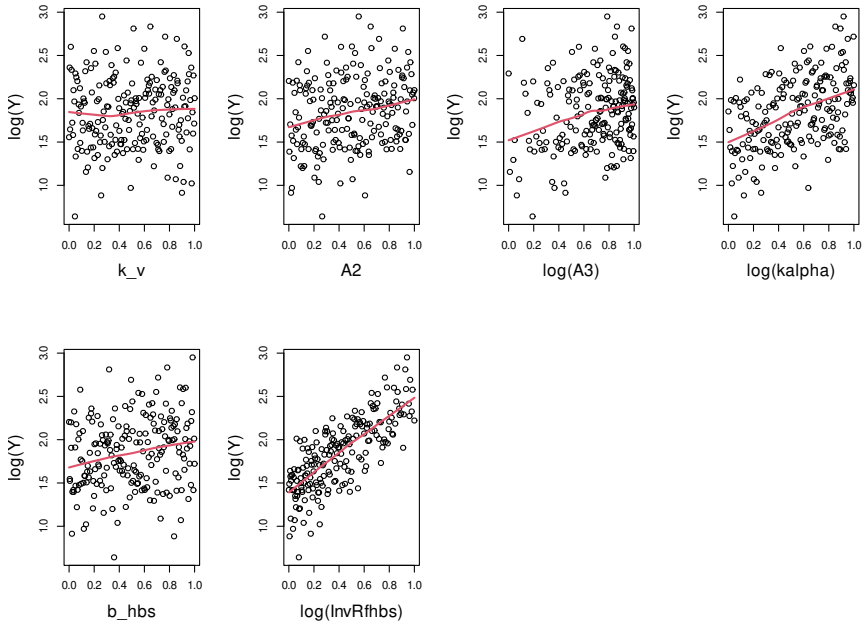


Figure 4.3: Scatter plots between the p inputs $\log(\widetilde{\boldsymbol{\theta}})$ and the output transformed to $\log \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ associated with $\lambda_k = 1.099$. The red curve represents the local polynomial approximation.

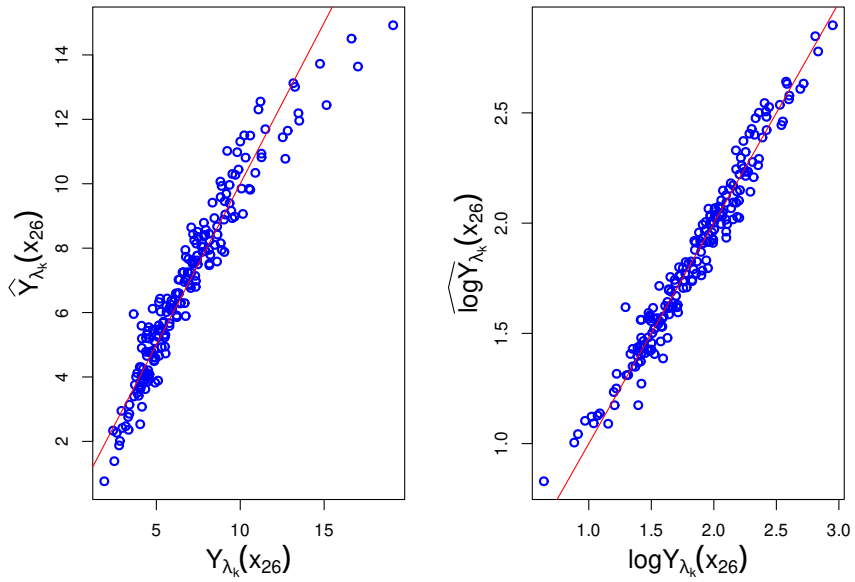


Figure 4.4: On the left, the scatter plot between $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ and $\widehat{\boldsymbol{Y}}_{\lambda_k}(\boldsymbol{x}_{26})$, and on the right, the scatter plot between $\log \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_{26})$ and $\widehat{\log \boldsymbol{Y}}_{\lambda_k}(\boldsymbol{x}_{26})$ with $\lambda_k = 1.099$.

4.3 Linear approximation by Bayesian regression

Bayesian linear regression combines the concepts of classical linear regression with Bayesian estimation. Unlike classical linear regression that provides a point estimate for the model coefficients, Bayesian linear regression provides for them to be treated as random variables. In other words, instead of obtaining a single set of values for the model coefficients, we obtain multiple plausible values, allowing us to account for the uncertainty associated with these estimates. Bayesian inference relies on two key elements:

- Prior knowledge of the coefficients before observing the data, quantified by a prior distribution.
- Update of the prior knowledge using the observed data. Indeed, the prior distribution is updated using Bayes' theorem to obtain a much more informative posterior distribution. This posterior distribution represents the residual epistemic uncertainty of the model coefficients after taking into account the observed data.

In our framework, Bayesian inference consists of three steps:

i. **Specification of the prior distribution:** we must specify a prior distribution that represents our initial knowledge of the model coefficients. Without any prior information, a commonly used prior distribution on the coefficients $(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k})$ is the objective prior [Kass and Wasserman, 1996], defined as:

$$\pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right) \propto \frac{1}{\sigma^2_{i,\lambda_k}}. \tag{4.8}$$

ii. **Specification of the data likelihood:** the likelihood $\mathscr{L}(Y_{\lambda_k}(\boldsymbol{x}_i)|g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k})$ quantifies the probability of observing the data $Y_{\lambda_k}(\boldsymbol{x}_i)$ given the values of the model coefficients $g_{\lambda_k}(\boldsymbol{x}_i)$ and the residual variance σ^2_{i,λ_k}.

iii. **Computation of the posterior distribution:** Using Bayes' theorem, the prior distribution is combined with the likelihood to derive the posterior distribution of $(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k})$. From this posterior distribution, we can retrieve some optimal estimators of the coefficients and residual variance, which can be, for example, the maximum a posteriori estimator (MAP) or the posterior mean.

4.3.1 Posterior and posterior predictive probability distributions

The posterior distribution $\pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right)$ is given by:

$$\pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k} \mid \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right) \propto \mathscr{L}\left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)|g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right) \pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right).$$

It can be shown that the marginals of the posterior distribution $\pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right)$ respectively follow a multivariate Student distribution with $n-p-1$ degrees of freedom, denoted as $T_{n-p-1}(.,.)$, for the coefficients $g_{\lambda_k}(\boldsymbol{x}_i)$ and an inverse Gamma distribution, denoted as $IG(.,.)$, for the residual variance σ^2_{i,λ_k} [Bishop, 2006]:

$$\pi\left(g_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right) \sim T_{n-p-1}\left(\widehat{g}_{\lambda_k}(\boldsymbol{x}_i),\, \widehat{\sigma}^2_{i,\lambda_k}\left(\boldsymbol{\Theta}^t_{\text{reg},k}\boldsymbol{\Theta}_{\text{reg},k}\right)^{-1}\right), \tag{4.9}$$

$$\pi\left(\sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right) \sim IG\left(\frac{n-p-1}{2}, \frac{n-p-1}{2}\widehat{\sigma}^2_{i,\lambda_k}\right), \tag{4.10}$$

where the estimators $\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)$ and $\widehat{\sigma}^2_{i,\lambda_k}$ are given by:

$$\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) = \left(\boldsymbol{\Theta}^t_{\text{reg},k}\boldsymbol{\Theta}_{\text{reg},k}\right)^{-1}\boldsymbol{\Theta}^t_{\text{reg},k}\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i), \tag{4.11}$$

$$\widehat{\sigma}^2_{i,\lambda_k} = \frac{1}{n-p-1}\left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t\left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right). \tag{4.12}$$

Appendix 4.7.1.1 details the calculation of these marginal posterior distributions.

From this posterior distribution $\pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right)$ (Equation (4.76)), we can derive a predictive distribution $\pi\left(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right)$ associated with any new numerical experimental design $\boldsymbol{\Theta}^{\text{new}}_k$ of size n_{new}. This predictive distribution provides a probabilistic estimate of the model response for the new numerical design, taking into account the uncertainty in the model coefficients. This enables us to assess the robustness of predictions and obtain uncertainty intervals for the predicted response. The posterior predictive distribution of the RGF of rod i associated with the thermal conductivity λ_k is derived by integrating the conditional distribution $\mathscr{L}\left(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)$ out $\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)$:

$$\pi\left(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right) = \int \mathscr{L}\left(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)\pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right) dg_{\lambda_k}(\boldsymbol{x}_i)d\sigma^2_{i,\lambda_k}. \tag{4.13}$$

It can be demonstrated that the probability density in Equation (4.13) follows a multivariate Student distribution with $n-p-1$ degrees of freedom:

$$\pi(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) \sim T_{n-p-1}\Bigg(\boldsymbol{\Theta}^{\text{new}}_k\widehat{g}_{\lambda_k}(\boldsymbol{x}_i), \widehat{\sigma}^2_{i,\lambda_k}\Bigg(I_{n_{\text{new}}} + \boldsymbol{\Theta}^{\text{new}}_k\left(\boldsymbol{\Theta}^t_{\text{reg},k}\boldsymbol{\Theta}_{\text{reg},k}\right)^{-1}(\boldsymbol{\Theta}^{\text{new}}_k)^t\Bigg)\Bigg). \tag{4.14}$$

This predictive distribution is used to predict the RGF of rod i associated with the thermal conductivity λ_k. The details of the calculation of this distribution are presented in Appendix 4.7.1.2. From all the posterior distributions

$$\pi\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right) \quad \text{for} \quad 1 \leq k \leq m \text{ and } 1 \leq i \leq n_{\text{exp}},$$

we can now calibrate the model parameters $\boldsymbol{\theta}$ of the fission gas behavior model conditionally on λ_k. Being estimated solely from the simulated data of the CARACAS model, these coefficients will play a crucial in this process.

4.3.2 Analysis of estimation uncertainty

As presented before, the coefficients

$$\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)$$

are also random wariables in a Bayesian linear regression model. We can be interested in assessing the impact of their posterior distribution on the output variable of the CARACAS code (i.e., the simulated RGF output). We then rewrite Equation (4.3) to represent the linear relationship between the RGF $Y_{\lambda_k}(\boldsymbol{x}_i)$ of rod i associated with the thermal conductivity λ_k as follows:

$$Y_{\lambda_k}(\boldsymbol{x}_i) = g_{\lambda_k,0}(\boldsymbol{x}_i) + g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta} + \epsilon_{i,\lambda_k}. \tag{4.15}$$

In this equation, $\boldsymbol{\theta}$ is the vector of parameters to be calibrated. The coefficients

$$g_{\lambda_k}(\boldsymbol{x}_i) = \begin{pmatrix} g_{\lambda_k,0}(\boldsymbol{x}_i) \\ g_{\lambda_k,1}(\boldsymbol{x}_i) \end{pmatrix} \in \mathbb{R}^{p+1}$$

represent the parameters of the linear regression, which are now assumed to follow the posterior distribution defined by Equation (4.9). Additionally, the term $\epsilon_{i,k} \mid \sigma^2_{i,\lambda_k} \sim \mathcal{N}(0, \sigma^2_{i,\lambda_k})$ quantifies the inherent uncertainty of the linear model, where σ^2_{i,λ_k} follows the posterior distribution of Equation (4.10).

When calibrating the parameters $\boldsymbol{\theta}$, the epistemic uncertainty of $(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k})$ (due to the estimation step) should ideally be taken into account. By propagating this epistemic uncertainty into the calibration process $\boldsymbol{\theta}$, the posterior distribution of $\boldsymbol{\theta}$ obtained will no longer be explicit (due to the Student distribution obtained in Equation (4.14) with $\boldsymbol{\Theta}_k^{\text{new}} = \begin{pmatrix} 1 & \boldsymbol{\theta}^t \end{pmatrix}$). Therefore, Markov Chain Monte Carlo (MCMC) algorithms must be used to estimate this posterior distribution [Chib and Greenberg, 1995; Robert et al., 1999; Andrieu and Thoms, 2008]. However, the use of these MCMC algorithms, such as the Metropolis-Hastings algorithm [Chib and Greenberg, 1995; Andrieu and Thoms, 2008], represents an extra computational burden espacially when the number of parameters to calibrate is large. Instead, we may neglect the epistemic uncertainty of the pair $\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)$ and hence set these parameters at their posterior mode values $\left(\widehat{g}_{\lambda_k}(\boldsymbol{x}_i), \widehat{\sigma}^2_{i,\lambda_k}\right)$ (called plug-in approach), which are given in Equations (4.11) and (4.12), respectively. In this way, we would obtain an explicit posterior distribution for $\boldsymbol{\theta}$. However, it is essential to note that this approximation remains valid only if the uncertainty of the pair $\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)$ remains negligible compared to the uncertainty of $\boldsymbol{\theta}$. Altough expected to be true we must justify it properly. In the following, we are thus carrying out a Global Sensitivity Analysis to quantify the relative importance of the epistemic uncertainty of the pair $\left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)$ compared to the uncertainty of $\boldsymbol{\theta}$ in order to confirm whether the former can be neglected in the calibration process.

Reminder of ANOVA decomposition and Sobol' indices

Global sensitivity analysis aims to quantify the impact of uncertainties in variables (here, $g_{\lambda_k}(\boldsymbol{x}_i)$, $\boldsymbol{\theta}$, σ^2_{i,λ_k}) on the output variability (here, $Y_{\lambda_k}(\boldsymbol{x}_i)$). To perform this analysis, Sobol' sensitivity indices [Sobol, 1993; Saltelli et al., 2009] are often used because they are easily interpretable. For mutually independent uncertain input variables (or group of input variables), these indices rely on a variance decomposition of the output into several terms, where each term represents the contribution of an input or a group of inputs to the variance of the output. This decomposition, known as ANOVA (ANalysis Of VariAnce), was introduced in [Hoeffding, 1992] and the Sobol' indices are the normalized terms of this ANOVA.

More precisely, let us consider a model with a scalar output $Y = f(X)$[1] where $X = (X_1, \cdots, X_d)^t$ is the vector of d uncertain and independent input variables. According to the Hoeffding decomposition [Hoeffding, 1992], the function f has a decomposition of the following form:

$$f(X) = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{i<j}^{d} f_{ij}(X_i, X_j) + \cdots + f_{1,2,\cdots,d}(X_1, \cdots, X_d). \tag{4.16}$$

Furthermore, this decomposition is unique under the orthogonality constraint, given by:

$$\mathbb{E}\left[f_J(\boldsymbol{x}_J) | X_{I_s}\right] = \int f_{i_1,\cdots,i_s}(X_{i_1}, \cdots, X_{i_s}) d\mathbb{P}_{X_J} = 0, \; \forall J \in I_s = \{i_1, \cdots, i_s\}, \; 1 \leq s \leq d, \tag{4.17}$$

[1] The function f is assumed to be square-integrable.

where $d\mathbb{P}_{X_J}$ denotes the probability measure associated with X_J. Under the orthogonality constraint, the terms of the decomposition are given by:

$$f_0 = \mathbb{E}\left[f(X)\right], \tag{4.18}$$
$$f_i(X_i) = \mathbb{E}\left[f(X)|X_i\right] - f_0, \tag{4.19}$$
$$f_J(X_J) = \mathbb{E}\left[f(X)|X_J\right] - \sum_{u \subsetneq J} f_u(X_u),\ \forall\, J \subset \{1, \cdots, d\}. \tag{4.20}$$

The functions $\{f_i\}_{1\leq i\leq d}$ represent the main effects, the functions $\{f_{i,j}\}_{1\leq i<j\leq d}$ represent the main effects of second order interactions, and so on for higher-order effects. Using the orthogonality property, we obtain the ANOVA decomposition of the variance of $Y = f(X)$ into a sum of partial variances:

$$\mathbb{V}(Y) = \sum_{i=1}^{d} \mathbb{V}_i + \sum_{i<j}^{d} \mathbb{V}_{ij} + \cdots + \mathbb{V}_{12\cdots d}, \tag{4.21}$$

where

$$\mathbb{V}_i = \mathbb{V}\left(f_i(\boldsymbol{x}_i)\right) = \mathbb{V}\big(\mathbb{E}\left[f(Y)|X_i\right]\big), \tag{4.22}$$
$$\mathbb{V}_{ij} = \mathbb{V}\left(f_{i,j}(\boldsymbol{x}_i, X_j) = \mathbb{V}\left(\mathbb{E}\left[f(Y)|X_i,\, X_j\right]\right) - \mathbb{V}_i - \mathbb{V}_j, \right. \tag{4.23}$$
$$\mathbb{V}_J = \mathbb{V}\left(f_J(\boldsymbol{x}_J)\right) = \mathbb{V}\left(\mathbb{E}\left[f(X)|X_J\right]\right) - \sum_{J' \subset J} (-1)^{|J|-|J'|} \mathbb{V}\left(\mathbb{E}\left[f(X)|X_{J'}\right]\right), \tag{4.24}$$

with $|J|$ denoting the cardinal of J. We deduce the Sobol' indices [Sobol, 1993], which are the normalized terms of the ANOVA decomposition:

$$S_i = \frac{\mathbb{V}_i}{\mathbb{V}(Y)}, \quad S_{ij} = \frac{\mathbb{V}_{ij}}{\mathbb{V}(Y)}, \quad S_J = \frac{\mathbb{V}_J}{\mathbb{V}(Y)}. \tag{4.25}$$

The total sum of all Sobol' indices is equal to 1. The indices S_i, called first-order indices (also called main effects), represent the part of variance of Y explained only by the variable X_i. The second-order indices

$$(S_{ij})_{1\leq i<j\leq d}$$

quantify the part of variance of Y due to the interaction of the pair (X_i, X_j) regardless of their main effects, and so on for higher-order indices. For each X_i, you can define the total index S_i^T, which measures the influence of X_i by including its main effect and its posssible interaction with the other variables. It is given by:

$$S_i^T = S_i + \sum_{i<j} S_{ij} + \sum_{j\neq i, k\neq i, j<k} S_{ijk} + \cdots = \sum_{i\in J} S_J, \tag{4.26}$$

where J is the set of all indices from $\{1, \cdots, d\}$ containing i. A zero total index means that the variable X_i has no (variance-wise) influence on the model output (i.e., Y is independent of X_i).

We aim to demonstrate that the effect of the uncertainty affecting

$$\vartheta_k := \left(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}\right)$$

is negligible on $Y_{\lambda_k}(\boldsymbol{x}_i)$ compared to the uncertainty associated with $\boldsymbol{\theta}$ by considering two groups of random variables, ϑ_k and $\boldsymbol{\theta}$. This comes down to making a Sobol' decomposition in a two-dimensional space with two independent groups. As a reminder, the prior distribution of $\boldsymbol{\theta}$ is given by:

$$\pi(\boldsymbol{\theta}) \sim \prod_{l=1}^{p} \mathscr{U}\left[\boldsymbol{\theta}_l^{min}, \boldsymbol{\theta}_l^{max}\right]. \tag{4.27}$$

The distribution of ϑ_k is given in Appendix 4.7.1.1. We first perform an ANOVA decomposition of $Y_{\lambda_k}(\boldsymbol{x}_i)$. Then, we estimate the second-order Sobol' indices associated with $(\vartheta_k, \boldsymbol{\theta})$, as well as the second-order Sobol' index associated with the interaction between ϑ_k and $\boldsymbol{\theta}$. The ANOVA decomposition of each output $Y_{\lambda_k}(\boldsymbol{x}_i)$ is as follows:

$$\mathbb{V}(Y_{\lambda_k}(\boldsymbol{x}_i)) = \mathbb{V}_{\boldsymbol{\theta}}^{i,k} + \mathbb{V}_{\vartheta_k}^{i,k} + \mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,k}, \tag{4.28}$$

$$\mathbb{V}_{\boldsymbol{\theta}}^{i,k} = \mathbb{V}(\mathbb{E}\left[Y_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{\theta}\right]), \tag{4.29}$$

$$\mathbb{V}_{\vartheta_k}^{i,k} = \mathbb{V}(\mathbb{E}\left[Y_{\lambda_k}(\boldsymbol{x}_i)|\vartheta_k\right]), \tag{4.30}$$

$$\mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,k} = \mathbb{V}(\mathbb{E}\left[Y_{\lambda_k}(\boldsymbol{x}_i)|\vartheta_k, \boldsymbol{\theta}\right]) - \mathbb{V}_{\boldsymbol{\theta}}^{i,k} - \mathbb{V}_{\vartheta_k}^{i,k}. \tag{4.31}$$

The associated Sobol' indices are defined as:

$$S_{\boldsymbol{\theta}}^{i,k} = \frac{\mathbb{V}_{\boldsymbol{\theta}}^{i,k}}{\mathbb{V}(Y_{\lambda_k}(\boldsymbol{x}_i))}, \tag{4.32}$$

$$S_{\vartheta_k}^{i,k} = \frac{\mathbb{V}_{\vartheta_k}^{i,k}}{\mathbb{V}(Y_{\lambda_k}(\boldsymbol{x}_i))}, \tag{4.33}$$

$$S_{\vartheta_k,\boldsymbol{\theta}}^{i,k} = \frac{\mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,k}}{\mathbb{V}(Y_{\lambda_k}(\boldsymbol{x}_i))}. \tag{4.34}$$

The computation of these indices allows us to determine whether the total influence of ϑ_k (consisting of its main effect and the possible interaction with $\boldsymbol{\theta}$) is negligible compared to that of $\boldsymbol{\theta}$. This is expressed by the following inequality:

$$S_{\vartheta_k}^{i,k} + S_{\vartheta_k,\boldsymbol{\theta}}^{i,k} \ll S_{\boldsymbol{\theta},k}^{i,k}, \ \forall\ i,k. \tag{4.35}$$

To compute the explained variances, we utilize the distributions of the components of ϑ_k and $\boldsymbol{\theta}$, as well as Equation (4.15). We then obtain the following analytical expressions:

$$\mathbb{V}_{\boldsymbol{\theta}}^{i,\ k} = \mathbb{E}\left[g_{\lambda_k,1}(\boldsymbol{x}_i)\right]^t \mathbb{V}(\boldsymbol{\theta})\mathbb{E}\left[g_{\lambda_k,1}(\boldsymbol{x}_i)\right] = \mathbb{E}\left[g_{\lambda_k}(\boldsymbol{x}_i)\right]^t B^t\mathbb{V}(\boldsymbol{\theta})B\mathbb{E}\left[g_{\lambda_k}(\boldsymbol{x}_i)\right], \tag{4.36}$$

$$\mathbb{V}_{\vartheta_k}^{i,k} = A^t\mathbb{V}\left(g_{\lambda_k}(\boldsymbol{x}_i)\right)A + \mathbb{E}\left[\boldsymbol{\theta}\right]^t B\mathbb{V}\left(g_{\lambda_k}(\boldsymbol{x}_i)\right)B^t\mathbb{E}\left[\boldsymbol{\theta}\right], \tag{4.37}$$

$$\mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,\ k} = Tr(\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{V}(\boldsymbol{\theta})) = Tr\left(B\mathbb{V}\left(g_{\lambda_k}(\boldsymbol{x}_i)\right)B^t\mathbb{V}(\boldsymbol{\theta})\right), \tag{4.38}$$

where

- $Tr(\cdot)$ is the trace operator on square matrices,
- $A = (1, 0, \cdots, 0)$ is a row vector of size $p+1$,
- $B = \begin{pmatrix}\mathbf{0} & I_p\end{pmatrix} \in \mathbb{R}^{p\times(p+1)}$ is a matrix of size $p \times (p+1)$.

We refer to Appendix 4.7.2 for details of the calculations. It is worth noting that we have $n_{\text{exp}} = 40$ values of RGF for the n_{exp} fuel rods, and for each rod, we build a linear model conditionally on a given thermal conductivity value λ_k $(1 \leq k \leq m)$. In total, we thus constructed $m \times n_{\text{exp}}$ linear models. For each linear regression, the Sobol' indices given in Equations (4.32), (4.33) and (4.34) were exactly computed in order to validate the inequality in Equation (4.35). For $m = 20$ thermal conductivity values λ, we estimated the coefficients

$$\left\{\widehat{g}_{\lambda_k}(\boldsymbol{x}_i),\ \widehat{\sigma}^2_{i,\lambda_k}\right\}_{1\leq i\leq n_{\text{exp}}},\ 1 \leq k \leq m,$$

using simulated data $\{\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\}_{1\leq i\leq n_{\text{exp}}}$, where each $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)$ is a vector of size $n = 200$. The Sobol' indices given in Equations (4.32), (4.33) and (4.34) were then calculated analytically

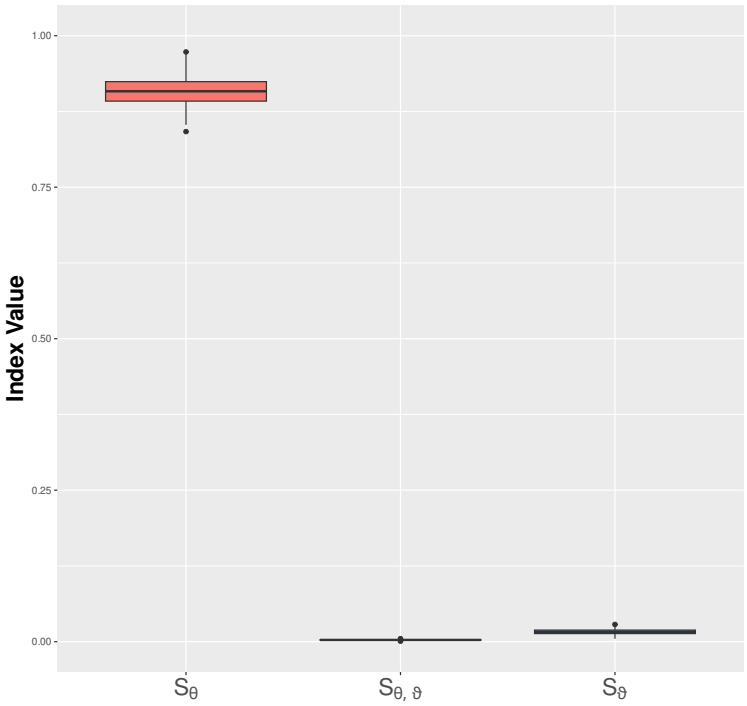


Figure 4.5: Boxplot of the first and second-order Sobol' indices calculated analytically for the $m \times n_{\text{exp}} = 20 \times 40 = 800$ linear models and estimated for each on $n = 200$ simulated data points.

and in Figure 4.5, we present the box plots of these indices. We can see that the total influence of

$$\left\{ g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k} \right\}_{1 \leq i \leq n_{\text{exp}},\, 1 \leq k \leq m}$$

is consistently negligible compared to that of $\boldsymbol{\theta}$. In other words, the inequality reported in Equation (4.35) is satisfied and therefore, the plug-in approximation is suitable for parameter calibration of $\boldsymbol{\theta}$ conditioned on λ_k. We note that the plug-in approximation also remains valid in the case of Bayesian log-linear regression (see Figure 4.6).

Remark 7

The prior distribution chosen on $\widetilde{\boldsymbol{\theta}}$ (resp. $\log(\widetilde{\boldsymbol{\theta}})$) is a uniform on the hyperrectangle obtained after applying all the transformations that enabled us to obtain $\tilde{\boldsymbol{\theta}}$ (resp. $\log(\tilde{\boldsymbol{\theta}})$) on $\mathscr{D}_{\boldsymbol{\Theta}} := \prod_{l=1}^{p} [\boldsymbol{\theta}_l^{min}, \boldsymbol{\theta}_l^{max}]$.

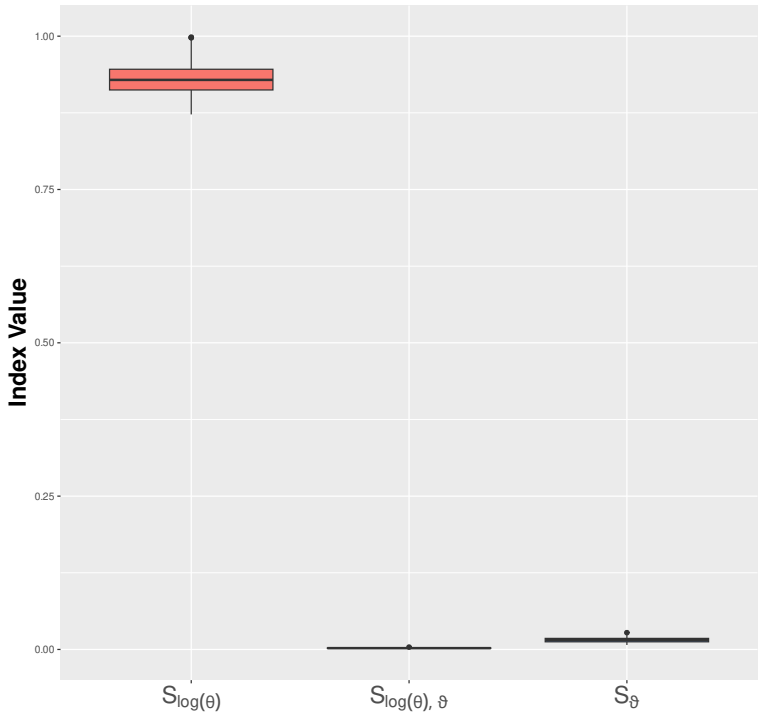


Figure 4.6: Boxplot of the first and second-order Sobol' indices calculated analytically for the $m \times n_{\text{exp}} = 20 \times 40 = 800$ log-linear models whose its coefficients are estimated with $\{\log \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\}_{1 \leq i \leq n_{\text{exp}}}$ and $\{\log \widetilde{\boldsymbol{\Theta}}_{\text{reg},k}\}_{1 \leq k \leq m}$ and estimated for each on $n = 200$ simulated data points.

4.4 Validation of the linear models

4.4.1 Residuals normality analysis

In this section, we aim to check the validity of the chosen linear model to ensure the relevance of the estimations of $\boldsymbol{\theta}(\lambda)$. The validation of the linear model relies on verifying the associated assumptions, such as the independence and normality of the linearization uncertainties

$$\{\epsilon_{i,\,k}\}_{1 \leq i \leq n_{\text{exp}}, 1 \leq k \leq m}.$$

Since the linearization error is not directly observable, we turn to the study of the predicted residuals after the estimation stage from simulated data. The n_{exp} standardized residuals for each fuel rod $\{i\}_{1 \leq i \leq n_{\text{exp}}}$ and each λ_k $(1 \leq k \leq m)$ are defined by the following equation:

$$\frac{\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^{(l)} - \widehat{g}_{\lambda_k,0}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k,1}(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda_k)^{(l)}}{\sqrt{\widehat{\sigma}^2_{i,\lambda_k}}},\ 1 \leq l \leq n = 200, \tag{4.39}$$

where each $\boldsymbol{\theta}(\lambda_k)^{(l)}$ is an element of $\boldsymbol{\Theta}_k = \left\{\boldsymbol{\theta}(\lambda_k)^{(1)}, \cdots, \boldsymbol{\theta}(\lambda_k)^{(n)}\right\}$.

To check the normality of the standardized residuals for each fuel rod $\{i\}_{1 \leq i \leq n_{\text{exp}}}$, we proceeded with a normality analysis by evaluating if, for each $1 \leq l \leq n$:

$$r_{i,\,\lambda_k}\left(\boldsymbol{\theta}(\lambda_k)^{(l)}\right) := \frac{\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^{(l)} - \widehat{g}_{\lambda_k,0}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k,1}(\boldsymbol{x}_i)^t \boldsymbol{\theta}(\lambda_k)^{(l)}}{\sqrt{\widehat{\sigma}^2_{i,\lambda_k}}} \sim \mathcal{N}(0,\,1), \quad \text{for } 1 \leq k \leq m. \tag{4.40}$$

To perform this analysis, we may use graphical visualization tools such as:

1. A Q-Q plot comparing the empirical quantiles associated with the standardized residuals with the quantiles of the normal distribution $\mathcal{N}(0,\,1)$.

2. An histogram that compares the standardized residuals with the density of the standard normal distribution $\mathcal{N}(0,\,1)$.

3. A plot comparing the empirical distribution function of the standardized residuals with that of the standard normal distribution $\mathcal{N}(0,\,1)$.

In this study, we needed to construct $m \times n_{\text{exp}} = 20 \times 40 = 800$ linear models, which gives us a substantial number of residuals to analyze. Therefore, it was also relevant to use quantitative tools such as goodness-of-fit tests. Such tests allow us to quantify the fit of the empirical histogram of the data to the desired theoretical distribution. In practice, the most commonly non-parametric tool for assessing the adequacy of samples to a specified theoretical distribution (here Gaussian distribution) is the Kolmogorov-Smirnov test. This test compares the empirical cumulative distribution function of the data (in this case, the standardized residuals in Equation (4.39)) with the theoretical cumulative distribution function by calculating the test statistic and its associated p-value. The p-value is the probability of observing a test statistic as extreme as the one calculated under the null hypothesis H_0 (here, H_0: adequacy to the theoretical Gaussian distribution). If the p-value is less than a chosen level of the test (usually set at a 5% significance level of *type I error*), the null hypothesis is rejected, indicating that the data are not likely to follow the assumed theoretical distribution. Conversely, if the p-value is greater than the level of the test, the null hypothesis is retained, suggesting that it is quite possible for the data to follow this theoretical distribution.

For $m = 20$ thermal conductivity values, we constructed $n_{\text{exp}} \times m = 40 \times 20 = 800$ linear models. For each model, we followed two steps hereafter:

1. Calculation of the standardized residuals from Equation (4.39).

2. Implementation of the Kolmogorov-Smirnov test on each model to obtain the associated p-values
$$\{\text{p-val}_{i,k}\}_{1\leq i\leq n_{\text{exp}},\,1\leq k\leq m}.$$

To assess whether the standardized residuals follow a Gaussian distribution (null hypothesis H_0), we compare those p-values to the threshold $\alpha = 5\%$:

$$\text{H}_0 \text{ acceptable if and only if } \text{p-val}_{i,\,k} > \alpha, \quad \forall\, 1 \leq i \leq n_{\text{exp}},\, \forall\, 1 \leq k \leq m. \tag{4.41}$$

For each fuel rod i ($1 \leq i \leq n_{\text{exp}}$), we calculated the acceptance rate of the null hypothesis H_0 over all the thermal conductivities $\{\lambda_k\}_{1\leq k\leq m}$:

$$\overline{\text{p-val}}_i := \frac{1}{m}\sum_{k=1}^{m} \mathbf{1}_{\text{p-val}_{i,\,k}>\alpha}. \tag{4.42}$$

Figure 4.7 illustrates the quantities

$$\left\{\overline{\text{p-val}}_i\right\}_{1\leq i\leq n_{\text{exp}}}$$

for the two types of regressions: linear and log-linear. We observe that the null hypothesis H_0 is generally accepted for the majority of the fuel rods regardless of the regression type. Note that the fuel rods are ordered in decreasing order of burnup. It appears that the hypothesis H_0 is much more accepted for the linear model than for the log-linear model. We have also calculated the adjusted coefficient of determination denoted R^2_{adjusted} [Hastie et al., 2009; Cornillon and Matzner-Løber, 2011]. The expression of R^2_{adjusted} for the i-th fuel rod and for the thermal conductivity λ_k is given by:

$$1 - \frac{n-1}{n-p-1}\left(1 - \frac{\frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^{(l)} - \widehat{g}_{\lambda_k,0}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda_k)^{(l)}\right)^2}{\frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^{(l)} - \frac{1}{n}\sum_{l'=1}^{n}\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^{(l')}\right)^2}\right). \tag{4.43}$$

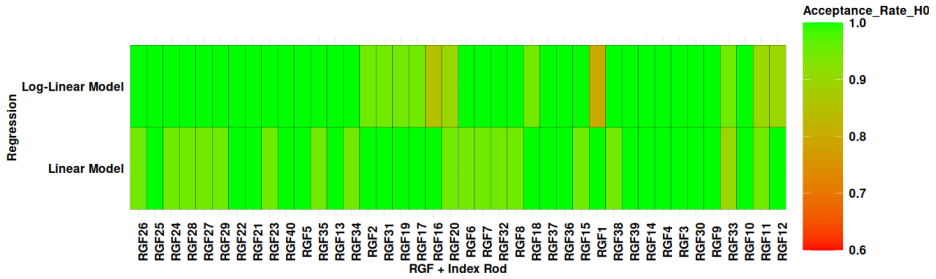


Figure 4.7: Illustration of the acceptance rates of the null hypothesis for the two linear regressions. Fuel rods are sorted in descending order of combustion rate.

Figure 4.8: The R^2_{adjusted} values by fuel rod and by thermal conductivity value for the two linear regressions. Fuel rods are sorted by decreasing combustion rate, and thermal conductivity values are also sorted in descending order.

The adjusted coefficient R^2_{adjusted} represents the proportion of variance explained by the linear model over the total variance of the data $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)$ [Hastie et al., 2009]. The results are reported in Figure 4.8, where we see that the log-linear regression appears to be superior. We also computed the Mean Squared Error (MSE) between the RGF simulations and the predictions of both simple linear and log-linear models. The MSE is expressed as the mean squared difference between observed and predicted values. The smaller MSE, the closer the predictions to the observed values [Steurer et al., 2021]. For a fuel rod i and the thermal conductivity λ_k, the MSE is given respectively by:

$$\frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^{(l)} - \widehat{g}_{\lambda_k,0}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda_k)^{(l)}\right)^2. \tag{4.44}$$

Figure 4.9 presents these MSE values and we can see again that the log-linear model is better.

4.4.2 Residual structure analysis

Analysis of the residual structure in linear regression is an important step to evaluate the goodness of fit of a linear model and check the validity of the underlying assumptions. Residuals measure the gap between the observed values and those predicted by the model. In practice, this analysis may be carried out by visualizing residuals in various ways to detect any structure that may indicate a poor model fit or lack of independence and homoscedasticity. Specifically, for each linear model (i.e., each pair (i,k) with $1 \leq i \leq n_{\text{exp}}$ and $1 \leq k \leq m$), we can create scatter plots of the residuals in different ways:

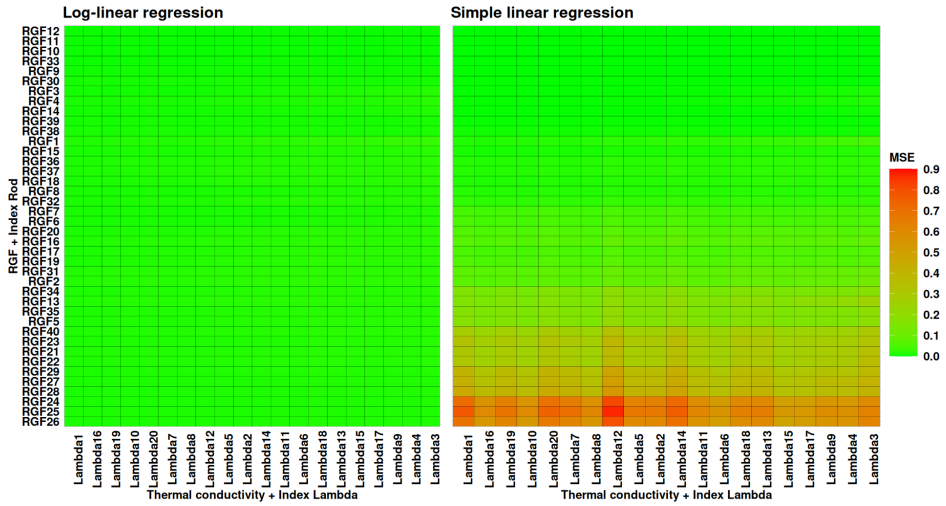


Figure 4.9: Computation of the MSE criterion for the two linear regressions. Fuel rods are sorted by decreasing combustion rate and thermal conductivity values are also sorted in descending order.

1. Plot the pairs
$$\left\{l, r_{i,\lambda_k}\left(\boldsymbol{\theta}(\lambda_k)^{(l)}\right)\right\}$$
to inspect the presence of some pattern(s).

2. Plot the pairs
$$\left\{\left(\widehat{Y}_{\lambda_k}(\boldsymbol{x}_i)^{(l)},\, r_{i,\lambda_k}\left(\boldsymbol{\theta}(\lambda_k)^{(l)}\right)\right)\right\}_{l=1}^{n}$$
to scan the presence of some pattern(s).

3. For the j-th explanatory variable $\boldsymbol{\theta}_j(\lambda_k)$, plot the pairs
$$\left\{\boldsymbol{\theta}_j(\lambda_k)^{(l)}, r_{i,\lambda_k}\left(\boldsymbol{\theta}(\lambda_k)^{(l)}\right)\right\}_{1\leq l\leq n} \quad \text{with} \quad j\in\{1,\ldots,p\}$$
to look at any potential pattern(s).

To simplify equations, we will henceforth use $\boldsymbol{\theta}$ in place of $\boldsymbol{\theta}(\lambda_k)$. Such residual plots can be created simultaneously for all the $n_{\text{exp}} \times m$ linear models. In these plots, it is essential to scan the presence of any pattern which might include polynomial trends, sinusoidal patterns or decreasing trends. The presence of such patterns in residuals could indicate a poor model fit which may result from missing explanatory variables or improper transformations of one or more explanatory variables or the output of interest. Identifying any unusual pattern in the residuals provides some opportunities for model improvement by adjusting explanatory variables, applying appropriate transformations, or considering other models more suitable for the data.

As we deal with a large number of residuals, using visualization tools to detect patterns can be challenging. To address this issue, we propose using a quantitative tool based on independence tests to detect potential dependencies between the residuals and the explanatory variables. This approach replaces the scatter plots mentioned earlier.

An independence test between two variables X_j and Y is a statistical procedure aimed at testing the following two hypotheses:

- The null hypothesis (H_0), which states that the random variables X_j and Y are independent of each other.

- The alternative hypothesis (H_1), which states that the random variables X_j and Y are dependent on each other.

To decide between these hypotheses, a test statistic is calculated from the available data, reflecting the probabilistic dependence of the variables X_j and Y. Subsequently, the p-value associated with the estimated test statistic is compared to the *first-kind error* (or *level of test*), usually set at 5% or 10%. If the p-value is lower than the level of test, the null hypothesis is rejected, which is a strong evidence that the variables X_j and Y are dependent on each other. Otherwise, it is considered that the independence hypothesis is acceptable.

Here we want to employ these tests to detect the presence of structures in residuals that may be related to one or more explanatory variables. The absence of any pattern between the j-th explanatory variable $\boldsymbol{\theta}_j$ and the standardized residuals r_{i,λ_k} must be characterized by a scatter plot close to the horizontal line $y = 0$. This implies that the explanatory variable $\boldsymbol{\theta}_j$ and the standardized residuals r_{i,λ_k} are independent of each other. Conversely, the presence of any structure in this scatter plot indicates some form of dependency.

To test the independence between the j-th explanatory variable and the residuals, we performed the HSIC-based independence testing, taking into account the uncertainty of the thermal conductivity λ. This allows us to apply the methodology presented in Section 3.5 of Chapter 3 to detect the presence/absence of structure in the residuals. In particular, the following steps are carried out:

1. **Definition of the quantity of interest:** $r_{i,\lambda_k},\ 1 \leq i \leq n_{\text{exp}}$ (i.e., Approach A of Section 3.4.2 with $\widetilde{f}_i(\boldsymbol{\theta}, \lambda) = r_{i,\lambda_k}$),
2. **Choice of the HSIC estimator:** $\widehat{\widehat{\mathscr{H}_j}}^2$ (mean of V-statistics).
3. **Choice of the sampling strategy (strategy (S1)):**
$$D_{m,n} = \left\{ \left(\boldsymbol{\Theta}^{(k,l)}, R^{(l)}_{i,\lambda_k} \right) \right\}_{1 \leq l \leq n} \quad \text{with} \quad R^{(l)}_{i,\lambda_k} = r_{i,\lambda_k} \left(\boldsymbol{\Theta}^{(k,l)} \right).$$
4. **Estimation of the following p-value (with the permutation-based test procedure of Section 3.5.2.1):**
$$p_j = \mathbb{P}_{\text{H}_0} \left(\widehat{\widehat{\mathscr{H}_j}}^2 (D_{m,n}) > \widehat{\widehat{\mathscr{H}_j}}^2 \left(D^{\text{obs}}_{m,n} \right) \right),$$
with
$$D^{\text{obs}}_{m,n} = \left\{ \left(\boldsymbol{\theta}(\lambda_k)^{(l)}, r_{i,\lambda_k} \left(\boldsymbol{\theta}(\lambda_k)^{(l)} \right) \right) \right\}_{1 \leq l \leq n}.$$

Therefore, for $m = 20$ realizations of the thermal conductivity, we have estimated these p-values for each r_{i,λ_k} for $1 \leq i \leq n$ and $1 \leq k \leq m$, and the results are presented in Figures 4.10 and 4.11, respectively for the simple and log-linear models. For all linear models, in these figures, we do not observe any pattern in the residuals for the parameters $\boldsymbol{\theta}_5$ (i.e., physical parameter b_{hbs}) and $\boldsymbol{\theta}_1$ (i.e., physical parameter k_{v}) across all RGF fuel rods. For the remaining three parameters, the behavior varies depending on the regression model. For example, for the parameter $\boldsymbol{\theta}_4$ (i.e., physical parameter k_α), we observe the absence of patterns for only 6 of the $n_{\text{exp}} = 40$ fuel rods in the linear model and for 15 fuel rods in the log-linear model for $\log(\boldsymbol{\theta}_4)$.

As the deployment of our GP-LinCC approach is facilitated by using a single regression model for all fuel rods, our choice has leaned towards the log-linear model, which, overall, provides better results. However, as this model may not be entirely satisfactory for some fuel rods, other non-linear statistical modeling approaches could be considered. We leave this perspective for future work (see Chapter 5).

4.4.3 Comparison between the log-linear model predictions and the fission gas behavior model simulations

In this section, we are concentrating on the predictions of the log-linear model which are compared with the CARACAS simulations using a new numerical design denoted by $\boldsymbol{\Theta}^{\text{new}}$

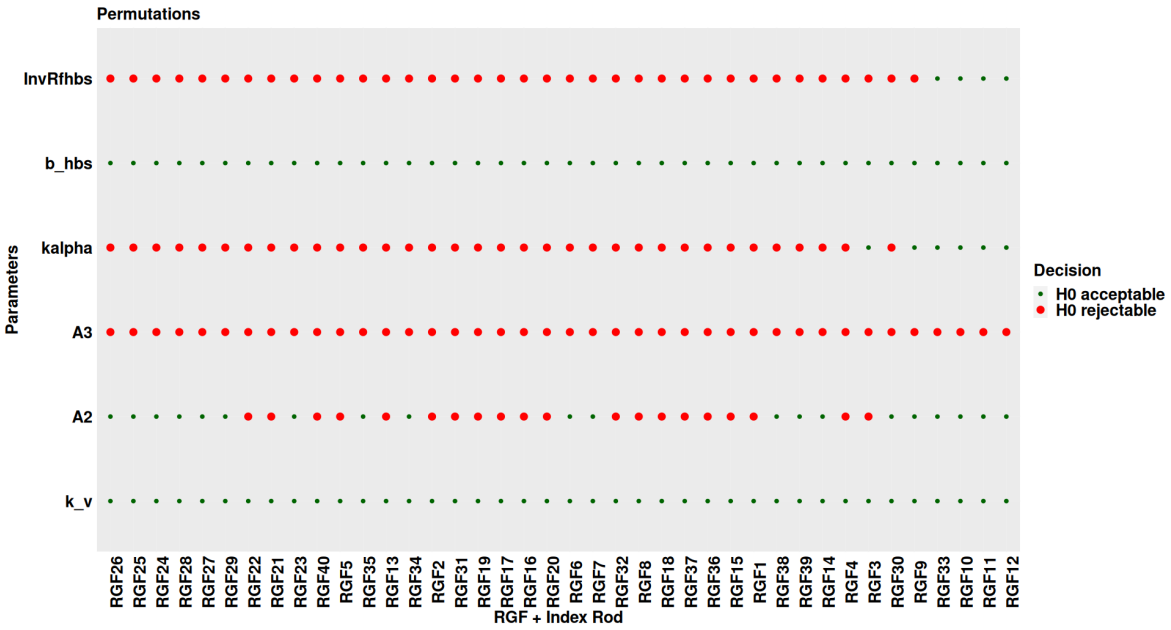


Figure 4.10: Detection of structure in residuals of **the linear model** constructed for each RGF fuel rod, based on each parameter, using an HSIC-based independence test integrated over the entire thermal conductivity uncertainty. Fuel rods are ordered in descending order of burnup.

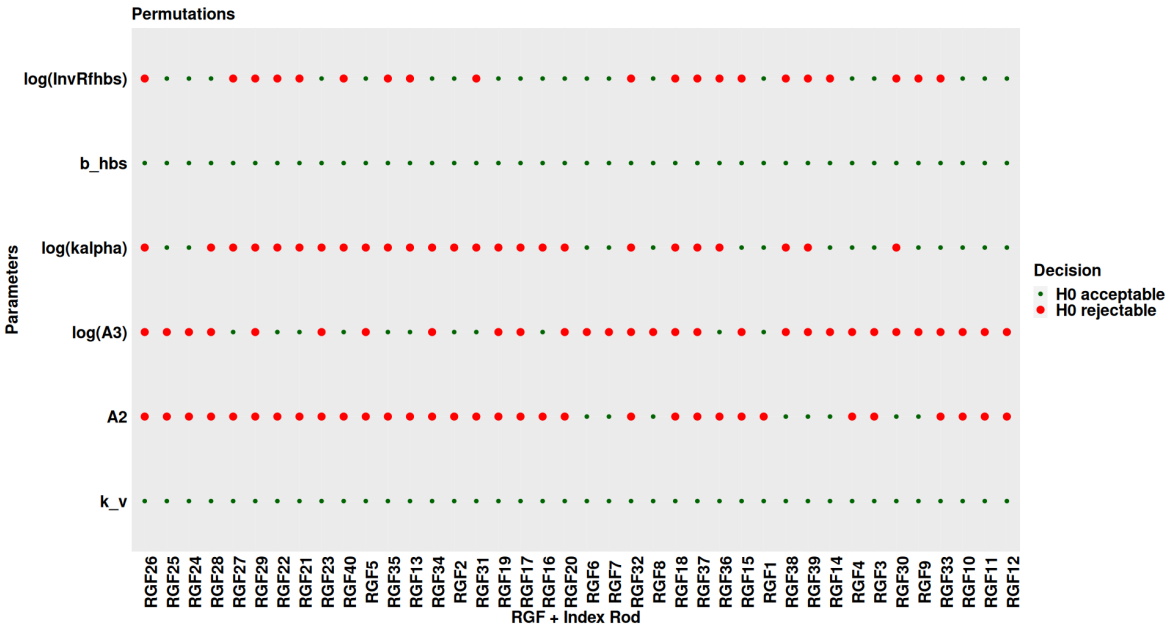


Figure 4.11: Detection of structure in residuals of **the log-linear model** constructed for each RGF fuel rod, based on each parameter, using an HSIC-based independence test integrated over the entire thermal conductivity uncertainty. Fuel rods are ordered in descending order of burnup.

of size n_{new}. The approximated posterior predictive distribution (associated with $\boldsymbol{\Theta}^{\text{new}}$) of the RGF for the i-th fuel rod associated with the thermal conductivity λ_k (Equation (4.53)) is given by:

$$\pi(Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i) \mid \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) \sim \mathcal{N}_{n_{\text{new}}}\left(\boldsymbol{\Theta}^{\text{new}}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i),\, \widehat{\sigma}^2_{i,\lambda_k} I_{n_{\text{new}}}\right). \tag{4.45}$$

From this posterior predictive distribution, we can compare the predictions of the log-linear model $\boldsymbol{\Theta}^{\text{new}}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)$ to those predictions of the numerical model $Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)$. To do this, for all $1 \leq k \leq m$, we can perform a Kolmogorov-Smirnov test to check the normality of the standardized residuals below:

$$\frac{Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)^{(l)} - (\boldsymbol{\Theta}^{\text{new}})^{(l)}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)}{\sqrt{\widehat{\sigma}^2_{i,\lambda_k}}} \sim \mathcal{N}(0,\, 1), \quad \text{for } 1 \leq l \leq n_{\text{new}}, \tag{4.46}$$

where $(\boldsymbol{\Theta}^{\text{new}})^{(l)}$ is a row vector of size $p+1$. We have a total of 200 simulations of the RGF for each fuel rod $\{i\}_{1\leq i\leq n_{\text{exp}}}$ for $m = 20$ different values of thermal conductivity, allowing us to construct $n_{\text{exp}} \times m = 40 \times 20 = 800$ log-linear models. However, to test the normality of the standardized residuals in Equation (4.46), we will need new simulations of the RGF.

A common practice in the statistical learning community is to split the available data into two sets: a training dataset, which is generally $\frac{2}{3}$ of the available data, and a test dataset, used to evaluate the predictive performance of the regression models constructed [Hastie et al., 2009]. In our case, we split the total available simulations into two equal parts of size $n = 100$ and $n_{\text{new}} = 100$, representing the training and test data, respectively. This choice was motivated by our wish to have a sufficient test dataset to evaluate the normality of the standardized residuals in Equation (4.46) and to calculate the predictive coefficient denoted by Q^2. This coefficient Q^2 is equivalent to the coefficient of determination R^2 (except the former is computed on a test data) and represents the proportion of variance explained by the regression model with respect to the total variance of $\boldsymbol{Y}_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)$ [Demay et al., 2022]. A higher Q^2 (close to 1) indicates a better fit of regression. For the i-th fuel rod and the thermal conductivity λ_k, the predictive coefficient is given by:

$$1 - \frac{\frac{1}{n_{\text{new}}}\sum_{l=1}^{n_{\text{new}}}\left(\boldsymbol{Y}_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)^{(l)} - \left(\widetilde{\boldsymbol{\Theta}}^{\text{new}}\right)^{(l)}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^2}{\frac{1}{n_{\text{new}}}\sum_{l=1}^{n_{\text{new}}}\left(\boldsymbol{Y}_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)^{(l)} - \frac{1}{n_{\text{new}}}\sum_{l'=1}^{n_{\text{new}}}\boldsymbol{Y}_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)^{(l')}\right)^2}. \tag{4.47}$$

In Figure 4.12, we present the results of the Kolmogorov-Smirnov test conducted on a test dataset of size $n_{\text{new}} = 100$ associated with the $m = 20$ values of thermal conductivity in order to check the normality of the standardized residuals in Equation (4.46) (Figure 4.12 (a)). We observe that the majority of the standardized residuals are Gaussian, except for λ_{15}. We have also calculated the predictive coefficient (Figure 4.12 (b)). We can see that all the Q^2 values are above 84%. In other words, the log-linear models explain more than 84% of the output variance, which is acceptable for the calibration process.

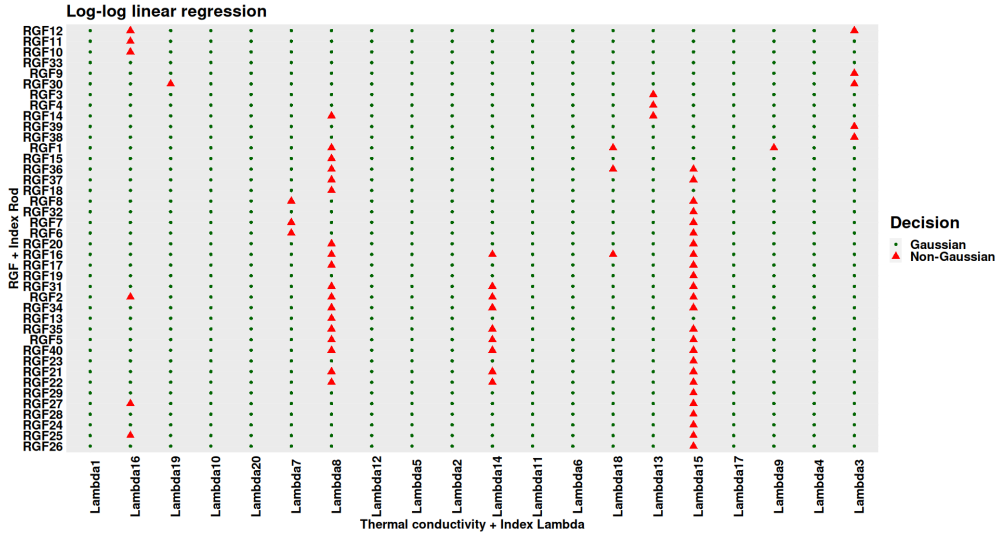
4.4.4 Examining the monotonicity of the log-linear model regressors

We are interested in the possible monotonicity of the regressors

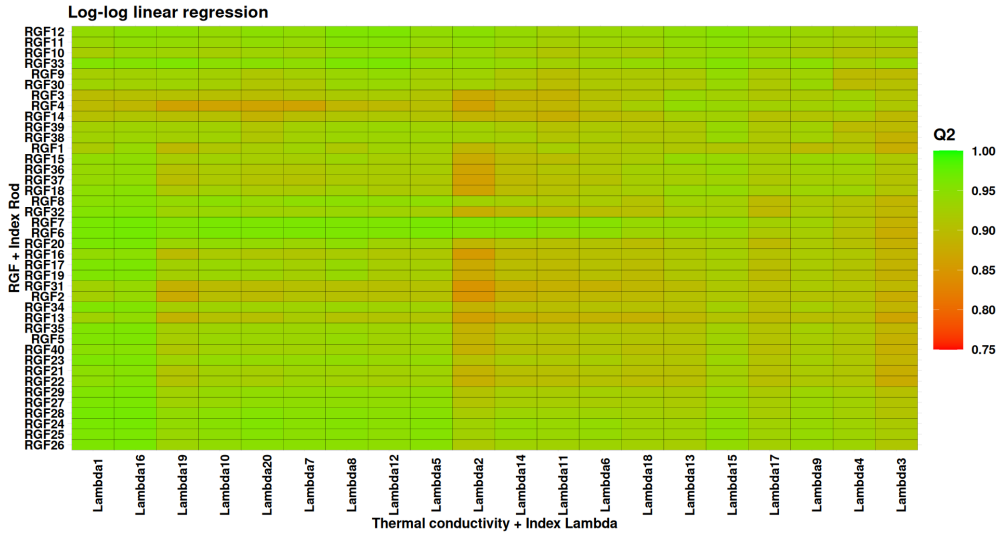
$$\{g_{\lambda_k,1}(\boldsymbol{x}_i)\}_{1\leq i\leq n_{\text{exp}},1\leq k\leq m}$$

resulting from a physical guess. These unknown regressors are estimated by the least-squares estimator:

$$\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) = \left(\boldsymbol{\Theta}_{\text{reg},k}^t\boldsymbol{\Theta}_{\text{reg},k}\right)^{-1}\boldsymbol{\Theta}_{\text{reg},k}^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) \quad \text{and} \quad \widehat{g}_{\lambda_k,1}(\boldsymbol{x}_i) = B\widehat{g}_{\lambda_k}(\boldsymbol{x}_i), \tag{4.48}$$



(a) Kolmogorov-Smirnov test conducted on a test dataset of size $n_{\text{new}} = 100$.



(b) Calculation of Q^2 on a test dataset of size $n_{\text{new}} = 100$.

Figure 4.12: (a) Inspection of the normality of the standardized residuals in Equation (4.46). (b) The predictive coefficient Q^2 of the $n \times m$ log-linear models constructed.

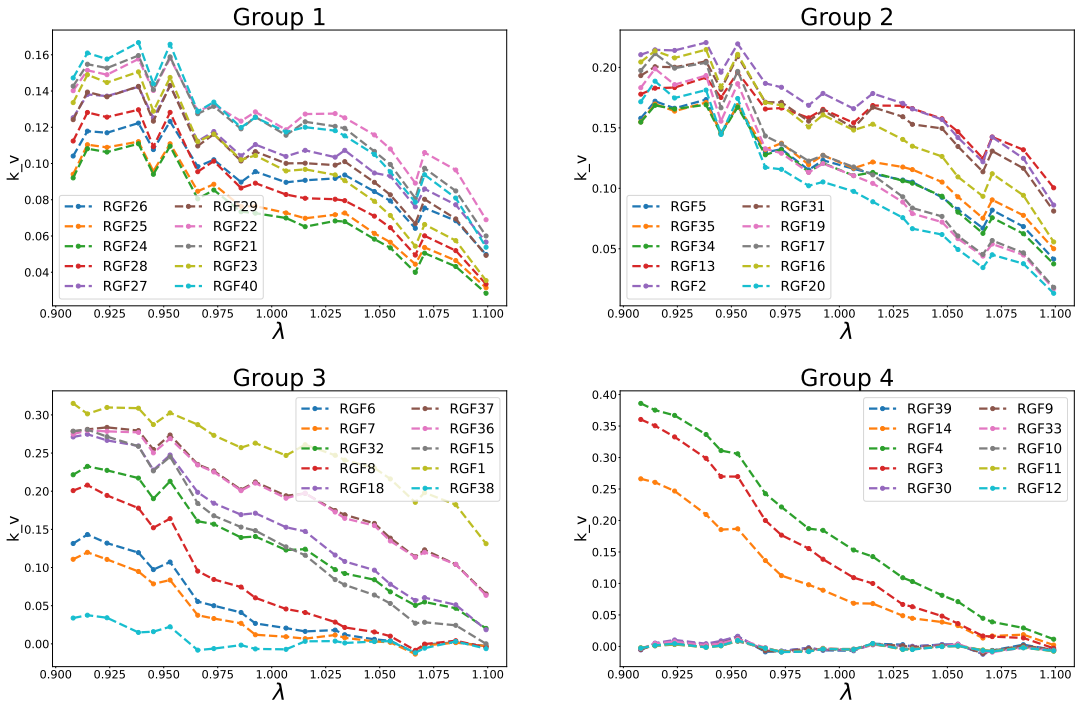


Figure 4.13: Plots of coefficient $g_{\lambda_k,1}(\boldsymbol{x}_i)$ associated with $\boldsymbol{\theta}_1$ estimated for $i = 1, \cdots, n = 40$ as a function of λ_k with $k = 1, \cdots, m = 20$.

with $B = \begin{pmatrix} \mathbf{0} & I_p \end{pmatrix} \in \mathbb{R}^{p\times(p+1)}$ is a matrix of size $p \times (p+1)$.

In Figures 4.13 to 4.18, we have plotted the $p = 6$ components of the regressors $g_{\lambda_k,1}(\boldsymbol{x}_i)$ as functions of the $m = 20$ values of thermal conductivity, only for the log-linear model. We observe monotonicity for the parameters $\boldsymbol{\theta}_2, \log(\boldsymbol{\theta}_4)$ and $\log(\boldsymbol{\theta}_6^{-1})$ (i.e., physical parameter A_2, log-physical parameter $\log(k_\alpha)$, and log-inverse physical parameter R_{fhbs} denoted by log(InvRhbs) respectively). Note that in these six plots, we have grouped the rods into 4 groups of ten based on decreasing combustion rates. Furthermore, we have also calculated the sensitivity indices of the linear regression models, equal to the square of the standardized regression coefficients denoted by SRC^2 [Iooss and Lemaître, 2015], for each $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)$ with respect to each component of $\boldsymbol{\theta}(\lambda)$. For all $1 \leq k \leq m$, the SRC^2 coefficient between the j-th parameter $\boldsymbol{\theta}_j(\lambda_k)$ and the i-th rod is given by:

$$\text{SRC}^2\left(\boldsymbol{\theta}_j(\lambda_k), \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)\right) = \left(\left(\widehat{g}_{\lambda_k,1}(\boldsymbol{x}_i)\right)_j\right)^2 \frac{\mathbb{V}(\boldsymbol{\theta}_j(\lambda_k))}{\mathbb{V}(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))},\ 1 \leq i \leq n_{\text{exp}},\ 1 \leq j \leq p. \tag{4.49}$$

The SRC^2 coefficient gives the proportion of the total variance of $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))$ explained by the j-th parameter $\boldsymbol{\theta}_j(\lambda_k)$. For $m = 20$ values of thermal conductivity λ, we have computed all the SRC^2 coefficients for all rods $\{i\}_{1\leq i\leq n_{\text{exp}}}$. The results are shown in Figure 4.19, where boxplots of these coefficients are presented for the 4 groups of RGF rods. We can see that the parameters contributing the most to the total variance of the majority of rods are, from the most to the less influential $\log(\boldsymbol{\theta}_6^{-1})$, $\boldsymbol{\theta}_2$, $\log(\boldsymbol{\theta}_4)$, $\log(\boldsymbol{\theta}_3)$, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_5$. We also stress that there is more variability in the SRC^2 coefficients, depending on λ, associated with $\boldsymbol{\theta}_2$, followed by $\log(\boldsymbol{\theta}_6^{-1})$, whereas the variabilities of the SRC^2 coefficients associated with $\log(\boldsymbol{\theta}_4)$ and $\log(\boldsymbol{\theta}_3)$ are less significant. For the others, there is little variability.

In the context of conditional calibration by the GP-LinCC method (see Section 4.5), it therefore seems reasonable to start by calibrating only the two parameters $\boldsymbol{\theta}_2$ and $\log(\boldsymbol{\theta}_6^{-1})$ while fixing the other parameters at their maximum a posteriori (MAP). These MAP values are computed by maximizing the posterior density on samples of the posterior distribution of equation (4.55) generated by a Gibbs algorithm using a specific value of λ (we choose

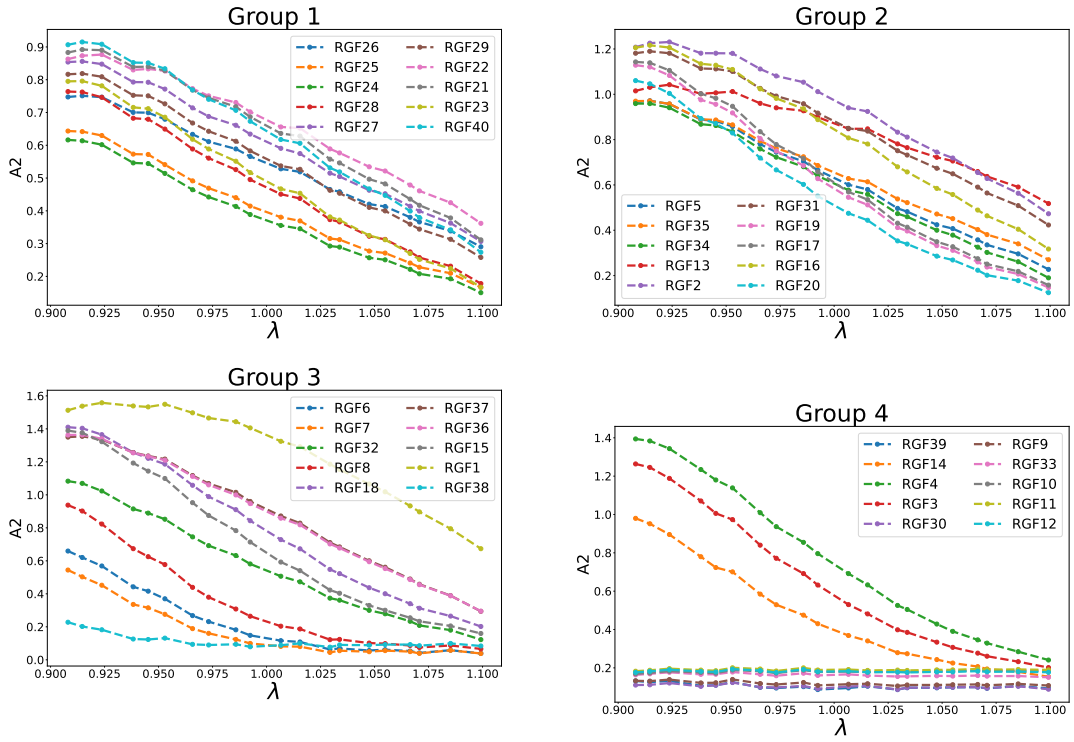


Figure 4.14: Plots of coefficient $g_{\lambda_k,1}(\boldsymbol{x}_i)$ associated with $\boldsymbol{\theta}_2$ estimated for $i = 1, \cdots, n = 40$ as a function of λ_k with $k = 1, \cdots, m = 20$.

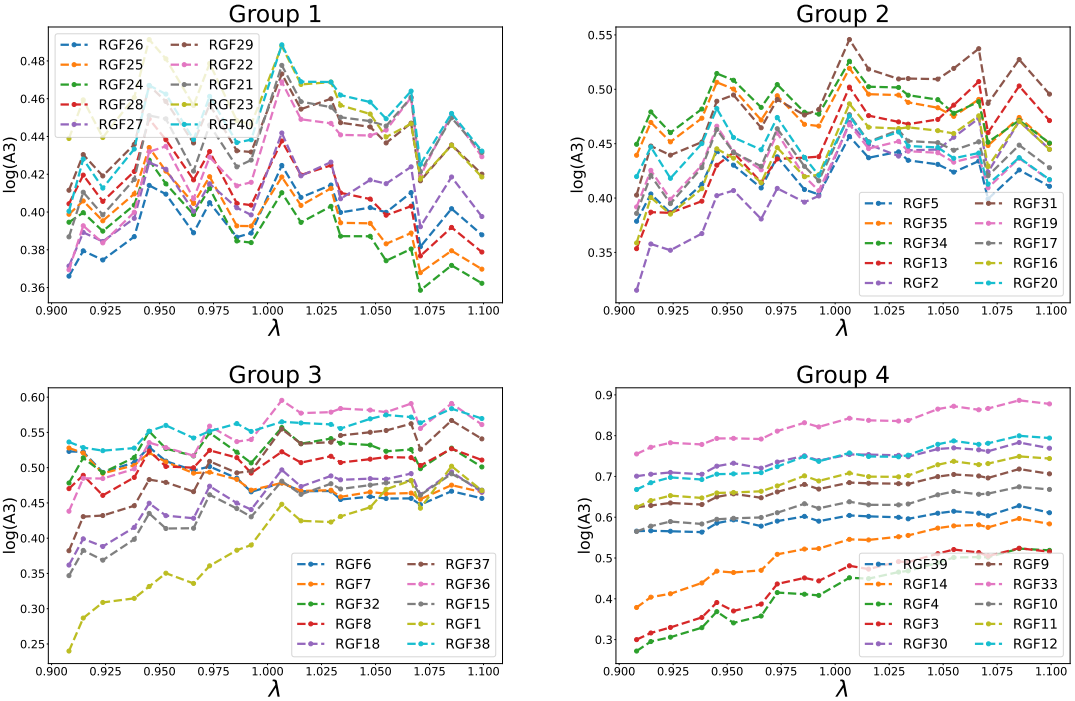


Figure 4.15: Plots of coefficient $g_{\lambda_k,1}(\boldsymbol{x}_i)$ associated with $\log(\boldsymbol{\theta}_3)$ estimated for $i = 1, \cdots, n = 40$ as a function of λ_k with $k = 1, \cdots, m = 20$.

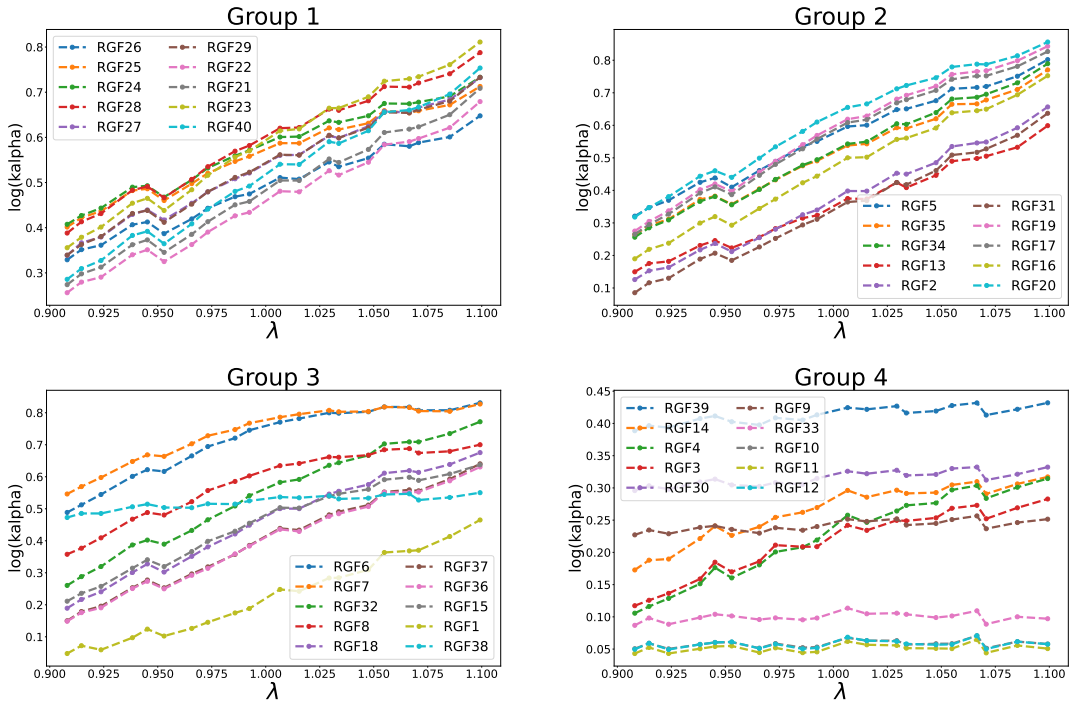


Figure 4.16: Plots of coefficient $g_{\lambda_k,1}(\boldsymbol{x}_i)$ associated with $\log(\boldsymbol{\theta}_4)$ estimated for $i = 1, \cdots, n = 40$ as a function of λ_k with $k = 1, \cdots, m = 20$.

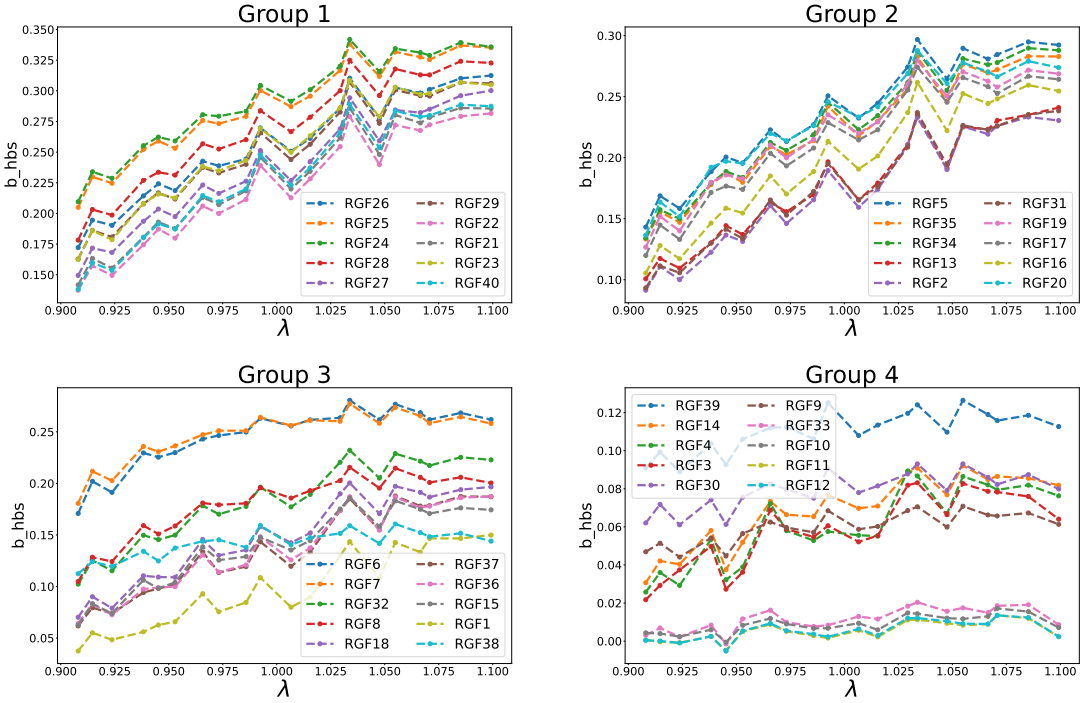


Figure 4.17: Plots of coefficient $g_{\lambda_k,1}(\boldsymbol{x}_i)$ associated with $\boldsymbol{\theta}_5$ estimated for $i = 1, \cdots, n = 40$ as a function of λ_k with $k = 1, \cdots, m = 20$.

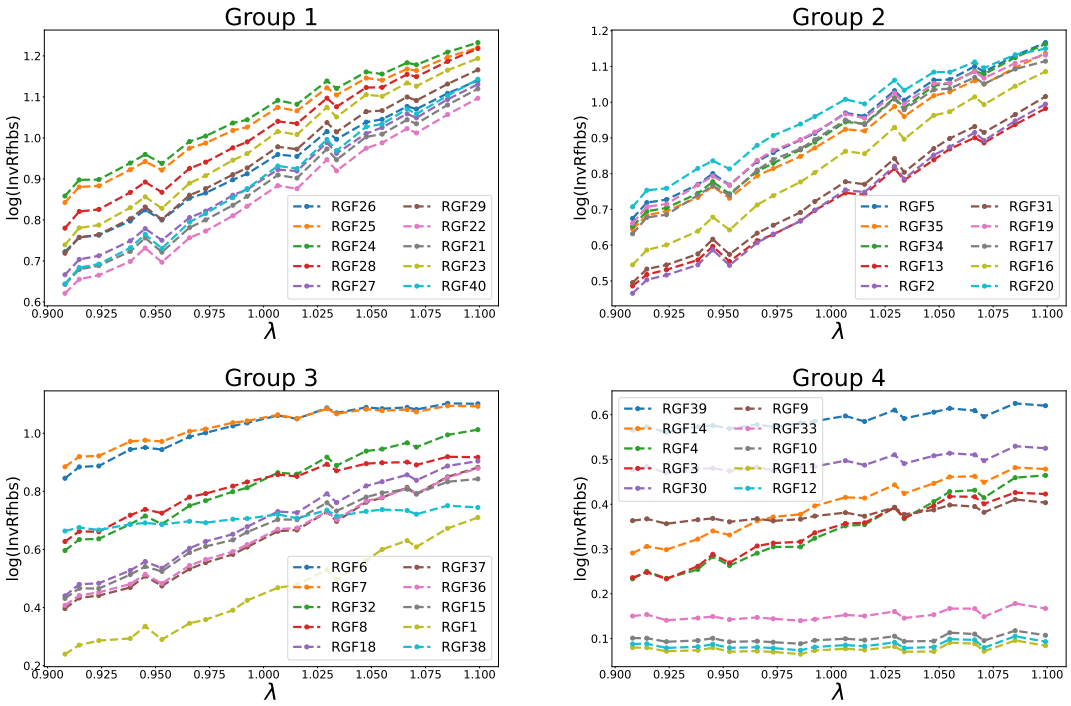


Figure 4.18: Plots of coefficient $g_{\lambda_k,1}(\boldsymbol{x}_i)$ associated with $\log(\boldsymbol{\theta}_6^{-1})$ estimated for $i = 1, \cdots, n = 40$ as a function of λ_k with $k = 1, \cdots, m = 20$.

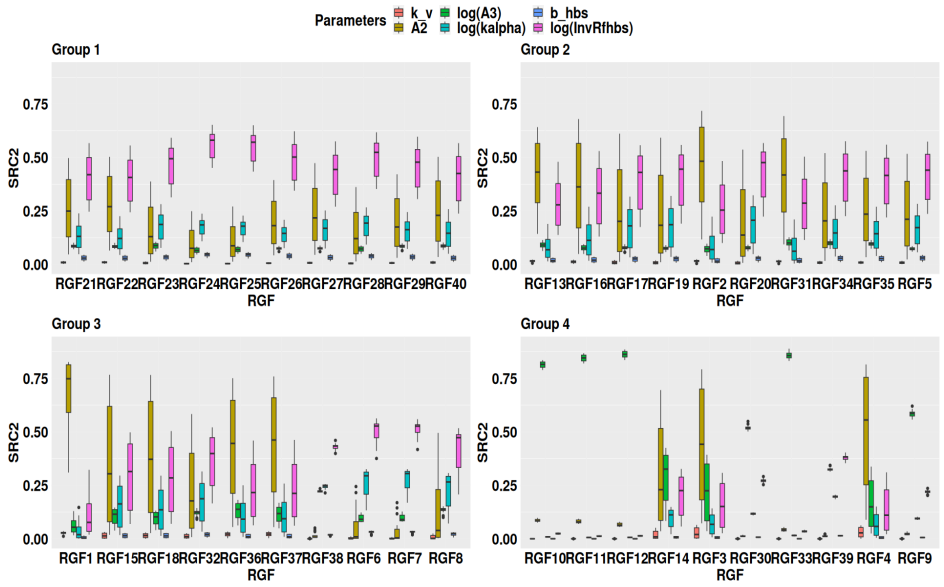


Figure 4.19: Boxplot of the SRC^2 of the 4 RGF groups against burnup.

$\lambda = 1.099$). The number of variables to be calibrated can be increased in a later stage by considering the four parameters: $\boldsymbol{\theta}_2$, $\log(\boldsymbol{\theta}_6^{-1})$, $\log(\boldsymbol{\theta}_4)$ and $\log(\boldsymbol{\theta}_3)$ with the other parameters still fixed at their MAP values. Both cases are dealt with in the following section.

4.5 GP-LinCC approach: application on ALCYONE test case

In the sequel, we will use the notation $\boldsymbol{\theta}(\lambda)$ instead of $\boldsymbol{\theta}$. This new notation implies that we will calibrate $\boldsymbol{\theta}$ given λ. Recall that the probabilistic equation connecting an experimental RGF $\boldsymbol{z}_i$ for the i-th fuel rod to a simulation of RGF $Y_\lambda(\boldsymbol{x}_i)$ associated with a realization λ of the thermal conductivity is given by:

$$\boldsymbol{z}_i = Y_\lambda(\boldsymbol{x}_i) + \epsilon_{\text{exp},i}, \tag{4.50}$$

where $\epsilon_{\text{exp},i} \sim \mathcal{N}(0, \sigma^2_{\epsilon_i})$ is the realization of the experimental uncertainty with $\sigma^2_{\epsilon_i}$ being known. Combining Equations (4.15) and (4.50), we obtain the probabilistic equation relating the experimental data $\boldsymbol{z}_i$ to the corresponding output of the linear model:

$$\boldsymbol{z}_i = g_{\lambda,0}(\boldsymbol{x}_i) + g_{\lambda,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda) + \epsilon_{i,\lambda} + \epsilon_{\text{exp},i}. \tag{4.51}$$

Furthermore, the Bayesian linear regression analysis carried out in Section 4.3 has yielded the following three results:

1. First, the boxplots of the Sobol' indices in Figure 4.5 show that the uncertainty of
$$\left\{g_\lambda(\boldsymbol{x}_i), \sigma^2_{i,\lambda}\right\}_{1\leq i\leq n_{\text{exp}}}$$
is negligible compared to that of $\boldsymbol{\theta}(\lambda)$. It is therefore acceptable to simplify the model by setting these parameters to their MAP
$$\left\{\widehat{g}_\lambda(\boldsymbol{x}_i), \widehat{\sigma}^2_{i,\lambda}\right\}_{1\leq i\leq n_{\text{exp}}}$$

2. Second, the posterior predictive distribution $\pi(Y^{\text{new}}_\lambda(\boldsymbol{x}_i)|\boldsymbol{Y}_\lambda(\boldsymbol{x}_i))$ for new inputs $\boldsymbol{\Theta}^{\text{new}}$ of size n_{new} is a multivariate Student distribution with $n - p - 1 = 193$ degrees of freedom. However, given the high number of degrees of freedom, this distribution can be approximated by a normal distribution. This approximation is valid since $n-p-2 = 192 > 30$.

Thus, for a new design $\boldsymbol{\Theta}^{\text{new}} = \begin{pmatrix}1 & \boldsymbol{\theta}(\lambda)^t\end{pmatrix}$, the associated predictive distribution is approximated as follows:

$$\pi\left(Y^{\text{new}}_\lambda(\boldsymbol{x}_i)|\boldsymbol{Y}_\lambda(\boldsymbol{x}_i)\right) \approx \mathcal{N}\Big(\widehat{g}_{\lambda,0}(\boldsymbol{x}_i)+\widehat{g}_{\lambda,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda), \widehat{\sigma}^2_{i,\lambda}\left(1 + \boldsymbol{\Theta}^{\text{new}}\left(\boldsymbol{\Theta}^t_{\text{reg},\lambda}\boldsymbol{\Theta}_{\text{reg},\lambda}\right)^{-1}(\boldsymbol{\Theta}^{\text{new}})^t\right)\Big), \tag{4.52}$$

where $\boldsymbol{\Theta}_{\text{reg},\lambda}$ is the numerical design associated with λ. In addition, the sensitivity analysis allows for a plug-in approximation of the predictive distribution (4.52), as follows:

$$\pi(Y^{\text{new}}_\lambda(\boldsymbol{x}_i)|\boldsymbol{Y}_\lambda(\boldsymbol{x}_i)) \approx \mathcal{N}\left(\widehat{g}_{\lambda,0}(\boldsymbol{x}_i) + \widehat{g}_{\lambda,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda), \widehat{\sigma}^2_{i,\lambda}\right). \tag{4.53}$$

Thus, an approximation of the likelihood equal to the probability of the data $\boldsymbol{z}$ given $\boldsymbol{\theta}(\lambda)$ can be written as:

$$\mathscr{L}(\boldsymbol{z}|\boldsymbol{\theta}(\lambda)) \approx \prod_{i=1}^{n_{\text{exp}}} \frac{1}{\sqrt{2\pi(\widehat{\sigma}^2_{i,\lambda} + \sigma^2_{\epsilon_i})}} \exp\left(-\frac{1}{2}\frac{\left(\boldsymbol{z}_i - \widehat{g}_{\lambda,0}(\boldsymbol{x}_i) - \widehat{g}_{\lambda,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}(\lambda)\right)^2}{\widehat{\sigma}^2_{i,\lambda} + \sigma^2_{\epsilon_i}}\right). \tag{4.54}$$

The corresponding posterior distribution $\pi(\boldsymbol{\theta}(\lambda)|\boldsymbol{z})$ is obtained using Bayes' formula. This posterior distribution is a truncated Gaussian distribution over $\mathscr{D}_{\boldsymbol{\Theta}}$:

$$\pi(\boldsymbol{\theta}(\lambda) \mid \boldsymbol{z}) \propto \mathscr{N}\left(\boldsymbol{\theta}(\lambda); \mathbb{E}\left[\boldsymbol{\theta}(\lambda)|\boldsymbol{z}\right], \Sigma_{\lambda}\right) \mathbf{1}_{\{\boldsymbol{\theta}(\lambda)\in\mathscr{D}_{\boldsymbol{\Theta}}\}}, \tag{4.55}$$

with a mean $\mathbb{E}\left[\boldsymbol{\theta}(\lambda)|\boldsymbol{z}\right]$ and a covariance matrix Σ_{λ} given respectively by:

$$\mathbb{E}\left[\boldsymbol{\theta}(\lambda)|\boldsymbol{z}\right] = \Sigma_{\lambda}\widehat{g}_{\lambda,1}(\boldsymbol{x})^{t}\Sigma_{\epsilon,\lambda}^{-1}\left(\boldsymbol{z} - \widehat{g}_{\lambda,0}(\boldsymbol{x})\right), \tag{4.56}$$

$$\Sigma_{\lambda} = \left(\widehat{g}_{\lambda,1}(\boldsymbol{x})^{t}\Sigma_{\epsilon,\lambda}^{-1}\widehat{g}_{\lambda,1}(\boldsymbol{x})\right)^{-1}, \tag{4.57}$$

with:

$$\Sigma_{\epsilon,\lambda} = \operatorname{diag}\left(\widehat{\sigma}_{1,\lambda}^{2} + \sigma_{\epsilon_1}^{2}, \cdots, \widehat{\sigma}_{n_{\exp},\lambda}^{2} + \sigma_{\epsilon_{n_{\exp}}}^{2}\right) \in \mathbb{R}^{n_{\exp}\times n_{\exp}}, \tag{4.58}$$

$$\widehat{g}_{\lambda,1}(\boldsymbol{x}) = \left[\widehat{g}_{\lambda,1}(\boldsymbol{x}_1), \cdots, \widehat{g}_{\lambda,1}(\boldsymbol{x}_{n_{\exp}})\right] \in \mathbb{R}^{n_{\exp}\times p}, \tag{4.59}$$

$$\widehat{g}_{\lambda,0}(\boldsymbol{x}) = \left(\widehat{g}_{\lambda,0}(\boldsymbol{x}_1), \cdots, \widehat{g}_{\lambda,0}(\boldsymbol{x}_{n_{\exp}})\right)^{t} \in \mathbb{R}^{n_{\exp}}. \tag{4.60}$$

Finally, the estimation of the posterior distribution of $\boldsymbol{\theta}(\lambda)$ can be efficiently carry out using approximation methods. This truncated Gaussian distribution can in fact be approximated by Gibbs sampling [Robert, 1995] to eventually derive a MAP estimator of $\boldsymbol{\theta}(\lambda)$. However, this approach has two drawbacks. On the one hand, it is computationally expensive, requiring us to run as many independent Markov chains as the number of thermal conductivity values λ of interest. Furthermore, for each considered value of λ, a preliminary step of linearization of the CARACAS code is required, relying on a number n of simulations. As the CARACAS code is computationally expensive, this can be problematic when one needs to estimate $\boldsymbol{\theta}(\lambda)$ for numerous λ values. On the other hand, this approach does not take into account the fact that, under certain regularity conditions, the posterior distribution of $\boldsymbol{\theta}(\lambda)$ could provide information about that of $\boldsymbol{\theta}(\lambda')$ if λ is not too far from λ'.

An alternative is to perform the scheme (linearization, sensitivity analysis, followed by calibration) while considering only a limited number m of well-chosen values $\{\lambda_k\}_{1\leq k\leq m}$ of λ. Using these selected values, we can build a regression model to derive a posterior predictive distribution of $\boldsymbol{\theta}(\lambda)$ for any other values of λ. More precisely, we can provide a truncated Gaussian predictive posterior distribution (over $\mathscr{D}_{\boldsymbol{\Theta}}$) of $\boldsymbol{\theta}(\lambda)$ for any value of λ without rerunning the CARACAS code for that λ. This approach, called GP-LinCC for Gaussian Process and Linearization-based Conditional Calibration has been developed in Chapter 2 of the present manuscript. Note that, due to the parameter bound constraints on $\boldsymbol{\theta}$, the Gaussian predictive distribution provided by the GP-LinCC method transforms into a truncated Gaussian distribution over $\mathscr{D}_{\boldsymbol{\Theta}}$.

We applied the GP-LinCC approach only to the log-linear models constructed for the purpose. First, we estimated the hyperparameters ϕ of the prior of the Gaussian processes of the GP-LinCC method over the full set of $m = 20$ thermal conductivity values. Let $\widehat{\phi}$ denote these estimated hyperparameters (given in Equation (2.31)). Then, we compared the predictive distribution of GP-LinCC with the posterior distribution of Equation (4.55) (called target distribution) for some thermal conductivity values. After min-max scaling, both the target and predictive distributions were truncated on $[0,1]^p$.

We recall that the GP-LinCC method (see Chapter 2) provides a Gaussian predictive distribution of

$$\log(\widetilde{\boldsymbol{\theta}}(\lambda)) \quad \text{with} \quad \widetilde{\boldsymbol{\theta}}(\lambda) = \left(\exp(\boldsymbol{\theta}_1(\lambda)), \exp(\boldsymbol{\theta}_2(\lambda)), \boldsymbol{\theta}_3(\lambda), \boldsymbol{\theta}_4(\lambda), \exp(\boldsymbol{\theta}_5(\lambda)), \boldsymbol{\theta}_6^{-1}(\lambda)\right)^{t}.$$

which is transformed into a truncated multivariate Gaussian on $[0,1]^p$[2], easily simulated by a Gibbs algorithm [Robert et al., 1999]. **An estimate of the posterior distribution of $\boldsymbol{\theta}$ is**

[2] Note that this truncation is obtained by truncation of the Gaussian predictive posterior distribution from Theorem 2 in Chapter 2.

then computed by applying all the necessary inverse transformations (inverse cumulative distribution function of uniform distribution, exponential function (for $\log(\boldsymbol{\theta}_3), \log(\boldsymbol{\theta}_4)$ **and** $\log(\boldsymbol{\theta}_6^{-1})$**)) on the posterior samples of** $\log(\widetilde{\boldsymbol{\theta}}(\lambda))$ **and finally applying inverse funtion on the posterior samples associated with** $\boldsymbol{\theta}_6^{-1}$ **for each** $\lambda \in \{\lambda_1, \cdots, \lambda_m\}$. We also did the same for the samples of the target distribution. Combining those transformations with the leave-one-out (LOO) cross-validation procedure on λ_k for $1 \leq k \leq m$ [Bachoc, 2013], we tested the predictive ability of the GP-LinCC approach in the two following situations:

1. calibration of $\boldsymbol{\theta} := (\mathrm{A}_2, \mathrm{R}_{\mathrm{fhbs}})^t$
2. calibration of $\boldsymbol{\theta} := (\mathrm{A}_2, \mathrm{A}_3, \mathrm{k}_\alpha, \mathrm{R}_{\mathrm{fhbs}})^t$.

4.5.1 Calibration of $\boldsymbol{\theta} := (\mathbf{A}_2, \mathbf{R}_{\mathbf{fhbs}})^t$

Figure 4.20 presents the outcomes of the previous procedure implemented for each

$$\lambda \in \{\lambda_1, \cdots, \lambda_m\}.$$

In general, we observe a close alignment between the predictive and target posterior means for every λ value. Nevertheless, the GP-LinCC method tends to exhibit an underestimation of the prediction uncertainty, as shown by the fact that the predicted variances are lower than the corresponding target posterior variances. This underestimation can partly be attributed to the use of marginal likelihood maximization technique (also referred to as empirical Bayes) for estimating the hyperparameters ϕ [Rasmussen et al., 2006; Reich and Ghosh, 2019] and/or the choice of the covariance function for each component of

$$\log(\widetilde{\boldsymbol{\theta}}(\lambda)) \quad \text{with} \quad \widetilde{\boldsymbol{\theta}}(\lambda) := \left(\exp(\boldsymbol{\theta}_1), \boldsymbol{\theta}_6^{-1}\right)^t \in \mathbb{R}^2.$$

In this study, the Matérn 5/2 covariance function was used which is the most popular choice in practice [Gu et al., 2018].

Physical interpretation of the results in Figure 4.20

To some extent, the results obtained are consistent with the physics modeled in the ALCYONE application and the fission gas behavior model. In fact, most of the physical phenomena related to the fission gas are activated by the temperature T. This is the case for diffusion. In the model and in the operating regime considered in this study, the gas diffusion coefficient in UO_2 can be written as follows:

$$D(\dot{F}, T) \sim A_2 \sqrt{\dot{F}} e^{-\frac{E_2}{T}} + A_3 \dot{F}, \tag{4.61}$$

where

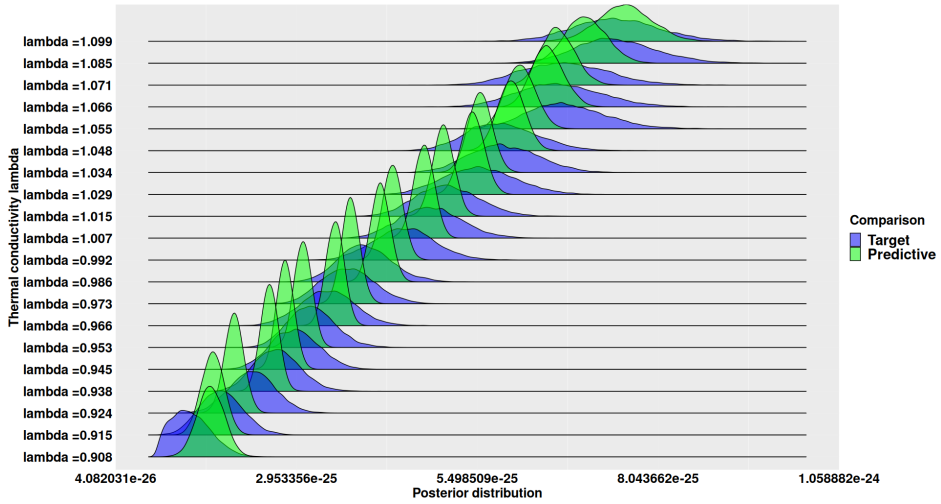
- $\dot{F}$ is the fission rate (fission/m^3/s),
- E_2 is an activation energy.

The fission rate $\dot{F}$ is independent of the temperature T. The first term is a thermal diffusion coefficient (Arrhenius-type law) accelerated by irradiation and the second is an athermal diffusion coefficient related only to irradiation. In the diffusion equation, the gas flow out of the grains (of radius R_{grain}) by diffusion is proportional to this diffusion coefficient:

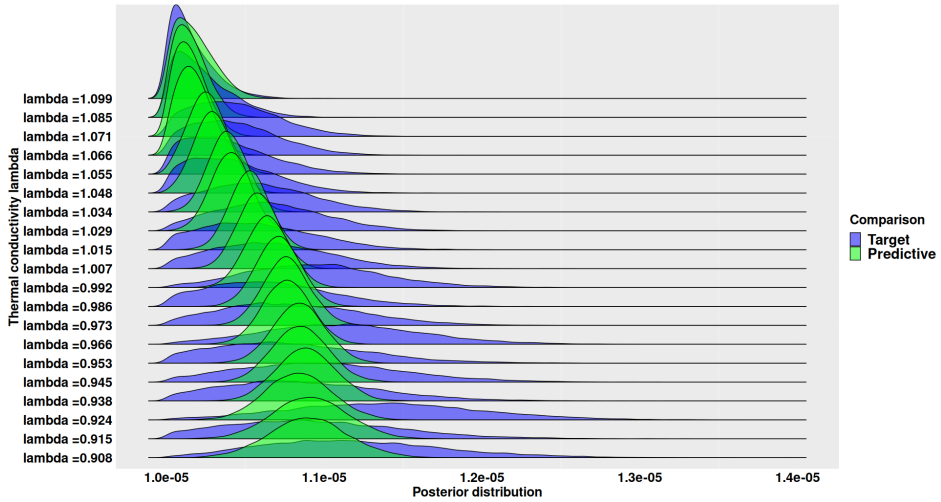
$$\Phi = 4\pi R_{\mathrm{grain}}^2 D(\dot{F}, T) \|\nabla c(r)\|_{r=R_{\mathrm{grain}}},$$

where $c(r)$ is the fission gas concentration profile in the grain (obeys Fick law).

Regardless of λ, for a given fuel rod, we aim to calibrate the parameters $\boldsymbol{\theta}$ to calculate the same value for the released gas fraction. By greatly simplifying the representation, we need to



(a) Posterior distributions (target and predictive) relative to the parameter A_2.



(b) Posterior distributions (target and predictive) relative to the parameter R_{fhbs}.

Figure 4.20: Comparison of the posterior distributions predictive (provided by the GP-LinC + all inverse transformations) and target (given in Equation (4.55) with $p = 2$) of the couple (A_2, R_{fhbs}) obtained by LOO on the $m = 20$ values of λ.

ensure that for any λ, the flow of gas leaving the grain is the same. Given the mathematical form of D, when the temperature T is high, corresponding to low values of λ, the value of A_2 after calibration must be lower than that obtained when the temperature is lower (high λ case). This is shown in Figure 4.20(a) where the posterior distribution of A_2 increases as a function of the thermal conductivity λ.

At high burnup rates, in the peripheral zones of the pellet, the microstructure of the nuclear fuel is restructured: the initial metallographic structure (polycrystalline microstructure with grain sizes of around 10µm) is transformed into a microstructure with very small grains (of the order of 0.5µm) known as the High Burnup Structure (HBS). In the fission gas behavior model, after this restructuring, the diffusion equation is no longer solved at the scale of a metallographic grain, but at the scale of a fuel fragment, represented in the form of a sphere of radius R_{fhbs} and comprising several micro-grains. This equation uses the same formulation of the diffusion coefficient D presented above, possibly accelerated by a constant coefficient. The parameter R_{fhbs} therefore corresponds to the distance that the fission gas atoms must travel to leave the fuel and be released. At the periphery of the pellet, given the local temperature/fission rate regime, the diffusion coefficient is practically independent of temperature. It depends only on the fission rate $\dot{F}$ which is independent of λ:

$$D(\dot{F}, T) \sim A_3 \dot{F}$$

The outgoing gas flow from the sphere of radius R_{fhbs} by diffusion is given by:

$$\Phi = 4\pi R_{fhbs}^2 D(\dot{F}, T) \|\nabla c(r)\|_{r=R_{fhbs}}.$$

Then, we have:

$$\Phi \sim R_{fhbs}^2 D.$$

Since D does not depend on λ, for the outgoing gas flow from the sphere not to depend on λ, the parameter R_{fhbs} must also vary very little with λ. This happens, as we can see in Figure 4.20 (b).

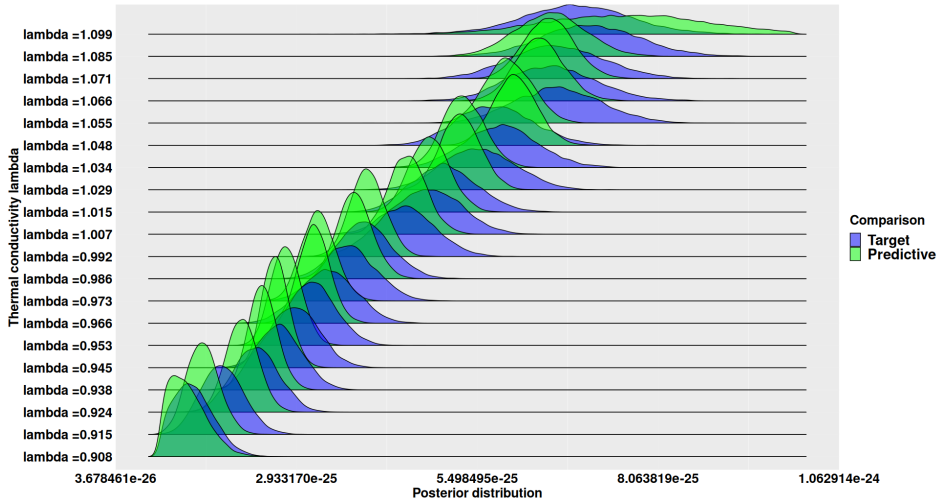
4.5.2 Calibration of $\boldsymbol{\theta} := (\mathbf{A}_2, \mathbf{A}_3, \mathbf{k}_\alpha, \mathbf{R}_{\mathbf{fhbs}})^t$

Using the same procedure that gave us the results shown in Figure 4.20 (see Section 4.5), Figures 4.21 and 4.22 present the results obtained for this 4D calibration problem. For the predictive posterior distributions of the parameters A_3, k_α and R_{fhbs}, more questions arise. Indeed, for all $m = 20$ values of λ, these three marginal distributions are concentrated at the lower bounds of the prior variation ranges, which are $[10^{-40}, 10^{-39}]$, $[0.8, 1.2]$ and $[10^{-5}, 3 \times 10^{-5}]$, respectively. This trend raises the possibility that the limited performance of the GP-LinCC method in learning these two components (i.e., A_3 and k_α) of

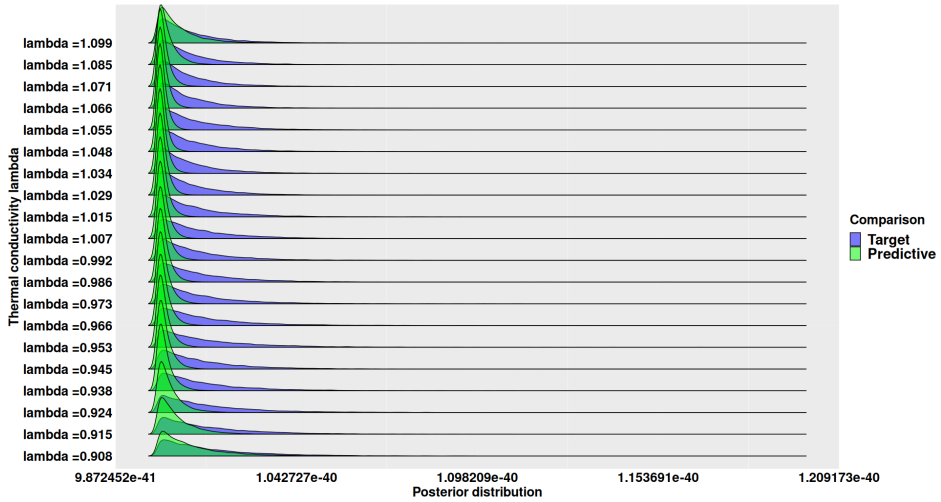
$$\boldsymbol{\theta}(\lambda) = (\boldsymbol{\theta}_1(\lambda), \boldsymbol{\theta}_3(\lambda), \boldsymbol{\theta}_4(\lambda), \boldsymbol{\theta}_6(\lambda))^t$$

could be attributed to a poor estimation of the hyperparameters ϕ. This is probably due to the specific shape of the posterior means of the marginal distributions associated with $\boldsymbol{\theta}_3(\lambda)$ and $\boldsymbol{\theta}_4(\lambda)$. The poor estimation of these hyperparameters might have been propagated to the fourth component.

A possible solution would be to follow the approach in Subsection 4.5.1, setting $\log(\boldsymbol{\theta}_4)$ and $\log(\boldsymbol{\theta}_3)$ to their MAP and to calibrate only the other two parameters or to extend the prior lower and upper bounds of these parameters. Another solution would be to improve the GP-LinCC method by, for example, replacing the Gaussian process prior assumption for $\log(\boldsymbol{\theta}_4)$ and $\log(\boldsymbol{\theta}_3)$ with linear statistical models whose coefficients could be jointly estimated with the hyperparameters ϕ of the remaining Gaussian processes.

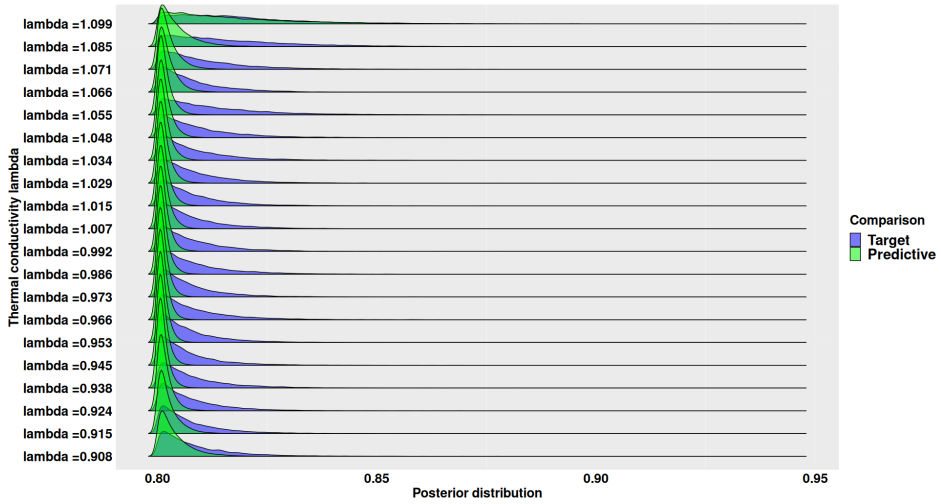


(a) Posterior distributions (target and predictive) relative to the parameter A_2.

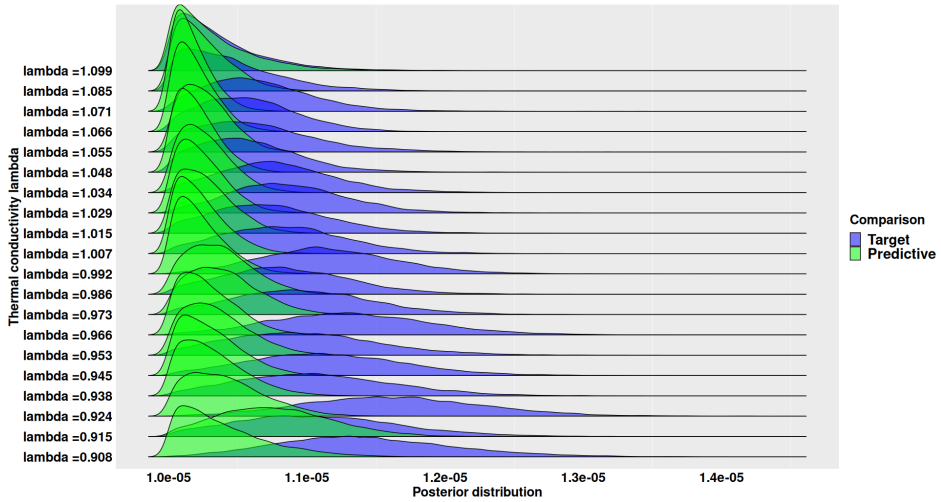


(b) Posterior distributions (target and predictive) relative to the parameter A_3.

Figure 4.21: Comparison of the marginal posterior distributions predictive (provides by the GP-LinC + all inverse transformations) and target (given in Equation (4.55) with $p = 4$) of the couple $(A_2, A_3, k_\alpha, R_{\text{fhbs}})$ obtained by LOO on the 20 values of λ.



(a) Posterior distributions (target and predictive) relative to the parameter k_α.



(b) Posterior distributions (target and predictive) relative to the parameter R_{fhbs}.

Figure 4.22: Comparison of the marginal posterior distributions predictive (given in Theorem 2 with $p = 4$) and target (given in Equation (4.55) with $p = 4$) of the couple $(A_2, A_3, k_\alpha, R_{fhbs})$ obtained by LOO on the 20 values of λ.

Physical interpretation of the results in Figures 4.21 and 4.22

The physical analysis of the results obtained for the 2D case remains relevant (Figure 4.21 (a) and Figure 4.22 (b)), and can even be extended to the parameters A_3 (i.e., $\boldsymbol{\theta}_3$) and k_α (i.e., $\boldsymbol{\theta}_4$). In the equation describing the fission gas diffusion coefficient, the second temperature independent term is directly related to the gas release from the peripheral zones of the pellet, before and after the HBS restructuring described above. Since the fission rate $\dot{F}$ is not temperature dependent, the posterior distribution of A_3 should be relatively insensitive to λ, as can be seen in Figure 4.21 (b). Similarly, the parameter k_α controls the kinetics of irradiation defect production. When the number of defects exceeds a certain threshold, HBS restructuring is triggered. In high temperature zones (center of the pellet), defect production is ineffective because the irradiation defects are annealed by the temperature. The parameter k_α therefore only plays a role in the cold zones of the pellet (periphery) and its posterior distribution k_α should be very little dependent on λ, as shown in Figure 4.22 (a).

From a statistical point of view, in both 2D and 4D cases, only one of the four parameters identified in the HSIC sensitivity analysis depends on λ after calibration (Figures 4.20, 4.21 and 4.22). However, the dispersion of the SRC^2 associated with the parameter R_{fhbs} (i.e., $\boldsymbol{\theta}_6$) for different values of λ and for all the fuel rods suggested a more global dependence of the parameters on λ. Broadly speaking, the RGF in the CARACAS code comprises two contributions:

- A temperature-activated contribution whose identified influential parameter is A_2 (i.e., $\boldsymbol{\theta}_2$).
- An athermal contribution involving A_3, k_α and R_{fhbs} (i.e., $\boldsymbol{\theta}_3$, $\boldsymbol{\theta}_4$ and $\boldsymbol{\theta}_6$ respectively).

The part played by these two contributions to the total RGF is different for each of the fuel rods in the calibration database. From a physical point of view, there was nothing to suggest a priori that the athermal component would be virtually independent of λ. It was to be expected that the share of each contribution would depend on λ.

4.5.3 Comparison between the RGF predicted by the CARACAS code and the log-linear-based calibrated model predictions

For any new value of the thermal conductivity $\lambda^\star$ and for the i-th fuel rod, the predictive distribution of the log-RGF at the MAP, denoted by $\widehat{\log \widetilde{\boldsymbol{\theta}}}_{\text{pred}}(\lambda^\star)$ is approximately a Gaussian distribution:

$$\pi\left(\log y_{\lambda^\star}(\boldsymbol{x}_i) \mid \log \boldsymbol{Y}_{\lambda^\star}(\boldsymbol{x}_i)\right) \approx \mathscr{N}\left(\log y^{\text{pred}}_{\widehat{\log \widetilde{\boldsymbol{\theta}}}_{\text{pred}}(\lambda^\star)}(\boldsymbol{x}_i),\, \widehat{\sigma}^2_{i,\lambda^\star}\right), \tag{4.62}$$

with

$$\log y^{\text{pred}}_{\widehat{\log \widetilde{\boldsymbol{\theta}}}_{\text{pred}}(\lambda^\star)}(\boldsymbol{x}_i) := \widehat{g}_{\lambda^\star,0}(\boldsymbol{x}_i) + \widehat{g}_{\lambda^\star,1}(\boldsymbol{x}_i)^t \widehat{\log \widetilde{\boldsymbol{\theta}}}_{\text{pred}}(\lambda^\star). \tag{4.63}$$

To simplify the equations in the rest of the chapter, we will use $\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)$ which is defined by

$$\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star) := \widehat{\log \widetilde{\boldsymbol{\theta}}}_{\text{pred}}(\lambda^\star)$$

instead of $\widehat{\log \widetilde{\boldsymbol{\theta}}}_{\text{pred}}(\lambda^\star)$. By applying the exponential transformation to $\log y_{\lambda^\star}(\boldsymbol{x}_i)$, then the predicted RGF obtained follows approximately the following log-normal distribution:

$$\mathscr{LN}\left(\log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}_i),\, \widehat{\sigma}^2_{i,\lambda^\star}\right) \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}} = 40. \tag{4.64}$$

The median of the above log-normal distribution is given by:

$$y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}_i) := \exp\left(\log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}_i)\right) \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}. \tag{4.65}$$

Thus, we can compare the simulations of CARACAS code run at the MAP

$$\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star) := \exp\left(\widehat{\log \widetilde{\boldsymbol{\theta}}}_{\text{pred}}(\lambda^\star)\right)^{3}$$

denoted by

$$\left\{ y_{\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)}(\boldsymbol{x}_i) \right\}_{1\leq i\leq n_{\text{exp}}},$$

with the predictions

$$\left\{ y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}_i) \right\}_{1\leq i\leq n_{\text{exp}}}, \tag{4.66}$$

of the exponential model (Equation (4.65)). The associated prediction interval of 95% is obtained by applying the exponential transformation of the prediction interval of 95% to the Gaussian distribution in Equation (4.62):

$$IF_{95\%}(y_{\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)}(\boldsymbol{x}_i)) := \left[\exp\left(\log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}_i) \pm 1.96\sqrt{\widehat{\sigma}^2_{i,\lambda^\star}}\right)\right] \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}. \tag{4.67}$$

Figure 4.23 presents in (a) the scatter plots between the log-RGF predicted by the CARACAS model run at $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and reference value of $\boldsymbol{\theta}$ denoted by $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$[4] and the predictions of the log-linear model (Equation (4.63)) computed with

$$\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star) \quad \text{and} \quad \log\left(\widetilde{\boldsymbol{\theta}}_{\text{ref}}(\lambda^\star)\right),$$

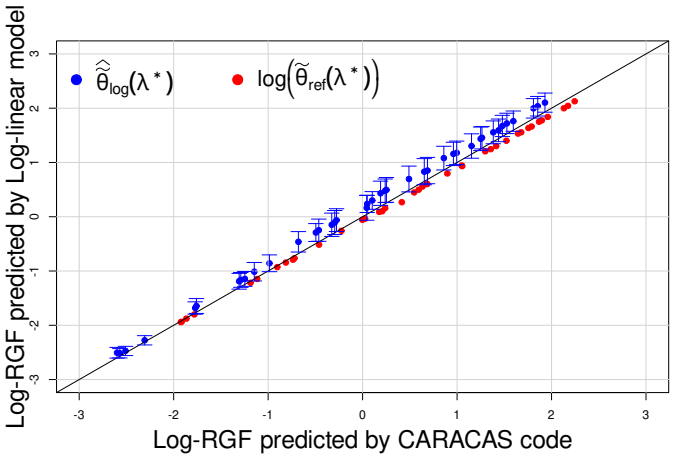
with

$$\widetilde{\boldsymbol{\theta}}_{\text{ref}}(\lambda^\star) = \left(\exp(\boldsymbol{\theta}_{\text{ref},1}(\lambda^\star)), \boldsymbol{\theta}_{\text{ref},3}(\lambda^\star), \boldsymbol{\theta}_{\text{ref},4}(\lambda^\star), \boldsymbol{\theta}^{-1}_{\text{ref},6}(\lambda^\star)\right)^t \quad \text{and} \quad \lambda^\star = \lambda_{\text{nom}} = 1.$$

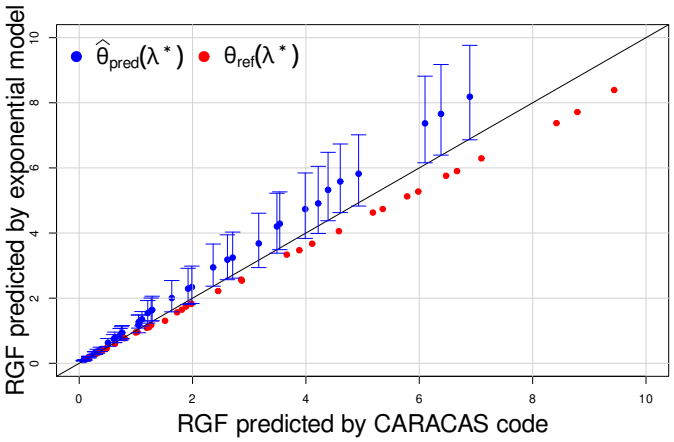
respectively and in (b) the exponential model predictions (Equation (4.65)) and RGF predicted by CARACAS code. In this way, we have also plotted the scatter plot between the reference values and the CARACAS code predictions. We can see that the log-linear model systematically overpredicts the predictions of the CARACAS code. This becomes increasingly observable when the RGF is high due to the different scales of the RGF from 6×10^{-2} to 10. In other words, this discrepancy is positively correlated with the burnup of each fuel rod. This could be due to the fact that the log-linear model would tend to overpredict the log-RGF simulated in the MAP zone, and/or the absence of a model discrepancy that would be due to the fuel rod burnups and thus the thermal power $P(t)$. The latter was considered in this study as a deterministic quantity for each fuel rod. Note that the linearization uncertainty, whose amplitude is estimated by the $\widehat{\sigma}^2_{i,\lambda^\star}$, is compatible with these figures (even if the bisector crosses the ends of the blue intervals).

[3] Note that $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ represents the MAP of $\boldsymbol{\theta}$ in $\mathscr{D}_{\boldsymbol{\Theta}}$ after applying all the necessary inverse transformations (inverse cumulative distribution function of uniform distribution, inverse and exponential functions).

[4] $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$ corresponds to the reference parameter values of the CARCAS code associated with $\lambda^\star = 1$ before using the GP-LInCC calibration method.



(a) Comparison between the log-RGF predicted by Equation (4.63) (with $\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)$ and $\log(\widetilde{\boldsymbol{\theta}}_{\text{ref}}(\lambda^\star))$ respectively) and the CARACAS code at $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$. Error bar is the prediction interval at 95% with $\lambda^\star = 1$.



(b) Comparison between the RGF predicted by Equation (4.65) and CARACAS code at $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$. Error bar is the prediction interval at 95% (Equation (4.67)).

Figure 4.23: Comparison of (a) the log-RGF (b) the RGF predicted with $\lambda^\star = \lambda_{\text{nom}} = 1$.

4.5.4 Comparison between the experimental RGF and the log-linear-based calibrated model predictions

For any new value of the thermal conductivity $\lambda^\star$ and for the i-th fuel rod, the predictive distribution of the experimental log-RGF at $\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)$ of $\log(\widetilde{\boldsymbol{\theta}}(\lambda^\star))$ is approximately a Gaussian distribution:

$$\pi(\log \boldsymbol{z}_{\lambda^\star}^{\text{pred}}(\boldsymbol{x}_i)|\log \boldsymbol{z}) \approx \mathscr{N}\left(\log y^{\text{pred}}_{\underset{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}{\approx}}(\boldsymbol{x}_i),\, \widehat{\sigma}^2_{i,\lambda^\star} + \sigma^2_{\text{exp},i}\right) \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}. \tag{4.68}$$

By applying the exponential transformation to $\log \boldsymbol{z}_{\lambda^\star}(\boldsymbol{x}_i)$, then the predicted RGF obtained follows approximately the log-normal distribution below:

$$\mathscr{LN}\left(\log y^{\text{pred}}_{\underset{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}{\approx}}(\boldsymbol{x}_i),\, \widehat{\sigma}^2_{i,\lambda^\star} + \sigma^2_{\text{exp},i}\right) \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}. \tag{4.69}$$

The median of the above log-normal distribution is given by:

$$\boldsymbol{z}_{\lambda^\star}^{\text{pred}}(\boldsymbol{x}_i) := \exp\left(\log y^{\text{pred}}_{\underset{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}{\approx}}(\boldsymbol{x}_i)\right) \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}. \tag{4.70}$$

Thus, we can compare the experimental RGF $\boldsymbol{z}$ to

$$\boldsymbol{z}_{\lambda^\star}^{\text{pred}} := \left\{z_{\lambda^\star}^{\text{pred}}(\boldsymbol{x}_i)\right\}_{1\leq i\leq n_{\text{exp}}}$$

The associated prediction interval of 95% is obtained by applying the exponential transformation of the prediction interval of 95% to the Gaussian distribution of Equation (4.68):

$$IF_{95\%}(\boldsymbol{z}_i) := \left[\exp\left(\log y^{\text{pred}}_{\underset{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}{\approx}}(\boldsymbol{x}_i) \pm 1.96\sqrt{\widehat{\sigma}^2_{i,\lambda^\star} + \sigma^2_{\text{exp},i}}\right)\right] \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}. \tag{4.71}$$

For $\lambda^\star = \lambda_{\text{nom}} = 1$, we first adjusted the parameters of the fission gas behavior model by taking their MAP after using the GP-LinCC method (i.e, $\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)$), according to the scenario where the parameters

$$\log(\widetilde{\boldsymbol{\theta}}(\lambda^\star)) \quad \text{with} \quad \widetilde{\boldsymbol{\theta}}(\lambda^\star) := \left(\exp(\boldsymbol{\theta}_1(\lambda^\star)), \boldsymbol{\theta}_3(\lambda^\star), \boldsymbol{\theta}_4(\lambda^\star), \boldsymbol{\theta}_6^{-1}(\lambda^\star)\right),$$

were calibrated, while the other two parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_5$ were kept at their MAP values.

Figure 4.24(a) presents the scatter plots between the experimental log-RGF and the predictions of the log-linear model (Equation (4.63)) respectively at $\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)$ and

$$\log\left(\widetilde{\boldsymbol{\theta}}_{\text{ref}}(\lambda^\star)\right) \quad \text{with} \quad \widetilde{\boldsymbol{\theta}}_{\text{ref}}(\lambda^\star) = \left(\exp(\boldsymbol{\theta}_{\text{ref},1}(\lambda^\star)), \boldsymbol{\theta}_{\text{ref},3}(\lambda^\star), \boldsymbol{\theta}_{\text{ref},4}(\lambda^\star), \boldsymbol{\theta}^{-1}_{\text{ref},6}(\lambda^\star)\right)^t.$$

4.24(b) compares the predictions of Equation (4.70) to the experimental RGF. It can be seen that the log-linear model tends to overestimate the log-RGF for most of the fuel rods. This is partly (or maybe fully) explained by the fact that at the MAP values, the log-linear models outperform the log-RGF simulations. However, we cannot exclude the existence of a discrepancy function between the CARACAS code and physical truth in the burnup and the thermal conductivity [Kennedy and O'Hagan, 2001; Higdon et al., 2004; Gu et al., 2018]. This is why non-linear calibration should eventually be carried out in further work (mathematical developments in progress). Superimposing the histogram of the standardized residuals

$$\frac{\log \boldsymbol{z}_i - \log y^{\text{pred}}_{\underset{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}{\approx}}(\boldsymbol{x}_i)}{\sqrt{\widehat{\sigma}^2_{i,\lambda^\star} + \sigma^2_{\text{exp},i}}} \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}, \tag{4.72}$$

on the standard normal density, we see that the variance of these residuals is strictly greater than 1 (Figure 4.25). In other words, the variances $\widehat{\sigma}^2_{i,\lambda^\star} + \sigma^2_{\text{exp},i}$ are too small relative to

$$\log \boldsymbol{z}_i - \log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}_i) \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}.$$

Once again this could be due to the imperfect accuracy of the log-linear regression models and/or to a discrepancy term that was not accounted for [Higdon et al., 2004; Bachoc et al., 2014; Brynjarsdottir and O'Hagan, 2014]. These further researches are left for the future.

Finally, we computed the MSE and Q^2 associated with the log-predictor

$$\log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}) = \left\{ \log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)}(\boldsymbol{x}_i) \right\}_{1 \leq i \leq n_{\text{exp}}}.$$

These indicators are respectively equal to 0.181 and 0.915. We also calculated the MSE and Q^2 associated with the median predictor $z^{\text{pred}}_{\lambda^\star}$. These indicators are approximately equal to 0.36 and 0.927. Finally, we calculated the mean absolute error between these two quantities. It is equal to 0.418. The mean absolute error measures the mean absolute deviation between observed and predicted values [Steurer et al., 2021; Tofallis, 2015].

4.5.5 Comparison between the experimental RGF and the CARACAS predictions

In this section, we compare the experimental RGF with both predictions of the CARACAS code obtained with $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$. The results of this section are much more important than those of the previous two sections, from the point of view of the CARACAS code. In fact, we aim to compare the experimental RGF with both predictions of the CARACAS code run respectively at the MAP ($\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$).

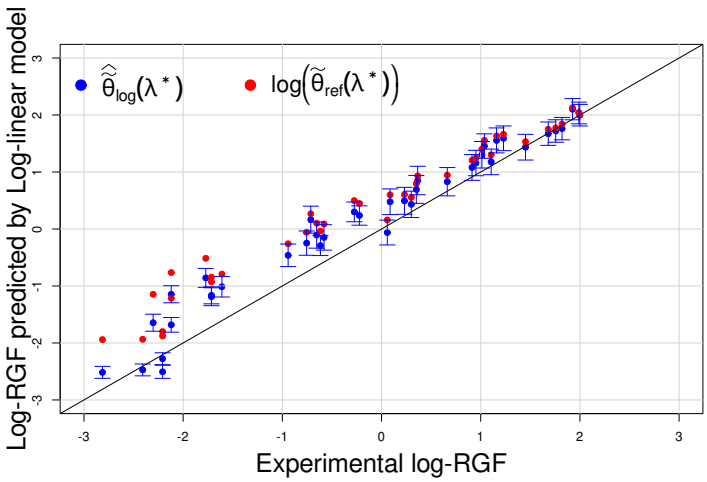
Figure 4.26 presents the scatter plots of the experimental log-RGF (a) and RGF (b) against the predictions of the CARACAS code at the parameter values $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ (with $\lambda^\star = 1$) and $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$ respectively. We can see that the scatter plots obtained with the RGF predictions obtained with $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ are closer to the first bisector than the scatter plots obtained with $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$. In other words, the predictions obtained with $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ appear to be more accurate than those obtained with $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$. Indeed, we calculated the MSE, MAE and Q^2 in both cases. The MSE and the MAE obtained with $\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ are lower and the Q^2 is significantly improved (Table 4.1).

Measured and predicted	MSE	MAE	Q^2
$\left(\log \boldsymbol{z}, \log y_{\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)}(\boldsymbol{x})\right)$	**0.11**	**0.26**	**0.95**
$\left(\log \boldsymbol{z}, \log y_{\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)}(\boldsymbol{x})\right)$	0.46	0.60	0.78
$\left(\boldsymbol{z}, y_{\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)}(\boldsymbol{x})\right)$	**0.26**	**0.34**	**0.95**
$\left(\boldsymbol{z}, y_{\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)}(\boldsymbol{x})\right)$	1.29	0.89	0.74

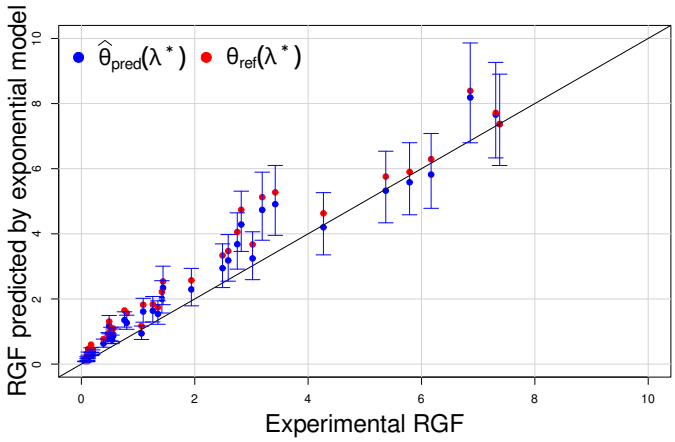
Table 4.1: Computation of MSE, MAE and Q^2 with $\lambda^\star = 1$.

Remark 8

The results in Table 4.1 were obtained from the experimental RGF used for the calibration stage. However, we have not yet validated the accuracy of CARACAS simulations for new RGF data. Another way would be to perform a leave-one-out validation. This solution would involve optimizing the hyperparameters of the GP-LinCC approach at each stage of the leave-one-out process. This may be investigated in future work.



(a) Comparison of the log-RGF predicted of Equation (4.63) (with $\hat{\tilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)$ and $\log(\tilde{\boldsymbol{\theta}}_{\text{ref}}(\lambda^\star))$ respectively) to the log-RGF measured. Error bar is the prediction interval at 95%.



(b) Comparison of RGF predicted of Equation (4.70) (with $\hat{\tilde{\boldsymbol{\theta}}}_{log}(\lambda^\star)$ and $\log(\tilde{\boldsymbol{\theta}}_{\text{ref}}(\lambda^\star))$ respectively) to RGF measured. Error bars is the prediction interval of 95% (Equation (4.71)).

Figure 4.24: Comparison of predicted and measured RGF with $\lambda^\star = \lambda_{\text{nom}} = 1$.

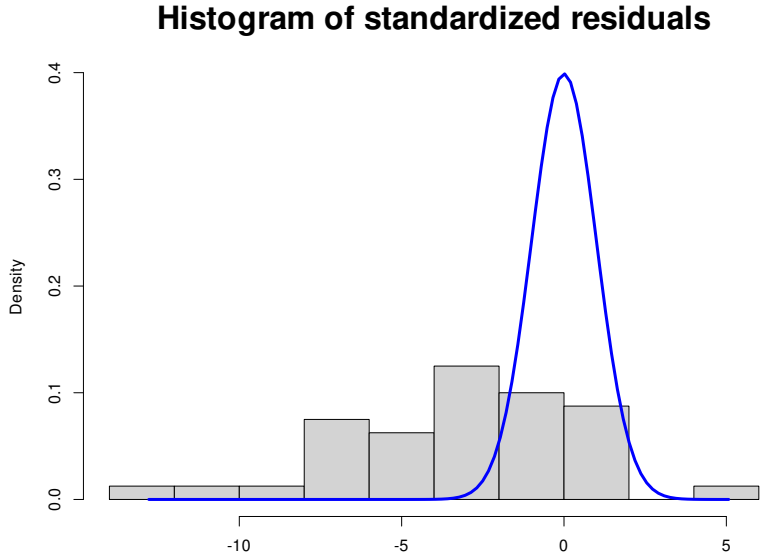


Figure 4.25: Histogram of standardized residuals of Equation (4.72) with $\lambda^\star = 1$.

4.5.6 Examination of the compensation hypothesis (after calibration)

In this section, we inspect the compensation hypothesis for the calibrated log-linear model and the calibrated CARACAS code. This hypothesis states that the RGF are uninformative on the uncertainty of the thermal conductivity. In other words, whatever the value taken by the thermal conductivity, the RGF should remain unchanged. For this purpose, we can examine whether the predictions of the log-linear model (resp. the CARACAS code) are quasi-invariant regardless of the value of the thermal conductivity for all fuel rods. To do this, for the i-th rod, we can perform a statistical test to check whether the support of the probability distribution of the random variable

$$\log y_{\lambda_1}(\boldsymbol{x}_i) - \log y_{\lambda_2}(\boldsymbol{x}_i) \mid \log \boldsymbol{Y}_{\lambda_1}(\boldsymbol{x}_i), \log \boldsymbol{Y}_{\lambda_2}(\boldsymbol{x}_i) \tag{4.73}$$

contains 0 for every $\lambda_1 \neq \lambda_2$. We can also test this hypothesis by constructing a coverage rate at 95% of 0. This rate would be computed from the quantities

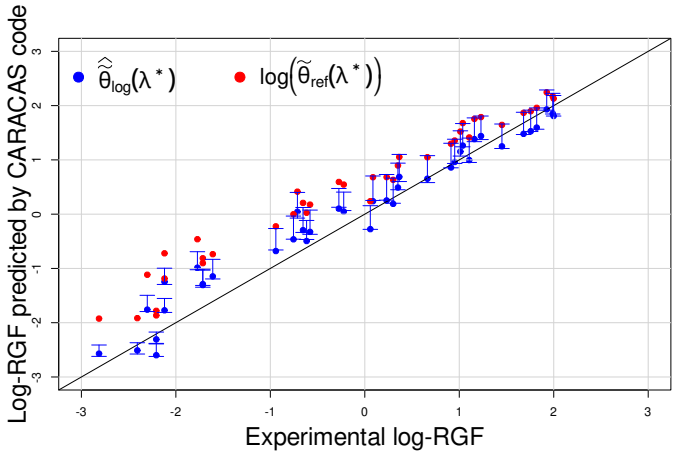
$$\mathbf{1}_{\{0 \in IC_{95\%}(\boldsymbol{x}_i, \lambda_1^{(j)}, \lambda_2^{(j)})\}}$$

associated with different realizations $\{\lambda_1^{(j)}\}_{j=1}^N$ and $\{\lambda_2^{(j)}\}_{j=1}^N$ as

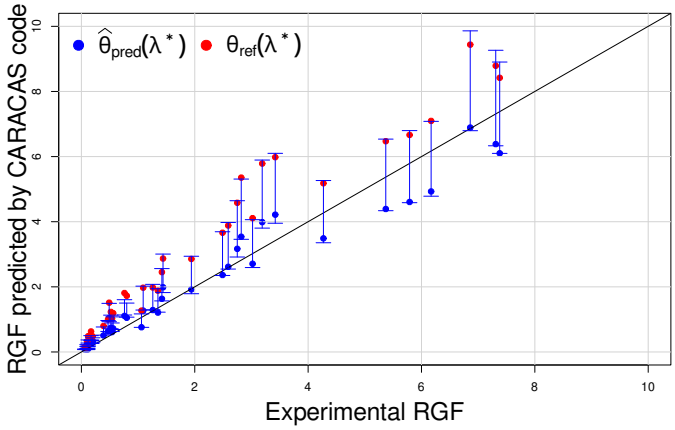
$$\widehat{\Delta}(5\%, \boldsymbol{x}_i) := \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{\{0 \in IC_{95\%}(\boldsymbol{x}_i, \lambda_1^{(j)}, \lambda_2^{(j)})\}}, \tag{4.74}$$

where the 95% predictive interval denoted by $IC_{95\%}(\boldsymbol{x}_i, \lambda_1^{(j)}, \lambda_2^{(j)})$ is derived from the distribution of the random variable in Equation (4.73) (with $\lambda_1 = \lambda_1^{(j)}$ and $\lambda_2 = \lambda_2^{(j)}$). In this way, we can state that the compensation hypothesis is not questioned when $\widehat{\Delta}(5\%, x_i)$ is greater than 95%. This possibility is presented in Chapter 2. However, as a first step we opted for a graphical check. This solution first consists of sampling some values for the thermal conductivity λ denoted by $\{\lambda_k^\star\}_{k=1}^{m^\star}$ then computing $\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda_k^\star)$ for each $\lambda_k^\star$ and finally showing the scatter plot

$$\left(\log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda_{\text{nom}})}(\boldsymbol{x}_i), \log y^{\text{pred}}_{\widehat{\widetilde{\boldsymbol{\theta}}}_{log}(\lambda_k^\star)}(\boldsymbol{x}_i) \right) \quad \text{with} \quad \lambda_{\text{nom}} = 1 \quad \text{and} \quad 1 \leq i \leq n_{\text{exp}}.$$

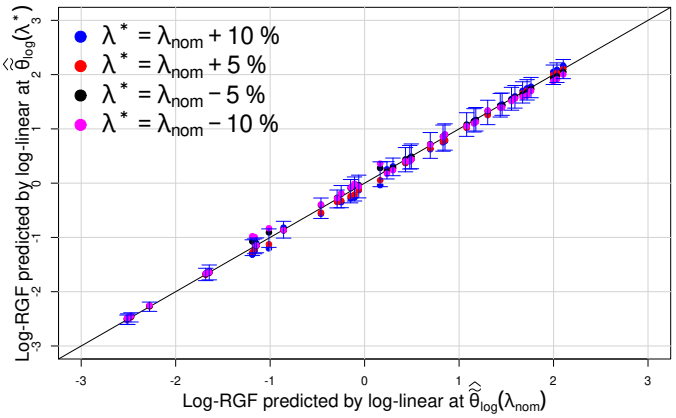


(a) Comparison of log-RGF predicted by the CARACAS code at ($\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star$ respectively) to the log-RGF measured. Error bar is the prediction interval at 95%.

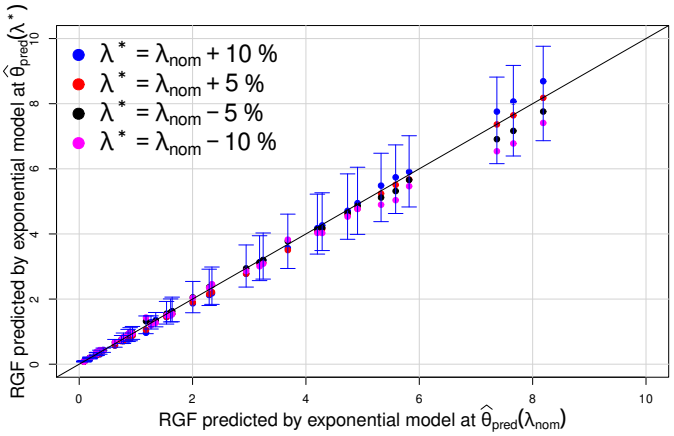


(b) Comparison of RGF predicted by the CARACAS code at ($\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda^\star)$ and $\boldsymbol{\theta}_{\text{ref}}(\lambda^\star)$ respectively) to RGF measured. Error bar is the prediction interval of 95% (Equation (4.71)).

Figure 4.26: Comparison of (a) log-RGF predicted by the CARACAS code and experimental and (b) RGF predicted by the CARACAS code and experimental with $\lambda^\star = 1$.



(a) Inspection of the compensation hypothesis on the log-RGF (Equation (4.63)). Error bars are the 95 % predictive intervals associated with $\lambda_{\text{nom}} = 1$.



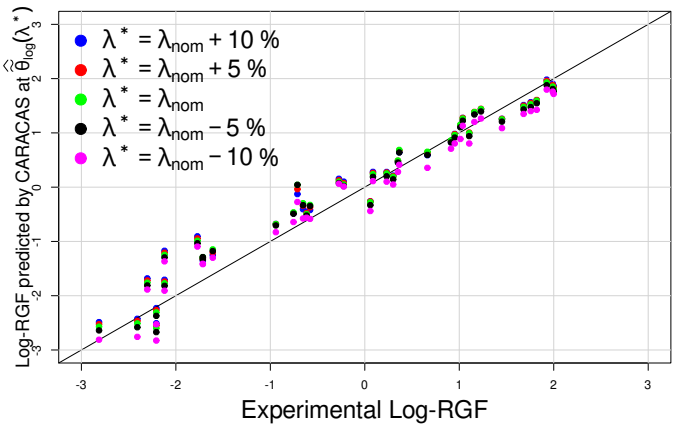
(b) Inspection of the compensation hypothesis on the RGF (Equation (4.65)). Error bars are the 95% predictive intervals associated with $\lambda_{\text{nom}} = 1$.

Figure 4.27: Examination of **the compensation hypothesis for the predictions of the log-linear model** (Equation (4.63)) and **the exponential model** (Equation (4.65)) with $\lambda^\star \in \{1-10\%, 1-5\%, 1+5\%, 1+10\%\}$ and $\lambda_{\text{nom}} = 1$.

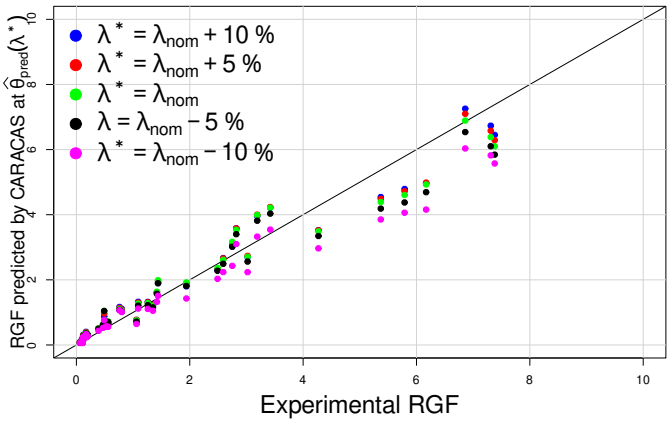
The compensation hypothesis will be said to be satisfied if the points lie within 95% uncertainty interval of the predictive distribution $\pi\left(\log y_{\lambda_{\text{nom}}}(\boldsymbol{x}_i) \mid \log \boldsymbol{Y}_{\lambda_{\text{nom}}}(\boldsymbol{x}_i)\right)$. For $\lambda_k^\star \in \{1-10\%, 1-5\%, 1+5\%, 1+10\%\}$, the results of this graphical solution are shown in Figure 4.27. We can see that the points are quite close to the first bisector and are covered by this interval. Therefore, the compensation hypothesis is unquestioned. The same is true for the calibrated CARACAS code (see Figure 4.28). Figure 4.28 shows the scatter plot

$$\left(y_{\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda_{\text{nom}})}(\boldsymbol{x}_i), y_{\widehat{\boldsymbol{\theta}}_{\text{pred}}(\lambda_k^\star)}(\boldsymbol{x}_i)\right) \quad \text{with} \quad 1 \leq i \leq n_{\text{exp}}.$$

In summary, the compensation hypothesis appears to be verified a posteriori by the calibrated predictions of the log-linear model (Figure 4.27) and also by the corresponding calibrated CARACAS code (Figure 4.28).



(a) Inspection of the compensation hypothesis on the log-RGF predicted by the calibrated CARACAS code.



(b) Inspection of the compensation hypothesis on the RGF predicted by the calibrated CARACAS code.

Figure 4.28: Examination of **the compensation hypothesis for the RGF predictions of the calibrated CARACAS code** obtained with $\lambda^\star \in \{1-10\%, 1-5\%, 1+5\%, 1+10\%\}$ and $\lambda_{\text{nom}} = 1$.

4.6 Conclusions and perspectives

In this study, we have used a Bayesian calibration approach to quantify parameter uncertainties in fuel simulation. More specifically, we aimed to quantify the uncertainty associated with the parameters $\boldsymbol{\theta}$ of the fission gas behavior model as a function of the uncertainty in the thermal conductivity λ of the thermal model. These two models are integrated into the ALCYONE application of the PLEIADES simulation platform dedicated to PWR fuel modeling at CEA.

To achieve this goal, we have developed a new method called GP-LinCC (for Gaussian Process and Linearization-based Conditional Calibration). This method provides a posterior predictive distribution of the parameters $\boldsymbol{\theta}$ as a function of every possible values of the thermal conductivity λ. To achieve this, the GP-LinCC method relies on the use of Gaussian process priors, a linear approximation of the fission gas behavior model (i.e., CARACAS model), and the compensation hypothesis. The linearization of the fission gas behavior model facilitates the use of this methodology. By this approach, we avoided any use of Metropolis-Hastings algorithm to calibrate the parameters $\boldsymbol{\theta}$ conditional on λ.

In our work, we implemented the GP-LinCC approach and observed several significant results. **First, it is important to stress that the GP-LinCC method succeeds in compensating in λ, confirming that the compensation hypothesis formulated by experts is validated by this approach**. The compensation hypothesis states that the experimental data of the fractions of gas released provide little or no information on the uncertainty of the thermal conductivity. **Second, we also found that the calibrated fission gas behavior model shows a tendency to underestimate high burnup fuel rods.** This underestimation could be attributed to the imperfect linear models a posteriori and/or the non-inclusion of a model discrepancy function.

Therefore, our results open the way to some possible improvements of the GP-LinCC method. Further work may be devoted to two major improvements. First, the linear models currently used could be replaced in the GP-LinCC method by non-linear statistical models, such as Gaussian process. Second, we are considering the possibility of introducing a model discrepancy component into our approach.

4.7 Supplementary material

4.7.1 Calculation of posterior and predictive distributions

4.7.1.1 Marginal posterior distributions

Proof: We start from Equation (4.3):

$$\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) = \boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i) + \varepsilon_{i,\,\lambda_k}.$$

The likelihood of the data $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)$ conditional on the pair $(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k})$ is given by:

$$\begin{aligned}\mathscr{L}(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)|g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}) \;\propto\; (\sigma^2_{i,\lambda_k})^{-\frac{n}{2}} \exp -\Big\{ &\tfrac{1}{2}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i)\big)^t \\ &\times \tfrac{1}{\sigma^2_{i,\lambda_k}}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i)\big)\Big\}.\end{aligned}$$

The prior distribution $\pi(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k})$ is given in Equation (4.8). Therefore, the posterir distribution $\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))$ is obtained using Bayes' formula:

$$\begin{aligned}\pi(g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) \;&\propto\; \mathscr{L}(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)|g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k})\pi(g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k}) \\ &\propto\; \exp -\Big\{ \tfrac{1}{2}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i)\big)^t \tfrac{1}{\sigma^2_{i,\lambda_k}}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) \\ &\qquad -\boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i)\big)\Big\}(\sigma^2_{i,\lambda_k})^{-\frac{n}{2}-1}.\end{aligned} \tag{4.75}$$

Let us put $W := \boldsymbol{\Theta}_{\text{reg},k}{}^t \boldsymbol{\Theta}_{\text{reg},k}$, $\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) := W^{-1}\boldsymbol{\Theta}_{\text{reg},k}{}^t \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)$ and

$$\boldsymbol{a} := \left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i)\right)^t \left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i)\right).$$

We can easily show that:

$$\begin{aligned}\boldsymbol{a} \;=\; &\left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t \left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right) \\ &+ \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t W \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right).\end{aligned}$$

In fact,

$$\begin{aligned}\boldsymbol{a} \;&=\; \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - 2\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i) + g_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{\Theta}_{\text{reg},k}{}^t\boldsymbol{\Theta}_{\text{reg},k} g_{\lambda_k}(\boldsymbol{x}_i) \\ &=\; \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - 2\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) W g_{\lambda_k}(\boldsymbol{x}_i) + g_{\lambda_k}(\boldsymbol{x}_i)^t W g_{\lambda_k}(\boldsymbol{x}_i) \\ &=\; \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - 2\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W g_{\lambda_k}(\boldsymbol{x}_i) + g_{\lambda_k}(\boldsymbol{x}_i)^t W g_{\lambda_k}(\boldsymbol{x}_i) + \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W \widehat{g}_{\lambda_k}(\boldsymbol{x}_i) \\ &\qquad -\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W \widehat{g}_{\lambda_k}(\boldsymbol{x}_i) \\ &=\; \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W \widehat{g}_{\lambda_k}(\boldsymbol{x}_i) + \big(\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) - g_{\lambda_k}(\boldsymbol{x}_i)\big)^t W \big(\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) - g_{\lambda_k}(\boldsymbol{x}_i)\big).\end{aligned}$$

Furthermore, we have:

$$\begin{aligned}\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W \widehat{g}_{\lambda_k}(\boldsymbol{x}_i) \;&=\; \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - 2\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W \widehat{g}_{\lambda_k}(\boldsymbol{x}_i) \\ &\qquad + \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W \widehat{g}_{\lambda_k}(\boldsymbol{x}_i) \\ &=\; \boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - 2\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) \\ &\qquad + \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t\boldsymbol{\Theta}_{\text{reg},k}{}^t\boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) \\ &=\; \big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t \big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big).\end{aligned}$$

Then, one has:

$$\begin{aligned}\boldsymbol{a} &= \big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\\ &\quad + \big(\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) - g_{\lambda_k}(\boldsymbol{x}_i)\big)^t W\big(\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) - g_{\lambda_k}(\boldsymbol{x}_i)\big).\end{aligned}$$

As a result, Equation (4.75) becomes:

$$\begin{aligned}\pi(g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) \;\propto\; (\sigma^2_{i,\lambda_k})^{-\frac{n}{2}-1}\exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\Big\{&\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\\ &\times\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\\ &+ \big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t W\big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\Big\}.\end{aligned} \tag{4.76}$$

The marginal posterior distribution $\pi(\sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))$ is obtained by integrating the joint posterior distribution $\pi(g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))$ with respect to $dg_{\lambda_k}(\boldsymbol{x}_i)$:

$$\begin{aligned}\pi(\sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) &= \int \pi(g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))dg_{\lambda_k}(\boldsymbol{x}_i)\\ &\propto \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\\ &\quad\times\int \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t W\big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)dg_{\lambda_k}(\boldsymbol{x}_i)\\ &\quad\times(\sigma^2_{i,\lambda_k})^{-\frac{n}{2}-1}\\ &\propto \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\\ &\quad\times\sqrt{2\pi}^{p+1}|\sigma^2_{i,\lambda_k}W^{-1}|^{\frac{1}{2}}\frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n}{2}+1}}\\ &\propto \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\\ &\quad\times\Big(\frac{1}{\sigma^2_{i,\lambda_k}}\Big)^{\frac{n-p-1}{2}+1}.\end{aligned} \tag{4.77}$$

In (4.77), we recognize the expression of an Inverse-Gamma distribution:

$$\pi(\sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) \sim IG(\frac{n-p-1}{2}, \frac{n-p-1}{2}\widehat{\sigma}^2_{i,\lambda_k}),$$

where

$$\widehat{\sigma}^2_{i,\lambda_k} = \frac{1}{n-p-1}\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big).$$

The marginal posterior distribution $\pi(g_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))$ is obtained by integrating the joint posterior distribution $\pi(g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))$ with respect to $d\sigma^2_{i,\lambda_k}$:

$$\begin{aligned}\pi(g_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) &= \int \pi(g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))d\sigma^2_{i,\lambda_k}\\ &\propto \int \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\Big\{\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\\ &\quad+\big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t W\big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)\Big\}\frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n}{2}+1}}d\sigma^2_{i,\lambda_k}.\end{aligned}$$

Let us put

$$T := \big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t\big(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big) + \big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big)^t W\big(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\big).$$

As a result, we have:

$$\begin{aligned}\pi(g_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) &\propto \int \frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n}{2}+1}} \exp -\frac{1}{2}\frac{T}{\sigma^2_{i,\lambda_k}} d\sigma^2_{i,\lambda_k} \\ &\propto \left(\frac{T}{2}\right)^{-\frac{n}{2}}\Gamma(\frac{n}{2}) \times \int \left(\frac{T}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(\frac{n}{2})} \frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n}{2}+1}} \exp -\frac{1}{2}\frac{T}{\sigma^2_{i,\lambda_k}} d\sigma^2_{i,\lambda_k}.\end{aligned}$$

Therefore, one has:

$$\begin{aligned}\pi(g_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) &\propto \left(\frac{T}{2}\right)^{-\frac{n}{2}}\Gamma(\frac{n}{2}) \\ &\propto T^{-\frac{n}{2}}.\end{aligned} \tag{4.78}$$

Now, the expression T is equal to:

$$\begin{aligned}T &= \left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t \left(\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}_{\text{reg},k}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right) \\ &\quad + \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t W \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right) \\ T &= (n-p-1)\widehat{\sigma}^2_{i,\lambda_k} + \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t W \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right).\end{aligned}$$

By replacing T by its expression in Equation (4.78), we obtain:

$$\begin{aligned}\pi(g_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) &\propto \left((n-p-1)\widehat{\sigma}^2_{i,\lambda_k} + \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t W \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)\right)^{-\frac{n}{2}} \\ &\propto \left(1 + \frac{1}{(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}} \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t W \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)\right)^{-\frac{n-p-1+p+1}{2}}.\end{aligned} \tag{4.79}$$

In (4.79), we recognize to within one constant the expression of a density of a multivariate Student distribution with $n-p-1$ degrees of freedom. ■

4.7.1.2 Predictive Distribution

We start from Equation (4.13):

Proof:

$$\begin{aligned}\pi(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) &= \int \mathscr{L}(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i) \mid g_{\lambda_k}(\boldsymbol{x}_i), \sigma^2_{i,\lambda_k}) \pi(g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k}|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) dg_{\lambda_k}(\boldsymbol{x}_i) d\sigma^2_{i,\lambda_k} \\ &\propto \int \left[\frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n_{\text{new}}}{2}}} \exp -\left\{\frac{1}{2\sigma^2_{i,\lambda_k}} \left(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}^{\text{new}}_k g_{\lambda_k}(\boldsymbol{x}_i)\right)^t \right.\right. \\ &\qquad \left.\left. \times \left(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i) - \boldsymbol{\Theta}^{\text{new}}_k g_{\lambda_k}(\boldsymbol{x}_i)\right)\right\} \frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n}{2}+1}} \right. \\ &\qquad \exp -\frac{1}{2\sigma^2_{i,\lambda_k}} \left(\left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right)^t W \left(g_{\lambda_k}(\boldsymbol{x}_i) - \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\right) \right. \\ &\qquad \left.\left. + (n-p-1)\widehat{\sigma}^2_{i,\lambda_k} \right) dg_{\lambda_k}(\boldsymbol{x}_i) \right] d\sigma^2_{i,\lambda_k}.\end{aligned} \tag{4.80}$$

To prove (4.14), we will first expand the expression of the exponential of Equation (4.80) to bring out the following two quantities:

- $\boldsymbol{\Theta}^{\text{new}}_k \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)$,
- $\widehat{\sigma}^2_{i,\lambda_k}(I_{n_{\text{new}}} + \boldsymbol{\Theta}^{\text{new}}_k (\boldsymbol{\Theta}_{\text{reg},k}{}^t \boldsymbol{\Theta}_{\text{reg},k})^{-1} (\boldsymbol{\Theta}^{\text{new}}_k)^t)$,

and then apply the Fubini-Tonelli theorem to the integral. First, let us develop the expression of the exponential:

$$\begin{aligned}Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)^t Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i) - 2\left(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) + (\boldsymbol{\Theta}^{\text{new}}_k)^t Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)\right)^t g_{\lambda_k}(\boldsymbol{x}_i) + g_{\lambda_k}(\boldsymbol{x}_i)^t \left((\boldsymbol{\Theta}^{\text{new}}_k)^t \boldsymbol{\Theta}^{\text{new}}_k + W\right) g_{\lambda_k}(\boldsymbol{x}_i) \\ + \widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t W \widehat{g}_{\lambda_k}(\boldsymbol{x}_i) + (n-p-1)\widehat{\sigma}^2_{i,k}.\end{aligned} \tag{4.81}$$

In addition, we have:

$$-2\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)^t g_{\lambda_k}(\boldsymbol{x}_i)+g_{\lambda_k}(\boldsymbol{x}_i)^t\Sigma^{-1}g_{\lambda_k}(\boldsymbol{x}_i)=\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)^t\Sigma^{-1}\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)-\Big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\Big)^t\Sigma\Big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\Big).$$

By replacing in Equation (4.81) and putting the terms together, we obtain:

$$Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)^t\Big[I_{n_{\text{new}}}-\boldsymbol{\Theta}_k^{\text{new}}\Sigma\,(\boldsymbol{\Theta}_k^{\text{new}})^t\Big]Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)+\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)^t\Big[W-W\Sigma W\Big]\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)$$
$$-2Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)^t\boldsymbol{\Theta}_k^{\text{new}}\Sigma W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)^t\Sigma^{-1}\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)+(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}\quad(4.82)$$

where $\Sigma^{-1}:=(\boldsymbol{\Theta}_k^{\text{new}})^t\,\boldsymbol{\Theta}_k^{\text{new}}+W$. According to the Woodbury-Shermann-Morrisson identity, we have:

$$\begin{aligned}I_{n_{\text{new}}}-\boldsymbol{\Theta}_k^{\text{new}}\Sigma\,(\boldsymbol{\Theta}_k^{\text{new}})^t&=\big(I_{n_{\text{new}}}+\boldsymbol{\Theta}_k^{\text{new}}W^{-1}\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)^{-1},\\ W-W\Sigma W&=W-W\big\{W^{-1}-W^{-1}\big(I_{n_{\text{new}}}+\boldsymbol{\Theta}_k^{\text{new}}W^{-1}\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)^{-1}W^{-1}\big\}W\\ &=\big(I_{n_{\text{new}}}+\boldsymbol{\Theta}_k^{\text{new}}W^{-1}\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)^{-1}.\end{aligned}$$

In addition, we have:

$$\boldsymbol{\Theta}_k^{\text{new}}\Sigma W=\big(I_{n_{\text{new}}}+\boldsymbol{\Theta}_k^{\text{new}}W^{-1}\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)^{-1}\boldsymbol{\Theta}_k^{\text{new}}.$$

Because,

$$\begin{aligned}\Sigma\Sigma^{-1}=I_{p+1}\Rightarrow\Sigma W&=I_{p+1}-\Sigma(\boldsymbol{\Theta}_k^{\text{new}})^t\boldsymbol{\Theta}_k^{\text{new}},\\ \boldsymbol{\Theta}_k^{\text{new}}\Sigma W&=\boldsymbol{\Theta}_k^{\text{new}}-\boldsymbol{\Theta}_k^{\text{new}}\Sigma\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\boldsymbol{\Theta}_k^{\text{new}}\\ \boldsymbol{\Theta}_k^{\text{new}}\Sigma W&=\big(I_{n_{\text{new}}}-\boldsymbol{\Theta}_k^{\text{new}}\Sigma\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)\boldsymbol{\Theta}_k^{\text{new}}\\ \boldsymbol{\Theta}_k^{\text{new}}\Sigma W&=\big(I_{n_{\text{new}}}+\boldsymbol{\Theta}_k^{\text{new}}W^{-1}\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)^{-1}\boldsymbol{\Theta}_k^{\text{new}}.\end{aligned}$$

Consequently, Equation (4.82) becomes:

$$\Big(Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)-\boldsymbol{\Theta}_k^{\text{new}}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\Big)^t\big(I_{n_{\text{new}}}+\boldsymbol{\Theta}_k^{\text{new}}W^{-1}\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)^{-1}\Big(Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)-\boldsymbol{\Theta}_k^{\text{new}}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\Big)+\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^tY_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)^t\Sigma^{-1}\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)+(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}.\quad(4.83)$$

By replacing the expression for the exponential of the integral (4.80) by the expression (4.83), we obtain:

$$\pi(Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))\propto\int\frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n_{\text{new}}+n}{2}+1}}\exp-\frac{1}{2\sigma^2_{i,\lambda_k}}\Big\{C+(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}+\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)^t\Sigma^{-1}\Big(g_{\lambda_k}(\boldsymbol{x}_i)-\Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)+(\boldsymbol{\Theta}_k^{\text{new}})^t\,Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)\big)\Big)\Big\}dg_{\lambda_k}(\boldsymbol{x}_i)d\sigma^2_{i,\lambda_k},\quad(4.84)$$

where

$$C:=\Big(Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)-\boldsymbol{\Theta}_k^{\text{new}}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\Big)^t\big(I_{n_{\text{new}}}+\boldsymbol{\Theta}_k^{\text{new}}W^{-1}\,(\boldsymbol{\Theta}_k^{\text{new}})^t\,\big)^{-1}\Big(Y_{\lambda_k}^{\text{new}}(\boldsymbol{x}_i)-\boldsymbol{\Theta}_k^{\text{new}}\widehat{g}_{\lambda_k}(\boldsymbol{x}_i)\Big).$$

By Fubini-Tonelli, we integrate the integral first with respect to $dg_{\lambda_k}(\boldsymbol{x}_i)$ and then with respect to $d\sigma^2_{i,\lambda_k}$:

$$\pi(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) \propto \int \frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n_{\text{new}}+n}{2}+1}} \exp -\frac{1}{2\sigma^2_{i,\lambda_k}} \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\{C+(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}\}$$
$$\times \left\{ \int \exp -\frac{1}{2\sigma^2_{i,\lambda_k}} \Big(g_{\lambda_k}(\boldsymbol{x}_i) - \Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) + \boldsymbol{\Theta}^{\text{new}\,t}_k Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)\big)\Big)^t \Sigma^{-1} \Big(g_{\lambda_k}(\boldsymbol{x}_i) - \right.$$
$$\left. \Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) + (\boldsymbol{\Theta}^{\text{new}}_k)^t\, Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)\big)\Big) dg_{\lambda_k}(\boldsymbol{x}_i) \right\} d\sigma^2_{i,\lambda_k}. \quad (4.85)$$

Now the integral with respect to $dg_{\lambda_k}(\boldsymbol{x}_i)$ is equal to:

$$\int \exp -\frac{1}{2\sigma^2_{i,\lambda_k}} \Big(g_{\lambda_k}(\boldsymbol{x}_i) - \Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) + (\boldsymbol{\Theta}^{\text{new}}_k)^t\, Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)\big)\Big)^t \Sigma^{-1} \Big(g_{\lambda_k}(\boldsymbol{x}_i) - \Sigma\big(W\widehat{g}_{\lambda_k}(\boldsymbol{x}_i) + \right.$$
$$\left. (\boldsymbol{\Theta}^{\text{new}}_k)^t\, Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)\big)\Big) dg_{\lambda_k}(\boldsymbol{x}_i) = \sqrt{2\pi}^{p+1} |\sigma^2_{i,\lambda_k}\Sigma|^{\frac{1}{2}}.$$

The expression (4.85) then becomes:

$$\begin{aligned}
\pi(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i)) &\propto \int \frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n_{\text{new}}+n}{2}+1}} \exp -\frac{1}{2\sigma^2_{i,\lambda_k}} \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\{C+ \\
&\quad (n-p-1)\widehat{\sigma}^2_{i,\lambda_k}\} \times \sqrt{2\pi}^{p+1} |\sigma^2_{i,\lambda_k}\Sigma|^{\frac{1}{2}} d\sigma^2_{i,\lambda_k} \\
&\propto \int \frac{1}{(\sigma^2_{i,\lambda_k})^{\frac{n_{\text{new}}+n-p-1}{2}+1}} \exp -\frac{1}{2\sigma^2_{i,\lambda_k}}\{C+(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}\} d\sigma^2_{i,\lambda_k} \\
&\propto \left(C+(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}\right)^{\frac{n_{\text{new}}+n-p-1}{2}} \\
&\propto \left(1+\frac{C}{(n-p-1)\widehat{\sigma}^2_{i,\lambda_k}}\right)^{\frac{n_{\text{new}}+n-p-1}{2}}.
\end{aligned}$$

As a result, $\pi(Y^{\text{new}}_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{Y}_{\lambda_k}(\boldsymbol{x}_i))$ follows a multivariate Student distribution with $n-p-1$ degrees of freedom. ■

4.7.2 Computation of Sobol' indices

Note that the variables $\boldsymbol{\theta}$ and $\vartheta_k = (g_{\lambda_k}(\boldsymbol{x}_i),\, \sigma^2_{i,\lambda_k})$ are independent. The variable $\boldsymbol{\theta}$ follows the distribution given in Equation (4.27), $g_{\lambda_k}(\boldsymbol{x}_i)$, σ^2_{i,λ_k} are assumed to follow the probability distributions given in Equations (4.7), (4.3) and $\epsilon_{i,\,k}|\sigma^2_{i,\lambda_k} \sim \mathcal{N}(0, \sigma^2_{i,\lambda_k})$ respectively. For the analytical calculation of the Sobol' indices, we start from Equation (4.12) :

$$Y_{\lambda_k}(\boldsymbol{x}_i) = g_{\lambda_k,0}(\boldsymbol{x}_i) + g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta} + \epsilon_{i,\lambda_k}.$$

In addition, we have the following two equalities:

$$g_{\lambda_k,0}(\boldsymbol{x}_i) = (1,0,\cdots 0) \begin{pmatrix} g_{\lambda_k,0}(\boldsymbol{x}_i) \\ g_{\lambda_k,1}(\boldsymbol{x}_i) \end{pmatrix} = A g_{\lambda_k}(\boldsymbol{x}_i), \quad (4.86)$$

$$g_{\lambda_k,1}(\boldsymbol{x}_i) = \begin{pmatrix} \boldsymbol{0} & I_p \end{pmatrix} g_{\lambda_k}(\boldsymbol{x}_i) = B g_{\lambda_k}(\boldsymbol{x}_i), \quad (4.87)$$

where $\boldsymbol{0} = (0,\cdots,0)^t \in \mathbb{R}^p$.

4.7.2.1 Calculation of $\mathbb{V}_{\boldsymbol{\theta}}^{i,k}$

Proof:

$$\begin{aligned}
\mathbb{V}_{\boldsymbol{\theta}}^{i,k} &= \mathbb{V}\big(\mathbb{E}\big(Y_{\lambda_k}(\boldsymbol{x}_i)|\boldsymbol{\theta}\big)\big) = \mathbb{V}\big(\mathbb{E}\big(g_{\lambda_k,0}(\boldsymbol{x}_i) + g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta} + \epsilon_{i,\lambda_k}|\boldsymbol{\theta}\big)\big)\\
\mathbb{V}_{\boldsymbol{\theta}}^{i,k} &= \mathbb{V}\big(\mathbb{E}(g_{\lambda_k,0}(\boldsymbol{x}_i))\big) + \mathbb{V}\big(\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}|\boldsymbol{\theta})\big) + \mathbb{V}\big(\mathbb{E}(\epsilon_{i,\lambda_k})\big)\\
\mathbb{V}_{\boldsymbol{\theta}}^{i,k} &= \mathbb{V}\big(\mathbb{E}(Ag_{\lambda_k}(\boldsymbol{x}_i))\big) + \mathbb{E}(g_{\lambda_k}(\boldsymbol{x}_i))^t B^t\mathbb{V}\big(\boldsymbol{\theta}\big)B\mathbb{E}(g_{\lambda_k}(\boldsymbol{x}_i)) + \mathbb{V}\big(\mathbb{E}(\mathbb{E}(\epsilon_{i,\lambda_k}|\sigma^2_{i,\lambda_k}))\big)\\
\mathbb{V}_{\boldsymbol{\theta}}^{i,k} &= 0 + \mathbb{E}(g_{\lambda_k}(\boldsymbol{x}_i))^t B^t\mathbb{V}\big(\boldsymbol{\theta}\big)B\mathbb{E}(g_{\lambda_k}(\boldsymbol{x}_i)) + \mathbb{V}\big(\mathbb{E}(0)\big) = \mathbb{E}(g_{\lambda_k}(\boldsymbol{x}_i))^t B^t\mathbb{V}\big(\boldsymbol{\theta}\big)B\mathbb{E}(g_{\lambda_k}(\boldsymbol{x}_i)).
\end{aligned}$$

■

4.7.2.2 Calculation of $\mathbb{V}_{\vartheta_k}^{i,k}$

Proof:

$$\begin{aligned}
\mathbb{V}_{\vartheta_k}^{i,k} &= \mathbb{V}\big(\mathbb{E}\big(Y_{\lambda_k}(\boldsymbol{x}_i)|\vartheta_k\big)\big) = \mathbb{V}\big(\mathbb{E}\big(g_{\lambda_k,0}(\boldsymbol{x}_i) + g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta} + \epsilon_{i,\lambda_k}|g_{\lambda_k,0}(\boldsymbol{x}_i), g_{\lambda_k,1}(\boldsymbol{x}_i)\,\sigma^2_{i,\lambda_k}\big)\big)\\
\mathbb{V}_{\vartheta_k}^{i,k} &= \mathbb{V}\big(g_{\lambda_k,0}(\boldsymbol{x}_i) + \mathbb{E}(\boldsymbol{\theta})^t g_{\lambda_k,1}(\boldsymbol{x}_i) + \mathbb{E}(\epsilon_{i,\lambda_k}|\sigma^2_{i,\lambda_k})\big) = \mathbb{V}\big(g_{\lambda_k,0}(\boldsymbol{x}_i)\big) + \mathbb{V}\big(\mathbb{E}(\boldsymbol{\theta})^t g_{\lambda_k,1}(\boldsymbol{x}_i)\big)\\
\mathbb{V}_{\vartheta_k}^{i,k} &= A\mathbb{V}(g_{\lambda_k}(\boldsymbol{x}_i))A^t + \mathbb{E}(\boldsymbol{\theta})^t B\mathbb{V}(g_{\lambda_k}(\boldsymbol{x}_i))B^t\mathbb{E}(\boldsymbol{\theta}).
\end{aligned}$$

■

4.7.2.3 Calculation of $\mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,k}$

Proof:

$$\begin{aligned}
\mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,k} &= \mathbb{V}\big(\mathbb{E}\big(Y_{\lambda_k}(\boldsymbol{x}_i)|\vartheta_k,\,\boldsymbol{\theta}\big)\big) - \mathbb{V}_{\boldsymbol{\theta}}^{i,k} - \mathbb{V}_{\vartheta_k}^{i,k}\\
\mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,k} &= \mathbb{V}\big[\mathbb{E}(g_{\lambda_k,0}(\boldsymbol{x}_i)|g_{\lambda_k,0}(\boldsymbol{x}_i)) + \mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}|\boldsymbol{\theta},\,g_{\lambda_k,1}(\boldsymbol{x}_i)) + \mathbb{E}\big(\epsilon_{i,\lambda_k}|\sigma^2_{i,\lambda_k}\big)\big] - \mathbb{V}_{\boldsymbol{\theta}}^{i,k} - \mathbb{V}_{\vartheta_k}^{i,k}\\
\mathbb{V}_{\vartheta_k,\boldsymbol{\theta}}^{i,k} &= \mathbb{V}\big(g_{\lambda_k,0}(\boldsymbol{x}_i)\big) + \mathbb{V}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big) - \mathbb{V}_{\boldsymbol{\theta}}^{i,k} - \mathbb{V}_{\vartheta_k}^{i,k}. \qquad (4.88)
\end{aligned}$$

The expression $\mathbb{V}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big)$ is equal to:

$$\begin{aligned}
\mathbb{V}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big) &= \mathbb{E}\big((g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta})^2\big) - (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2\\
&= \mathbb{E}\big((g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta})^t g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big) - (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2\\
&= \mathbb{E}\big(\boldsymbol{\theta}^t g_{\lambda_k,1}(\boldsymbol{x}_i)g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big) - (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2. \qquad (4.89)
\end{aligned}$$

Furthermore, we have:

$$\boldsymbol{\theta}^t g_{\lambda_k,1}(\boldsymbol{x}_i)g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta} = \sum_{l=1}^{p}\sum_{j=1}^{p}\boldsymbol{\theta}_l(g_{\lambda_k,1}(\boldsymbol{x}_i))_l(g_{\lambda_k,1}(\boldsymbol{x}_i))_j\boldsymbol{\theta}_j = Tr(\boldsymbol{\theta}\boldsymbol{\theta}^t g_{\lambda_k,1}(\boldsymbol{x}_i)g_{\lambda_k,1}(\boldsymbol{x}_i)^t).$$

By replacing in (4.89), we obtain:

$$\begin{aligned}
\mathbb{V}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big) &= \mathbb{E}\big(Tr(\boldsymbol{\theta}\boldsymbol{\theta}^t g_{\lambda_k,1}(\boldsymbol{x}_i)g_{\lambda_k,1}(\boldsymbol{x}_i)^t)\big) - (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2\\
&= Tr\big(\mathbb{E}\big(\boldsymbol{\theta}\boldsymbol{\theta}^t g_{\lambda_k,1}(\boldsymbol{x}_i)g_{\lambda_k,1}(\boldsymbol{x}_i)^t\big)\big) - (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2\\
&= Tr\big(\mathbb{E}\big(\boldsymbol{\theta}\boldsymbol{\theta}^t\big)\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)g_{\lambda_k,1}(\boldsymbol{x}_i)^t\big)\big) - (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2\\
&= Tr\left[(\mathbb{V}(\boldsymbol{\theta}) + \mathbb{E}(\boldsymbol{\theta})\mathbb{E}(\boldsymbol{\theta})^t)(\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i)) + \mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t)\right] - (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2\\
&= Tr\big(\mathbb{V}(\boldsymbol{\theta})\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\big) + Tr\big(\mathbb{V}(\boldsymbol{\theta})\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\big) + Tr\big(\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(\boldsymbol{\theta})\mathbb{E}(\boldsymbol{\theta})^t\big).
\end{aligned} \qquad (4.90)$$

We obtain (4.90) because:

$$\begin{aligned}
(\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2 &= Tr\big(\mathbb{E}(\boldsymbol{\theta})\mathbb{E}(\boldsymbol{\theta})^t\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\big)\\
&= \sum_{l=1}^{p}\sum_{j=1}^{p}\mathbb{E}(\boldsymbol{\theta}_l)\mathbb{E}(\boldsymbol{\theta}_j)\mathbb{E}((g_{\lambda_k,1}(\boldsymbol{x}_i))_l)\mathbb{E}((g_{\lambda_k,1}(\boldsymbol{x}_i))_j) = (\mathbb{E}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big))^2.
\end{aligned}$$

We show that:

$$Tr\big(\mathbb{V}(\boldsymbol{\theta})\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\big) = \mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\mathbb{V}(\boldsymbol{\theta})\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i)),$$
$$Tr\big(\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(\boldsymbol{\theta})\mathbb{E}(\boldsymbol{\theta})^t\big) = \mathbb{E}(\boldsymbol{\theta})^t\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(\boldsymbol{\theta}).$$

In fact, we have:

$$\begin{aligned} Tr\big(\mathbb{V}(\boldsymbol{\theta})\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\big) &= \sum_{l=1}^{p}\sum_{j=1}^{p}\mathbb{V}(\boldsymbol{\theta})_{l,\,j}\big(\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\big)_{j,\,l} \\ &= \sum_{l=1}^{p}\sum_{j=1}^{p}\mathbb{V}(\boldsymbol{\theta})_{l,\,j}\big(\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\big)_j\big(\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))\big)_l \\ &= \mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\mathbb{V}(\boldsymbol{\theta})\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i)). \end{aligned}$$

We do the same for $Tr\big(\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(\boldsymbol{\theta})\mathbb{E}(\boldsymbol{\theta})^t\big)$. As a result, we have:

$$\mathbb{V}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big) = Tr\big(\mathbb{V}(\boldsymbol{\theta})\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\big) + \mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\mathbb{V}(\boldsymbol{\theta})\mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i)) + \mathbb{E}(\boldsymbol{\theta})^t\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(\boldsymbol{\theta}). \tag{4.91}$$

By replacing the expression for $\mathbb{V}\big(g_{\lambda_k,1}(\boldsymbol{x}_i)^t\boldsymbol{\theta}\big)$ in (4.88), we obtain:

$$\begin{aligned} \mathbb{V}^{i,k}_{\vartheta_k,\boldsymbol{\theta}} &= Tr\big(\mathbb{V}(\boldsymbol{\theta})\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\big) + \mathbb{E}(g_{\lambda_k,1}(\boldsymbol{x}_i))^t\mathbb{V}(\boldsymbol{\theta})\mathbb{E}(g_{\lambda_k}(\boldsymbol{x}_i)) + \mathbb{E}(\boldsymbol{\theta})^t\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\mathbb{E}(\boldsymbol{\theta}) - \mathbb{V}^{i,k}_{\boldsymbol{\theta},k} - \mathbb{V}^{i,k}_{\vartheta_k} \\ &= Tr\big(\mathbb{V}(\boldsymbol{\theta})\mathbb{V}(g_{\lambda_k,1}(\boldsymbol{x}_i))\big). \end{aligned}$$

■

Chapter 5

Conclusion et perspectives (français)

Résumé des deux contributions

Le sujet principal de cette thèse est le calage bayésien des paramètres de modèle en simulation multiphysique du combustible nucléaire. Le calage de modèle est un processus qui permet de quantifier et de réduire les incertitudes paramétriques à partir des données disponibles. Le calage bayésien quantifie ces incertitudes par des distributions de probabilité. Les outils et méthodes développés dans le cadre de cette thèse ont été appliqués au simulateur chaîné ALCYONE-CARACAS dans le but de quantifier les incertitudes des paramètres d'entrée $\boldsymbol{\theta}$ du modèle de comportement des gaz de fission (code CARACAS) conditionnellement à toute l'incertitude de la conductivité thermique λ du modèle thermique en utilisant des données expérimentales des fractions de gaz relâchées (RGF pour *Release Gas Fraction*).

Pour cela, nous avons proposé une nouvelle méthode appelée GP-LinCC (pour *Gaussian Process and Linearization-based Conditonal Calibration*) qui apprend la relation entre $\boldsymbol{\theta}$ et λ à travers la distribution *a posteriori* conditionnelle de $\boldsymbol{\theta} \mid \lambda$ obtenue à l'aide des données expérimentales de RGF disponibles (Chapitre 2). Cette nouvelle méthode non-paramétrique basée sur des processus gaussiens et sur la linéarité du modèle numérique par rapport à $\boldsymbol{\theta} \mid \lambda$ fournit une distribution *a posteriori* prédictive des paramètres incertains $\boldsymbol{\theta}$ conditionnellement à n'importe quelle réalisation de λ échantillonnée dans sa distribution de probabilité. Ainsi, à partir de cette distribution prédictive, nous pouvons déduire l'estimateur MAP (maximum *a posteriori*) de $\boldsymbol{\theta}$ sachant λ sans avoir besoin de relancer le code de calcul pour la ou (les) nouvelle (s) valeur(s) de λ considérée(s). En pratique, la méthode GP-LinCC s'appuie d'une part sur un ensemble de données simulées de la quantité de sortie d'intérêt du modèle associées à différentes réalisations du couple $(\boldsymbol{\theta}, \lambda)$, et d'autre part sur des données expérimentales de la sortie d'intérêt. Les données simulées vont servir à construire les modèles linéaires tandis que les données expérimentales seront utilisées dans le processus de calage de la méthode GP-LinCC. Cette approche a été testée sur des exemples analytiques (Chapitre 2) et elle a montré de bonnes performances prédictives. Cependant, son déploiement sur le cas applicatif ALCYONE-CARACAS exposé au Chapitre 4 a nécessité le développement d'une seconde contribution. En effet, cette méthode de calage n'échappe pas au problème de la dimension qui est un problème courant en apprentissage statistique et quantification des incertitudes. Pour faciliter la mise en œuvre de la méthode GP-LinCC sur le modèle de comportement des gaz de fission qui comporte une dizaine de paramètres incertains, nous avons procédé à une sélection au préalable des paramètres de calage $\boldsymbol{\theta}$ significativement influents tout en prenant en compte toute l'incertitude de λ. Cette présélection a été faite via une analyse de sensibilité globale basée sur le critère de Hilbert-Schmidt d'indépendance HSIC.

Le critère HSIC (pour *Hilbert-Schmidt Independence Criterion*) est un outil de plus en plus populaire pour l'analyse de sensibilité globale d'un modèle numérique. Il se base sur des noyaux qui induisent des espaces de Hilbert spéciaux, appelés espaces de Hilbert à noyau reproduisant, afin de détecter l'indépendance statistique entre deux objets aléatoires à support

dans un espace métrique séparable. Cette détection se fait via une statistique de test basée sur la distance entre les représentants des distributions de probabilité de ces deux objets aléatoires dans les espaces de Hilbert à noyau reproduisant. Notre seconde contribution qui fait l'objet du Chapitre 3 a ainsi consisté à utiliser le cadre des indices HSIC pour sélectionner les paramètres $\boldsymbol{\theta}$ significativement influents tout en intégrant toute l'incertitude de λ dans le processus d'estimation des indices HSIC. Cette seconde contribution peut se résumer en quatre étapes. D'abord, nous avons proposé des indices HSIC adaptés à la présence des deux niveaux d'incertitude : l'incertitude de $\boldsymbol{\theta}$ et celle de λ. Ensuite, nous avons montré que ces nouvelles mesures de sensibilité caractérisent bien l'indépendance entre chaque entrée (c'est-à-dire chaque composante de $\boldsymbol{\theta}$) et la/les sortie(s) d'intérêt considérée(s), et ce, en présence de l'incertitude de la conductivité thermique. Puis, nous avons proposé quatre estimateurs possibles issus de deux stratégies d'échantillonnage Monte Carlo (stratégies (S1) et (S2)) et nous avons également montré que ces estimateurs bénéficiaient de bonnes propriétés statistiques (biais, consistance, vitesse de convergence) semblables aux estimateurs Monte Carlo classiques. Enfin, nous avons établi deux procédures de test (une non-paramétrique et une paramétrique) afin de tester efficacement l'hypothèse nulle d'indépendance qui stipule que chaque composante de $\boldsymbol{\theta}$ et la/les sortie(s) d'intérêt considérée(s) sont indépendantes.

In fine, ces deux contributions méthodologiques ont été appliquées afin d'une part de sélectionner les paramètres incertains influents du modèle de comportement des gaz de fission tout en incorporant toute l'incertitude de la conductivité thermique λ du modèle thermique, et d'autre part, de caler les paramètres retenus à l'issue de l'analyse de sensibilité globale.

Cas applicatif: simulateur ALCYONE

Nous avons tout d'abord appliqué l'analyse de sensibilité par HSIC en support au calage conditionnel de $\boldsymbol{\theta}$ sachant λ afin d'identifier les paramètres pertinents pour déployer l'approche GP-LinCC sur les données simulées du code ALCYONE-CARACAS. La seconde contribution nous a ainsi permis de réduire le nombre de paramètres à caler de 11 paramètres initiaux à 4 paramètres. De plus, elle nous a permis de construire les différents modèles linéaires uniquement sur les paramètres identifiés lors de la préselection. Après ces étapes, la méthode GP-LinCC a été appliquée pour caler les paramètres incertains identifiés. Les résultats obtenus nous montrent que :

- la propagation *a posteriori* des incertitudes paramétriques du modèle de comportement doit être faite conjointement avec la propagation de l'incertitude de la conductivité thermique λ afin d'obtenir des fractions gaz relâchées simulées comparables. En d'autres termes, **ces résultats montrent qu'il y a bien un effet de compensation lorsque ces deux incertitudes paramétriques sont propagées conjointement.**

- La dépendance à λ des distributions des paramètres *a posteriori* est cohérente avec la physique du modèle de comportement des gaz de fission. C'est le cas en particulier pour les paramètres relatifs à la diffusion des gaz dans le combustible.

- Une possible discrépance de modèle non prise en compte dans cette étude est présente. En effet, les résidus standardisés obtenus après calage ne suivent pas une distribution gaussienne standard et la variance de ces résidus est strictement supérieure à 1. Ceci pourrait s'expliquer par les choix de modélisation de la méthode GP-LinCC els que l'utilisation de métamodèles linéaires, la sélection des quantités retenues pour l'étude (la conductivité thermique) ou encore à la négligence de l'incertitude de la puissance thermique $P(t)$ du combustible nucléaire.

Liste des contributions

- **Publications dans des revues internationales à comité de lecture**
  - ◇ O. Baldé, G. Damblin, A. Marrel, A. Bouloré et L. Giraldi, 2023.
    "Nonparametric Bayesian approach for quantifying the conditional uncertainty of input parameters in chained numerical models".
    Soumis au *SIAM/ASA Journal on Uncertianty Quantification (JUQ)*, Preprint https://arxiv.org/abs/2307.01111.
  - ◇ O. Baldé, G. Sarazin, A. Marrel, G. Damblin et A. Bouloré, 2024.
    "Kernel-based parameter screening for Bayesian calibration of chained numerical models: application to fuel performance simulation of pressurized water reactors".
    En cours de finalisation pour soumission d'ici fin février.
- **Communications orales**
  - ◇ O. Baldé, G. Damblin, A. Marrel, A. Bouloré et L. Giraldi, 2023.
    "Bayesian calibration for the quantification of conditional uncertainty of input parameters in chained numerical models".
    École Thématique sur les Incertitudes en Calcul Scientifique (ETICS), 08-13 octobre 2023, Lège-Cap-Ferret, France.
  - ◇ O. Baldé, G. Damblin, A. Marrel, A. Bouloré et L. Giraldi, 2023.
    "Bayesian calibration for the quantification of conditional uncertainty of input parameters in chained numerical models".
    European Network for Business and Industrial Statistics (ENBIS), 10-14 septembre 2023, Valence, Espagne.
  - ◇ O. Baldé, G. Damblin, A. Marrel, A. Bouloré et L. Giraldi, 2023.
    "Calage conditionnel bayésien d'un modèle numérique".
    Café Thésard du Département d'Études des Combustibles (DEC) du CEA du centre Cadarache, janvier 2023, Saint-Paul-lez-Durance, France.
  - ◇ O. Baldé, G. Damblin, A. Marrel, A. Bouloré et L. Giraldi, 2022.
    "Calage conditionnel bayésien d'un modèle numérique".
    53èmes Journées de Statistique de la Société Française de Statistique (SFdS), 13-17 juin 2022, Lyon, France.
  - ◇ O. Baldé, G. Damblin, A. Marrel, A. Bouloré et L. Giraldi, 2022.
    "Bayesian conditional calibration of a numerical model". *(Poster) Journées scientifiques du Consortium Industrie Recherche pour l'Optimisation et la QUantification d'incertitude pour les données Onéreuses (CIROQUO), 23-25 mai 2022, Grenoble, France.*
  - ◇ O. Baldé, G. Damblin, A. Marrel, A. Bouloré et L. Giraldi, 2022.
    "Bayesian conditional calibration of a numerical model". *Journées scientifiques de l'Institut des Sciences appliquées et de la Simulation (ISAS) du CEA, 16-17 mai 2022, Gif-sur-Yvette, France.*

Perspectives

Nos résultats ouvrent plusieurs pistes d'amélioration. Une première piste d'amélioration de la modélisation de la méthode GP-LinCC serait de remplacer le métamodèle linéaire par un métamodèle non-linéaire. On pourrait par exemple considérer un métamodèle de type régression par processus gaussien. Une autre piste pour le cas applicatif serait d'incorporer l'incertitude de la puissance thermique $P(t)$ dans l'étude. En effet, dans cette thèse elle a été négligée, ce qui pourrait être une hypothèse trop forte.

Extension au cas non-linéaire de la méthode GP-LinCC

La modélisation de la dépendance entre $\boldsymbol{\theta}$ et λ dans la méthode GP-linCC repose sur des processus gaussiens indépendants avec des hyperparamètres ϕ inconnus. Ces processus modélisent *a priori* chaque composante de la fonction de calage $\boldsymbol{\theta}(\lambda)$. De plus, l'approche GP-LinCC s'appuie sur sur la linéarité des fractions de gaz relâchée simulée par le modèle de comportement des gaz de fission par rapport à $\boldsymbol{\theta}(\lambda)$. La combinaison de ces deux éléments à travers la formule de Bayes nous permet d'obtenir des distributions *a posteriori* explicites. Cependant, en l'absence de linéarité de la sortie d'intérêt, l'obtention de distributions

a posteriori explicites n'est en général plus possible. L'estimation des hyperparamètres ϕ par maximisation de la vraisemblance marginale, qui ne serait plus explicite, pourrait aussi poser alors problème. Dans ces circonstances, des méthodes d'approximation telles que les méthodes MCMC ou les approches variationnelles peuvent être utilisées. **Nous avons choisi d'opter pour les approches variationnelles pour adapter la methode GP-LinCC au cadre non-lineaire. Les developpements mathematiques de cette extension, appelée approche GP-VICC (pour *Gaussian Process and Variational Inference based Conditional Calibration*), sont actuellement en cours.** Ce choix est motivé par notre souhait, comme pour GP-LinCC, d'obtenir des distributions *a posteriori* prédictives gaussiennes multivariées tronquées sur le support *a priori* des paramètres de calage. De plus, le support de la distribution *a posteriori* jointe utilisé pour apprendre la relation entre $\boldsymbol{\theta}$ et λ vit dans $\mathbb{R}^{pm}$ où m est le nombre de valeurs prises par la conductivité thermique. L'échantillonnage dans $\mathbb{R}^{pm}$ par un algorithme de Metropolis-Hastings pourrait s'avérer infaisable lorsque p et/ou m est grand. En outre, les approches variationnelles, bien qu'elles sous-estiment l'incertitude *a posteriori*, convergent plus rapidement que les algorithmes MCMC.

En finalisant ce présent manuscrit, nous avons pris connaissance des récents travaux de [Ye et al., 2022] traitant du calage conditionnel de deux modèles physiques chaînés. Leur approche de calage conditionnel bien que différente de notre approche GP-LinCC se base également sur les modèles de coupure [Plummer, 2015] afin de partitionner adéquatement les données disponibles des deux modèles. Par ailleurs, leur méthode traite le cas non-linéaire en remplaçant les deux modèles chaînés par deux processus gaussiens indépendants. Cependant, cette approche ne propage pas toute l'incertitude *a posteriori* des paramètres du premier modèle dans le second modèle. En effet, ils ne considèrent que le MAP des paramètres du premier modèle dans l'estimation de la distribution *a posteriori* des paramètres du second modèle. **Dans la suite de ces travaux de thèse, il serait intéressant dans un premier temps de proposer une généralisation de la méthode de [Ye et al., 2022] afin de pouvoir propager toute l'incertitude des paramètres du premier modèle dans le second modèle. Dans un second temps, il serait pertinent de l'adapter au cas applicatif afin de comparer ses performances à l'approche GP-VICC.**

Simulateur ALCYONE

Dans cette étude, nous avons uniquement pris en compte l'incertitude associée à la conductivité thermique, tandis que celles associées aux autres paramètres du modèle thermique ont été négligées (Figure 1.8 du Chapitre 1). Les résultats obtenus montrent l'existence possible d'une discrépance de modèle, probablement attribuable à l'omission de certaines quantités, notamment la puissance thermique $P(t)$. Il est important de souligner que l'indice de Sobol' du premier ordre, relatif à la puissance thermique et quantifiant son impact sur la température calculée en sortie du modèle thermique, atteint environ 0.25 (appelée *linear heat rate* sur la Figure 1.8 du Chapitre 1). Cela suggère que négliger cette incertitude pourrait constituer une hypothèse trop forte. Par conséquent, il serait intéressant, en plus de considérer l'incertitude liée à la conductivité thermique, d'intégrer également celle associée à $P(t)$ lors du processus de calage des paramètres incertains du modèle de comportement des gaz de fission. L'objectif serait alors d'inférer la distribution de $\boldsymbol{\theta} \mid \lambda, P(t)$. Cependant, nous ne disposons pas d'une distribution de probabilité sur $P(t)$ mais seulement d'un historique de puissance pour les différents crayons combustibles. Il faudrait donc préalablement inférer un modèle de distribution de probabilité pour $P(t)$ à partir de l'historique de puissance. Une première piste serait d'ajuster une distribution paramétrique (log-normale, exponentielle etc.) ou de modéliser $P(t)$ par un processus gaussien.

Bibliography

Alex, Mara, T. and Rakoto Joseph, O. (2008). Comparison of some efficient methods to evaluate the main effect of computer model factors. *Journal of Statistical Computation and Simulation*, 78(2):167–178. 15

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and computing*, 18:343–373. 8, 11, 12, 61, 64, 65, 126

Arendt, P. D., Apley, D. W., and Chen, W. (2012). Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability. *Journal of Mechanical Design*, 134(10):100908. 9

Aronszajn, N. (1950). Theory of Reproducing Kernels. *Transactions of the American mathematical society*, 68(3):337–404. 16, 69

Auder, B. and Iooss, B. (2008). Global sensitivity analysis based on entropy. In *Safety, reliability and risk analysis-Proceedings of the ESREL 2008 Conference*, pages 2107–2115. 68

Baccou, J., Zhang, J., Fillion, P., Damblin, G., Petruzzi, A., Mendizabal, R., Reventos, F., Skorek, T., Couplet, M., Iooss, B., et al. (2020). Sapium: A Generic Framework for a Practical and Transparent Quantification of Thermal-Hydraulic Code Model Input Uncertainty. *Nuclear Science and Engineering*, 194(8-9):721–736. 5

Bachoc, F. (2013). *Estimation paramétrique de la fonction de covariance dans le modèle de krigeage par processus gaussiens: application à la quantification des incertitudes en simulation numérique*. PhD thesis, Paris 7. 8, 36, 37, 144

Bachoc, F., Bois, G., Garnier, J., and Martinez, J.-M. (2014). Calibration and Improved Prediction of Computer Models by Universal Kriging. *Nuclear Science and Engineering*, 176(1):81–97. 7, 9, 30, 60, 63, 153

Barratt, S. (2018). A Matrix Gaussian Distribution. *arXiv preprint arXiv:1804.11010*. 35

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007). A Framework for Validation of Computer Models. *Technometrics*, 49(2):138–154. 7, 9

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*, 22:773–793. 33

Berlinet, A. and Thomas-Agnan, C. (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media. 69

Binois, M. and Wycoff, N. (2022). A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization. *ACM Transactions on Evolutionary Learning and Optimization*, 2(2):1–26. 12, 64

Bishop, C. M. (2006). Machine learning. *Machine learning*, 128(9). 124

Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784. 16, 68

Borgonovo, E. and Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3):869–887. 16

Bouloré, A. (2019). Importance of uncertainty quantification in nuclear fuel behaviour modelling and simulation. *Nuclear Engineering and Design*, 355:110311. 4

Bouloré, A., Struzik, C., Bouineau, V., Gaudier, F., Damblin, G., and Bernaud, S. (2023). Modelling of uo2 thermal conductivity: Improvement of the irradiation defects contribution and uncertainty quantification. *Nuclear Engineering and Design*, 407:112304. 19, 20, 21, 62, 79, 80, 91

Boulore, A., Struzik, C., and Gaudier, F. (2009). Modelling of the Uncertainty of Nuclear Fuel Thermal Behaviour Using the URANIE Framework. In *2009 First International Conference on Advances in System Simulation*, pages 50–55. IEEE. 22

Bouloré, A., Struzik, C., and Gaudier, F. (2010). Modeling of the Uncertainty of Nuclear Fuel Thermal Behavior. *Transactions*, 102(1):231–232. 20

Bouloré, A., Struzik, C., and Gaudier, F. (2012). Uncertainty and sensitivity analysis of the nuclear fuel thermal behavior. *Nuclear Engineering and Design*, 253:200–210. 13, 19, 20

Boulore, A., Struzik, C., Masson, R., Mailhe, P., and Largenton, R. (2017). Approach to better assess fission gas behaviors, applicable to fuels with complex microstructures. In *WRFPM 2017-2017 Water Reactor Fuel Performance Meeting*. WRFPM 2017-2017 Water Reactor Fuel Performance Meeting. 21

Broto, B., Bachoc, F., and Depecker, M. (2020). Variance Reduction for Estimation of Shapley Effects and Adaptation to Unknown Input Distribution. *SIAM/ASA Journal on Uncertainty Quantification*, 8(2):693–716. 15, 67

Brown, D. A. and Atamturktur, S. (2018). Nonparametric Functional Calibration of Computer Models. *Statistica Sinica*, pages 721–742. 10, 11, 26, 29

Brynjarsdottir, J. and O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11):114007. 9, 153

Campbell, K. (2006). Statistical calibration of computer simulations. *Reliability Engineering & System Safety*, 91(10-11):1358–1363. 4, 5, 7, 28, 60

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental modelling & software*, 22(10):1509–1518. 6, 61

Campolongo, F., Saltelli, A., and Cariboni, J. (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer physics communications*, 182(4):978–988. 6, 61

Carmassi, M., Barbillon, P., Keller, M., Parent, E., and Chiodetti, M. (2019). Bayesian calibration of a numerical code for prediction. 9

Chatterjee, S. (2021). A New Coefficient of Correlation. *Journal of the American Statistical Association*, 116(536):1–16. 15, 67

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast Parallel Kriging-Based Stepwise Uncertainty Reduction With Application to the Identification of an Excursion Set. *Technometrics*, 56(4):455–465. 33

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The american statistician*, 49(4):327–335. 8, 11, 31, 61, 64, 126

Ciuffo, B. and Azevedo, C. L. (2014). A Sensitivity-Analysis-Based Approach for the Calibration of Traffic Simulation Models. *IEEE Transactions on Intelligent Transportation Systems*, 15(3):1298–1309. 13

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368):829–836. 120

Clouvel, L., Iooss, B., Chabridon, V., Idrissi, M. I., and Robin, F. (2023). Variance-based importance measures in the linear regression context: Review, new insights and applications. 62

Cocci, R., Damblin, G., Ghione, A., Sargentini, L., and Lucor, D. (2022). Extension of the CIRCE methodology to improve the Inverse Uncertainty Quantification of several combined thermal-hydraulic models. *Nuclear Engineering and Design*, 398:111974. 20

Cole, D. (2020). *Parameter Redundancy and Identifiability*. CRC Press. 12, 34

Cornillon, P.-A. and Matzner-Løber, E. (2011). Régression spline et régression à noyau. In *Régression avec R*, pages 211–228. Springer. 131

Crestaux, T., Le Maıtre, O., and Martinez, J.-M. (2009). Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172. 15, 67

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments. *Journal of the American Statistical Association*, 86(416):953–963. 8, 11, 33, 61, 64

Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85(7):1283–1305. 16, 26, 62, 69, 74

Da Veiga, S. (2021). Kernel-based ANOVA decomposition and Shapley effects–Application to global sensitivity analysis. *arXiv preprint arXiv:2101.05487*. 16

Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021). *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. SIAM. 6, 12, 16, 62

Da Veiga, S., Gamboa, F., Lagnoux, A., Klein, T., and Prieur, C. (2023). New estimation of Sobol' indices using kernels. *arXiv preprint arXiv:2303.17832*. 15, 16, 67

Da Veiga, S. and Marrel, A. (2020). Gaussian process regression with linear inequality constraints. *Reliability Engineering & System Safety*, 195:106732. 49

Damblin, G. (2015). *Contributions statistiques au calage et à la validation des codes de calcul*. PhD thesis, Paris, AgroParisTech. 9, 28, 60

Damblin, G., Barbillon, P., Keller, M., Pasanisi, A., and Parent, É. (2018). Adaptive Numerical Designs for the Calibration of Computer Codes. *SIAM/ASA Journal on Uncertainty Quantification*, 6(1):151–179. 10, 12, 30, 63, 64

Damblin, G. and Gaillard, P. (2020). Bayesian inference and non-linear extensions of the CIRCE method for quantifying the uncertainty of closure relationships integrated into thermal-hydraulic system codes. *Nuclear Engineering and Design*, 359:110391. 12, 20, 64

Damblin, G. and Ghione, A. (2021). Adaptive use of replicated Latin Hypercube Designs for computing sobol'sensitivity indices. *Reliability Engineering & System Safety*, 212:107507. 15, 67

Damblin, G., Keller, M., Barbillon, P., Pasanisi, A., and Parent, E. (2016). Bayesian model selection for the validation of computer codes. *Quality and Reliability Engineering International*, 32(6):2043–2054. 9

De Lozzo, M. and Marrel, A. (2016). New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation*, 86(15):3038–3058. 16, 62, 69, 72, 74, 86, 87

De Lozzo, M. and Marrel, A. (2017). Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators. *Stochastic environmental research and risk assessment*, 31:1437–1453. 66

De Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in Industrial Practice: A guide to quantitative uncertainty management*. John Wiley & Sons. 5, 60

Delipei, G. (2019). *Development of an Uncertainty Quantification methodology for Multi-Physics Best Estimate analysis and application to the Rod Ejection Accident in a Pressurized Water Reactor*. PhD thesis, Université Paris Saclay (COmUE). 4, 17

Delipei, G.-K., Garnier, J., Le Pallec, J., and Normand, B. (2019). Uncertainty analysis methodology for multi-physics coupled rod ejection accident. In *International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C 2019)*. 69

Demay, C., Iooss, B., Le Gratiet, L., and Marrel, A. (2022). Model selection based on validation criteria for gaussian process regression: An application with highlights on the predictive variance. *Quality and Reliability Engineering International*, 38(3):1482–1500. 136

Derennes, P., Morio, J., and Simatos, F. (2018). Estimation of moment independent importance measures using a copula and maximum entropy framework. In *2018 Winter Simulation Conference (WSC)*, pages 1623–1634. IEEE. 68

Dighe, V. V., Becker, M., Göçmen, T., Sanderse, B., and van Wingerden, J.-W. (2022). Sensitivity analysis and Bayesian calibration of a dynamic wind farm control model: FLORIDyn. In *Journal of Physics: Conference Series*, volume 2265, page 022062. IOP Publishing. 12, 13, 61, 64, 65, 66

El Amri, M. R. and Marrel, A. (2022). Optimized HSIC-based tests for sensitivity analysis: Application to thermalhydraulic simulation of accidental scenario on nuclear reactor. *Quality and Reliability Engineering International*, 38(3):1386–1403. 69, 72, 86, 87

El Amri, M. R. and Marrel, A. (2024). More Powerful HSIC-Based Independence Tests, Extension to Space-Filling Designs and Functional Data. *to appear in 2024 in International Journal for Uncertainty Quantification*, 14(2):69–98. 16, 26, 49, 69, 73, 74, 86, 87, 112

Fellmann, N., Blanchet-Scalliet, C., Helbert, C., Spagnol, A., and Sinoquet, D. (2023). Kernel-based sensivity analysis for (excursion) sets. *arXiv preprint arXiv:2305.09268*. 84

Fiszeder, P. and Orzeszko, W. (2021). Covariance matrix forecasting using support vector regression. *Applied intelligence*, 51(10):7029–7042. 32

Fukumizu, K., Gretton, A., Schölkopf, B., and Sriperumbudur, B. K. (2008). Characteristic Kernels on Groups and Semigroups. *Advances in neural information processing systems*, 21. 16, 70

Gamboa, F., Gremaud, P., Klein, T., and Lagnoux, A. (2022). Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374. 15, 67

Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014). Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575 – 603. 66

Gamboa, F., Janon, A., Klein, T., Lagnoux, A., and Prieur, C. (2016). Statistical inference for Sobol pick-freeze Monte Carlo method. *Statistics*, 50(4):881–902. 14

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical Methods for Eliciting Probability Distributions. *Journal of the American statistical Association*, 100(470):680–701. 5

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). Bayesian Data Analysis (vol. 2). 5, 9, 11, 12

Ghanem, R., Higdon, D., and Owhadi, H. (2017). *Handbook of Uncertainty Quantification*. Springer. 5

Ghosh, J. K., Delampady, M., and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*, volume 725. Springer. 5

Gou, J., Miao, C., Duan, Q., Tang, Q., Di, Z., Liao, W., Wu, J., and Zhou, R. (2020). Sensitivity Analysis-Based Automatic Parameter Calibration of the VIC Model for Streamflow Simulations Over China. *Water Resources Research*, 56(1):e2019WR025968. 13, 61, 65

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. Chapman and Hall/CRC. 8, 11, 33, 64

Gretton, A. (2015). A simpler condition for consistency of a kernel independence test. *arXiv preprint arXiv:1501.06103*. 70

Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., and Smola, A. (2006). A Kernel Method for the Two-Sample-Problem. *Advances in neural information processing systems*, 19. 16, 69

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005a). Measuring statistical dependence with Hilbert-Schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer. 69, 70

Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., and Smola, A. (2007). A Kernel Statistical Test of Independence. *Advances in neural information processing systems*, 20. 16, 70, 72, 86

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005b). Kernel Methods for Measuring Independence. *Journal of Machine Learning Research*, 6(70):2075–2129. 16, 49

Gu, M. and Wang, L. (2018). Scaled Gaussian Stochastic Process for Computer Model Calibration and Prediction. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1555–1583. 5, 7, 9, 60, 63

Gu, M., Wang, X., and Berger, J. O. (2018). Robust Gaussian stochastic process emulation. *The Annals of Statistics*, 46(6A):3038–3066. 33, 144, 152

Guillet, J.-L., Guerin, Y., and Bonin, B. (2008). Combustibles nucléaires. vii, 17, 19, 20, 24

Haario, H., Saksman, E., and Tamminen, J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli*, pages 223–242. 12, 64, 65

Hamby, D. (1995). A Comparison of Sensitivity Analysis Techniques. *Health physics*, 68(2):195–204. 6, 61

Hamby, D. M. (1993). A Review of Sensitivity Analysis Techniques. https://www.osti.gov/biblio/10127564. 6, 61

Hamby, D. M. (1994). A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Environmental monitoring and assessment*, 32:135–154. 6, 61

Hart, J. L., Alexanderian, A., and Gremaud, P. A. (2017). Efficient computation of Sobol' indices for stochastic models. *SIAM Journal on Scientific Computing*, 39(4):A1514–A1530. 81

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer. 8, 31, 37, 131, 132, 136

Herin, M., Idrissi, M. I., Chabridon, V., and Iooss, B. (2022). Proportional marginal effects for global sensitivity analysis. *arXiv preprint arXiv:2210.13065*. 104

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008). Computer Model Calibration Using High-Dimensional Output. *Journal of the American Statistical Association*, 103(482):570–583. 9

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004). Combining Field Data and Computer Simulations for Calibration and Prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466. 7, 9, 152, 153

Hoeffding, W. (1992). A Class of Statistics with Asymptotically Normal Distribution. *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 308–334. 13, 66, 126

Hutter, F., Hoos, H., and Leyton-Brown, K. (2014). An Efficient Approach for Assessing Hyperparameter Importance. In *International conference on machine learning*, pages 754–762. PMLR. 8

Iooss, B. and Le Gratiet, L. (2019). Uncertainty and sensitivity analysis of functional risk curves based on Gaussian processes. *Reliability Engineering & System Safety*, 187:58–66. 13

Iooss, B. and Lemaître, P. (2015). A Review on Global Sensitivity Analysis Methods. *Uncertainty management in simulation-optimization of complex systems: algorithms and applications*, pages 101–122. 14, 67, 138

Iooss, B. and Marrel, A. (2019). Advanced Methodology for Uncertainty Propagation in Computer Experiments with Large Number of Inputs. *Nuclear Technology*, 205(12):1588–1606. 12, 64

Iooss, B. and Prieur, C. (2019). Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol' indices, numerical estimation and applications. *International Journal for Uncertainty Quantification*, 9(5). 62

Jacob, P. E., Murray, L. M., Holmes, C. C., and Robert, C. P. (2017). Better together? Statistical learning in models made of modules. *arXiv preprint arXiv:1708.08719*. 30

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2014). Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364. 14, 67

Jomard, G., Struzik, C., Boulore, A., Mailhé, P., Auret, V., and Largenton, R. (2014). CARACAS. An industrial model for the description of fission gas behavior in LWR-UO_2 fuel. In *Proc. WRFPM, 2014, Paper No. 100154*, Sendai, Japan. 21

Kamary, K., Keller, M., Barbillon, P., Gœury, C., and Parent, É. (2019). Computer code validation via mixture model estimation. *arXiv preprint arXiv:1903.03387*. 9

Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. K. (2018). Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences. *arXiv preprint arXiv:1807.02582*. 96

Kass, R. E. and Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules. *Journal of the American statistical Association*, 91(435):1343–1370. 5, 61, 124

Kazi-Aoual, F., Hitier, S., Sabatier, R., and Lebreton, J.-D. (1995). Refined approximations to permutation tests for multivariate inference. *Computational statistics & data analysis*, 20(6):643–656. 73, 86, 87, 112

Kendall, M. G. (1949). Rank correlation methods.]. *Journal of the Institute of Actuaries*, 75(1):140–141. 16, 68

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. 5, 7, 8, 9, 10, 28, 29, 30, 60, 61, 63, 152

Kennedy, M. C. and O'Hagan, A. (2001). Supplementary details on Bayesian Calibration of Computer Models. Technical report, Internal Report. URL http://www. shef. ac. uk/˜ st1ao/ps/calsup. ps. 9

Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., and Shah, N. (2009). Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering & System Safety*, 94(7):1135–1148. 62

Le Gratiet, L., Cannamela, C., and Iooss, B. (2014). A Bayesian Approach for Global Sensitivity Analysis of (Multifidelity) Computer Codes. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):336–363. 33

Lebel, D., Soize, C., Fünfschilling, C., and Perrin, G. (2019). Statistical inverse identification for nonlinear train dynamics using a surrogate model in a Bayesian framework. *Journal of Sound and Vibration*, 458:158–176. 64

Lee, Y. and Park, J.-S. (2020). Generalized Nonlinear Least Squares Method for the Calibration of Complex Computer Code Using a Gaussian Process Surrogate. *Entropy*, 22(9):985. 7

Lefebvre, L., Segond, M., Spaggiari, R., Le Gratiet, L., Deri, E., Iooss, B., and Damblin, G. (2023). Improving the Predictivity of a Steam Generator Clogging Numerical Model by Global Sensitivity Analysis and Bayesian Calibration Techniques. *Nuclear Science and Engineering*, pages 1–14. 13

Leoni, N. (2022). *Bayesian inference of model error for the calibration of two-phase CFD codes*. PhD thesis, Ecole polytechnique. 9, 10

Liu, F., Bayarri, M., and Berger, J. (2009). Modularization in Bayesian Analysis, with Emphasis on Analysis of Computer Models. *Bayesian Analysis*, 4(1):119–150. 8, 9, 12, 30, 61, 64, 65

López-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2018). Finite-Dimensional Gaussian Approximation with Linear Inequality Constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1224–1255. 49

Marelle, V., Goldbronn, P., Bernaud, S., Castelier, É., Julien, J., Nkonga, K., Noirot, L., and Ramière, I. (2016). New developments in ALCYONE 2.0 fuel performance code. In *Top Fuel 2016-Light Water Reactor (LWR) Fuel Performance Meeting*. vii, 5, 17, 21

Marie, N., Li, S., Marrel, A., Marquès, M., Bajard, S., Tosello, A., Perez, J., Grosjean, B., Gerschenfeld, A., Anderhuber, M., et al. (2021). Vvuq of a thermal-hydraulic multi-scale tool on unprotected loss of flow accident in SFR reactor. *EPJ N-Nuclear Sciences & Technologies*, 7:3. 69

Marque-Pucheu, S., Perrin, G., and Garnier, J. (2016). Calibration of Nested Computer Models. In *VII European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress), Crete Island, Greece, 5-10 June 2016*. 29, 33

Marque-Pucheu, S., Perrin, G., and Garnier, J. (2017). Calibration and prediction of two nested computer codes. preprint. 33

Marrel, A. and Iooss, B. (2023). Probabilistic surrogate modeling by Gaussian process: A new estimation algorithm for more reliable prediction. https://cea.hal.science/cea-04322818. 64

Marrel, A., Iooss, B., and Chabridon, V. (2022). The ICSCREAM Methodology: Identification of Penalizing Configurations in Computer Experiments Using Screening and Metamodel—Applications in Thermal Hydraulics. *Nuclear Science and Engineering*, 196(3):301–321. 12, 33, 64

Marrel, A., Iooss, B., Jullien, M., Laurent, B., and Volkova, E. (2011). Global sensitivity analysis for models with spatially dependent outputs. *Environmetrics*, 22(3):383–397. 13

Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009). Calculations of Sobol' indices for the Gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742–751. 11, 15, 33, 67

Marrel, A., Iooss, B., Van Dorpe, F., and Volkova, E. (2008). An efficient methodology for modeling complex computer codes with gaussian processes. *Computational Statistics & Data Analysis*, 52(10):4731–4744. 33

Mathieu, A., Vidal, T., Jullien, A., Wu, Q., Chambon, C., Bayol, B., and Cournède, P.-H. (2018). A new methodology based on sensitivity analysis to simplify the recalibration of functional–structural plant models in new conditions. *Annals of botany*, 122(3):397–408. 13

McKay, M., Beckman, R., and Conover, W. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. *Technometrics* 21 (2): 239–245. ISSN 0040-1706. 32, 38, 66, 94

McKay, M. D. (1997). Nonparametric variance-based methods of assessing uncertainty importance. *Reliability Engineering & System Safety*, 57(3):267–279. 13

McKay, M. D., Morrison, J. D., and Upton, S. C. (1999). Evaluating prediction uncertainty in simulation models. *Computer Physics Communications*, 117(1-2):44–51. 15

Meynaoui, A. (2019). *New developments around dependence measures for sensitivity analysis: application to severe accident studies for generation IV reactors*. PhD thesis, Toulouse, INSA. 26

Michel, B., Ramière, I., Viallard, I., Introini, C., Lainet, M., Chauvin, N., Marelle, V., Boulore, A., Helfer, T., Masson, R., et al. (2021). Two fuel performance codes of the PLEIADES platform: ALCYONE and GERMINAL. In *Nuclear Power Plant Design and Analysis Codes*, pages 207–233. Elsevier. vii, 4, 5, 17, 18, 19, 21, 28, 62, 74

Minunno, F., van Oijen, M., Cameron, D., and Pereira, J. (2013). Selecting Parameters for Bayesian Calibration of a Process-Based Model: A Methodology Based on Canonical Correlation Analysis. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):370–385. 12, 13, 68

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174. 6, 61, 67

Morris, M. D. and Mitchell, T. J. (1995). Exploratory Designs for Computational Experiments. *Journal of statistical planning and inference*, 43(3):381–402. 6, 61

Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B., et al. (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141. 70

Murphy, K. P. (2022). *Probabilistic Machine Learning: an introduction*. MIT press. 5, 31

Nagel, J. B., Rieckermann, J., and Sudret, B. (2020). Principal component analysis and sparse polynomial chaos expansions for global sensitivity analysis and model calibration: Application to urban drainage simulation. *Reliability Engineering & System Safety*, 195:106737. 12, 13, 61, 64, 65, 66

Neal, R. M. et al. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2. 12

Oakley, J. E. and O'Hagan, A. (2007). Uncertainty in prior elicitations: a nonparametric approach. *Biometrika*, 94(2):427–441. 5

Owen, A. B. and Prieur, C. (2017). On Shapley Value for Measuring Importance of Dependent Inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):986–1002. 62

O'Hagan, A. (2019). Expert knowledge Elicitation: Subjective but Scientific. *The American Statistician*, 73(sup1):69–81. 5

Paulo, R. (2005). Default priors for Gaussian processes. *The Annals of Statistics*, 33(2):556 – 582. 8

Perret, G., Wicaksono, D., Clifford, I. D., and Ferroukhi, H. (2022). Global Sensitivity Analysis and Bayesian Calibration on a Series of Reflood Experiments with Varying Boundary Conditions. *Nuclear Technology*, 208(4):711–722. 12, 61, 65

Plischke, E. and Borgonovo, E. (2015). Copula-based Sensitivity Measures of Computer Experiments. *Safety and Reliability of Complex Engineered Systems*. 68

Plischke, E., Borgonovo, E., and Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3):536–550. 68

Plumlee, M. (2017). Bayesian Calibration of Inexact Computer Models. *Journal of the American Statistical Association*, 112(519):1274–1285. 9

Plumlee, M., Joseph, V. R., and Yang, H. (2016). Calibrating Functional Parameters in the Ion Channel Models of Cardiac Cells. *Journal of the American Statistical Association*, 111(514):500–509. 10, 11

Plummer, M. (2015). Cuts in Bayesian graphical models. *Statistics and Computing*, 25:37–43. 30, 170

Qian, P. Z. and Wu, C. J. (2008). Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments. *Technometrics*, 50(2):192–204. 9

Rahman, S. (2016). The f-sensitivity index. *SIAM/ASA journal on uncertainty quantification*, 4(1):130–162. 68

Rasmussen, C. E., Williams, C. K., et al. (2006). *Gaussian Processes for Machine Learning*, volume 1. Springer. 8, 11, 32, 33, 36, 64, 144

Ratto, M., Tarantola, S., and Saltelli, A. (2001). Sensitivity analysis in model calibration: GSA-GLUE approach. *Computer Physics Communications*, 136(3):212–224. 13

Reich, B. J. and Ghosh, S. K. (2019). *Bayesian Statistical Methods*. Chapman and Hall/CRC. 5, 34, 144

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and computing*, 5:121–125. 143

Robert, C. P., Casella, G., and Casella, G. (1999). *Monte Carlo Statistical Methods*, volume 2. Springer. 8, 9, 11, 29, 31, 61, 64, 126, 143

Rollón de Pinedo, Á., Couplet, M., Iooss, B., Marie, N., Marrel, A., Merle, E., and Sueur, R. (2021). Functional Outlier Detection by Means of h-Mode Depth and Dynamic Time Warping. *Applied Sciences*, 11(23):11475. 69

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical science*, 4(4):409–423. 11, 33, 61, 64

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer physics communications*, 145(2):280–297. 14, 62, 67

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons. 6, 16, 61, 62

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., et al. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, volume 1. Wiley Online Library. 6, 12

Saltelli, A., Vidoni, D., and Mascherini, M. (2009). Recommended Practices in Global Sensitivity Analysis. In *Uncertainties in Environmental Modelling and Consequences for Policy Making*, pages 183–202. Springer. 20, 126

Santner, T. J., Williams, B. J., and Notz, W. I. (2018). *The Design and Analysis of Computer Experiments*. Springer. 5, 6, 9, 11, 12, 15, 28, 60

Santner, T. J., Williams, B. J., Notz, W. I., and Williams, B. J. (2003). *The Design and Analysis of Computer Experiments*, volume 1. Springer. 60, 61, 63, 64, 66

Sarazin, G., Marrel, A., da Veiga, S., and Chabridon, V. (2023). New insights into the feature maps of Sobolev kernels: application in global sensitivity analysis. 103

Saveleva, E., Svitelman, V., Blinov, P., and Valetov, D. (2021). Sensitivity analysis and model calibration as a part of the model development process in radioactive waste disposal safety assessment. *Reliability Engineering & System Safety*, 210:107521. 13

Serfling, R. (1980). Approximation Theorems of Mathematical Statistics. *Wiley Series in Probability and Statistics*. 68, 71, 99, 100

Simon-Gabriel, C.-J. and Schölkopf, B. (2018). Kernel Distribution Embeddings: Universal Kernels, Characteristic Kernels and Kernel Metrics on Distributions. *The Journal of Machine Learning Research*, 19(1):1708–1736. 70

Sobol, I. M. (1993). Sensitivity Analysis for Non-linear Mathematical Models. *Math. Modeling Comput. Exp.*, 1:407–414. 13, 20, 65, 66, 126, 127

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280. 13, 61, 66

Song, L., Smola, A., Gretton, A., Bedo, J., and Borgwardt, K. (2012). Feature Selection via Dependence Maximization. *Journal of Machine Learning Research*, 13(5). 71, 84

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101. 16, 68

Sriperumbudur, B., Fukumizu, K., and Lanckriet, G. (2010a). On the relation between universality, characteristic kernels and RKHS embedding of measures. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 773–780. JMLR Workshop and Conference Proceedings. 74, 96, 97

Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. R. (2010b). Hilbert Space Embeddings and Metrics on Probability Measures. *The Journal of Machine Learning Research*, 11:1517–1561. 95, 96, 97

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer Science & Business Media. 69

Steurer, M., Hill, R. J., and Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research*, 38(2):99–129. 132, 153

Struzik, C. and Boulore, A. (2010). Contribution of the Uncertainty Analysis Methods to the Interpretation of Nuclear Fuel Experimental Irradiations, invited. *Transactions of the American Nuclear Society*, 102:233–234. 22

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7):964–979. 15, 67

Surjanovic, S. and Bingham, D. (2013). Virtual Library of Simulation Experiments: Test Functions and Datasets. http://www.sfu.ca/~ssurjano. 40

Szabó, Z. and Sriperumbudur, B. K. (2017). Characteristic and Universal Tensor Product Kernels. *Journal Machine Learning Research*, 18(233):1–29. 97

Tissot, J.-Y. and Prieur, C. (2015). A randomized orthogonal array-based procedure for the estimation of first-and second-order Sobol' indices. *Journal of Statistical Computation and Simulation*, 85(7):1358–1381. 15, 67

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66:1352–1362. 153

Trucano, T. G., Swiler, L. P., Igusa, T., Oberkampf, W. L., and Pilch, M. (2006). Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering & System Safety*, 91(10-11):1331–1357. 7, 60

Tuo, R. and Wu, C. F. J. (2015). Efficient Calibration for Imperfect Computer Models. *The Annals of Statistics*, 43(6):2331 – 2352. 5, 7, 9, 28

Van Oijen, M., Cameron, D., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.-E., Kiese, R., Rahn, K.-H., Werner, C., and Yeluripati, J. (2011). A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest. *Agricultural and Forest Meteorology*, 151(12):1609–1621. 12, 61, 64, 65

Wagner, P.-R., Fahrni, R., Klippel, M., Frangi, A., and Sudret, B. (2020). Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels. *Engineering structures*, 205:110063. 13

Wei, P., Lu, Z., and Song, J. (2014). Moment-independent sensitivity analysis using copula. *Risk Analysis*, 34(2):210–222. 68

Willmann, H., Nitzler, J., Brandstäter, S., and Wall, W. A. (2022). Bayesian calibration of coupled computational mechanics models under uncertainty based on interface deformation. *Advanced Modeling and Simulation in Engineering Sciences*, 9(1):24. 28, 60, 63

Wong, R. K., Storlie, C. B., and Lee, T. C. (2017). A Frequentist Approach to Computer Model Calibration. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(2):635–648. 5, 7, 28, 60

Wu, X., Kozlowski, T., Meidani, H., and Shirvan, K. (2018). Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, Part 1: Theory. *Nuclear Engineering and Design*, 335:339–355. 9, 12, 28, 29, 60, 63, 64

Wu, X., Mui, T., Hu, G., Meidani, H., and Kozlowski, T. (2017). Inverse uncertainty quantification of TRACE physical model parameters using sparse gird stochastic collocation surrogate model. *Nuclear Engineering and Design*, 319:185–200. 12, 61, 64, 65

Wu, X., Shirvan, K., and Kozlowski, T. (2019). Demonstration of the relationship between sensitivity and identifiability for inverse uncertainty quantification. *Journal of computational physics*, 396:12–30. 12, 13, 61, 65, 66

Xu, X., Sun, C., Huang, G., and Mohanty, B. P. (2016). Global sensitivity analysis and calibration of parameters for a physically-based agro-hydrological model. *Environmental Modelling & Software*, 83:88–102. 12, 61, 65, 67

Ye, J., Mahmoudi, M., Karayagiz, K., Johnson, L., Seede, R., Karaman, I., Arroyave, R., and Elwany, A. (2022). Bayesian calibration of multiple coupled simulation models for metal additive manufacturing: A bayesian network approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 8(1):011111. 170

Yi, D. H., Kim, D. W., and Park, C. S. (2019). Parameter identifiability in Bayesian inference for building energy models. *Energy and Buildings*, 198:318–328. 10

Zambrano-Bigiarini, M., Zając, Z., and Tarantola, S. (2013). Global Sensitivity Analysis for the calibration of a fully-distributed hydrological model. https://www.gdr-mascotnum.fr/media/mascot13zambrano-poster.pdf. 12, 61, 65, 67

Zhang, Q., Filippi, S., Gretton, A., and Sejdinovic, D. (2018). Large-scale kernel methods for independence testing. *Statistics and Computing*, 28:113–130. 72, 86

Zhou, C., Shi, Z., Kucherenko, S., and Zhao, H. (2022). A unified approach for global sensitivity analysis based on active subspace and Kriging. *Reliability Engineering & System Safety*, 217:108080. 12, 64

Ziegel, J., Ginsbourger, D., and Dümbgen, L. (2022). Characteristic kernels on Hilbert spaces, Banach spaces, and on sets of measures. *arXiv preprint arXiv:2206.07588*. 74, 96, 97

Résumé

Dans le cadre des études des comportements des systèmes physiques complexes tels que les réacteurs nucléaires, les simulateurs numériques sont devenus des outils essentiels pour la modélisation, l'analyse, la compréhension et la prévision des phénomènes physiques impliqués. Ces simulateurs prennent souvent un grand nombre de paramètres en entrée, lesquels sont souvent entachés d'incertitudes, ce qui se traduit par des sorties également incertaines. Ainsi, il est crucial avant toute utilisation dans un contexte industriel, de quantifier et de réduire les différentes sources d'incertitude. Le processus de calage de modèle vise ainsi à réduire et à quantifier au mieux les incertitudes des paramètres en entrée, en se basant sur les données expérimentales et simulées disponibles. Il existe deux types de calage de modèle : le calage déterministe et le calage bayésien. Le calage bayésien est une méthode qui repose sur une approche probabiliste pour quantifier les incertitudes paramétriques par des distributions de probabilité. Dans cette thèse, nous nous sommes intéressés au calage bayésien conditionnel de deux modèles numériques chaînés simulant le comportement du combustible dans un réacteur à eau pressurisée. Plus précisément, l'objectif est de réaliser un calage bayésien des paramètres incertains du second modèle conditionnellement à toute l'incertitude *a posteriori* des paramètres incertains du premier modèle numérique. Pour ce faire, nous avons proposé une nouvelle méthodologie d'inférence bayésienne basée sur des processus gaussiens et appelée GP-LinCC (pour *Gaussian Process and Linearization-based Conditional Calibration*). La mise en œuvre pratique de cette nouvelle approche a nécessité le développement d'une méthode d'analyse de sensibilité afin de sélectionner préalablement les paramètres à caler du second modèle tout en prenant en compte toute l'incertitude des paramètres du premier modèle. Cette méthode d'analyse de sensibilité globale en support au calage conditionnel est basée sur des mesures de dépendance de type HSIC (*Hilbert-Schmidt Independence Criterion*). Enfin, ces deux contributions méthodologiques ont été appliquées au simulateur chaîné ALCYONE-CARACAS afin de quantifier les incertitudes paramétriques du code CARACAS simulant le comportement des gaz de fission conditionnellement à l'incertitude de la conductivité thermique du modèle thermique.

Mots clés: Calage bayésien, Processus gaussiens, Analyse de sensibilité globale, Critère d'indépendance Hilbert-Schmidt, Test d'indépendance, simulation combustible nucléaire.

Abstract

Nowadays, numerical models have become essential tools for modeling, understanding, analyzing and predicting the physical phenomena involved in complex physical systems such as nuclear power plants. Such numerical models often take a large number of uncertain input parameters, thus leading to uncertain outputs as well. Before any industrial use of those numerical models, an important step is therefore to reduce and quantify these uncertainties as much as possible. In this context, the goal of model calibration is to reduce and quantify the uncertainties of the input parameters based on available experimental and simulated data. There are two types of model calibration: deterministic calibration and Bayesian calibration. The latter quantifies parameter uncertainties by probability distributions. This thesis deals with the conditional Bayesian calibration of two chained numerical models. The objective is

to calibrate the uncertain parameters of the second model while taking into account the uncertain parameters of the first model. To achieve this, a new Bayesian inference methodology called GP-LinCC (Gaussian Process and Linearization-based Conditional Calibration) was proposed. In practice, the deployment of this new approach has required a preliminary step of global sensitivity analysis to identify the most significant input parameters to calibrate in the second model, while considering the uncertainty of the parameters of the first model. To do this, an integrated version of the HSIC (Hilbert-Schmidt Independence Criterion) was used to define well-suited sensitivity measures and the theoretical properties of their nested Monte Carlo estimators were investigated. Finally, these two methodological contributions have been applied to the multi-physics application called ALCYONE, to quantify the uncertain parameters of the CARACAS code (second model) simulating the behavior of fission gases in the pressurized water reactor conditionally on the uncertainty of the parameter conductivity of the thermal model (first model).

Keywords: Bayesian calibration, Gaussian process, Global sensitivity analysis, Hilbert-Schmidt independence criterion, Independence testing, Nuclear fuel simulation.