A Spectral Approach of Multi-Scale Metamodelling Applied to Acoustic Propagation

Alexandre Goupy

université
PARIS-SACLAY

Thèse de doctorat

NNT: 2021UPASM010

Une approche spectrale de la métamodélisation multi-échelle appliquée à la propagation acoustique

A Spectral Approach of Multi-Scale Metamodelling Applied to Acoustic Propagation

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 574, Mathématiques Hadamard (EDMH)
Spécialité de doctorat: Mathématiques appliquées

Unité de recherche: Université Paris-Saclay, CNRS,
ENS Paris-Saclay, Centre Borelli,
91190, Gif-sur-Yvette, France.
Référent: ENS Paris-Saclay

Thèse présentée et soutenue à Paris-Saclay,
le 26 mars 2021, par

Alexandre GOUPY

Composition du jury:

Frédéric Dias Professeur - UC Dublin & ENS Paris-Saclay	Président
Régis Marchiano Professeur - Sorbonne Université	Rapporteur
Julien Diaz Directeur de recherche - INRIA	Rapporteur

Direction de la thèse:

Didier Lucor Directeur de recherche - LIMSI, CNRS	Directeur
Christophe Millet Ingénieur chercheur - CEA, DAM	Directeur

Résumé

De nombreuses méthodes permettent de simuler numériquement la propagation d'une onde dans un milieu complexe avec une excellente précision. Cependant, la prise en compte des fluctuations du milieu de propagation requiert un traitement statistique nécessitant un grand nombre d'appel à des codes de calcul souvent coûteux. Afin de rendre accessible ces études nous proposons la construction d'un métamodèle basé sur une décomposition en polynômes de chaos des modes normaux. Cette approche permet de restituer les statistiques des signaux se propageant dans un milieu aléatoire avec un coût de calcul moindre.
Les applications proposées dans cette thèse concernent la propagation d'ondes acoustiques dans l'atmosphère terrestre. En effet, les fluctuations météorologiques modifiant considérablement les conditions de propagation, leur prise en compte est indispensable. Le coût numérique de la simulation sur un domaine de plusieurs centaines de milliers de kilomètres carrés justifie pleinement l'utilisation d'un métamodèle. Une application à la localisation de source couplant ces techniques de métamodèlisation avec une approche bayésienne est aussi proposée. En effet, le cadre bayésien permet une résolution du problème inverse dans un cadre probabiliste capable de prendre en compte les fluctuations du milieu et l'incertitude sur la localisation de la source.

Abstract

There exists many numerical methods to simulate wave propagation through complex media with a very good precision. However, taking into account the fluctuations of the propagation medium necessitates a statistical approach implying a prohibitive numerical cost. To have those studies affordable, we propose the construction of a metamodel based on a polynomial chaos decomposition of normal modes. This approach presents the great advantage to give statistics of signals propagating in a random medium at an affordable numerical cost.
Those results are illustrated with acoustic propagation in the atmosphere. In fact, meteorological fluctuations have a critical impact on the propagation, it is therefore essential to take them into account. The numerical cost of a simulation over thousands of kilometers fully justifies the use of a metamodel. An application to source localization is proposed to illustrate the joint use of a metamodel and a bayesian inversion. The bayesian framework allows a resolution of the inverse problem in a probabilistic context able to take into account the fluctuations of the medium and uncertainties due to unknown source localization.

Table des matières

Chapitre 0

Motivations et synthèse des résultats obtenus

Ce premier chapitre, en français, présente les motivations ayant conduites au projet de recherche ainsi que les problèmes présentés dans les différents chapitres qui constituent le manuscrit. Il s'agit ici de donner un rapide aperçu du travail mené pendant la thèse en le replaçant dans son contexte applicatif.

1 Origines du projet de recherche doctorale

1.1 Les infrasons dans le système de surveillance international

Ouvert à la signature en 1996, le traité d'interdiction complète des essais nucléaires (TICE) interdit tout essai nucléaire quelle que soit l'énergie dégagée. Un système de surveillance international (SSI) a donc été développé pour enregistrer les explosions à la surface de la Terre, qu'elles soient souterraines, atmosphériques ou sous-marines. Pour cela, le SSI dispose de quatre types de capteurs : sismique (explosions souterraines), hydroacoustique (explosions sous-marines), infrasonores (explosions atmosphériques) ainsi que des capteurs de concentrations en radionucléides permettant quant à eux de distinguer une explosion conventionnelle d'une explosion nucléaire. Nous nous intéresserons dans ce manuscrit à la technologie infrasonore qui repose sur un réseau de soixante stations réparties dans 35 pays. Les stations infrasons sont équipées de microbarographes capables de mesurer et d'enregistrer au niveau du sol les variations de pression produites par la propagation des ondes infrasonores ou infrasons. Les infrasons sont des ondes acoustiques possédant une fréquence plus basse que la borne inférieure de la bande de fréquence audible par l'oreille humaine qui varie en général de 20 à 20 000 Hz. Ils sont produits aussi bien par l'activité humaine (explosions, tirs de carrière, éoliennes, avions supersoniques, fusées, etc.) que par des sources d'origine naturelle (éruptions volcaniques, orages, topographie, houle océanique, écoulements atmosphériques, séismes, météorites, etc.).

Les objectifs du centre international de données (CID) de Vienne qui traite l'ensemble des données reçues sont : 1 – de détecter et distinguer tous les événements malgré les différents bruits atmosphériques, 2 – d'associer les signaux reçus par plusieurs stations et correspondant au même événement, 3 – localiser l'origine de chacun des événements et 4 – caractériser les sources de ces signaux c'est à dire calculer l'energie associée [Mialle et al., 2019]. La détection est basée sur un algorithme de corrélation croisée [Cansi, 1995] qui utilise les mesures des différents capteurs de chaque station. Les différentes détections sont ensuite associées automatiquement pour former des événements qui seront ensuite étudiés par des analystes. En 2010, Arora *et al.* ont proposé un algorithme d'inférence bayésienne pour associer les détections et localiser les sources [Arora et al., 2010]. Cet algorithme était initialement développé pour les détections sismiques mais au vu des performances de cet algorithme des travaux ont été initiés dès 2015 pour adapter cet algorithme aux spécificités des infrasons. En effet, les stations infrasonores détectent de nombreux événements d'origine naturelle qui présentent des arrivées supplémentaires. Par ailleurs, les incertitudes liées au milieu de propagation sont beaucoup plus importantes dans le cas de la propagation infrasonore que dans le cas sismique ou hydro-acoustique.
En effet, pour les infrasons, les conditions de propagation sont déterminées par les conditions météorologiques : l'évolution de la température et des vents en fonction de l'altitude. Les conditions usuelles de températures favorisent la propagation infrasonores via la création de guides d'onde mais les profils de vent peuvent ensuite accentuer ou empêcher cette propagation. Dans les deux cas, les caractéristiques de l'onde reçue dépendront fortement des conditions météorologiques entre la source et la station.

1.2 Variabilité atmosphérique et propagation infrasonore

Les ondes infrasonores peuvent se propager sur des centaines voire des milliers de kilomètres. Sur de telles distances, la variabilité des conditions météorologiques se répercute directement sur les signaux reçus. La prise en compte des données météorologiques s'appuie sur les données atmosphériques transmises par le CEPMMT (Centre Européen de Prévision Météorologique à Moyen Terme) et des codes de propagation ou de mécanique des fluides selon les cas. L'état de l'atmosphère est calculé au CEPMMT à partir d'une méthode d'assimilation de données qui affine au mieux la représentation en combinant modèles (lois physiques, etc.) et observations. Il subsiste des incertitudes qui s'étendent sur plusieurs échelles spatiales et temporelles et qu'il est difficile de quantifier. Aux échelles planétaires, ces incertitudes sont directement liées à la variabilité naturelle du climat. À plus petite échelle, les incertitudes peuvent être décrites sous la forme d'une composante aléatoire censée se substituer à la physique non résolue dans les modèles du CEPMMT. Cette composante est relativement complexe à modéliser puisqu'elle dépend de l'altitude, des sources qui en sont à l'origine et des caractéristiques de l'écoulement à grande échelle.

Dans les codes de propagation, les données atmosphériques entreront en compte via le profil de la vitesse du son dans l'atmosphère. En effet, celle-ci dépend des profils

de température et de vent dans la direction de propagation :

$$c(z) = \sqrt{\gamma RT\,(z)} + U\,(z) \tag{1}$$

où γ est le coefficient de Laplace, R la constante des gaz parfaits, $T(z)$ le profil de température en fonction de l'altitude et $U(z)$ le profil de vent dans la direction de propagation. La figure 1 présente (à gauche) la variabilité de cette vitesse du son au cours de l'année 2018 entre la station norvégienne I37NO et le site militaire de Hukkakero. Les profils des 1^{er} janvier et juin montrent une différence marquée à partir de 40 km d'altitude. Cette différence s'explique par la présence d'un jet stratosphérique à ces altitudes dont l'orientation change en fonction des saisons. Le site militaire de Hukkakero est utilisé tous les étés pour détruire des munitions ; en 2018 ces explosions se sont étalées du 16 au 27 août et la variabilité des profils de vitesse du son sur cette période est indiquée en rouge sur la figure. La source est donc connue, elle est localisée à 320 km de la station et correspond à une énergie de l'ordre de 20-30 t[1]. À droite de la figure 1, l'enveloppe des signaux reçus à la station est représentée pour les différentes années. Les sources étant identiques chaque jour, la variabilité observée traduit uniquement la variabilité des conditions atmosphériques. Au delà de la variabilité des états atmosphériques décrits par les données du CEPMMT, les perturbations non résolues par les modèles météorologiques ont un impact important sur la propagation des infrasons [Waxler and Assink, 2019]. Parmi ces phénomènes physiques, les ondes de gravités sont particulièrement actives autour de 50 km d'altitude [Chunchuzov and Kulichkov, 2019].

Afin d'associer, localiser et caractériser précisément les événements explosifs à l'aide des infrasons, il est donc central de savoir comment les signaux ont été modifiés lors de la propagation. Pour cela, il faudrait être capable de calculer la fonction de transfert liant le signal source et les signaux reçus à la station mais cette fonction dépend des propriétés atmosphériques entre la source et la station au moment de l'événement. Il est bien sûr illusoire d'espérer connaître précisément les conditions météorologiques à tout temps et en tout lieu à la surface du globe[2]. Il convient donc, en tâchant de la quantifier assez précisément, de prendre en compte la variabilité du milieu de propagation comme une incertitude dont on souhaite mesurer l'impact sur la propagation des infrasons.

1.3 Les différents modèles de propagation

Les techniques numériques de résolution de l'équation d'onde sont bien connues et documentées depuis les années soixante-dix. Le développement de méthodes d'ordre arbitrairement élevé a permis des simulations numériques avec des domaines comportant jusqu'à 10^{10} mailles. En dépit d'un coût élevé, les techniques issues de la CFD

[1] L'énergie libérée lors d'une explosion est donnée en tonne de TNT (t), c'est-à-dire l'énergie libérée par l'explosion d'environ une tonne de TNT.

[2] Cette impossibilité se comprend d'autant mieux si l'on prend en compte des phénomènes aléatoires par nature : les couches turbulentes dans la basse atmosphère ou le déferlement des ondes de gravité dans la moyenne atmosphère.

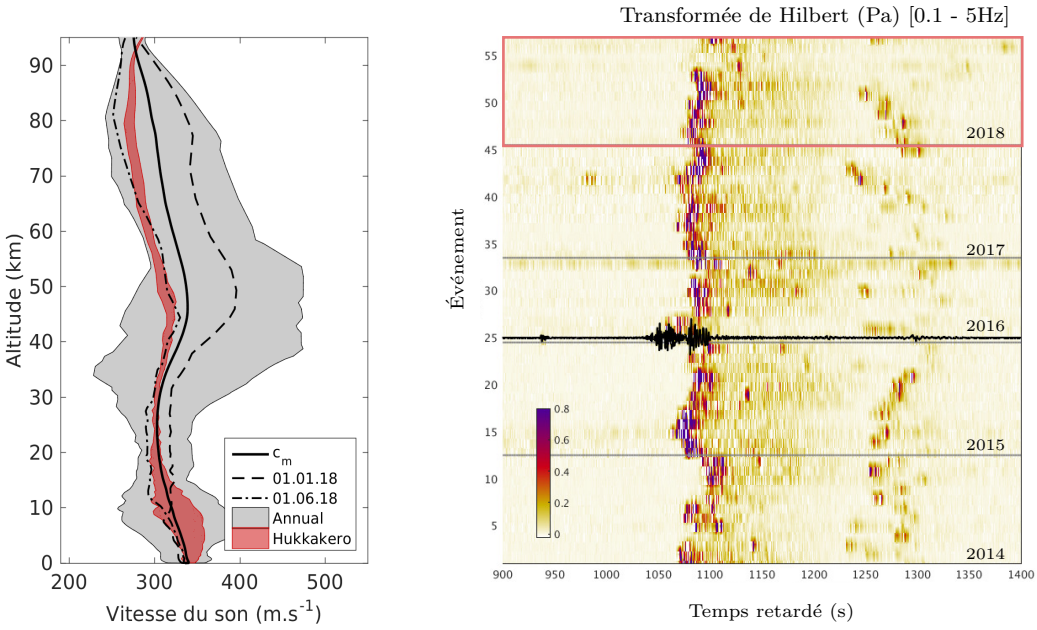

FIGURE 1 : Variabilité de la vitesse du son entre la station I37NO et la base militaire d'Hukkakero au cours de l'année 2018 (en gris) et du 16 au 27 août en rouge. Le profil moyen c_m présente une forme classique en W. À droite, les enveloppes des signaux reçus à I37NO sont représentées pour les différentes expériences de 2014 à 2018. Le cadre rouge correspond aux signaux reçus entre les 16 et 27 août.

(Computational Fluid Dynamics) sont peu adaptés à la modélisation de la propagation acoustique [Millet et al., 2007]. Dans ces techniques, la principale difficulté est de maintenir l'équilibre dynamique de l'état atmosphérique, souvent imposé sous la forme d'une condition initiale, tout en y propageant une onde acoustique. Une onde acoustique peut en effet interagir avec l'écoulement atmosphérique pour produire des ondes de différentes natures et un écoulement à grande échelle différent de celui qui est imposé en condition initiale. Or, il est difficile de distinguer les comportements dus à la perturbation acoustique de ceux qui sont la conséquence d'un déséquilibre de l'état atmosphérique, en particulier lorsqu'on s'intéresse à des domaines qui s'étendent sur plusieurs milliers de kilomètres. La communauté des géophysiciens utilise donc plutôt des méthodes asymptotiques telles que le tracé de rayons [Drob et al., 2013, Hedlin and Drob, 2014], la sommation de modes normaux [Millet et al., 2007, Bertin et al., 2014, Lalande and Waxler, 2016, Assink et al., 2017] ou la résolution d'une approximation parabolique de l'équation d'onde [Klyatskin and Tatarskii, 1970, Ostashev et al., 2019, Blanc-Benon et al., 2001].

Dans cette thèse nous utiliserons la décomposition en modes normaux qui s'appuie sur le calcul des éléments propres (k_n, ϕ_n) de l'équation d'onde :

$$\left[\frac{\partial^2}{\partial z^2} + \frac{\omega^2}{c^2}\right] \phi_n = k_n^2 \phi_n \tag{2}$$

où $c(z)$ est la vitesse du son définie par l'équation 1. Ces modes sont les éléments

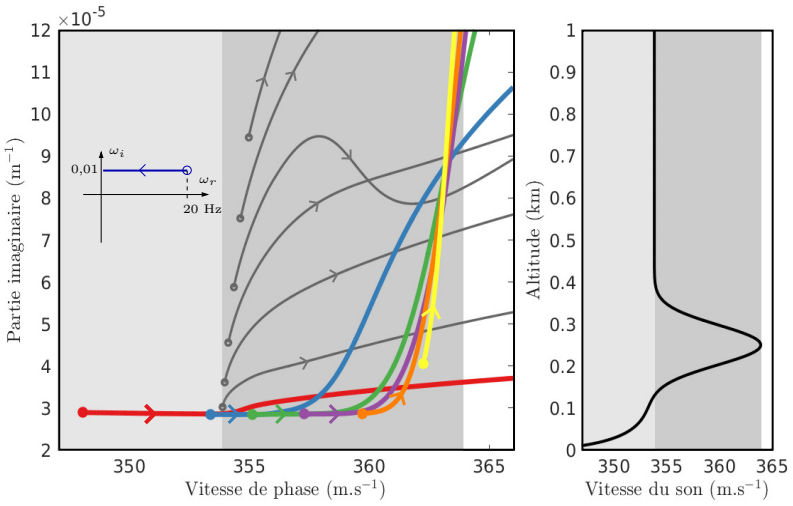

FIGURE 2 : Pour une évolution en fréquence donnée en bleu, les différents modes évoluent dans le plan complexe selon les trajectoires colorées. Les trajectoires en gris correspondent aux modes du spectre continu (ou plutôt de sa discrétisation). Les zones grisées correspondent aux différentes zones sur le profil donnant la vitesse du son.

du spectre de l'opérateur différentiel défini ci-dessus[3]. Ce spectre se sépare en deux parties : une partie discrète (les modes représentés en couleurs sur la figure 2) et les modes issus de la partie continue (en gris sont représentés les valeurs propres de l'opérateur discrétisé correspondant au spectre continu de l'opérateur différentiel). Sur la figure 2, l'évolution des modes dans le plan complexe en représenté pour des fréquences inférieures à 20 Hz. Les modes se situent dans le plan complexe car ω possède une partie imaginaire permettant de souligner les différentes structures correspondant au profil étudié [Bertin et al., 2014]. La figure 3 donne l'évolution en fréquence des fonctions propres des modes représentés ci-dessus. Il est intéressant de noter que ceux-ci sont confinés pour les fréquences les plus hautes avant d'osciller sur tout le domaine. D'un point de vue acoustique, c'est la valeur de la fonction propre au sol qui quantifie la contribution du mode. On peut donc définir une fréquence de coupure – appelée *cut-off frequency* – en deçà de laquelle le mode ne contribue plus acoustiquement.

Une fois calculés, ces modes acoustiques permettent de calculer le champ de pression dans le domaine fréquentiel :

$$p(\omega) = \sum_{n=1}^{N} p_n(\omega) = \frac{\widehat{s}(\omega)\, e^{i\pi/4}}{\sqrt{8\pi}} \sum_{n=1}^{N} \frac{\phi_{n0}^2(0) e^{ik_{n0}r}}{\sqrt{Rk_{n0}}} \tag{3}$$

où R est la distance de propagation et $\widehat{s}(\omega)$ le spectre de la source. Pour obtenir un signal temporel, il reste à effectuer une transformée de Fourier. On peut aussi

[3]Pour définir un opérateur différentiel il ne suffit pas de donner son expression mais il faut aussi donner son domaine de définition. Ici le domaine sera $\mathbb{R}^+$ avec les conditions aux limites considérées dans le problème de propagation.

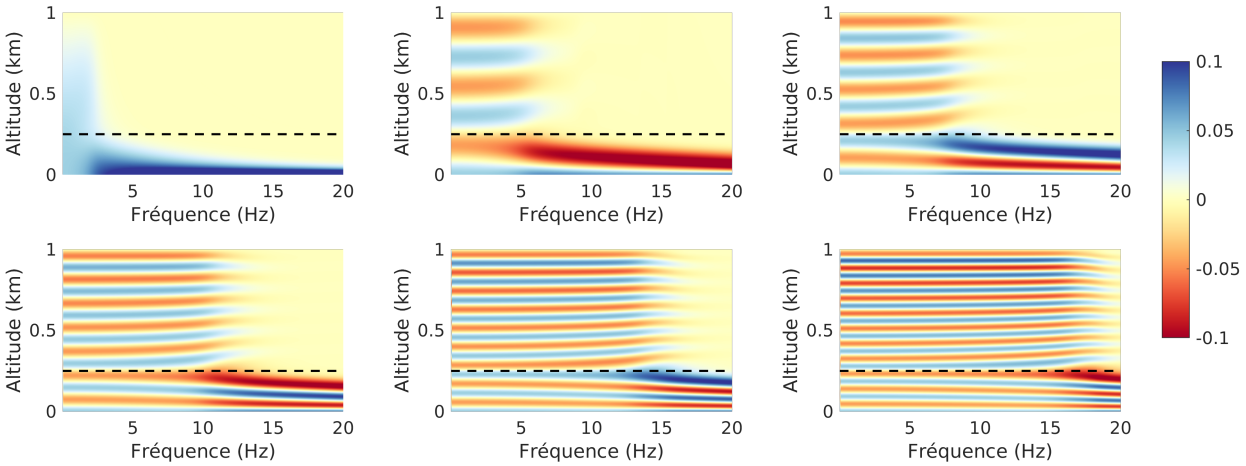

FIGURE 3 : Évolution fréquentielle des différentes fonctions propres associées aux modes tracés en couleur sur la figure 2. Le trait en pointillé indique l'altitude du maximum de vitesse du son. Sur la première ligne sont représentés les modes rouge, bleu et vert et les modes violet, orange puis jaune sur la deuxième.

faire la transformée de Fourier de chacun des $p_n(\omega)$ pour obtenir les *paquets d'onde* associés aux différents modes. Ceux-ci sont représentés sur la figure 4 en conservant le même code couleur que pour la figure 2. En sommant les différents paquets d'onde on retrouve le signal obtenu en calculant la transformée de Fourier de (ω) comme le montre la dernière ligne où les deux signaux sont superposés.

1.4 Sujet de la thèse

Afin de tirer profit des modèles de propagation dans les processus de localisation, nous proposons d'en construire un *métamodèle*. Un métamodèle peut se définir comme un modèle de modèle, moins coûteux à évaluer numériquement mais capable de restituer les statistiques des grandeurs calculées par le premier modèle. Il existe différentes méthodes pour construire un métamodèle mais dans tous les cas une première étape consiste à calibrer le métamodèle en utilisant un nombre restreint d'appels au modèle originel. Une fois calibré, le métamodèle peut être utilisé pour calculer des grandeurs statistiques liées à la sortie du modèle : moments, probabilité de dépassement d'un seuil par exemple.

L'objectif de cette thèse est donc de construire un métamodèle stochastique, continu, capable de restituer, à moindre coût, des signaux infrasonores et leurs statistiques en tenant compte de la variabilité du milieu atmosphérique.

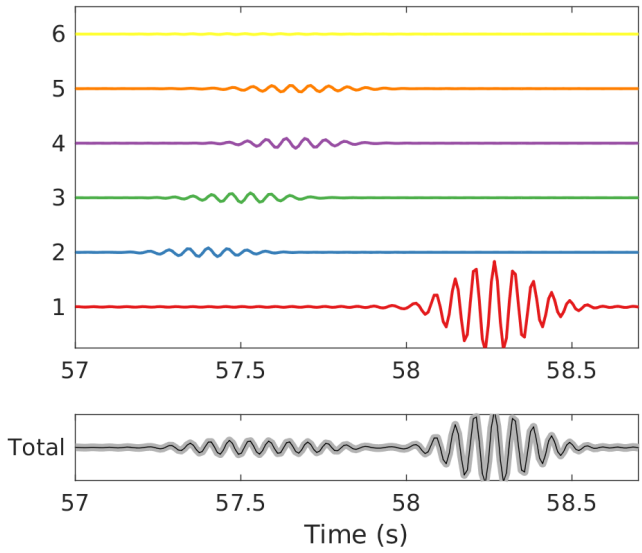

FIGURE 4 : Calcul des différents paquets d'onde correspondant à chacun des modes en couleur sur la figure 2. Obtenus pour une source située entre 15 et 20 Hz à 20 km du récepteur. En noir le signal correspondant à la somme des 6 modes est représenté avec la somme des six paquets d'onde superposée en gris.

2 Métamodèle de signaux propagés

2.1 Métamodélisation par polynômes de chaos

Il existe différentes méthodes pour construire un métamodèle, nous utiliserons ici les polynômes de chaos [Wiener, 1938, Ghanem and Spanos, 1991]. L'idée est de décomposer la sortie d'un modèle $Y = F(X)$ sur une base de polynômes en la variable (aléatoire) d'entrée X :

$$Y = \sum_{j \in J} a_j H_j(X) \tag{4}$$

où J est un ensemble fini et $(H_j)_{j \in \mathbb{N}}$ est la famille de polynômes orthogonaux par rapport à la mesure de probabilité de X. L'orthogonalité des polynômes permet de calculer les coefficients $(a_j)_{j \in J}$ de deux manières différentes. Soit par projection en calculant le produit scalaire :

$$a_j = \mathbb{E}_X \left[F(X) H_j(X) \right] \tag{5}$$

ou alors par régression, le projeté orthogonal minimisant la distance entre la fonction et l'espace d'approximation :

$$(a_j)_{j \in J} = \underset{(a_j) \in \mathbb{R}^{|J|}}{\operatorname{argmin}} \, \mathbb{E}_X \left[\left(F(X) - \sum_{j \in J} a_j H_j(X) \right)^2 \right]. \tag{6}$$

La valeur du métamodèle réside dans la capacité à estimer les coefficients $(a_j)_{j\in J}$ en utilisant peu d'appels à F. Ces estimations peuvent se faire de différentes manières mais restent dans tous les cas sensibles à la dimension en entrée du problème.

2.2 Modélisation des incertitudes

Afin de simuler la propagation des infrasons dans une atmosphère réaliste, plusieurs approches sont possibles. La première serait de partir de données historiques pour simuler une variabilité représentative de celle observée précédemment [Assink et al., 2013]. Une autre possibilité consiste à utiliser des modèles théoriques développés dans le cadre d'études sur la dynamique de l'atmosphère [Cugnet et al., 2019]. Dans les deux cas la démarche est la même, partir de l'état de l'atmosphère au moment de l'évènement – fourni par les organismes météorologiques – et y ajouter une perturbation aléatoire. C'est cette perturbation qui sera pour nous la donnée d'entrée du métamodèle, ce sont donc ses propriétés qui nous ont guidées dans le choix des techniques de métamodélisation. Une propriété importante qui nous a orienté vers les polynômes de chaos est l'amplitude considérable que peut prendre cette perturbation et qui écarte toute approche perturbative. Par exemple, le déferlement d'ondes de gravité dans la stratosphère (i.e. entre 20 km et 60 km) peut conduire à des vents de l'ordre de 50m.s^{-1}. Toutefois, les perturbations ne peuvent se résumer à ces grandes déviations et comportent aussi des structures plus fines (dues au caractère turbulent de la couche limite planétaire par exemple) qui auront, elles aussi, un impact important sur la propagation des infrasons.

Afin de concilier performance numérique et précision du métamodèle nous avons décidé de scinder en deux la perturbation. Les grandes amplitudes seront prises en compte par les polynômes de chaos et les structures plus petites par une analyse perturbative. Pour cela nous proposons de décomposer la perturbation (qui mathématiquement est un processus stochastique) en série de Karhunen-Loève [Kac and Siegert, 1947, Karhunen, 1947, Loeve, 1948]. Cette décomposition est basée sur une décomposition spectrale de la matrice de covariance du processus. Ainsi, elle sépare naturellement les différentes échelles et, en tronquant le développement, cela permet de décrire les grandes déviations en utilisant un nombre limité de variables aléatoires. En effet, les techniques de décomposition en polynômes de chaos restent sensibles à la dimension des données d'entrée et seront donc d'autant plus efficaces que le nombre de variables d'entrée sera petit. Les petites structures, représentées par les termes d'ordre élevé de la décomposition Karhunen-Loève, seront pris en compte dans un deuxième temps par une approche perturbative [Kato, 2013, Fouque et al., 2007]. Cette approche est plus classique et nous verrons uniquement dans ce manuscrit de quelle façon elle peut être associée avec l'utilisation des polynômes de chaos.

2.3 Le problème d'intégration en temps long

L'objet principal de cette thèse est donc la construction d'un métamodèle à base de polynômes de chaos capable de restituer des signaux infrasonores après leurs propagations dans une atmosphère perturbée. La principale difficulté vient de la restitution

de signaux aléatoires pour des distances importantes. En effet, si les polynômes de chaos sont très efficaces pour obtenir des métamodèles lorsque la quantité d'intéret est scalaire, la restitution de signaux aléatoires se heurte au problème d'*intégration en temps long* [Pettit and Beran, 2006, Le Maître et al., 2010]. Dans sa formulation la plus classique, le problème d'intégration en temps long concerne la convergence du développement en polynômes de chaos de signaux aléatoires pour des temps longs. En effet, pour des systèmes aléatoires qui dépendent aussi du temps, la convergence du développement en polynômes de chaos dépendra aussi du temps. Cela signifie qu'il n'existe pas de troncature du développement en polynômes de chaos qui permette de représenter correctement le système pour tous les temps. Dans notre cas, les événements mentionnés au paragraphe 1.1 sont sources de signaux impulsifs très courts mais les distances de propagation très importantes posent un problème similaire. Parmi les différents travaux proposés pour palier cette limitation nous avons notamment testé l'approche proposée par Chu Van Mai *et al.* nommée *time-warping* [Mai and Sudret, 2017]. L'idée est de considérer qu'il existe un temps propre à chaque réalisation, ce temps est donc une variable aléatoire qui dépend du temps physique :

$$\tau = kt + \phi. \tag{7}$$

Dans l'expression précédente, k et ϕ sont des variables aléatoires dont on peut calculer un métamodèle avec des polynômes de chaos. La transformation étant bijective on peut l'inverser pour exprimer les signaux en fonction du temps τ. Les signaux restent aléatoires mais en choisissant k et τ pour homogénéiser les différents signaux on peut accélérer la convergence des polynômes de chaos. Pour chaque réalisation $s_i(t)$ du signal aléatoire, les réalisations k_i de k et ϕ_i de ϕ sont ajustées pour maximiser la ressemblance entre le signal $s_i(k_i t + \phi_i)$ et une référence $s_r(t)$:

$$(k_i, \phi_i) = \underset{(k,\phi)\in\mathbb{R}^2}{\operatorname{argmin}} \frac{\langle s_i(kt+\phi), s_r(t)\rangle}{\|s_i(kt+\phi)\|\|s_r(t)\|}. \tag{8}$$

Malheureusement, il n'est pas toujours possible de déterminer k et ϕ permettant de transformer les signaux de manière adéquate. Pour illustrer ce phénomène, nous allons considérer le cas d'une perturbation localisée entre 100m et 400m d'altitude sur le profil utilisé dans la section précédente (figure 5-(c)). Cette perturbation étant localisée, elle n'aura d'impact que sur la première des deux arrivées du signal comme illustré sur la figure (a). Le deuxième mode étant un de ceux constituant cette première arrivée, il est impacté par la perturbation. Nous avons appliqué l'algorithme de time-warping au paquet d'onde associé au deuxième mode puis au signal total (figure 5-(b)). Dans le cas du paquet d'onde simple, l'algorithme parvient sans peine à trouver k et ϕ permettant de superposer les deux signaux. Pour le signal complet, en revanche, l'algorithme ne trouve pas de k et de ϕ convenables car il ne peut améliorer la première arrivée sans détériorer la deuxième.

Les signaux issus de la propagation atmosphérique étant constitués de plusieurs arrivées avec des incertitudes différentes pour chacune, nous avons écarté cette technique. Nous retenons tout de même de cette étude la possibilité de travailler sur les paquets d'onde isolément. C'est ce point de vue que nous avons adopté en travaillant

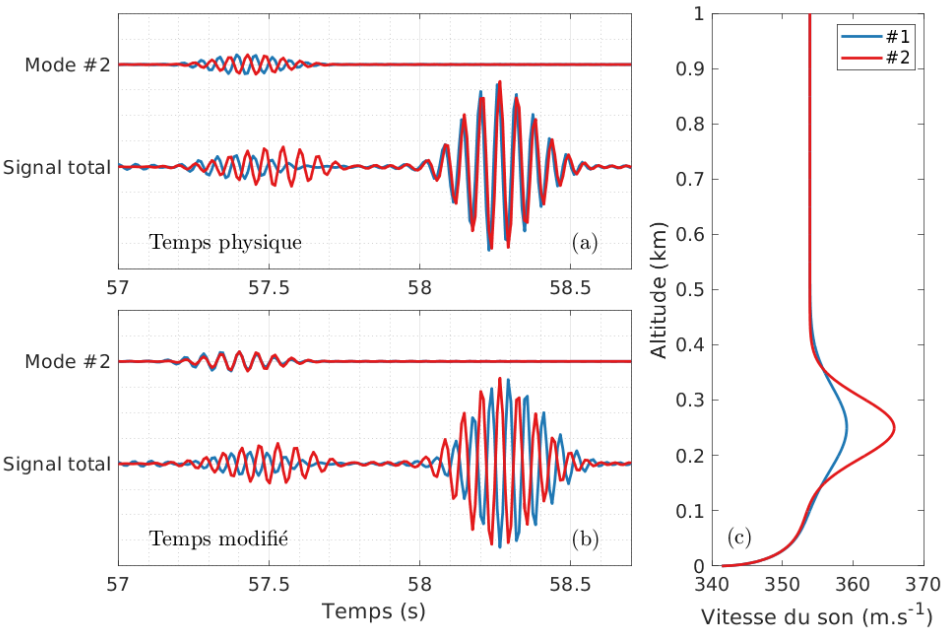

FIGURE 5 : En modifiant uniquement la portion du profil située entre 100 et 400m (c) seuls les modes concernés (dont le deuxième ici représenté) sont modifiés (a). Seule la première des deux arrivées s'en trouve donc modifiée (a). L'algorithme de time-warping permet de recaler les paquets d'ondes correspondant aux deux réalisations si le nombre d'extrema est le même (a) mais ne peut donc pas recaler le signal total (b) dont les deux réalisations sont trop différentes.

directement sur les modes acoustiques[4]. Ceux-ci portent en eux toute l'information relative à la propagation et ont l'avantage d'être des quantités scalaires. Une fois calculés, les développements en polynômes de chaos des modes acoustiques permettront de générer les signaux temporels associés à l'aide d'une simple transformée de Fourier en temps et d'une convolution avec le signal source. Cette approche présente également le double avantage d'obtenir un métamodèle indépendant de la distance de propagation et de la source utilisée. L'impact des petites structures se traduit également au niveau des modes acoustiques par un couplage inter-modes qui peut être calculé en utilisant les représentations en polynômes de chaos.

2.4 La dynamique des modes acoustiques

Les modes normaux semblent donc être les grandeurs pertinentes pour caractériser la propagation atmosphérique en tenant compte des incertitudes. Cependant, la construction d'un métamodèle de ces modes acoustiques suppose d'être capable de les isoler et de les suivre quelque soit la valeur de la perturbation [Ghosh and Ghanem, 2012, Kalaba et al., 1981, Georg et al., 2018]. La dynamique des modes acoustiques est inconnue

[4]aussi appelés modes normaux ils constituent une base de décomposition du champ de pression adaptée aux structures atmosphériques.

a priori et peut se révéler très accidentée ce qui rend difficile le suivi des modes. Il est par exemple courant que deux modes s'échangent continûment et une discretisation trop lâche pourrait manquer ce phénomène. Afin de suivre fidèlement les modes en fonction de la perturbation sans obérer les performances du métamodèle nous avons développé une technique de suivi adaptée à cette application en jouant sur le suivi en fréquence. Les modes acoustiques apparaissent comme les éléments propres d'un opérateur différentiel. Le spectre de cet opérateur possède une partie discrète et une partie continue qui est en général négligée dans les applications infrasonores [Waxler, 2002, Waxler, 2003]. Pour des perturbations différentes la taille du spectre discret peut être différente. Cependant, cesser de prendre en compte un mode induirait une discontinuité dans le domaine fréquentiel et un bruit sur les signaux calculés ensuite. En réalité, un mode ne disparaît pas, il passe de la partie discrète à la partie continue du spectre. Nous avons donc utilisé une couche artificielle appelée *Perfectly Matched Layer* (PML) [Bérenger, 1994] pour fermer le domaine. En effet, la PML modifie la partie continue en lui ajoutant une partie imaginaire [Olyslager, 2004] qui entraînera une atténuation du mode d'un point de vue acoustique. Cela nous permet donc de continuer de prendre en compte le mode considéré sans que cela ait un impact sur les signaux calculés.

Les échanges de modes mentionnés plus haut traduisent un changement physique de la propagation acoustique. L'altitude de réfraction des modes va changer et, en conséquence, leur sensibilité à la perturbation changera également. Pour rendre compte de ce changement de régime, nous avons également adapté le développement en polynômes de chaos pour capturer ces différents régimes. En effet, ces échanges de modes pour différentes réalisations de la perturbation se traduisent d'un point de vue probabiliste par une distribution bimodale pour les modes concernés. Cette distribution bimodale rend la convergence des développements en polynômes de chaos plus lente [Nouy, 2010, Soize, 2015], nous avons donc adapté notre métamodèle en utilisant des développements différents pour chacun des régimes.

3 Apport des gPC pour la localisation de sources

Les polynômes de chaos s'intègrent naturellement au cadre bayésien [Marzouk et al., 2007, Marzouk and Najm, 2009, Marzouk and Xiu, 2009]. En construisant une base de polynômes de chaos adaptée à l'a priori, il est possible d'estimer plusieurs paramètres liés à un évènement donné (localisation, énergie, etc.). Cette approche présente un intérêt dans le cadre des activités du Centre international de données (CID), dont le bulletin final incorpore des données sismiques, hydroacoustiques et acoustiques d'une grande qualité. Actuellement la fiabilité des renseignements qu'il est possible d'en déduire est difficile à quantifier, même si certaines méthodes [Arora et al., 2013, Mialle et al., 2019] ont récemment permis d'améliorer les performances des algorithmes automatiques. L'approche qui fait aujourd'hui référence repose sur la méthode du tracé de rayons, elle permet d'accroître le niveau de confiance via le *pointé* et l'identification dans les signaux des arrivées associées à l'événement. Mais à cause de l'absence de continuité dans la description géométrique, les arrivées sont souvent identifiées visuel-

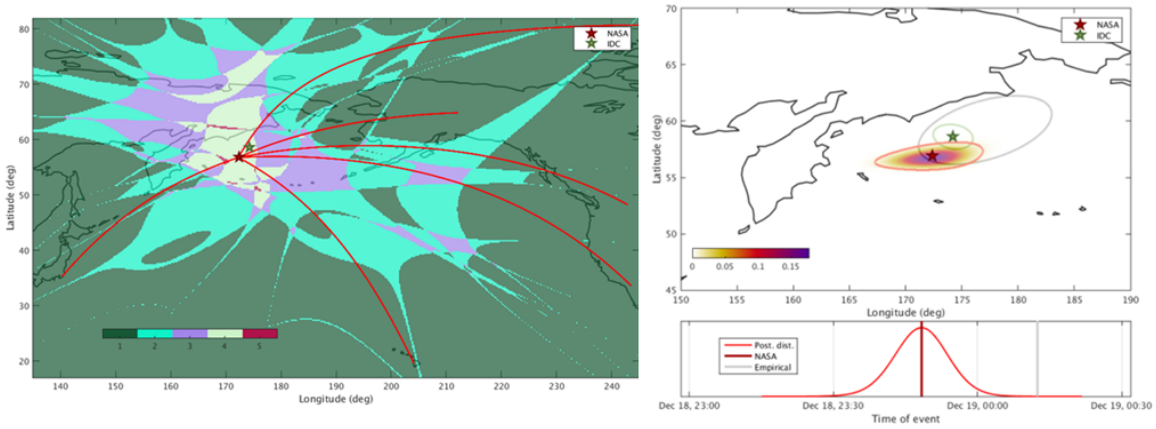

FIGURE 6 : Nombre de stations dont la vraisemblance est supérieure à la moyenne (à gauche) et vraisemblance totale (à droite). L'événement correspond à l'explosion d'une météorite au-dessus de la mer de Bering, le 18 décembre 2018 à 23h48 (mesure NASA). À droite, les ellipses grise et verte correspondent à l'estimation faite par le CID et en utilisant un modèle de propagation (empirique) de référence. À gauche, les lignes représentent les grands cercles liant la source aux six stations de mesure considérées ici.

lement, en extrapolant les résultats hors du domaine de validité[5]. Une représentation des arrivées par des polynômes de chaos ne présente pas ce type de limitation.

L'approche qui a été développée dans le cadre de cette thèse, a consisté à tirer parti d'une représentation des arrivées par des polynômes de chaos. En combinaison avec les algorithmes de localisation et de caractérisation, ces métamodèles permettent de tenir compte du contexte associé à chaque événement (état de l'atmosphère, incertitude liée au pointé automatique des arrivées).

La Figure 6 illustre le gain obtenu, lors de la ré-analyse d'un événement qui s'est produit le 18 décembre 2018, au-dessus du détroit du Béring. L'événement a été mesuré par plusieurs stations, nous avons donc calibré un métamodèle par station, chacun de ces métamodèles prenant en compte la variabilité due aux effets de projection. On parle en effet de projection car en fonction des positions possibles pour la source, les profils de vent et de température seront différents mais l'azimuth de propagation est également différent. La formulation bayésienne se base sur un a priori qui se réfère à un jugement d'expert – qui pour nous sera la première localisation fournie par le CID – et un terme de vraisemblance qui traduit l'adéquation entre les signaux simulés et ceux mesurés. Les ellipses de confiance à 95%, et la vraisemblance, ont été obtenues en modélisant les temps des arrivées par des polynômes de chaos, dont les coefficients ont été estimés à partir de simulations très précises.

[5]Pour pallier les limitations liées à la méthode géométrique, plusieurs équipes se sont engagées dans le développement de modèles de propagation empiriques. Ces modèles sont obtenus à partir de résultats de simulations de signaux, en fixant une statistique de configurations atmosphériques. Ainsi par construction, ces modèles ne permettent pas d'adapter les résultats à une configuration atmosphérique précise.

L'erreur associée à l'extraction automatique des temps d'arrivée est également prise en compte. La figure 6 montre que les localisations produites par le CID (ellipse verte), ou en utilisant un modèle empirique (en gris), ne sont pas compatibles avec l'observation faite par la NASA (instrument MISR du satellite Terra). En outre, le calcul des vraisemblances des différentes observables permet d'évaluer l'apport de chaque station à la vraisemblance totale, et on constate ainsi que quatre stations parmi six contribuent réellement à la localisation (Figure 6).

Chapitre 1

Introduction

This introductory chapter is devoted to a presentation of uncertainty quantification techniques which will be used to simulate efficiently the propagation of infrasound in the atmosphere, taking into account its natural variability. We introduce Karhunen-Loève decomposition to reduce the input dimension and illustrate how to decompose atmospheric variability in practice. Then, the notion of metamodel is presented from a general perspective, giving the objectives of building a metamodel and a description of its main characteristics. The next chapters will explore the construction of a metamodel adapted to infrasound propagation. We end this chapter with a presentation of sensitivity analysis tools (Sobol' indices and Shapley effects) that are useful to trace the impact of uncertainties in the input parameters on the quantities of interest.

1 Motivation

1.1 Propagation Technology & CTBTO

The Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) was created after the signature of a treaty banning all forms of nuclear explosion. The role of this organization is to monitor and detect all nuclear explosions on earth : in the air, underground or underwater. To this purpose, the organization has developed the international monitoring system which relies on four technologies : seismic (170 stations), hydroacoustic (11 stations), infrasound (60 stations) and radionuclide (120 stations). Among the three propagation techniques, infrasound is the most exposed to medium variability ; for this reason, we will focus in this work on the infrasound technology. A map of the infrasound stations is given on figure 1.1.

Infrasound is defined as sound at frequencies less than 20 Hz (i.e. below the audible band of human ear). It has the ability to propagate over large distances with little dissipation, for example the first historical observation of infrasound has been recorded after the 1883 eruption of the Krakatoa volcano in Indonesia : it traveled around the globe at least seven times, shattered windows hundreds of miles away, and were recorded worldwide. More recently, a meteor entered the earth's atmosphere near

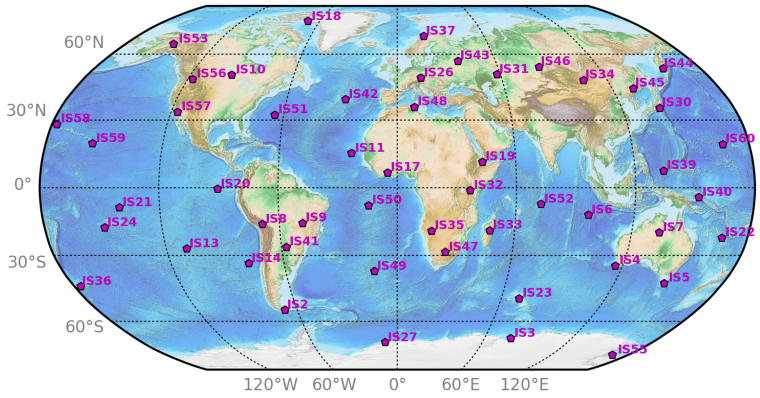

FIGURE 1.1 : Nearly 60 infrasound stations are implanted worldwide to register atmospheric explosions. *(figure : P. Vanderbecken)*

Chelyabinsk in Russia in 2013 and was detected in the US at 10 000 km [de Groot-Hedlin and Hedlin, 2014].

In practice, the sensor array is connected to the International Data Center (IDC) in Vienna, which operates continuously and in real time. The purpose of the IDC is to analyze the detections of the stations and infer the characteristics of the events (location, type of event for example). But the current IDC's automated system, a highly complex and well-tuned piece of software, still misses events, and about half of the reported events are spurious. A large team of expert analysts post-processes the automatic bulletins to improve their accuracy to acceptable levels. Difficulties in processing all the data can be classified in three categories [Arora et al., 2013] : (1) physical characteristics of a detection (time of arrival, duration of the signal ...) can change dramatically depending on atmospheric conditions ; (2) each detector is subject to local noise that may mask true signals and cause false detections ; and (3) many thousands of detections are recorded per day, so the problem of proposing and comparing possible events (sub-sets of detections) is daunting.

Due to those considerations, an approach based on probabilistic inference has been proposed in [Arora et al., 2013] to process seismic data. This method (called NET-VISA) achieves a reduction of around 60% in the number of missed events compared with the currently deployed system. It also detects some events missed by the human analysts post-processing the IMS output. Interestingly, NET-VISA is capable of dealing with those events without using any propagation model. To go one step further, we propose in this work to take into account the propagation effects in this probabilistic framework. For this purpose, we will use numerical techniques that have been developped in a context of uncertainty quantification.

1.2 Uncertainty Quantification

Recent development of computational facilities has increased the use of numerical simulation for complex physical phenomena. Those simulations often involve many parameters and the accuracy of such a model depends on the reliability of the input. To characterize the precision of the numerical simulation, one has to study the impact of the variability of the input parameters on the simulation results. Because the result of a simulation can be postprocessed, we usually define the *Quantity of Interest (QoI)* which is the final product we are interested in. A preliminary step is therefore to define the QoI. If this definition is most of the time straightforward, it is of prime importance because it will influence the entire study.

An *Uncertainty Quantification (UQ)* approach can be separated in three steps : 1 – characterization of the input variability, 2 – propagation of this uncertainty through the model in order to assess the impact on the quantity of interest, 3 – backward uncertainty propagation or Bayesian calibration. The last step can be conducted when physical measurements of the QoI are available in order to improve the knowledge on the input variability.

2 Input parameters and atmospheric uncertainties

Characterizing the variability of the input requires some knowledge on the physical parameters involved in the simulation. If data are available, a preliminary study can be conducted to fit a probability distribution, otherwise we rely on experts' opinion. We may mention here the existence of non parametrical estimation [Wegman, 1972], which allows to model the data variability without *a priori* informations on the shape of the distribution.

2.1 Atmospheric uncertainties

Meteorological observation tools (satellites and ground stations) provide a very large amount of data, which must be processed to produce the state of the atmosphere. Meteorological centers use data assimilation methods to produce an initial state used by numerical models to forecast the weather. To provide more precise data on the atmospheric state, a reanalysis is provided *a posteriori* by the European Center for Medium-Range Weather Forecasts (ECMWF). It consists in feeding the numerical weather processes with weather observations collected to recreate past atmospheric, sea- and land-surface conditions over specific time periods.

Figure 1.2 represents the annual variability of zonal and meridional winds in 2018. Profiles are given by the ECMWF at the station I37NO, which is located in Norway. This figure is useful to illustrate two simple facts : (1) fluctuations can be of great amplitude especially for the zonal wind, (2) there are two regions where fluctuations are more important : turbulence in the troposphere (at 8-18km) and the stratospheric wind around 60 km, which changes direction during the year.

These meteorological data give an illustration of the variations of winds and temperatures over one year. However, due to low resolution, both in space and time,

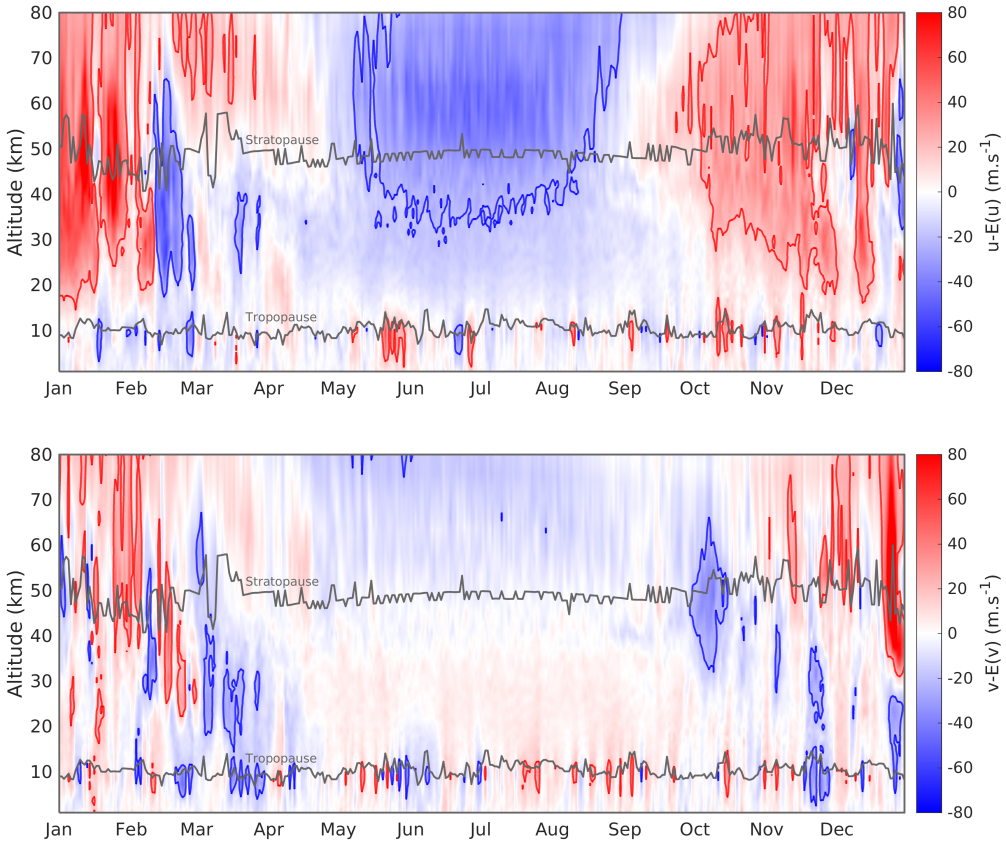

FIGURE 1.2 : Annual variations of zonal (u) and meridional (v) winds during 2018 at I37NO station in Norway (red and blue lines represent -20m.s^{-1} and 20m.s^{-1}). Tropopause and stratopause correspond to local minimum and maximum of the temperature profiles.

ECMWF data do not capture small structures, such as gravity waves (GW) [Le Pichon et al., 2010, Drob et al., 2013]. Although the most prevalent sources of atmospheric GW occur in the troposphere and initial wave amplitudes are relatively small, these gravity waves propagate upward and their amplitude tends to grow as the result of exponentially decreasing air density. Thus, at certain locations and seasons in the upper mesosphere and lower thermosphere (above 75 km), GW are the cause of random variations of great amplitude.

Thus, for the purpose of infrasound propagation modeling, the unresolved atmospheric gravity wave perturbations can be represented using stochastic models [Lott and Millet, 2010, Lott et al., 2012, de la Cámara et al., 2014] and superimposed on the resolved average background state, much in the same way that turbulence is parameterized in aerodynamic drag calculations.

In order to reduce the cost of a numerical study, especially when many stochastic input parameters are involved, we first need to reduce the input dimension d, which is *a priori* infinite if we take the whole fields of wind and temperature.

2.2 Reducing the input dimension

A problem with high-dimensional datasets is that, in many cases, only a small set of variables have a significant impact on Y. Moreover, many algorithms are sensitive to the input dimension and the *curse of dimensionality*[1] makes it computationally impossible to work with such data sets. As a consequence, many algorithms have been developped to reduce the dimension of the parameters space. To deal with continuous inputs (temperature and wind profiles in our case), one of the most widely used technique is the Karhunen-Loève decomposition [Kac and Siegert, 1947, Karhunen, 1947, Loeve, 1948].

The idea of Karhunen-Loève (KL) decomposition is to decompose the stochastic process on a basis formed by deterministic functions (coefficients will be random variables).

The KL decomposition is defined for centered, mean-square continuous[2] and square integrable (i.e. in L^2) processes $(X_t)_{t\in D}$ where D is the domain ($D = \mathbb{R}^+$ for us).

If we denote $K_X(s,t) = \mathbb{E}[X_s X_t]$ the covariance function, we can introduce the linear integral operator T_{K_X} :

$$\begin{aligned} T_{K_X} : L^2(D) &\to L^2(D) : \\ f &\mapsto \int_D K_X(s,\cdot) f(s)\, ds, \end{aligned} \tag{1.1}$$

note that this integral is completely deterministic as it is defined using $K_X(s,t)$ which is a deterministic function. From definition (1.1) it is easy to see that T_{K_X} is a linear operator. Let us denote $(\lambda_k, e_k(t))_{k\in\mathbb{N}}$ its spectrum :

$$\int_D K_X(s,t) e_k(s)\, ds = \lambda_k e_k(t). \tag{1.2}$$

Since the operator T_{K_X} is self-adjoint, the spectral theorem ensures that the eigenfunctions form a basis. Any stochastic process in $L^2(D)$ can be decomposed :

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t), \tag{1.3}$$

where the coefficients Z_k are given by the orthogonality property :

$$Z_k = \int_D X_t e_k(t)\, dt. \tag{1.4}$$

The orthogonality of the basis gives a decomposition in uncorrelated random variables :

$$\mathbb{E}[Z_k] = 0,\ \forall k \in \mathbb{N} \qquad \text{and} \qquad \mathbb{E}[Z_i Z_j] = \delta_{ij}\lambda_j,\ \forall i,j \in \mathbb{N}, \tag{1.5}$$

[1] Introduced by Richard Bellman, the curse of dimensionnality refers to all the phenomena arising when the input dimension increases.

[2] $(X_t)_{t\in D}$ is mean-square continuous iif $\lim_{\varepsilon\to 0} \mathbb{E}\left[(X_{t+\varepsilon} - X_t)^2\right] = 0$, which is equivalent to the continuity of the covariance function, see [Alexanderian, 2015] for technical details.

which allows computing directly the variance of the process :

$$\begin{aligned} \operatorname{var}[X_t] &= \sum_{k=0}^{\infty} e_k(t)^2 \operatorname{var}[Z_k] = \sum_{k=1}^{\infty} \lambda_k e_k(t)^2, \\ \int_D \operatorname{var}[X_t]\, dt &= \sum_{k=1}^{\infty} \lambda_k. \end{aligned} \tag{1.6}$$

This can be used to truncate the expansion (1.3) : to keep a proportion α of the variance, one have to find the order N such that :

$$\frac{\sum_{k=1}^{N} \lambda_k}{\operatorname{Var}[X_t]} \geq \alpha. \tag{1.7}$$

The KL decomposition is closely related to the Principal Component Analysis (PCA), which is widely used in data analysis. Assuming that we have a (sufficiently large) number of realizations, we can empirically approximate the covariance matrix of the process and compute its spectrum $\left(\widetilde{\lambda}_k, \widetilde{e_k}(z)\right)_{k=1,..,N}$ (we denote with a $\sim$ empirical values). The perturbation can be decomposed on this basis :

$$X_z = \sum_{k=1}^{N} \xi_k \sqrt{\widetilde{\lambda}_k} \widetilde{e_k}(z) \tag{1.8}$$

The difficulty of this approach is to characterize the random variables $(\xi_k)_{k=1,..,N}$. The different realizations of X_z can be used to sample those random variables, computing :

$$\xi_k(\omega) = \int_D X_z(\omega) \frac{\widetilde{e}_k(z)}{\sqrt{\widetilde{\lambda}_k}}\, dz, \tag{1.9}$$

which can be used as the new input parameters for our model F.

Then the probability density function can be computed empirically using a kernel method for instance. However, there is no reason for the sampling of X_z to ensure a good representation of ξ_k. Moreover, when the process is gaussian, the random variables $(Z_k)_{k\in\mathbb{N}}$ are independent but it will not be the case with empirical decomposition.

In the case of infrasound propagating in the atmosphere, sound ducts produced by the combination of temperature gradients and wind shear can be qualitatively determined using the effective sound speed approximation in which the influence of the wind is accounted for by adding the horizontal wind speed component in the direction of propagation to the sound speed [Godin, 2002]. This approximation allows to consider only a one dimensionnal input $c(z)$ on which a perturbation can be superimposed. Figure 1.3 shows a decomposition of the variability of an effective sound speed profile perturbed by a GW field using a PCA. To produce this decomposition, 1000 realizations of GW have been produced using the global climate model (GCM) LMDz ([de la Cámara et al., 2014]) and superimposed to an effective sound speed computed from

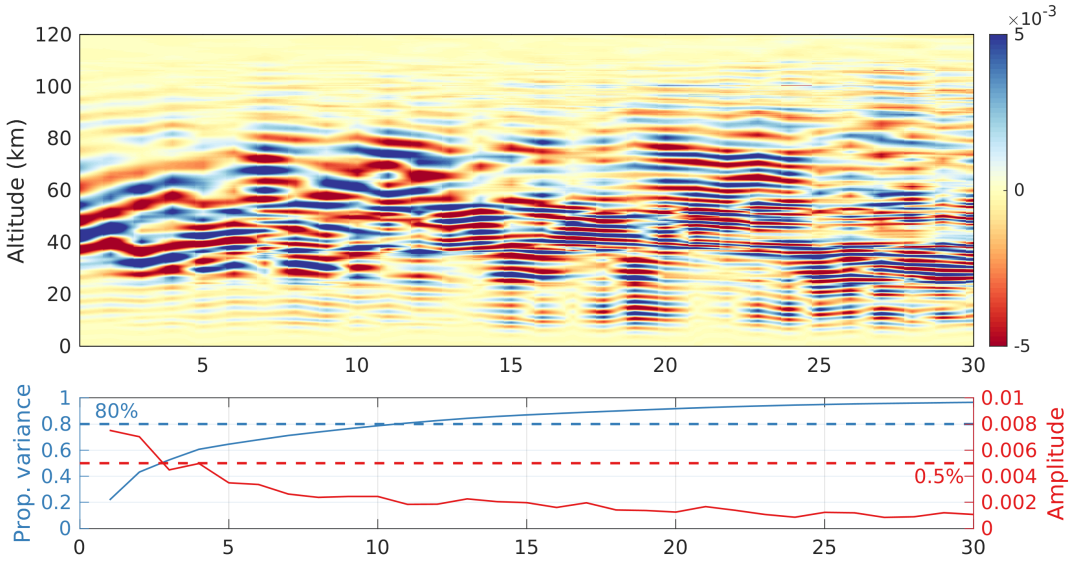

FIGURE 1.3 : PCA of the variability an effective sound speed profile perturbed by realisations of GW. The basis functions $\widetilde{e_k}(z)$ capture variability of different wavelengths (on the first line). On the second line, the amplitude of the different PCA modes $\|\widetilde{e_k}\|$ is shown in red and the cumulative proportion of the total variance is shown in blue. The ten first terms of the decomposition capture 80% of the variance and are the only ones whose amplitudes exceed 0.5% of the unperturbed profile amplitude.

real data. From the covariance matrix of these 1000 profiles, a PCA decomposition has been computed and the functions $\widetilde{e_k}(z)$ are plotted for $k \in \{1, .., 30\}$.

The truncation of this expansion can be done to preserve a fixed proportion (α) of the variance of the initial perturbation. In our case, to keep all the variance of the celerity profiles, we propose to treat the remaining terms as an additional perturbation. This additional perturbation will be treated using perturbative techniques. As an example on figure 1.3, the first 10 terms of the decomposition are of great amplitude and represent 80% of the variance, the other terms have a sufficiently small amplitude to be consistent with a perturbative approach.

3 Modelling and Metamodelling

3.1 Models of propagation

One primary purpose of a model is to characterize the input-output relationship of the system of interest—the input describes the relevant system properties and environmental conditions, and the output describes quantities of interest of the related problem. In this context, evaluating a model means performing a numerical simulation that implements the model, computes a solution, and thus maps an input onto an approximation of the output. For instance, the numerical simulation might involve

solving a partial differential equation (PDE), or solving a system of ordinary differential equations. Mathematically, we denote a model as a function $F : \mathcal{X} \to \mathcal{Y}$ that maps an input $X \in \mathcal{X}$ to an output $Y \in \mathcal{Y}$, where $\mathcal{X} \subseteq \mathbb{R}^d$ is the domain of the model inputs, with dimension $d \in \mathbb{N}$, and $\mathcal{Y} \subseteq \mathbb{R}^{d'}$ is the domain of the model outputs, with dimension $d' \in \mathbb{N}$. Model evaluations (i.e., evaluations of F) imply computational costs $c \in \mathbb{R}^+$ that typically increase with the accuracy of the output approximation.

For a given atmosphere, there exists different methods for simulating the propagation of an acoustic wave with different accuracies and different costs. The three basic approaches to model the long-range propagation of acoustic waves are : the geometric acoustics approximation or ray tracing [Drob et al., 2013, Hedlin and Drob, 2014, Lalande and Waxler, 2016], the resolution of the wave equation (using normal mode method [Millet et al., 2007, Bertin et al., 2014, Assink et al., 2017] or the parabolic approximation [Klyatskin and Tatarskii, 1970, Ostashev et al., 2019, Blanc-Benon et al., 2001]) and the resolution of Navier Stokes equations (with finite differences or a finite element method [Sabatini et al., 2019a, Sabatini et al., 2019b, Sabatini et al., 2016]). The ray tracing has the great advantage of reproducing the basic properties of infrasonic signals observed during experiments with a rather low computational cost. However, this approach is restricted to high frequency waves, and thus, fails to predict some important mean flow refraction effects such as mountainwakes or jet streams for example. For a downward refracting atmosphere, normal mode decomposition of the acoustic wave equation can give a good approximation of the solution [Waxler et al., 2017, Assink et al., 2017]. In all other cases the problem has to be solved with direct numerical simulations [Sabatini et al., 2019a, Sabatini et al., 2019b] or other computationally expensive techniques.

Every numerical method has its approximation error. In addition, the description of the propagation medium brings its own sources of uncertainties. These errors on the inputs X may be classified in two categories using the standard definition [Matthies, 2007, Der Kiureghian and Ditlevsen, 2009] :

- the *epistemic* (or reducible) uncertainties, which model mainly a lack of knowledge. In the case of atmospheric propagation, those uncertainties arise from atmospheric specifications products.

- the *aleatoric* (or stochastic) uncertainties, which measure the intrinsic variability of a system. The atmosphere is a dynamical system in which natural phenomena can evolve and change the propagation conditions.

3.2 Metamodels

Direct uncertainty propagation consists in propagating the variability of the inputs through the numerical model to assess the impact on the QoI. The simplest method to conduct such an analysis is Monte-Carlo simulation. This method is fairly simple to implement, can be easily parallelized and more importantly is insensitive to the dimension d of the input. Nevertheless, the convergence of this method is in $O\left(1/\sqrt{N}\right)$ where N is the number of sample, which is rather slow and only provides statistical

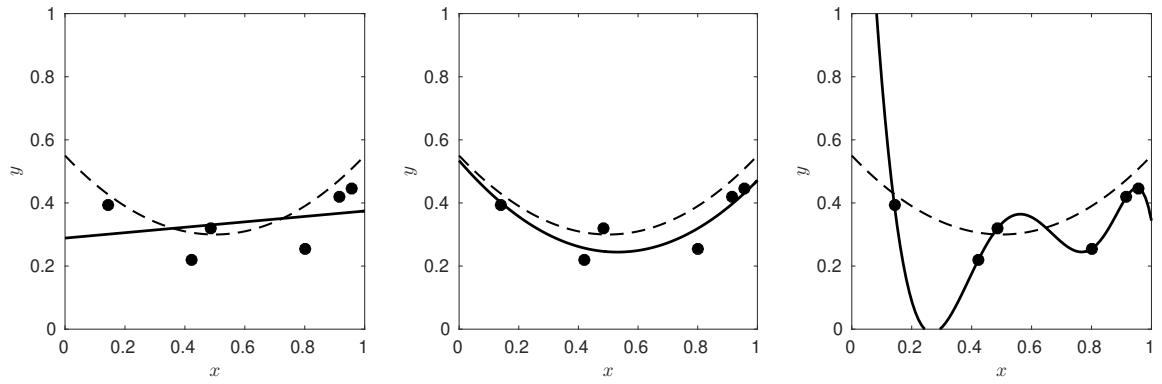

FIGURE 1.4 : Polynomial fit of data points $(x_i, y_i)_{i=1,..,6}$ defined by the expression : $y_i = x_i^2 - x_i + 0.5 + 0.05\eta_i, \eta_i \sim \mathcal{N}(0,1)$ (dashed line gives the original polynomial). A polynomial of order 1 under-fits the data (left), a good fit is obtained by a polynomial of degree 2 (center) and a polynomial of degree 9 overfits the data (right), although err_{emp} is small.

estimators.
There exists many refinements to the classical Monte-Carlo simulations. The simplest consists in using sampling strategies (such as *Latin Hypercube Sampling (LHS)* [McKay et al., 1979] or *Sobol' sequence* [Sobol', 1967]) to accelerate the convergence to a $O\left(\left(\log(N)^d\right)/N\right)$. In any cases those methods require many calls to the numerical model which becomes prohibitive in the case of a computationally demanding solver. For this reason, the development of alternative techniques has been an active field of research in the recent years. Those techniques (called *surrogate models* or *metamodels*) are designed to be more efficient than the classical Monte-Carlo approach but are this time very sensitive to the dimension d. The idea is to calibrate a new model $\hat{F} : \mathcal{X} \to \mathcal{Y}$ using a small number of runs of F in order to have a model $\hat{F}$ easier to assess numerically and reproducing the statistical properties of F.

The natural definition of the error of the metamodel $\hat{F}$ is called the *generalization error* in statistical learning [Vapnik, 1995] and defined by :

$$\text{err} = \mathbb{E}\left[\left(F(X_z(\xi)) - \hat{F}(X_z(\xi))\right)^2\right]. \tag{1.10}$$

In order to estimate this error, a natural method is to use the experimental design : $\left(\xi^{(1)}, \ldots, \xi^{(N)}\right)$, used to calibrate the metamodel :

$$\text{err}_{emp} = \frac{1}{N}\sum_{i=1}^{N}\left(F(X_z(\xi^{(i)})) - \hat{F}(X_z(\xi^{(i)}))\right)^2. \tag{1.11}$$

However, it is well-known that this estimation of the error is biased and conducts to over-fitting, as figure 1.4 shows. To circumvent this problem, the idea is to quantify the approximation of the metamodel on points that have not been used for the calibration. There exists mainly two methods for validation :

- The first idea is to use a test set $(\xi^{(t_1)}, ..., \xi^{(t_n)})$, and use the Q_2 criterion to quantify the behaviour of the metamodel on those points :

$$Q_2 = 1 - \frac{\sum_{i=1}^{n} (F(X_z\left(\xi^{(t_i)}\right)) - \hat{F}(X_z\left(\xi^{(t_i)}\right)))^2}{\sum_{i=1}^{n} (F(X_z\left(\xi^{(t_i)}\right)) - \mathbb{E}[F(X_z\left(\xi\right))])^2}, \tag{1.12}$$

  when the model perfectly estimates the test values, $Q_2 = 1$ and when $Q_2 = 0$ it means that a constant value $\mathbb{E}[F(\xi)]$ can do as well as the model. Q_2 can be negative, in this case it means that a constant model is doing a better job than the metamodel.

- In the case of a very expensive model F, a cross-validation technique can be used [Stone, 1974, Geisser, 1975], it consists in dividing the sample into two subsamples : one *training set* and one *validation set*. A refinement of this technique (called *ν-fold cross-validation*) consists in using a random partition of the sample : the learning set will contain in turn all but one of the subset which is considered as the test set. The generalization error is estimated for each of the sets and then averaged over all possibilities. In the case of subsets of size 1, this method is called *Leave-One-Out*.

3.3 Complexity of metamodels

Another interesting approach to analyze the performance of a metamodel can be found in computational learning theory, a subfield of machine learning focusing on the ability of algorithms to learn a model. To this purpose, the notion of *probably approximately correct* (PAC) learning has been introduced by Leslie Valiant in 1984 [Valiant, 1984] to characterize a concept which can be learned by an algorithm. Formally, a *concept* is an application from a set of inputs to a set of outputs (which can be either a collection of labels or a set of possible values) and a concept class $\mathcal{C}$ is a collection of concepts.

A concept class $\mathcal{C}$ is said to be PAC learnable if there exists an algorithm $\mathcal{A}$ that uses random learning sets $\mathcal{S}$ of size m and positive constants δ and ϵ such that : $\mathbb{P}(\text{err}_{\mathcal{S}}(\mathcal{A}) < \epsilon) \geq 1 - \delta$ where $\text{err}_{\mathcal{S}}(A)$ is the generalization error of the algorithm using sample $\mathcal{S}$. If m is polynomial in $1/\epsilon$ and $1/\delta$, $\mathcal{C}$ is said to be efficiently PAC learnable.

Changing the viewpoint, we can ask what class of concepts can be learned by a given algorithm. The *Vapnik-Chevronenkis dimension* (VC) for instance quantifies the ability of a classification algorithm to shatter a set of points [Blumer et al., 1989]. In other words, it quantifies the capacity of a space of functions to be learned by a statistical classification algorithm. The strength of the VC dimension is to give a probabilistic upper bound on the test error. However, it only applies to classification, which is not the classical use of metamodels.

The *Rademacher complexity* is a more modern notion of complexity [Koltchinskii, 2001, Bartlett and Mendelson, 2002] that is defined for real-valued functions – and

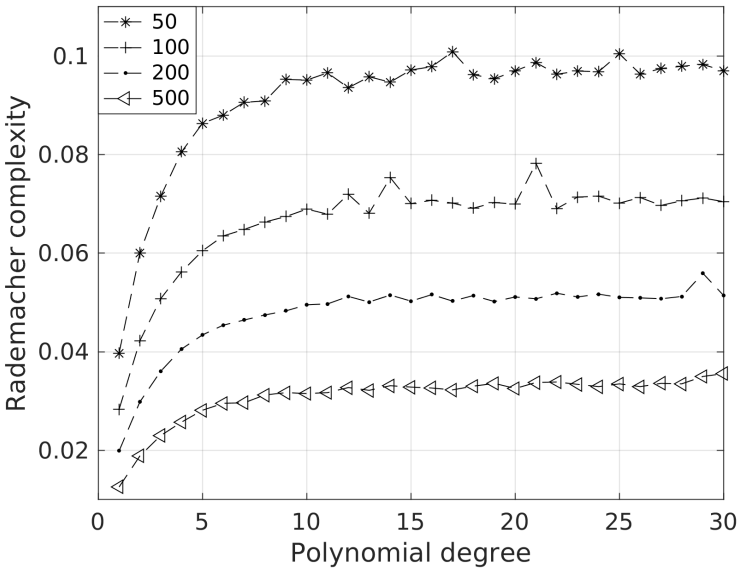

FIGURE 1.5 : Rademacher complexity of Polynomial Chaos Expansion for different sample sizes.

not only discrete-valued functions (as for VC dimension and PAC). Let $\mathcal{F}$ be a class of real functions and $\mathcal{S} = (z_1, .., z_m)$ a sample drawn of independents and identically distributed (from a distribution D) inputs, the empirical Rademacher complexity is defined by :

$$\hat{R}_m(\mathcal{F}) = \mathbb{E}_\sigma \left[\sup_{f \in \mathcal{F}} \left(\sum_{i=1}^{m} \sigma_i f(z_i) \right) \right], \tag{1.13}$$

where σ_i are random variables uniformly chosen over $\{-1, 1\}$. The Rademacher complexity is then defined by : $R_m(\mathcal{F}) = \mathbb{E}_D \left[\hat{R}_m(\mathcal{F}) \right]$.

Intuitively, in formula (1.13) the supremum measures, for a set $\mathcal{S}$ and a random vector σ, the maximum correlation between $f(z_i)$ and σ_i over all $f \in \mathcal{F}$. Taking the expectation over σ, we can then say that the empirical Rademacher complexity of $\mathcal{F}$ measures the ability of functions from $\mathcal{F}$ (when applied to a fixed set $\mathcal{S}$) to fit random noise. The Rademacher complexity of $\mathcal{F}$ then measures the expected noise-fitting-ability of $\mathcal{F}$ over all data sets $\mathcal{S}$ of size m.

The Rademacher complexity depends on the distribution attached to the sample $\mathcal{S}$. For a very general machine learning algorithm, this distribution is unknown but in uncertainty quantification, a characterization of the inputs has been done before.

Figure 1.5 shows the Rademacher complexity computed for generalized Polynomial Chaos (gPC) expansions using different sample size m. gPC is one of the techniques used to build a metamodel, it consists in decomposing the QoI on a polynomial basis adapted to the input distribution D. On figure 1.5, the Rademacher complexity increases with the polynomial degree to a certain value and then remain stable. In fact, the metamodel learns more from the sample when the polynomial degree increases,

but this learning ability is all the same limited. The fact that the Rademacher complexity decreases when the sample size increases is due to the definition : for a fixed polynomial order, the approximation gives better results on a sample of small size.

4 Sensitivity analysis

If the purpose of a metamodel is to propagate the uncertainties, one of its classical use is to conduct a sensitivity analysis of the QoI. It consists in studying the impact of uncertainties of each input on the QoI. This analysis can be useful to test the robustness of a result, to select the most relevant variables, to reduce the variance of the QoI by working on the input variables.

4.1 Sobol' indices

Variance-based decomposition has been introduced by the Russian mathematician Ilya M. Sobol' [Sobol', 1993] and consists in decomposing the variance of the QoI into the sum of the variances caused either by a single input or by the coupling between several inputs. This approach is also called Global Sensitivity Analysis [Saltelli et al., 2008], in opposition to Local Sensitivity Analysis, which consists in computing the partial derivatives of the output with respect to the different input variables.

We assume that our problem can be written $Y = F(\xi_1, \dots, \xi_d)$, where Y is the quantity of interest and $\xi_1, \xi_2, \dots, \xi_d$ are independent variables taking values on the unit hypercube of dimension d, Ω^d. This last assumption stands only for convenience and is not a limitation since every support can be transformed in the hypercube. Some generalizations to dependent input variables ξ exist [Chastaing et al., 2015] but are non-trivial and will not be discussed here.

The following decomposition is based on the work of Hoeffding [Hoeffding, 1948] on the projection of Y on orthogonal subspaces :

$$Y = F_0 + \sum_{i=1}^{d} F_i(\xi_i) + \sum_{i<j}^{d} F_{ij}(\xi_i, \xi_j) + \cdots + F_{1,2,\dots,d}(\xi_1, \xi_2, \dots, \xi_d), \tag{1.14}$$

with the following condition that implies orthogonality and thereby the unicity of the decomposition :

$$\int_0^1 F_{i_1 i_2 \dots i_s}(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_s}) d\xi_k = 0, \text{ for } k = i_1, ..., i_s. \tag{1.15}$$

The coefficients of the decomposition are defined as follows :

$$F_0 = \mathbb{E}(Y), \tag{1.16}$$
$$F_i(\xi_i) = \mathbb{E}(Y|\xi_i) - F_0, \tag{1.17}$$
$$F_{ij}(\xi_i, \xi_j) = \mathbb{E}(Y|\xi_i, \xi_j) - F_0 - F_i - F_j. \tag{1.18}$$

Once the function has been decomposed, if Y is squarely integrable, equation (1.14) gives thanks to the orthogonality :

$$\int_0^1 [F(\xi)]^2 \, d\xi - F_0^2 = \sum_{s=1}^{d} \sum_{i_1<\cdots<i_s}^{d} \int F_{i_1\ldots i_s}^2 d\xi_{i_1} \ldots d\xi_{i_s}. \tag{1.19}$$

This expression has to be understood in terms of variance, if we denote $V_i = \mathrm{Var}_{\xi_i}\left(\mathbb{E}_{\boldsymbol{\xi}_{\sim i}}(Y \mid \xi_i)\right)$ and $V_{ij} = \mathrm{Var}_{\xi_{ij}}\left(\mathbb{E}_{\boldsymbol{\xi}_{\sim ij}}(Y \mid \xi_i, \xi_j)\right) - \mathrm{V}_i - \mathrm{V}_j$, then (1.19) can be rewritten :

$$\mathrm{Var}(Y) = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + \cdots + V_{12\ldots d}. \tag{1.20}$$

This decomposition of the variance has the advantage to separate the effects of the different inputs. From this expression we can define the first order Sobol' index :

$$S_i = \frac{V_i}{\mathrm{Var}(Y)}. \tag{1.21}$$

The same definition can be derived for higher order Sobol' indices. However, there are 2^d indices and it is more convenient to use directly the total Sobol' index S_{T_i} which consists in summing all the contributions involving ξ_i. If the sum of all the Sobol' indices is normalized thanks to (1.20), the sum of all the total Sobol' indices is greater than one :

$$\sum_{i=1}^{d} S_i + \sum_{i<j}^{d} S_{ij} + \cdots + S_{12\ldots d} = 1, \tag{1.22}$$

$$\sum_{i=1}^{d} S_{T_i} \geq 1. \tag{1.23}$$

One can sometimes be interested in the second order Sobol' index S_{ij}, which characterizes the effect of the interaction between the two input variables ξ_i and ξ_j.

From a numerical point of view, the simplest way to compute the Sobol' indices is based on Monte-Carlo simulation [Sobol', 1993, Homma and Saltelli, 1996, Saltelli, 2002]. As we will see in the sequel, there exists more efficient methods based on metamodeling techniques [Sudret, 2008, Marrel et al., 2009].

To compute the first and total order Sobol' indices, one can proceed as follows :

- Generate two matrices A and B containing N realizations of the N inputs variables (A and B are matrices of size $N \times d$) and create matrix $C_i, i = 1, \ldots, d$ whose columns are taken from B except the i-th one which is taken from A.
- Compute the images using the model F : $y_A = F(A)$, $y_B = F(B)$, $y_{C_i} = F(C_i)$.
- The first-order and total order Sobol' indices are given by :

$$S_i = \frac{y_A.y_{C_i} - F_0^2}{y_A.y_A - F_0^2}, \tag{1.24}$$

$$S_{T_i} = 1 - \frac{y_B.y_{C_i} - F_0^2}{y_A.y_A - F_0^2}. \tag{1.25}$$

As in the case of a classical Monte-Carlo estimation, the convergence of this estimation is in $O\left(1/\sqrt{N}\right)$ but can be accelerated using space-filling sampling techniques. However, one advantage of Monte-Carlo simulations is the ability to obtain confidence intervals.

To reduce the dimension of the input parameter space, one can use the Sobol' indices to select the uncertain inputs which have the highest impact on the QoI.

4.2 Shapley effects

If Sobol' indices are perhaps the most popular measure of sensitivity, their fundamental assumption of independence among inputs can be problematic for some applications. In our case, the input variables are given by the Karhunen Loève decomposition and are not independent in the general case (independence is true only in the gaussian case). To overcome this limitation, some recent studies advocate for the use of Shapley effects [Song et al., 2016, Iooss and Prieur, 2017].

Shapley effects have been proposed in a context of sensitivity analysis by Owen in [Owen, 2014]. Originally used in economics, the concept of Shapley effect has been introduced in game theory. In this context, if we consider a game with a set of player $\mathcal{K} = \{1, 2, .., k\}$ the Shapley value v_i quantifies the incremental cost of including players i in set $\mathcal{J}$ averaged over all sets $\mathcal{J} \subseteq \mathcal{K} \setminus \{i\}$ for a given cost function c :

$$v_i = \sum_{\mathcal{J} \subseteq \mathcal{K} \setminus \{i\}} \frac{(k - |\mathcal{J}| - 1)!|\mathcal{J}|!}{k!} \left(c\left(\mathcal{J} \cup \{i\}\right) - c\left(\mathcal{J}\right)\right). \tag{1.26}$$

Notice that $\frac{(k-|\mathcal{J}|-1)!|\mathcal{J}|!}{k!} = \frac{1}{k}\binom{k-1}{|\mathcal{J}|}^{-1}$, this factor gives equal weight $1/k$ to all possible size of subsets $\mathcal{J}$ of $\mathcal{K} \setminus \{i\}$ and equal weight $\binom{k-1}{|\mathcal{J}|}^{-1}$ to all possible subsets of size $|\mathcal{J}|$. An important property of the Shapley values is its ability to sum to the total cost :

$$\sum_{i=1}^{k} v_i = c(\mathcal{K}). \tag{1.27}$$

For sensitivity analysis purposes, the set of players $\mathcal{K}$ can be seen as the set of inputs. For the cost function $c\left(\mathcal{J}\right)$, one would want to quantify the variance of Y caused by the uncertainties of the inputs in $\mathcal{J}$. A basic property required for the cost function is that : $c(\emptyset) = 0$ and $c\left(\mathcal{K}\right) = \mathrm{Var}\left(Y\right)$. The Shapley effects will then sum to the total variance thanks to (1.27). Owen proposed in [Owen, 2014] the following expression for the cost function :

$$\tilde{c}\left(\mathcal{J}\right) = \mathrm{Var}\left(\mathbb{E}[Y|X_{\mathcal{J}}]\right), \tag{1.28}$$

where $X_{\mathcal{J}}$ is the vector of input which indices are in $\mathcal{J}$. Song et al. proposed in [Song et al., 2016] another cost function c and proved that the Shapley value associated to c is equivalent to the one defined with $\tilde{c}$:

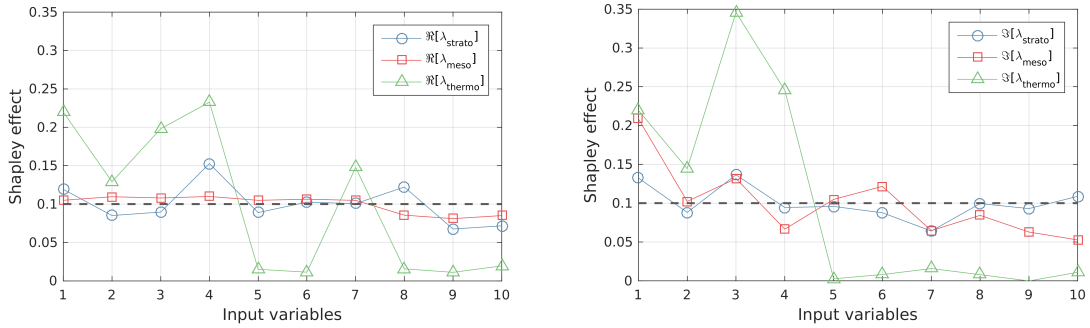

FIGURE 1.6 : Shapley effects of the real and imaginary part of three modes from stratosphere (in blue), mesosphere (in red) and thermosphere (in green).

$$c(\mathcal{J}) = \mathbb{E}[\mathrm{Var}(Y|X_{-\mathcal{J}})], \tag{1.29}$$

where $X_{-\mathcal{J}}$ is the vector of input which indices are not in $\mathcal{J}$. The advantage of this second cost function is that it allows more efficient practical computation of the Shapley effect. But in our case we cannot use directly the algorithm given by [Song et al., 2016] because we do not have explicit formulations for the input distributions but only a sample given by the PCA fitted on the data. In this context, a consistent estimator has been proposed by Broto et al.[3] [Broto et al., 2018].

On Figure 1.3, we have computed the decomposition of a wind velocity perturbation induced by gravity waves. We kept ten random variables emerging from this decomposition (the first eight modes and the 50^{th} and 100^{th}) as the input for our propagation model. The choice of those ten KL-modes allows to keep a good proportion of the perturbation variance and also to consider different spatial structures in the profiles. The idea here was to investigate the impact of the different scales of the perturbation on the different waveguides. As we will explain in chapter 1, our QoI will be the acoustic modes that characterize the different waveguides. Figure 1.6 shows the Shapley effects of three particular modes : one in the stratosphere (with a refraction altitude of around 40 km), one in the mesosphere (with a refraction altitude of around 80 km) and one in the thermosphere (with a refraction altitude of around 100 km) at 0.1Hz. The dashed line represents the Shapley effects in the case of equally important inputs. For the stratospheric and mesospheric modes, there is not any input which stands out. However, it seems that the first four modes have a greater impact than the six others on the thermospheric mode.

This preliminary study allowed us to introduce the kind of uncertainties at stake for atmospheric propagation and a first illustration of their impact on the acoustic propagation. However, this study remains incomplete and the purpose of this work will be to show how metamodelling techniques can be used to study the impact on the propagation (including the impact on the received signals) in a computationnally efficient way.

[3]This estimation is available in the R package *sensitivity* and has been used to produce Figure 1.6.

4.3 Sensitivity analysis applied to infrasound propagation

In the context of infrasound propagation, the inputs describe the fluctuation of the atmospheric state, mainly due to the meteorological conditions. The output can be at different stages, depending on the purpose of the simulation : it can be for instance the acoustic modes, quantities extracted from the temporal signal or another quantites computed using the result of the simulation. We give here three examples of application of sensitivity analysis to infrasound propagation computing the impact of fluctuations of the input on different outputs :

1. the decomposition of the perturbation according to paragraph 2.2 separates the different scales of the input, as a consequence, computing the Shapley effects (like in the previous paragraph) allows to distinguish which scales have an impact on the different acoustic modes.

2. in the case of other quantities of interest such as time of arrival or duration of the signal, a sensitivity analysis can be conducted to isolate the effect of the different scales of the perturbation on the signal.

3. if, as in chapter 4, the propagation is used to localize an event using infrasound, the sentivity analysis can give information on the critical resolution of the meteorological data necessary to a good localization.

Those three examples of application of sensitivity analysis are thought in the case of a decomposition of the input separating the different scales. If, on the contrary, one wants to study the impact of perturbations localized at a certain altitude, it could be done by designing a profil with shape parameters that will be the inputs. This is the case in chapter 3 where the impact of a nocturnal jet is studied.

5 Outline of the manuscript

The first chapter of this manuscript presents all the methods used in this work : the modal decomposition for the acoustical decomposition, its practical implementation and the framework of polynomial chaos, which is one of the methods used to build a metamodel.

The second chapter presents an application to the planetary boundary layer with a theoretical perturbation. The aim of this chapter is to show how the method can be used on a simple case for which comparison with Monte-Carlo estimation are affordable.

The last chapter deals with the problem of localization of a source detected by the IMS. In this context, the use of a metamodel of the propagation makes affordable a bayesian treatment.

Chapitre 2

Modal Expansions

1 Wave equation

We start this methodological chapter on the wave equation whose derivation from fluid mechanics is briefly recalled (with specifications used in the sequel) in a first paragraph. The normal mode decomposition of the wave operator is presented using spectral operator theory. Then, the numerical computation of the normal modes is discussed with a presentation of both finite differences and spectral collocation method. This section ends with a study on the numerical precision of the normal modes computation.

1.1 Atmospheric infrasound propagation

Infrasound generated by events of great amplitude can propagate over large distances (up to thousands of kilometres) in the atmosphere thanks to the ducts created by atmospheric temperature and wind speed gradients. The propagation is more efficient in the lower ducts of the atmosphere, for which losses as a result of geometrical spreading and absorption are minimal. In practice, tropospheric and stratospheric waveguides are most influential in determining long-range propagation conditions.

Figure 2.1 illustrates the directionality of sound propagation. A stratospheric jet (strong wind located at around 45 km) induces very different propagation conditions depending on which side of the source (located in 0 on the figure) the receiver is located. As the transmission loss representation suggests in figure 2.1, the received signal will result from the superposition of different wavepackets, each one coming from a different waveguide. Figure 2.2 shows the evolution of the propagation of an impulsive signal in such an atmosphere. The direct arrivals vanish as the distance increases but stratospheric (with a refracting altitude around 45 km) and thermospheric (with a refracting altitude around 90 km) waveguides lead to respectively five and two arrivals. The thermospheric arrival is attenuated by the absorption, which is much more significant for high altitudes.

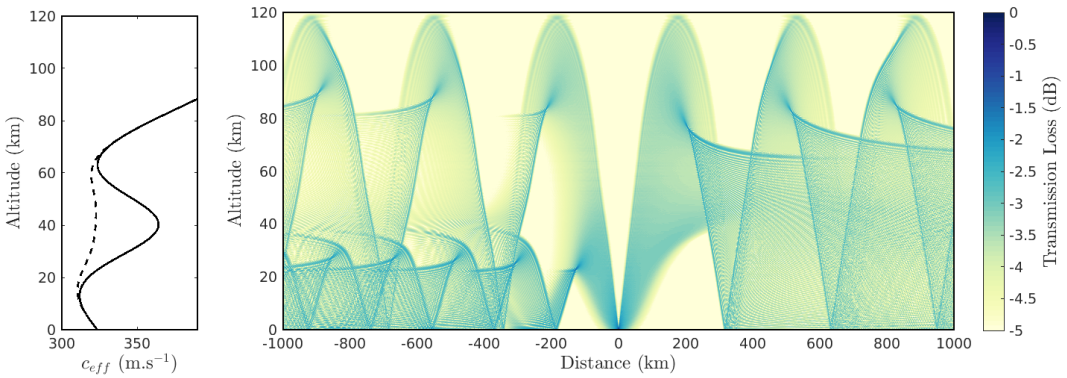

FIGURE 2.1 : Sound propagation in an atmosphere with a stratospheric wind propagating from the right to the left. The resulting effective sound speed profiles are presented (left, solid line for right to left profile and dashed line for left to right profile) and the transmission loss at a frequency of 1Hz (right) for a source on the ground.

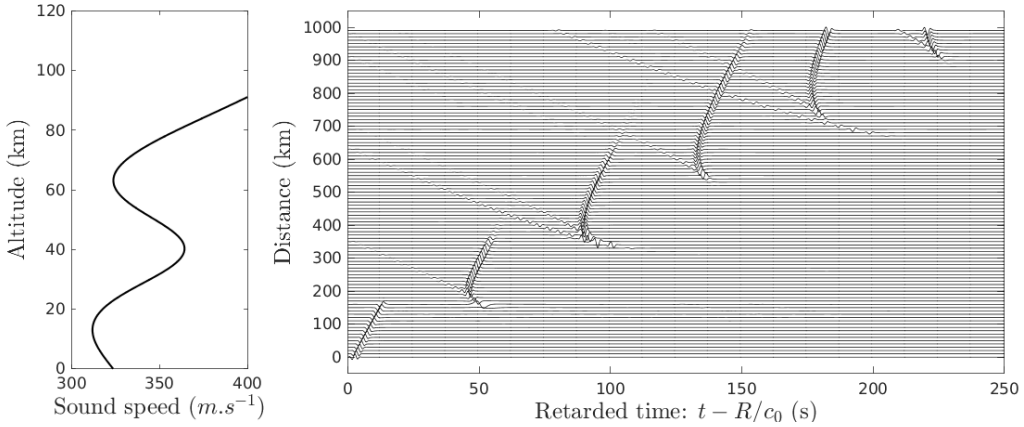

FIGURE 2.2 : Propagation of an impulsive signal in an atmosphere characterised by the sound speed profile given on the left. Time scale has been translated using a reference sound speed $c_0 = 330$ m.s^{-1}. The source signal has a central frequency at 0.4Hz.

Equation for the propagation

A classical framework for simulating wave propagation relies on the linear wave equation [Waxler et al., 2017, Assink et al., 2017, Lalande and Waxler, 2016] obtained from a linearization of the Euler equations. We are going to work in cylindrical coordinates and thus write the linear wave equation for a pressure field $p(r, z, t)$ in a stratified medium of density $\rho(z)$ with an horizontal wind velocity $v(z)$ [Brekhovskikh and Godin, 1999] :

$$\mathcal{D}_t \left[\frac{1}{c_a^2} \mathcal{D}_t^2 p - \Delta_\perp p - \rho \frac{\partial}{\partial z} \frac{1}{\rho} \frac{\partial p}{\partial z} \right] + 2 \left(\frac{dv}{dz} . \nabla_\perp \right) \frac{\partial p}{\partial z} = 0, \tag{2.1}$$

where $\mathcal{D}_t = [(\partial/\partial t) + v.\nabla_\perp]$, $\nabla_\perp$ (respectively $\Delta_\perp$) being the gradient (respectively the laplacian) in the horizontal direction and $c_a(z)$ is the adiabatic sound speed :

$$c_a(z) = \sqrt{\gamma r_g T(z)}, \tag{2.2}$$

where r_g is the gas constant for air and γ the adiabatic index. We will neglect the vertical wind shear, that cancels the second term in (2.1). Concerning the first term, if we consider that the mean flow is close to the direction of propagation $\vec{e_r}$ equation (2.1) can be approximated by :

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} - \Delta_\perp p - \rho \frac{\partial}{\partial z} \frac{1}{\rho} \frac{\partial p}{\partial z} = 0, \tag{2.3}$$

where $c(z) = c_a(z) + \overrightarrow{v(z)}.\vec{e_r}$ is the effective sound speed (with $\overrightarrow{v(z)}$ a vector in the direction of the mean flow and of magnitude $v(z)$). From this equation we can perform a Fourier transform in time and therefore use the variable $u(\omega, r, z)$, solution of :

$$\left[\frac{\omega^2}{c^2} + \Delta_\perp + \frac{\partial^2}{\partial z^2} + H \right] u = 0, \tag{2.4}$$

where H has the dimension of a frequency and writes :

$$H = \frac{1}{2} \frac{\rho''}{\rho} - \frac{3}{4} \left(\frac{\rho'}{\rho} \right)^2. \tag{2.5}$$

In the applications of this thesis we will have a celerity around 300m.s^{-1} and frequencies between 0.1Hz and 20Hz. With those conditions we have $Hc^2/\omega^2 \ll 1$ and will mainly use the following form of the wave equation :

$$\Delta u + \frac{\omega^2}{c^2} u = 0, \tag{2.6}$$

where $\Delta = \partial^2/\partial z^2 + \Delta_\perp$ is the laplacian.
To solve the propagation problem we will need some boundary conditions describing the behaviour of the pressure field when $z = 0$ and $z \to +\infty$. On the top of the domain we will only require some bounded solutions (this will exclude exponentially

increasing functions). At the ground, we will assume in this thesis that the ground is perfectly reflecting waves. This imposes a Neumann homogeneous condition :

$$\frac{du}{dz}(0) = 0. \tag{2.7}$$

A more realistic modelling would consist in assuming that reflections on the ground attenuate the propagation [Attenborough, 1988]. To take into account this attenuation, a complex impedance $Z(\omega)$[1] is often taken, leading to a Robin boundary condition [Waxler, 2002] :

$$\frac{du}{dz}(0) = -\frac{-i\omega\rho(0)}{Z(\omega)}(\omega)u(0). \tag{2.8}$$

1.2 Normal modes decomposition

We are interested in solving the wave equation (2.6) in the case of a point source $s(t)$ located at $r = 0, z = z_s$. This problem writes :

$$\Delta u + \frac{\omega^2}{c(z)^2}u = \delta(r)\delta(z - z_s)\hat{s}(\omega), \tag{2.9}$$

where $\hat{s}(\omega)$ is the Fourier transform of $s(t)$. We will consider here that the medium is invariant in the propagation direction and for this reason the effective sound speed depends only on the vertical coordinate $c = c(z)$. It is then natural to split the Laplacian in two terms $\Delta = \partial_z^2 + \Delta_\perp$ and to define the differential operator $\mathcal{L}$ that describes the vertical structures :

$$\mathcal{L} = \frac{d^2}{dz^2} + \frac{\omega^2}{c(z)^2}, \tag{2.10}$$

whose domain writes :

$$D(\mathcal{L}) = \left\{\psi \in H^2\left(\mathbb{R}^+\right) | \psi'(0) = 0, \lim_{z\to\infty}\psi(z) = 0\right\}. \tag{2.11}$$

With this definition, the left hand side of equation (2.9) becomes : $\mathcal{L}u + \Delta_\perp u$.

The spectral measure associated to $\mathcal{L}$

The idea of spectral decomposition is to build a unitary transformation $\mathcal{F}$ from the 'physical space' $D(\mathcal{L})$ to a 'spectral space' that contains functions of the spectral variable $\lambda \in \sigma(\mathcal{L})$, where $\sigma(\mathcal{L})$ is the spectrum of $\mathcal{L}$. The strength of this transformation is to *diagonalize* $\mathcal{L}$, in the sense that it converts the action of $\mathcal{L}$ into a simple multiplication by λ : $\mathcal{L} = \mathcal{F}^*\lambda\mathcal{F}$.

The spectral theorem guarantees the existence of such a transformation $\mathcal{F}$ for every self-adjoint operator, which is the case of $\mathcal{L}$:

[1] As explained in paragraph 1.2 an important property is that the differential operator associated with (2.6) is self-adjoint. With this complex impedance the operator is not a self-adjoint operator anymore.

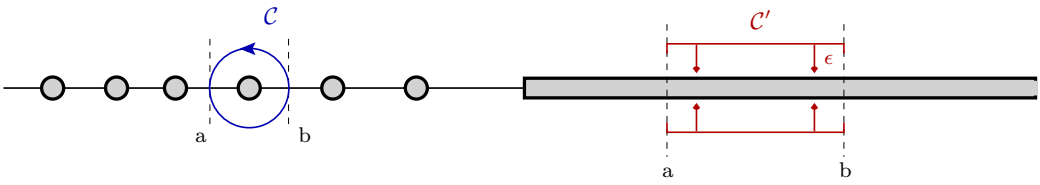

FIGURE 2.3 : Computation of the spectral measure using either Cauchy (contour $\mathcal{C}$) or Stone formula (contour $\mathcal{C}'$).

$$(\mathcal{L}\phi, \psi) = (\phi, \mathcal{L}\psi), \quad \forall \phi, \psi \in D(\mathcal{L}), \tag{2.12}$$

where $(,)$ is the classical inner product in $L^2(\mathbb{R}^+)$.

More precisely, the spectral theorem ensures the existence of a spectral measure E^2 (also called resolution of the identity) such that :

$$\mathcal{L}\phi = \int_{\mathbb{R}} \lambda \, dE_\lambda \phi \quad \forall \phi \in D(\mathcal{L}). \tag{2.13}$$

For practical computation of the spectral measure, Cauchy and Stone formulae give expressions using the resolvent $R(z) = (\mathcal{L} - zId)^{-1}$:

- *(Cauchy formula)* for $a, b \in \mathbb{R}, \forall \epsilon > 0, \sigma(\mathcal{L}) \cap]a-\epsilon; a+\epsilon[= \sigma(\mathcal{L}) \cap]b-\epsilon; b+\epsilon[= \emptyset$ (i.e. a and b are not in the neighbourhood of the spectrum of $\mathcal{L}$) :

$$E([a;b]) = E(]a;b[) = \frac{1}{2i\pi} \oint_{\mathcal{C}} R(s)\, ds, \tag{2.14}$$

  where $\mathcal{C}$ is a closed contour around $[a;b]$ which crosses the real axis in a and b ;

- *(Stone formula)* for $a, b \in \mathbb{R}$ such that $a < b$,

$$\frac{1}{2}\left(E(]a;b[) + E([a;b])\right) = \lim_{\epsilon \to 0} \frac{1}{2i\pi} \int_a^b \left(R(x+i\epsilon) - R(x-i\epsilon)\right) dx. \tag{2.15}$$

The Stone formula gives the general case, in particular it allows to compute the contribution of the continuous part spectrum. However, it requires the expression of the resolvent which is not known in the general case (a complete study has been done by Christophe Hazard in the particular case of the Pekeris waveguide, which can be derived analytically [Hazard, 2009]). For this reason, we will choose the Cauchy

[2] A spectral measure E is an application from the Borel subsets of $\mathbb{R}$ in the space of orthogonal projection on an Hilbert space $\mathcal{H}$ which fulfils the classical properties of real measures :

1. $E(\emptyset) = 0$ and $E(\mathbb{R}) = Id$
2. $E(\bigcup_{k=0}^{\infty} I_k) = \sum_{k=0}^{\infty} E(I_k)$, for $(I_k)_{k\in\mathbb{N}}$ disjoint Borel subsets of $\mathcal{X}$ (σ-additivity)
3. $E(I \cap J) = E(I)E(J)$ for I and J parts of $\mathcal{X}$.

formula using the residue theorem to compute the integral. It will only provide the contribution of the discrete part of the spectrum but it has been shown that the continuous contribution was negligible at the ground far from the source [Waxler, 2003].

In order to use those formulae, we need an expression of the resolvent $R(\zeta)$ for all $\zeta \in \mathbb{C}$. We will use the following integral expression :

$$(R(\zeta)\phi)(z) = \int_{\mathbb{R}^+} \gamma_\zeta(z, z')\,\phi(z')\,dz', \tag{2.16}$$

where γ_ζ is the Green function :

$$\begin{cases} \gamma_\zeta''(., z') + \left(\frac{\omega^2}{c(z)^2} - \zeta\right)\gamma_\zeta(., z') & = \delta_{z'}, \\ \gamma_\zeta'(0, z') & = 0, \\ \lim\limits_{z\to\infty} \gamma_\zeta(z, z') & = 0, \end{cases} \tag{2.17}$$

where $\delta_{z'}$ denotes the Dirac measure at $z = z'$. In order to compute the spectral measure associated to our operator $\mathcal{L}$ using the Cauchy or Stone formulae we need to know the dependencies of γ_ζ in ζ. If we cannot have a closed expression for γ_ζ because it will depend on the profile $c(z)$, we will still be able to express its contribution to integral (2.14) using the residue theorem.

For a fixed z' in $\mathbb{R}^+$, equation (2.17) can be written using the jump formula :

$$\{\gamma_\zeta''\} + \left(\frac{\omega^2}{c(z)^2} - \zeta\right) + \sigma_0\delta_{z'} + \sigma_1\delta_{z'}' = \delta_{z'}, \tag{2.18}$$

where $\{f\}$ stands for the regular distribution associated to the function f and σ_0 and σ_1 are respectively the jumps of γ_ζ and of its derivative in $z = z'$. Because $\delta_{z'}$ and $\delta_{z'}'$ are independent, equation (2.17) gives $\sigma_0 = 1$ and $\sigma_1 = 0$.

On $[0; z']$ and $[z'; +\infty]$, we can define two solutions γ_+ and γ_- and the solution on $\mathbb{R}^+$ will be a combination of those two solutions : $\gamma_\zeta = A\gamma_+ + B\gamma_-$. In $z = z'$, the jump formula gives :

$$\begin{cases} A\gamma_+(z') - B\gamma_-(z') & = 1, \\ A\gamma_+'(z') - B\gamma_-'(z') & = 0. \end{cases} \tag{2.19}$$

We can define the Wronskian : $W(z', \zeta) = \gamma_+(z')\gamma_-'(z') - \gamma_+'(z')\gamma_-(z')$. When $W(z', \zeta) \neq 0$, the solution of system (2.19) writes :

$$\begin{aligned} A &= \gamma_-(z')/W(z', \zeta), \\ B &= \gamma_+(z')/W(z', \zeta), \end{aligned}$$

and the solution of equation (2.17) is :

$$\begin{cases} \gamma_\zeta(z, z') = \gamma_-(z')\gamma_+(z)/W(z', \zeta) & \text{for} \quad z > z', \\ \gamma_\zeta(z, z') = \gamma_+(z')\gamma_-(z)/W(z', \zeta) & \text{for} \quad z < z'. \end{cases} \tag{2.20}$$

If we denote $(\zeta_i)_{i\in\mathcal{Z}}$ the zeros of the Wronskian W, then the former expression is not valid on those points. However, one can notice that for such a ζ_i the solutions γ_+ and γ_- are proportional ($\gamma_+ = \alpha\gamma_-$) and then the function defined by :

$$\begin{cases} \gamma_i(z) & = \gamma_+(z) \quad \text{for} \quad z > z', \\ \gamma_i(z) & = \alpha\gamma_-(z) \quad \text{for} \quad z < z', \end{cases} \tag{2.21}$$

is a solution of the problem (2.17) on $\mathbb{R}^+$ and then ζ_i is an eigenvalue of $\mathcal{L}$ associated to the eigenfunction γ_i. In fact, in this case the equality $\mathcal{L}\gamma_i = \zeta_i\gamma_i$ is valid on the entire domain above and below z'.

With this expression for the Green function γ_ζ, we can compute the spectral measure associated with operator $\mathcal{L}$ using (2.14), (2.16) and the residue theorem. Assuming that a and b fulfil the condition to use the Cauchy formula, and denoting $(\zeta_i)_{i\in I}$ the subgroup of eigenvalues encircled by the contour $\mathcal{C}$ [3] we have[4] :

$$\begin{aligned} \left(E\left([a;b]\right)\phi(z),\psi(z)\right) &= \frac{1}{2i\pi}\oint_{\mathcal{C}}\left(R(\zeta)\phi(z),\psi(z)\right)d\zeta, && (2.22)\\ &= \frac{1}{2i\pi}\oint_{\mathcal{C}}\left(\int_{\mathbb{R}^+}\gamma_\zeta\left(z,z'\right)\phi(z')\,dz',\psi(z)\right)\,d\zeta, && (2.23)\\ &= \frac{1}{2i\pi}\int_{(\mathbb{R}^+)^2}\left[\oint_{\mathcal{C}}\gamma_\zeta\left(z,z'\right)\,d\zeta\right]\,\phi(z')\,dz'\psi(z)\,dz, && (2.24)\\ &= \int_{(\mathbb{R}^+)^2}\left[\sum_{i\in I}Res\left(\gamma_\zeta,\zeta_i\right)\right]\phi(z')\,dz'\psi(z)\,dz, && (2.25) \end{aligned}$$

where $Res\left(\gamma_\zeta,\zeta_i\right)$ is the residue of γ_ζ in ζ_i. From expression (2.20) we deduce the expression of the residue :

$$Res\left(\gamma_\zeta,\zeta_i\right) = \frac{\gamma_i(z')\gamma_i(z)}{\partial_\zeta W\left(z',\zeta_i\right)}. \tag{2.26}$$

The last difficulty is to compute the derivative $\partial_\zeta W\left(z',\zeta_i\right)$, Michael Bertin has shown in his thesis [Bertin, 2014] that :

$$\partial_\zeta W\left(z',\zeta_i\right) = 2\zeta_i\int_0^\infty\gamma_i^2(z)\,dz. \tag{2.27}$$

We will consider (as in [Waxler, 2002]) that eigenfunctions are chosen in order to have this integral equal to one. This gives us the expression of the spectral measure :

[3] $\mathcal{I}$ is the subset of $\mathcal{Z}$ containing only the zeros of W circled by $\mathcal{C}$

[4] we work here with real functions, this is why we have dropped the conjugate bar in the scalar product in the following lines.

$$\left(E\left([a;b]\right)\phi(z),\psi(z)\right) = \sum_{i\in I}\int_{\mathbb{R}^+}\frac{\gamma_i(z')}{\sqrt{2\zeta_i}}\phi(z')\,dz'\int_{\mathbb{R}^+}\frac{\gamma_i(z)}{\sqrt{2\zeta_i}}\psi(z)\,dz, \tag{2.28}$$

$$= \sum_{i\in I}\mathcal{F}\phi(\zeta_i)\,\mathcal{F}^*\psi(\zeta_i). \tag{2.29}$$

This expression was valid on an interval $[a;b]$, an integration on $\mathbb{R}$ to include the whole spectrum gives :

$$\left(\int_{\mathbb{R}} dE(\lambda)\phi,\psi\right) = \sum_{i\in\mathcal{Z}}\mathcal{F}\phi(\zeta_i)\,\mathcal{F}^*\psi(\zeta_i). \tag{2.30}$$

As mentioned earlier, we have dropped here the contribution of the continuous part of the spectrum which should add to this first term. This expression of the spectral measure associated to $\mathcal{L}$ is useful to *diagonalize* the operator $\mathcal{L}$:

$$\mathcal{L}u = \int_{\mathbb{R}}\lambda dE(\lambda)u = \sum_{i\in\mathcal{Z}}\int_{\mathbb{R}^+}\zeta_i\frac{\gamma_i(z)u(z,r,\omega)}{\sqrt{2\zeta_i}}\,dz = \sum_{i\in\mathcal{Z}}\zeta_i\hat{u}\left(\zeta_i,r,\omega\right), \tag{2.31}$$

where $\hat{u}\left(\zeta_i,r,\omega\right) = \mathcal{F}u(z,r,\omega)$.

Resolution of equation (2.9)

We want to use the last expression (2.31) to solve equation (2.9). To this purpose, we remind that equation (2.9) writes :

$$\mathcal{L}u + \Delta_\perp u = \delta(r)\otimes\delta(z-z_s)\hat{s}(\omega). \tag{2.32}$$

Thus, using (2.31), $\hat{u}$ is a solution of :

$$\sum_{i\in\mathcal{Z}}\zeta_i\hat{u}\left(\zeta_i,r,\omega\right) + \Delta_\perp\hat{u}\left(\zeta_i,r,\omega\right) = \sum_{i\in\mathcal{Z}}\delta\left(r\right)\mathcal{F}\delta\left(z-z_s\right)\hat{s}(\omega), \tag{2.33}$$

$$= \sum_{i\in\mathcal{Z}}\delta\left(r\right)\frac{\gamma_i(z_s)\hat{s}(\omega)}{\sqrt{2\zeta_i}}. \tag{2.34}$$

We have applied the definition of $\mathcal{F}$ given in (2.29) to $\delta\left(z-z_s\right)$. On the left side we recognise a classical Helmholtz equation with a constant coefficient ζ_i. This equation solves using the cylindrical Hankel function of the first type $H_0^{(1)}$ and by linearity each term of the sum writes :

$$\hat{u}\left(\zeta_i,r,\omega\right) = \frac{i}{4}\frac{\gamma_i(z_s)\hat{s}(\omega)}{\sqrt{2\zeta_i}}H_0^{(1)}\left(\sqrt{\zeta_i}r\right). \tag{2.35}$$

One needs to apply the transformation $\mathcal{F}^*$ to find the solution in the 'physical space' :

$$u(z,r,\omega) = \frac{i}{8}\sum_{i\in\mathcal{Z}} \gamma_i(z_s)\gamma_i(z)\hat{s}(\omega)H_0^{(1)}\left(\sqrt{\zeta_i}r\right). \tag{2.36}$$

From expression (2.36) one only needs a Fourier transform in time to compute the signal in time at a given altitude z and a distance r from the source.

In practice, to compute an acoustic field at large distances we use an approximation of the Hankel function which behaves asymptotically like a complex exponential :

$$H_0^{(1)}(x) \equiv \sqrt{\frac{2}{\pi}}\frac{e^{i(x+\pi/4)}}{\sqrt{x}}. \tag{2.37}$$

1.3 Pseudo-spectrum

The aim of this work is to quantify the impact of a perturbation of the atmospheric specification on the propagation. To this purpose, we are interested in the impact of those perturbations on the spectrum of operator (2.10) and propose in this paragraph to illustrate this impact using the pseudo-spectrum [Trefethen and Embree, 2005]. The ϵ-pseudo-spectrum is defined for any $\epsilon \in \mathbb{R}_*^+$:

$$\sigma_\epsilon(\mathcal{L}) = \left\{ z \in \mathbb{C} \mid \left|\left|(zI-\mathcal{L})^{-1}\right|\right| \geq \frac{1}{\epsilon} \right\}, \tag{2.38}$$

with the convention that : $\left|\left|(zI-\mathcal{L})^{-1}\right|\right| = +\infty$ when $z \in \sigma(\mathcal{L})$. The pseudo-spectrum of an operator contains the values of the complex plane that are near the spectrum of the operator. At first sight there is no link with the perturbation of the atmospheric specification but the link will be clearer with the following characterisation :

$$\sigma_\epsilon(\mathcal{L}) = \bigcup_{\mathcal{L}'\in\mathcal{S}_\epsilon} \sigma(\mathcal{L}+\mathcal{L}'), \tag{2.39}$$

where $\mathcal{S}_\epsilon$ is the set of linear operators whose norms are smaller than ϵ. In the context of acoustic propagation, the perturbations will result in a modification of the sound speed profile and the perturbed operator $\mathcal{L}$ can be decomposed as a sum of two operators $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1$ where $\mathcal{L}_0$ is the classical operator (2.10) for the unperturbed profile c_0 and $\mathcal{L}_1$ defined by :

$$\mathcal{L}_1\Psi = \left(\frac{\omega^2}{c(z)^2} - \frac{\omega^2}{c_0(z)^2}\right)\Psi, \tag{2.40}$$

where $c(z)$ is the perturbed profile. Because $\mathcal{L}_1$ is a multiplicative operator, its norm is directly given by :

$$\eta = \left|\left|\frac{\omega^2}{c(z)^2} - \frac{\omega^2}{c_0(z)^2}\right|\right|. \tag{2.41}$$

Characterisation (2.39), gives the following result on the spectrum of the perturbed operator $\mathcal{L}$:

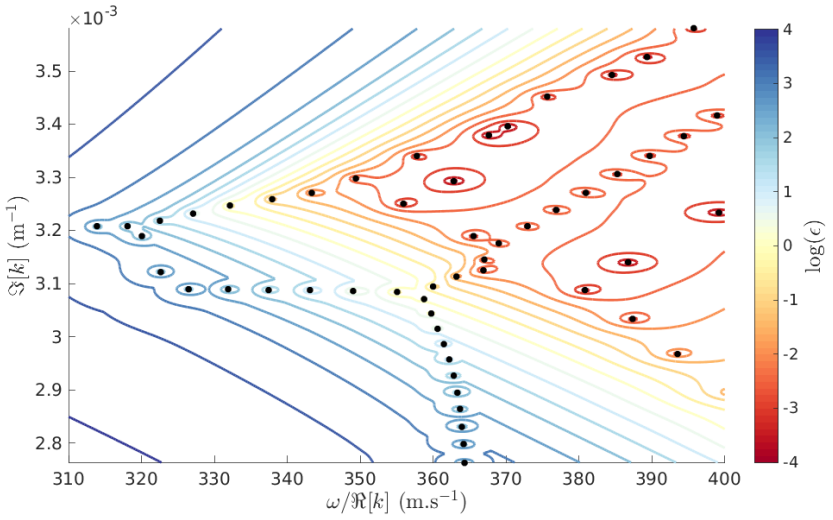

FIGURE 2.4 : Pseudo-spectrum computed for different values of ϵ. The spectrum (black dots) have been computed for the sound speed profile of figure 2.2 but for a frequency of 0.1Hz (less expensive than 1Hz numerically). A complex term is added to the frequency ($\Im[\omega] = 1$) in order to correct the aspect ratio.

$$\sigma\left(\mathcal{L}\right) \subset \sigma_\eta\left(\mathcal{L}_0\right). \tag{2.42}$$

To compute numerically the pseudo spectrum of a matrix (representing the operator) we give a last characterisation :

$$\sigma_\epsilon\left(\mathcal{L}\right) = \left\{z \in \mathbb{C} \mid s_{min}\left(zI - \mathcal{L}\right) \leq \epsilon\right\}, \tag{2.43}$$

where $s_{min}(A)$ is the smallest singular value of matrix A.

Figure 2.4 shows the contour of the pseudo-spectrum for different values of ϵ, highlighting the more sensitive eigenvalues. However, the pseudo-spectrum is a very general tool and does not give any information for a specific perturbation (imposed by data for instance). In fact, inclusion (2.42) is very large because $\mathcal{S}_\eta$ contains all the possible operators of norms smaller than η and we are only interested in perturbation by very specific operators of the form (2.40). Moreover, its computation requires to compute a singular value decomposition at each point of the domain which becomes unaffordable for large matrices.

1.4 Numerical computation of the spectrum

To compute numerically the spectrum of operator (2.10), the first operation consists in bounding the domain. This operation is not insignificant because it changes the nature of the operator whose resolvent becomes compact and the spectrum will only be discrete. On top of that, a numerical discretization of operator (2.10) will transform the problem in a computation of the spectrum of a matrix. To perform this discretization, there exist several techniques and many refinement, but we will only

present here the two most widely used techniques in a context of wave propagation : finite differences (FD) and spectral collocation (SC). We will look next at the impact of perturbations on the celerity profile on the spectrum. For this purpose, we need to investigate first the precision of the numerical computation of the spectrum.

Finite Differences

The numerical discretization of the operator (2.10) consists in computing the derivatives of an eigenfunction Ψ on equally spaced grid points $(z_j)_{j=1,..,N_d}$. The first idea that comes in mind is the finite differences method which consists in considering the Taylor expansions of $\Psi(z_{j+1})$ and $\Psi(z_{j-1})$ to approximate the first derivative by :

$$\Psi'(z_j) \approx \frac{\Psi(z_{j+1}) - \Psi(z_{j-1})}{2h}, \tag{2.44}$$

where $h = z_{j+1} - z_j$ is the step of the grid. By repeating the same operation, we obtain the second order derivative :

$$\Psi''(z_j) \approx \frac{\Psi(z_{j+1}) - 2\Psi(z_j) + \Psi(z_{j-1})}{h^2}. \tag{2.45}$$

Using those discretizations of the derivatives we can write a discrete form of operator $\mathcal{L}$ using a matrix-vector multiplication with constant entries :

$$\mathbb{L}^{(FD)} = \begin{pmatrix} \mathcal{B} + \frac{\omega^2}{c(z_0)^2} & \frac{1}{h^2} & 0 & \dots & & \\ \frac{1}{h^2} & -\frac{2}{h^2} + \frac{\omega^2}{c(z_1)^2} & \frac{1}{h^2} & 0 & & \\ 0 & \frac{1}{h^2} & -\frac{2}{h^2} + \frac{\omega^2}{c(z_2)^2} & \frac{1}{h^2} & & 0 \\ \vdots & & \ddots & & & \\ & 0 & \frac{1}{h^2} & -\frac{2}{h^2} + \frac{\omega^2}{c(z_{N_d})^2} & \frac{1}{h^2} & \\ & & 0 & \frac{1}{h^2} & -\frac{2}{h^2} + \frac{\omega^2}{c(z_{N_d})^2} & \end{pmatrix}. \tag{2.46}$$

The first line encodes the limit condition ($\mathcal{B}$), $\mathbb{L}^{(FD)}$ is thereby the discretization of operator (2.10) and its spectrum can be computed like a regular matrix. If this technique is very simple to implement, it converges relatively slowly compared with the spectral collocation.

Spectral Collocation

The idea of SC is a polynomial interpolation in unevenly spaced points that are chosen to be adapted to the family of polynomials [Trefethen, 2000, Canuto et al., 2006]. We will use here the Tchebychev polynomials T_j and the associated points :

$$s_j = \cos\left(j\pi/N'_d\right), j = 0, .., N'_d. \tag{2.47}$$

The eigenfunction are then decomposed on the T_j :

$$\Psi(s) = \sum_j a_j T_j(s). \tag{2.48}$$

To compute their derivative, we start by decomposing the derivative of the polynomial on the same basis :

$$T'_k(s) = \sum_j q_{jk} T_j(s). \tag{2.49}$$

This allows to decompose the derivatives of the eigenfunctions :

$$\Psi'(s) = \sum_j b_j T_j(s), \tag{2.50}$$

where

$$b_j = \sum_k a_k q_{jk}. \tag{2.51}$$

Like in the case of finite differences, we can write this decomposition using matrix expression. To this purpose, we introduce matrices $\mathbb{T}$ and $\mathbb{Q}$ that compute respectively the projection of Ψ and the derivation of the basis functions (the coefficients of $\mathbb{Q}$ are the $(q_{jk})_{j,k}$ defined at (2.49)) :

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \\ a_{N'_d} \end{pmatrix} = \mathbb{T} \begin{pmatrix} \Psi(s_1) \\ \Psi(s_2) \\ \vdots \\ \\ \Psi(s_{N'_d}) \end{pmatrix}, \tag{2.52}$$

and

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \\ b_{N'_d} \end{pmatrix} = \mathbb{Q} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \\ a_{N'_d} \end{pmatrix}. \tag{2.53}$$

With those two matrices, we can compute the differentiation matrix $\mathbb{D} = \mathbb{T}^{-1}\mathbb{Q}\mathbb{T}$ such that :

$$\begin{pmatrix} \Psi'(s_1) \\ \Psi'(s_2) \\ \vdots \\ \\ \Psi'(s_{N'_d}) \end{pmatrix} = \mathbb{D} \begin{pmatrix} \Psi(s_1) \\ \Psi(s_2) \\ \vdots \\ \\ \Psi(s_{N'_d}) \end{pmatrix}. \tag{2.54}$$

In the case of Tchebychev polynomials with Gauss-Lobatto-Tchebychev points (defined at (2.47)) we can give the expression of the derivation matrix (lines are indexed by i, columns by j) :

$$\mathbb{D} = \begin{pmatrix} \frac{2N'^2_d+1}{6} & \cdots & 2\frac{(-1)^j}{1-s_j} & \cdots & \frac{1}{2}(-1)^{N'_d} \\ \vdots & \ddots & \frac{(-1)^{i+j}}{s_i-s_j} & & \vdots \\ -\frac{1}{2}\frac{(-1)^i}{1-s_i} & & \frac{-s_i}{2(1-s_i^2)} & & \frac{1}{2}\frac{(-1)^{N'_d+i}}{1-s_i} \\ \vdots & \frac{(-1)^{i+j}}{s_i-s_j} & & \ddots & \vdots \\ -\frac{1}{2}(-1)^{N'_d} & \cdots & -2\frac{(-1)^{N'_d+j}}{1+s_j} & \cdots & -\frac{2N'^2_d+1}{6} \end{pmatrix}. \tag{2.55}$$

We have computed the differentiation matrix for the set of points defined at (2.47) that cover the interval $[-1;1]$. To solve the physical problem, we need to change the coordinates using the following expression :

$$z_j = \frac{1-s_j}{2} z_{max}. \tag{2.56}$$

This transformation impacts the derivation matrix and we define $\overline{\mathbb{D}} = \mathbb{J}\mathbb{D}$, where $\mathbb{J} = -\frac{2}{z_{max}}\mathbb{I}$ is the jacobian matrix of the transformation. The discretized operator will then write :

$$\mathbb{L}^{(SC)} = \overline{\mathbb{D}}^2 + \frac{\omega^2}{c(z_j)^2}\mathbb{I}. \tag{2.57}$$

1.5 Discretization error

Both methods presented in the former paragraph give a matrix whose spectrum is an approximation of the spectrum of the operator. However, the two methods are not equivalent from a computational point of view. An important difference between

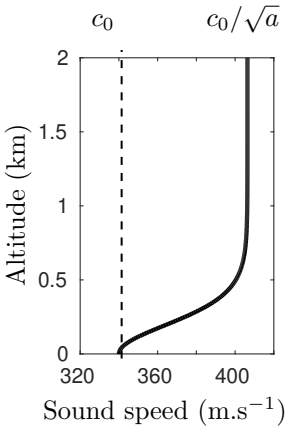

FIGURE 2.5 : Celerity profile of the boundary profile defined in (2.59). c_0 stands for the celerity at the ground, a sets the celerity for high altitudes and H the height of the boundary layer. Here, $c_0 = 340\text{m.s}^{-1}$, $a = 0.7$ and $H = 250\text{m}$.

finite-difference approximations to the eigenvalues and eigenfunctions of the operator $\mathcal{L}$ defined in (2.10) and the Tchebychev approximations advocated here is their order of accuracy. Finite-difference approximations give only a finite order of accuracy in the sense that errors behave asymptotically like $(\Delta x)^p$ for some finite p when the grid scale Δx approaches zero. On the other hand, if the celerity profile is infinitely differentiable, the Tchebychev polynomial approximations used here are of infinite order in the sense that errors decrease more rapidly than any power of $1/N$ as $N \to +\infty$, where N is the number of Tchebychev polynomials retained in the approximation [Canuto et al., 2006]. Another difference comes from the global character of the spectral collocation. Unlike finite differences, a singularity in the physical domain will affect the convergence of the series and thereby the solution on the entire domain and not only locally.

We are going to compute in this section the spectrum of the Helmoltz operator in the case of a boundary layer celerity profile that can be computed analytically. We use this example to validate the numerical methods and compare the performances of spectral collocation and finite differences.

We introduce a boundary layer profile, characterised by a height parameter H and a shape parameter a. The sound speed at the ground will be denoted c_0. Acoustic modes (k, Ψ) are characterised by the equation (2.10) :

$$\frac{d^2\Psi}{dz^2} + \left[\frac{\omega^2}{c(z)^2} - k^2\right]\Psi = 0, \tag{2.58}$$

where :

$$c(z) = \frac{c_0}{\sqrt{\frac{1-a}{\cosh^2\left(\frac{z}{H}\right)} + a}}. \tag{2.59}$$

If we perform the change of variable $\bar{z} = z/H$ (with $\frac{d^2}{dz^2} = H^2\frac{d^2}{d\bar{z}^2}$) and denote $k_0 = \frac{\omega}{c_0}$, (2.58) becomes :

$$\frac{d^2\Psi}{d\overline{z}^2} + \left[\frac{\nu\left(\nu-1\right)}{\cosh^2\left(\overline{z}\right)} - \mu^2\right]\Psi = 0, \tag{2.60}$$

with $\nu = \frac{1}{2} + \sqrt{\frac{1}{4} + \left(k_0 H\right)^2\left(1-a\right)}$ and $\mu = \sqrt{H^2\left(k^2 - k_0^2 a\right)}$.

If we transform this equation in $\overline{z}$ in an equation in $u = \tanh\left(\overline{z}\right)$, we obtain the classical Legendre differential equation :

$$\left(1-u^2\right)\frac{d^2\Psi}{du^2} - 2u\frac{d\Psi}{du} + \left[\nu\left(\nu-1\right) - \frac{\mu^2}{1-u^2}\right]\Psi = 0. \tag{2.61}$$

To obtain this equation we have transformed the derivatives in $\overline{z}$ in derivatives in u :

$$\begin{cases} \frac{d}{d\overline{z}} = \left(1-u^2\right)\frac{d}{du} \\ \frac{d^2}{d\overline{z}^2} = \left(1-u^2\right)\left[-2u\frac{d}{du} + \left(1-u^2\right)\frac{d^2}{du^2}\right] \end{cases} \tag{2.62}$$

and used the trigonometric relation : $\frac{1}{\cosh^2(\overline{z})} = 1 - \tanh^2(\overline{z})$.

Solution of equation 2.61 is given in [Abramowitz and Stegun, 1965, Olver, 1974] : $\Psi(u) = AP_{\nu-1}^{-\mu}(u)$, where P_b^a is the associated Legendre function of the first kind and A a constant of integration. To be a solution of the eigenvalue problem, function Ψ has to fulfil also the boundary conditions. We give the spectrum in the two classical case of Dirichlet and Neumann homogeneous.

Dirichlet homogeneous :

In the case of Dirichlet homogeneous conditions, we are looking for values of μ such that $P_{\nu-1}^{-\mu}(0) = 0$. The expression of $P_{\nu-1}^{-\mu}(0)$ is given in [Abramowitz and Stegun, 1965] and gives the following condition :

$$\cos\left(\frac{\pi}{2}\left(\nu-\mu-1\right)\right)\frac{\Gamma\left(\frac{\nu-\mu}{2}\right)}{\Gamma\left(\frac{\nu+\mu+1}{2}\right)} = 0, \tag{2.63}$$

but $\cos\left(\frac{\pi}{2}\left(\nu-\mu-1\right)\right) = \sin\left(\frac{\pi}{2}\left(\mu-\nu\right)\right)$ and using the relation $\Gamma\left(z\right)\Gamma\left(1-z\right) = \frac{\pi}{\sin(\pi z)}$, we can write equation (2.63) :

$$\frac{\pi}{\Gamma\left(1+\frac{\mu-\nu}{2}\right)\Gamma\left(\frac{1+\mu+\nu}{2}\right)} = 0. \tag{2.64}$$

Poles of the gamma function are located on the negative integers, hence :

$$\mu_n = \nu - 2n - 2,\ n \in \mathbb{N},$$
$$k_n = \left[\left(\frac{\nu-2n-2}{H}\right)^2 + k_0^2 a\right]^{1/2},\ n \in \mathbb{N}.$$

Neumann homogeneous

We can conduct exactly the same study in the case of Neumann homogeneous conditions. The value of the derivative at the ground of the associated Legendre function is also given in [Abramowitz and Stegun, 1965] and gives the following condition :

$$\sin\left(\frac{\pi}{2}\left(\nu-\mu-1\right)\right)\frac{\Gamma\left(\frac{\nu-\mu+1}{2}\right)}{\Gamma\left(\frac{\nu+\mu}{2}\right)}=0, \tag{2.65}$$

proceeding like in the Dirichlet case, we obtain the following condition :

$$\frac{\pi}{\Gamma\left(\frac{\mu-\nu-1}{2}\right)\Gamma\left(\frac{\nu+\mu}{2}\right)}=0, \tag{2.66}$$

and the eigenvalues are given by :

$$\begin{aligned}\mu_n &= \nu-2n+1,\ n\in\mathbb{N},\\ k_n &= \left[\left(\frac{\nu-2n+1}{H}\right)^2+k_0^2a\right]^{1/2},\ n\in\mathbb{N}.\end{aligned}$$

Numerical comparison

We can use this case to assess the numerical precision of the spectral collocation and finite differences scheme. We have used the profile (2.59) with parameters $c_0 = 340\text{m.s}^{-1}$, $a = 0.7$, $H = 250\text{m}$, $\omega = 2\pi\times 50\text{s}^{-1}$ and $z_{max} = 2\text{km}$. For the 30^{th} first eigenvalues, we have computed the eigenvalues with either a finite differences scheme or a spectral collocation. In each case, the size of the matrix is set using a number of points per vertical wavelength which is a characteristic length of the eigenfunctions. In practice, for the case we are studying here, four points per vertical wavelength corresponds to a matrix of size 1847 whereas twenty points per wavelength gives a matrix of size 9239. As shown on figure 2.6, the spectral collocation is at the machine precision even with only one point per wavelength whereas the finite difference method converges more slowly.

From now on, we will use the spectral collocation method with four points per vertical wavelength to compute the acoustic modes.

1.6 Numerical error

In the previous paragraph we were interested in the discretization error which is the error resulting from the fact that a function of a continuous variable is represented in the computer by a finite number of evaluations, in our case with vectors and matrices. As we have seen with figure 2.6 discretization error can usually be reduced by using a more finely spaced lattice, with an increased computational cost. In this paragraph we are going to focus on the numerical error.

In fact, floating-point numbers represent only a subset of real numbers. As such, floating-point arithmetic introduces approximations that can compound and have

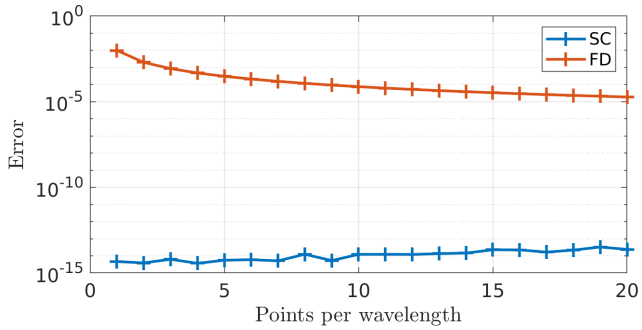

FIGURE 2.6 : Error of the computation of the spectrum in the case of a finite differences scheme (FD) and a spectral collocation approach (SC) for different number of points per wavelength.

a significant impact on numerical simulations. It is then natural to estimate the precision of our computations in order to ensure the validity of our approach. The method we have used [Demeure et al., 2021] consists in using pair arithmetic to assess floating-point accuracy, representing each number using a couple *(number , error)*. The main building block of this method is the use of Error-Free Transformations. To evaluate arithmetic operators and the square root function, it computes the error produced by the operation using an error-free transformation, and combines it with the error transmitted from the inputs using basic arithmetics. It gives the number of significant digits using the following formula :

$$err_{num}\left(number, error\right) = \left\lfloor -\log\left(\frac{error}{number}\right)\right\rfloor . \tag{2.67}$$

The implementation, called *Shaman*, only requires to change the types of the variables (in C++[5]) and execute the script normally. We first give an example on a very basic case : the computation of the integral of the cosine function between 0 and $\pi/2$ with the rectangle technique. The error, obtained comparing the numerical result with the exact value of the integral, is plotted on figure 2.7 (in blue) as a function of the number of rectangle. The red dots indicate the numerical error estimated by shaman. For a small number of rectangles (≤ 100) the numerical error is smaller than the discretization error of several orders of magnitude. But when the number of rectangles increases, the numerical error increases as the discretization error decreases ; with a discretization containing more than one hundred rectangles the error is only numerical. In the next chapters we are going to look at the impact of a perturbation on the celerity profile on the acoustic modes. Therefore, we will need to compute the modes for different celerity profiles and characterise their behaviours. Before developing sophisticated techniques to capture the variability of the modes, we need to check that the numerical error is not going to doom our efforts. We are going to consider in this paragraph that the modes are random variables computed using a celerity profile corresponding to a boundary layer profile with a random jet

[5]All the codes used in this thesis are in MATLAB but a part of them has been translated in C++ for this study.

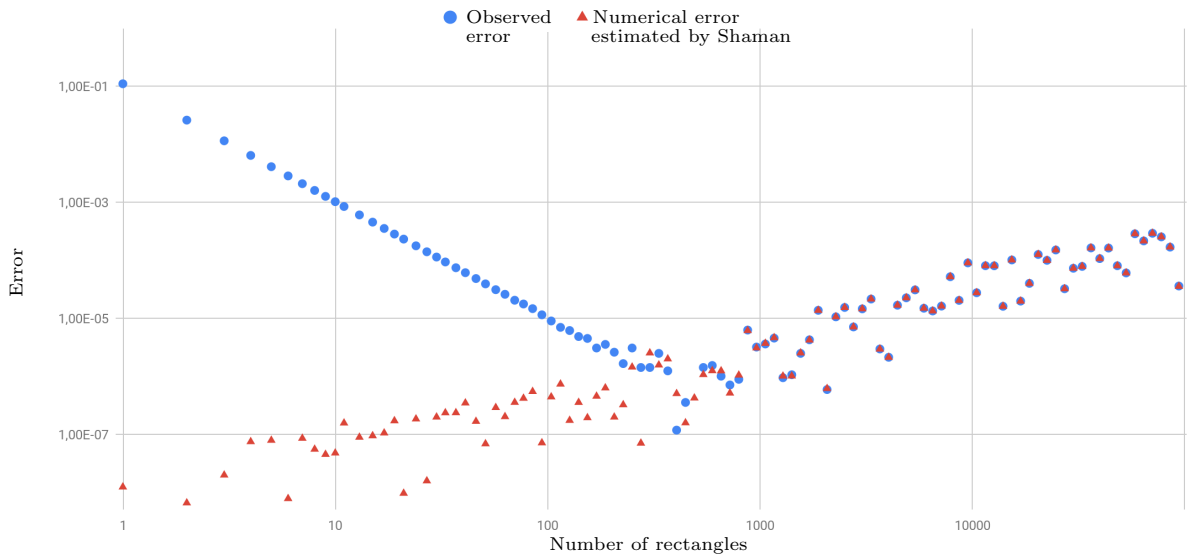

FIGURE 2.7 : Absolute value of the error as a function of the number of rectangles used for the integration of the cosine function between 0 and $\pi/2$ using the rectangle method. Both axes are displayed in a logarithmic scale (*figure* : Nestor Demeure PhD manuscript [Demeure, 2021]).

(see chapter 3 for a precise definition of the celerity profile and its perturbation). We have selected nine eigenvalues to estimate their probability densities, given a fixed perturbation of the profile. We have used two different methods to do so : Monte Carlo and Polynomial Chaos (cf next section for a definition of the generalized polynomial chaos (gPC)). Figure 2.8 shows the numerical error of our estimations of the densities compared with the standard deviation of each eigenvalue. As expected, the standard deviation estimated with gPC or Monte-Carlo is the same. However, the numerical error is higher in the case of Polynomial Chaos estimation. In any case, the numerical error is smaller than the standard deviation of each eigenvalue which ensures that our uncertainty quantification will be meaningful.

2 Polynomial Chaos Expansions for wave propagation

To take into account the physical fluctuations of the medium in our propagation model, we have chosen to build a metamodel using generalized Polynomial Chaos (gPC) expansions. This choice has been motivated by the necessity to capture a variability of great amplitude in a probabilistic context. We present here the global framework of gPC expansions and will give in the next chapter our strategy to couple it with acoustic propagation.

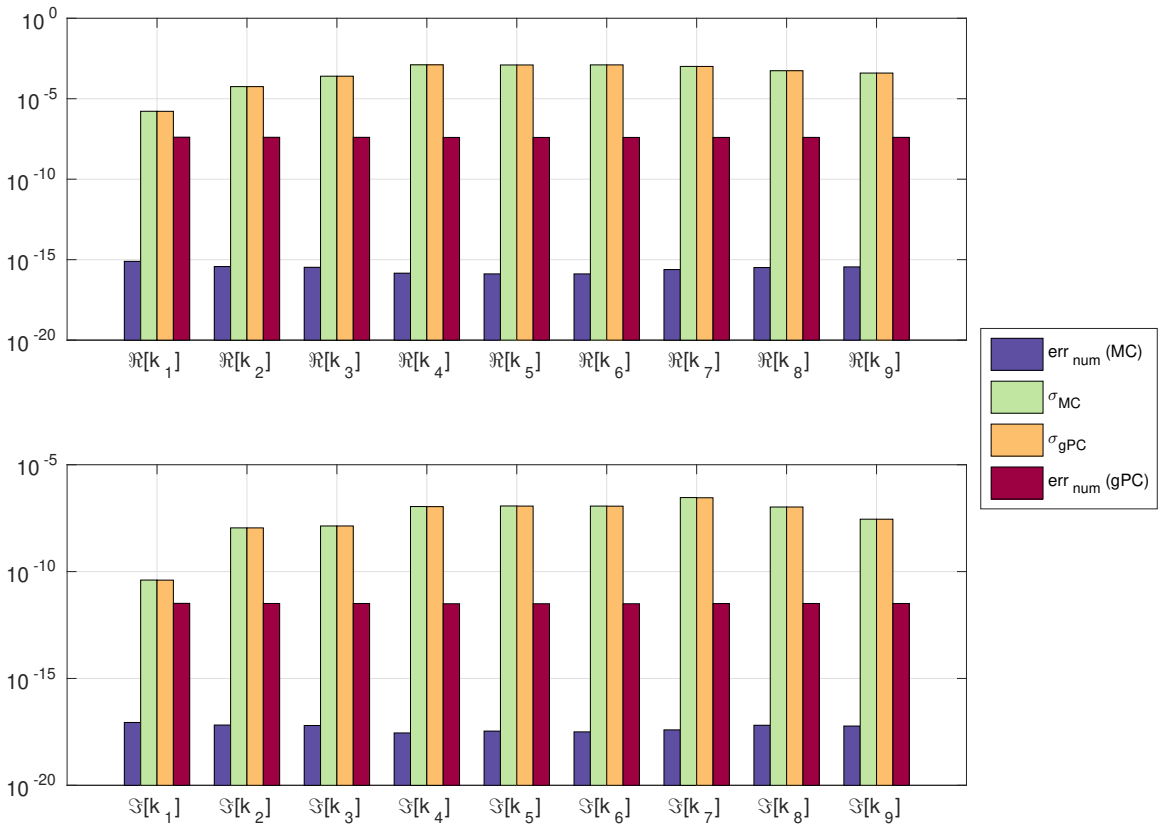

FIGURE 2.8 : Numerical error is computed for the different eigenvalues and compared to their standard deviation.

2.1 Polynomial Chaos expansion

The polynomial chaos expansion framework has been introduced in 1938 by Norbert Wiener [Wiener, 1938] and has known a regain of interest with the publication of the Ghanem and Spanos book [Ghanem and Spanos, 1991] in 1991. The modern use of polynomial chaos consists in decomposing a random variable (named Quantity of Interst, QoI) on a polynomial basis to conduct statistical studies when the model is expensive to evaluate. The strength of generalized Polynomial Chaos (gPC) relies in its ability to build a surrogate model of any second-order random field, as an infinite functional series based on multivariate polynomials of random variables ; these random variables parametrizing here the uncertain environmental information of the system. Once the surrogate model at hand, it is possible to extract statistical properties (mean, variance, global sensitivity indices, ...) of the QoI computed by the model, or predict its value at very low cost for a new particular environmental scenario within the range of variability.

The specific feature of polynomial chaos lies in the orthogonality of the polynomial basis with respect to the joint measure of the input parameters distributions. We first consider the case of one dimensional input ξ and a QoI Y depending on the input throught a model $F : Y = F(\xi)$, $F : \mathbb{R} \to \mathbb{R}$.

Orthogonal polynomials

If we denote by μ a probability measure on $\mathbb{R}$ and $(H_n)_{n\in\mathbb{N}}$ a family of polynomials, we will say that $(H_n)_{n\in\mathbb{N}}$ is orthogonal with respect to μ if :

$$\forall n, m \in \mathbb{N}^2, \int_{\mathbb{R}} H_n(x)H_m(x)d\mu(x) = \gamma_n \mathbb{1}_{\{n=m\}}. \tag{2.68}$$

If $\forall n \in \mathbb{N}, \gamma_n = 1$, the polynomials are called orthonormals.

The existence and the construction of such a family is ensured by the Gram-Schmidt process[6]. From a numerical point of view, this process is unstable and in practice recurrence relations can be found to compute the coefficients.

Once the family of orthonormal polynomials is computed, one still have to show that the family is complete (i.e. dense in $L^2(\mathbb{R}, \mu)$). In the case of a Gaussian variable, the validity of this decomposition has been originally proven in an article of Cameron and Martin in 1947 [Cameron and Martin, 1947]. It has been generalized to other types of random variables [Xiu and Karniadakis, 2002] and even arbitrary distributions [Soize and Ghanem, 2004]. We propose a proof here [Beck et al., 2005] in the case of a measure fulfilling the condition :

$$\exists \alpha \in \mathbb{R}_+^*, \int_{\mathbb{R}} e^{\alpha|x|}d\mu(x) < +\infty. \tag{2.69}$$

The demonstration we propose here is not the original one of Cameron and Martin but is using some complex analysis arguments and has the advantage of being quite

[6] A sufficient condition for the Gram-Schmidt process is that μ has an infinite support and $\int_{\mathbb{R}} x^n d\mu(x) < +\infty$. It is usually the case for classical probability laws.

synthetic. Let $g \in L^2(\mathbb{R}, \mu)$, we want to show that if :

$$\forall n, \int_{\mathbb{R}} x^n g(x) d\mu(x) = 0, \tag{2.70}$$

then $g = 0$. We will deduce that the family of $(X^n)_{n\in\mathbb{N}}$ polynomials is complete. Then a property of the Gram-Schmidt process is $\text{Vect}\,(X^i, 0 \leq j \leq n) = \text{Vect}\,(H_i, 0 \leq j \leq n)$; this will show the completeness of $(H_n)_{n\in\mathbb{N}}$.

We will assume here that the measure μ is absolutely continuous with respect to the Lebesgue measure, i.e. $\exists \rho, \forall A \subseteq \mathbb{R}\ \mu(A) = \int_A \rho(x)dx$.

We then define the function ϕ on $\mathbb{R}$ by $\phi(x) = g(x)\rho(x)$ and consider its Fourier transform on $\mathbb{R}$:

$$\widehat{\phi}(\omega) = \int_{\mathbb{R}} g(x) e^{-i\omega x} \rho(x) dx.$$

We want to build the analytic continuation of $\widehat{\phi}$ in a function Φ holomorphic on $B_\alpha = \{z \in \mathbb{C}/|\Im[z]| < \alpha/2\}$ (α given by condition (2.69)).
Let $h(x, z) = g(x)\rho(x)e^{-ixz}$, for $z \in B_\alpha$ we have $|h(x, z)| \leq e^{|x|\alpha/2}|g(x)|\rho(x)$ and by Cauchy-Schwarz inequality we have :

$$\int_{\mathbb{R}} e^{|x|\alpha/2}|g(x)|\rho(x)dx \leq \int_{\mathbb{R}} e^{|x|\alpha}\rho(x)dx \int_{\mathbb{R}} |g(x)|^2\rho(x)dx < +\infty. \tag{2.71}$$

$\Phi(z) = \int_{\mathbb{R}} h(x, z)dx$ is well-defined on B_α and holomorphic by integration of a dominated holomorphic function. We can compute the nth derivative of F :

$$\Phi^{(n)}(z) = (-i)^n \int_{\mathbb{R}} x^n g(x) e^{-ixz} \rho(x) dx, \tag{2.72}$$

which in $z = 0$ gives :

$$\Phi^{(n)}(0) = (-i)^n \int_{\mathbb{R}} x^n g(x) \rho(x) dx. \tag{2.73}$$

As a consequence of (2.70), $\forall n \in \mathbb{N}, \Phi^{(n)}(0) = 0$ and the Taylor series expansion gives $\Phi = 0$ on a neighbourhood of 0. The uniqueness of the analytic continuation (B_α being connected) implies that $\Phi = 0$ on B_α. In particular on the real axis : $\widehat{\phi} = 0$. The Fourier transform is injective so $\phi = 0$ and then $g = 0$ since ρ is positive because μ is a positive measure.

This result justifies the decomposition of all random variables of $L^2(\mathbb{R}, \mu)$ on the family of orthogonal polynomials.

Approximation

In practice, $Y = F(\xi)$ will be written as a truncated expansion on this polynomial basis. If we denote $\Pi_P Y$ the polynomial expansion of $Y \in L^2(\mathbb{R}, \mu)$ truncated at degree P the projection theorem gives :

$$\Pi_P Y = \underset{H \in \mathcal{F}_n}{\operatorname{argmin}} \|Y - H\|_{L^2(\mathbb{R},\mu)}, \tag{2.74}$$

where $\mathcal{F}_n$ is the set of polynomials with degree less than n. The practical computation of the coefficients of the decomposition will be detailed in the next chapter but the two possibilities rely on the orthogonality of the polynomials (projection) or on this result of optimality (regression).

An interesting result gives a bound on the error depending on the regularity of the function F. If we denote D^k the set of functions whose k first derivative exist and have a finite L^2-norm and $\|.\|_{D^k} = \sum_{i=0}^{k} \|.^{(i)}\|_2$ then there exist a constant C_k such that :

$$\|Y - \Pi_P Y\|_2 \leq \frac{C_k}{P^k} \|F\|_{D^k}. \tag{2.75}$$

Note that convergence is always in L^2-norm, which does not imply the uniform convergence. An illustration of this fact is the Gibbs phenomenon that occurs for non smooth functions.

Multivariate polynomials

In the case of a multidimensional input ($X \in \mathbb{R}^d$), we can work with multivariate polynomials. Those polynomials are linear combination of monomials $\boldsymbol{X}^{\boldsymbol{\beta}} = x_1^{\beta_1} x_2^{\beta_2} \dots x_d^{\beta_d}$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_d)$ is called multi-index and $|\boldsymbol{\beta}| = \beta_1 + \beta_2 + \ldots + \beta_d$ the degree of the monomial.

As in the one dimensional case, we need to build a orthogonal family of polynomials with respect to the joint measure of the input X. However, the Gram-Schmidt process does not give a unique family depending on the way the multi-indices are enumerated. Then, we need to weaken the definition of orthogonal family limiting relation (2.68) to polynomials of different degrees.

A simpler case arises when we look at independent inputs. In this case, the joint measure is the product of the measures in each dimension and there exists a family of strongly orthogonal polynomials. This family is obtained by taking the products of univariate polynomials in each direction.

In the next chapters, inputs will always be independent random variables in order to use this family of strongly orthogonal polynomials.

2.2 Practical computation of polynomial chaos expansion

Polynomial chaos [Wiener, 1938] and its generalization [Ghanem and Spanos, 1991] has known a regain of interest in the last two decades for quantifying parametric uncertainty propagation in a given model. The strength of generalized Polynomial Chaos (gPC) relies in its ability to build a surrogate model of any second-order random field, as an infinite functional series based on multivariate polynomials of random variables ; these random variables parametrizing here the uncertain environmental information of the system. Once the surrogate model at hand, it is possible to extract statistical

properties (mean, variance, global sensitivity indices, ...) of the acoustic field quantity of interest (QoI) computed by the model, or predict its value at very low cost for a new particular environmental scenario within the range of variability.
The numerical requirement in deploying gPC lies in the calibration of the surrogate model once the approximation space has been chosen. Two main classes of methods may be identified : – 1. Galerkin (*Intrusive*) methods : probabilistic model is introduced within the model partial differential equations and therefore the code needs to be modified in order to solve the obtained stochastic system of equations ; – 2. Direct (*Non-intrusive*) methods : the deterministic solver is seen as a black-box by the stochastic approximation. It is then only possible to operate by means of input and output data sampling. We will expose in this section the principle of non-intrusive polynomial chaos decomposition and how we propose to use it in the context of acoustic propagation.

Orthonormal polynomials

Let us consider a general probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and a model output $Y \in L_2(\Omega, \mathcal{F}, \mathcal{P})$ is the QoI. The idea is to represent, via iso-probabilistic transformations or model reduction such as Karhunen-Loève decomposition for instance, $Y = F(\xi; \theta)$ by a countable number of independent random variables $\xi \in \mathbb{R}^n$ representing the uncertain parameters and $\theta \in \mathbb{R}^{n'}$ the deterministic ones. The model F may be very elaborated but will only be considered as a mapping from the space of the inputs into the space of the outputs. The uncertain inputs $\xi = (\xi_1, ..., \xi_n)$ are modelled by independent random variables whose laws are given *a priori*. The aim of our gPC expansion is to characterise the random variable Y.

The existence of gPC expansion is guaranteed for all the squarely integrable random variables [Cameron and Martin, 1947] and relies on the construction of a polynomial basis $(H_k)_{k\in\mathbb{N}}$ orthogonal for the following scalar product using the probability density $\rho_\xi(x)$ of the input variable ξ :

$$\langle f, g\rangle = \mathbb{E}\left[f(\xi)g(\xi)\right] = \int_{\mathbb{R}^n} f(\xi)g(\xi)d\xi = \int_{\mathbb{R}^n} f(x)g(x)\rho_\xi(x)dx. \tag{2.76}$$

For classical random variables, families of orthogonal polynomials are well-known and part of the Askey-scheme [Xiu and Karniadakis, 2002] : Hermite polynomials for normal distribution, Legendre for uniform distribution among others.

To compute those polynomials, a natural method is the Gram-Schmidt orthonormalization procedure. Starting from the canonical polynomial family $(1, X, X^2, .., X^n)$ and computing the orthonormalization using the scalar product defined in (2.76). Unfortunately, this procedure is unstable numerically and another method is used in practice. It consists in looking for a recurrence relation to compute the coefficients. In fact, every family of orthogonals polynomials $(H_k)_{k\in\mathbb{N}}$ whose higher coefficient equals 1 fulfils a three term recurrence relation :

$$H_{n+1}(x) = (-\beta_n + x)H_n(x) - \alpha_n H_{n-1}(x), \tag{2.77}$$

with $H_0 = 1$, $H_{-1} = 0$ and :

$$\alpha_n = \frac{\int_{\mathbb{R}} H_n(x)^2 \, d\mu(x)}{\int_{\mathbb{R}} H_{n-1}(x)^2 \, d\mu(x)}, \ \alpha_0 = 1, \tag{2.78}$$

$$\beta_n = \frac{\int_{\mathbb{R}} x H_n(x)^2 \, d\mu(x)}{\int_{\mathbb{R}} H_n(x)^2 \, d\mu(x)}. \tag{2.79}$$

For instance we give the recurrence relations of two classical families of polynomials : the Hermite polynomials $(\mathrm{He}_n)_{n\in\mathbb{N}}$ associated to normal law and Legendre polynomials $(\mathrm{Le}_n)_{n\in\mathbb{N}}$ associated to uniform law.

$$\mathrm{He}_{n+1}(x) = x\mathrm{He}_n(x) - n\mathrm{He}_{n-1}(x), \tag{2.80}$$

$$\mathrm{Le}_{n+1}(x) = \frac{2n+1}{n+1} x\mathrm{Le}_n(x) - \frac{n}{n+1}\mathrm{Le}_{n-1}(x). \tag{2.81}$$

Tensorisation of the basis

Since ξ are independent random variables, we can write its measure as : $\rho_\xi(x) = \rho_{\xi_1}(x_1) \times ... \times \rho_{\xi_n}(x_n)$ and as a consequence, H_k will be the product of univariate polynomials $(P_k)_{k\in\mathbb{N}}$: $H_k(X) = \Pi_{j=1}^n P_{\alpha(k,j)}(X_j)$, where $\alpha(k,.)$ is a multi-index specifying the degree of each univariate polynomial in the product. The orthogonality property allows a simple definition of the coefficients of the gPC expansion (2.82) of Y :

$$Y = F(\xi;\theta) = \sum_{k=0}^{+\infty} a_k H_k(\xi) \ \text{ with} \tag{2.82}$$

$$a_k = \mathbb{E}[F(\xi;\theta)H_k(\xi)] = \int_{\mathbb{R}^n} F(x;\theta)H_k(x)\rho_\xi(x)dx. \tag{2.83}$$

Instead of indexing the expansion of equation (2.82) on a single integer, one can also make use of a multi-index notation that is equivalent. If Λ_p is an index set for multi-index $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n) \in \mathbb{N}_0^n$, then $\mathbb{P}_{\Lambda_p} \equiv \mathrm{span}\{H_{\boldsymbol{\gamma}} \mid \gamma \in \Lambda_p\}$ and we can then write $H_{\boldsymbol{\gamma}}(\xi) = \prod_{i=1}^n P_{\gamma_i}^{(i)}(\xi_i)$ where $P_{\gamma_i}^{(i)}$ is the γ_i^{th} order basis function in dimension (i). Using the notation introduced above, one can rewrite the truncated gPC expansion approximating Y as follows :

$$Y = \sum_{\boldsymbol{\gamma}\in\Lambda_p} a_{\boldsymbol{\gamma}} H_{\boldsymbol{\gamma}}(\xi) + e_T(\xi), \tag{2.84}$$

where e_T is the remainder due to the modal truncation. If the functional to approximate is a random process, it may also depend on space, time or some other parameters and in that case the gPC coefficients will be deterministic space-, time- or parameters-dependent fields. We will restrict ourselves to tensor-product polynomial spaces $\mathbb{P}_{\Lambda_p}$, where Λ_p is an index set of degree p, and where $P = \dim(\mathbb{P}_{\Lambda_p}) \equiv \#\Lambda_p$ will denote the cardinality of the selected polynomial space. There are different ways of constructing

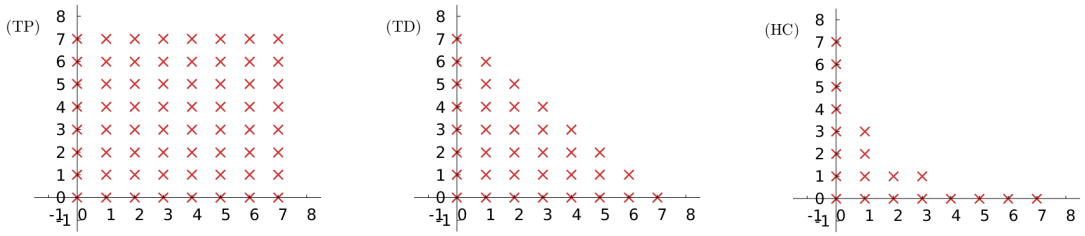

FIGURE 2.9 : Graphical representation of multi-index in dimension 2 for $p = 7$. The three differents strategies are : tensor product (TP) on the left, total degree (TD) at the center and Hyperbolic Cross (HC) on the right.

the approximating polynomial spaces that will impact their cardinality : – 1. Tensor Product (TP) : $\mathbb{P}^{\mathrm{TP}}_{\Lambda_p}$ with index set $\Lambda^{\mathrm{TP}}_p = \{\boldsymbol{\gamma} \in \mathbb{N}^n_0 : ||\boldsymbol{\gamma}||_\infty \leq p\}$, – 2. Total Degree (TD) : $\mathbb{P}^{\mathrm{TD}}_{\Lambda_p}$ with index set $\Lambda^{\mathrm{TD}}_p = \{\boldsymbol{\gamma} \in \mathbb{N}^n_0 : ||\boldsymbol{\gamma}||_1 \leq p\}$, or – 3. Hyperbolic Cross (HC) : $\mathbb{P}^{\mathrm{HC}}_{\Lambda_p}$ with index set $\Lambda^{\mathrm{HC}}_p = \{\boldsymbol{\gamma} \in \mathbb{N}^n_0 : \prod_{i=1}^n (\gamma_i + 1) \leq p + 1\}$. In this thesis, without any loss of generality, we will be using approximation spaces of total degree (TD), so Λ_p will refer to Λ^{TD}_p in the following. And in the case of total degree tensorization the size of the polynomial basis obtained is given by :

$$\#\Lambda^{\mathrm{TP}}_p = \frac{(n+p)!}{n!p!}. \tag{2.85}$$

A key aspect of the stochastic polynomial approximation is to take advantage of the regularity of the QoI in regards to the parametric uncertainty. Regression, (pseudospectral) projection and interpolation (or stochastic collocation) are three possible strategies to construct multivariate polynomial approximations. These approaches rely on discrete and uncoupled model evaluations and have no conceptual difficulties in treating non-linear problems. However, they face the same curse of dimensionality and lead to a number of function evaluations that scale exponentially with the number of random dimensions.

Computation of the coefficients

The first way of determining the coefficients is by use of a Galerkin projection. One can write, assuming the basis is *orthonormal* :

$$a_{\boldsymbol{\gamma}} = \mathbb{E}\left[H_{\boldsymbol{\gamma}}\, F(\xi;\theta)\right] \quad \text{with} \quad \boldsymbol{\gamma} \in \Lambda_p, \tag{2.86}$$

making use of a quadrature in the case of a pseudospectral implementation. Many quadrature families exist with different integration capabilities and (linear or exponential) nodal growth. One may distinguish two classes of grids : – *nested* grids : this is for instance the case of Newton-Cotes, Clenshaw-Curtis, Fejér and Gauss-Patterson-Kronrod formulae, or – *non-nested* grids such as standard Gauss-type formulae.
Another way is to rely on linear regressions, e.g. [Choi et al., 2004]. The least-squares

solution minimizes the residuals $\boldsymbol{r}$ in the L_2-norm, here written in vectorial form : $\boldsymbol{r} \equiv \boldsymbol{y} - \mathbf{H}_{\Lambda_p}\boldsymbol{a}$ and may be apprehended as an optimization problem :

$$\boldsymbol{a} = \underset{\mathbf{a}\in\mathbb{R}^P}{\operatorname{argmin}} \|\boldsymbol{y} - \mathbf{H}_{\Lambda_p}\mathbf{a}\|_2, \tag{2.87}$$

where $\mathbf{H}_{\Lambda_p}$ is the measurement matrix corresponding to the gPC expansion in the index set Λ_p evaluated on a sample $(\xi^{(k)})_{k\in\{1,..,K\}}$. The solution to (2.87) is obtained by computing the following system written in matrix form :

$$\boldsymbol{a} = \left(\mathbf{H}_{\Lambda_p}^{\mathsf{T}}\mathbf{H}_{\Lambda_p}\right)^{-1}\mathbf{H}_{\Lambda_p}^{\mathsf{T}}\mathbf{y}, \tag{2.88}$$

where $\mathbf{y}$ is a vector of observations of size $K \times 1$, $\mathbf{H}_{\Lambda_p}$ the *measurement matrix* of size $K \times P$ with $\mathrm{H}_{ij} = \mathrm{H_j}(\xi^{(i)})$, and $\boldsymbol{a}$ the vector of coefficients of size $P \times 1$. There has been a growing interest in understanding the conditions under which problem (2.87) leads to accurate and stable (multivariate) polynomial chaos approximations for data *randomly* and *independently* sampled according (or not) to their natural orthogonality measures [Cohen et al., 2013, Migliorati et al., 2014, Hampton and Doostan, 2015a]. More specifically, these studies focussed on the relation between the required number of samples and the cardinality of the approximation basis for different sampling measures. If one uses enough sampling points to be able to recover the orthonormality of the basis functions, then $(\mathbf{H}_{\Lambda_p}^{\mathsf{T}}\mathbf{H}_{\Lambda_p})$ is the identity matrix and the link between (2.88) and (2.86) becomes clear.
Other approaches propose to rely on adaptive greedier sampling strategies [Resmini et al., 2016] or address the robustness issue of surrogate modeling in the presence of noisy data samples [Langenhove et al., 2016].

2.3 The Long-term Integration Problem

For wave propagation in a random medium, a natural approach would be to decompose the temporal signals on a gPC basis. However, capturing the variability of a time-dependent problem has always been a challenge when using gPC [Wan and Karniadakis, 2006, Pettit and Beran, 2006, Le Maître et al., 2010].

In fact, a classical example is the case of the free vibration of a linear undamped oscillator [Mai, 2016, Le Maître et al., 2010] : let us consider $y\,(t,\xi)$ the time-dependent displacement of the oscillator for a system with a random stiffness $k = k_0 + k_1\xi$, $\xi \sim \mathcal{U}[-1,1]$ and deterministic initial conditions :

$$\begin{cases} y''(t,\xi) + k\,(\xi)^2\,y(t,\xi) = 0, \\ y(t=0) = 1, \\ y'(t=0) = 0. \end{cases} \tag{2.89}$$

The exact solution is given by :

$$y(t,\xi) = \cos(k(\xi)t). \tag{2.90}$$

Figure 2.10 presents the result of a classical time-frozen (i.e. one gPC development at each time step) performed by Chu Van Mai in his PhD thesis. At every $0.5s$ until $50s$, i.e. at $t = 0, 0.5, 1, ..., 50s$, he conducted $N = 500$ numerical simulations. Then he computed gPC expansions of degree up to 10 and 30 in order to judge the effect of the polynomial degree on the accuracy of the method.
Figure 2.10 shows that gPC of degree 10 are capable of predicting accurately the expected values of the response at considered time instants. Regarding the standard deviation, degree-10 gPC fail to capture the trend of the trajectory from $t = 20s$ onwards. Using polynomials of degree 30 only improves the estimation for the first instants but finally fails in representing the standard deviation too.

The case of wave propagation is similar to this problem but with the distance playing the role of time. In fact, our signal has always a compact support but the stochasticity of the medium will impact the time of arrival and create the same kind of uncertainty even with a simple signal. Moreover, in the case of sound propagation in an inhomogeneous medium the signal can be severely distorted and the variability due to a random medium will be even harder to capture with a simple time frozen gPC.

There have been many attempts to tackle this long-term integration problem and we can divide them in two categories :

- Work on the basis : the first attempt has been to use high degree polynomials to capture high order non linearities. But in order to maintain the gPC approach relevant, one has to use techniques to limit the size of the basis [Lucor and Karniadakis, 2004, Blatman and Sudret, 2010, Hampton and Doostan, 2015b]. Another approach has been to enrich the basis, introducing non polynomials functions : absolute value, step function and inverse function [Ghosh and Ghanem, 2008], sine and cosine [Ghosh and Iaccarino, 2007], wavelets [Pettit and Beran, 2006, Le Maître et al., 2007].

- Try to capture the dynamics : this second approach tries to take advantage of information about the dynamics (for instance for periodic systems see [Witteveen and Bijl, 2008, Le Maître et al., 2010]). To adapt the decomposition to the dynamics, time dependent polynomials have been introduced with different approaches to set the time dependence [Gerritsma et al., 2010, Heuveline and Schick, 2014, Luchtenburg et al., 2014]. Another interesting approach has been to consider the time as a stochastic process [Mai and Sudret, 2017].
  More recently, the use of autoregressive processes to capture the dynamics has also been investigated [Mai et al., 2016, Spiridonakos and Chatzi, 2015].

In this work, we will adopt a strategy that could be classified in the second category : use the acoustic modes to characterize the dynamics of wave propagation in a random medium. With this strategy, we transform a time-dependent problem into a random eigenvalue problem.

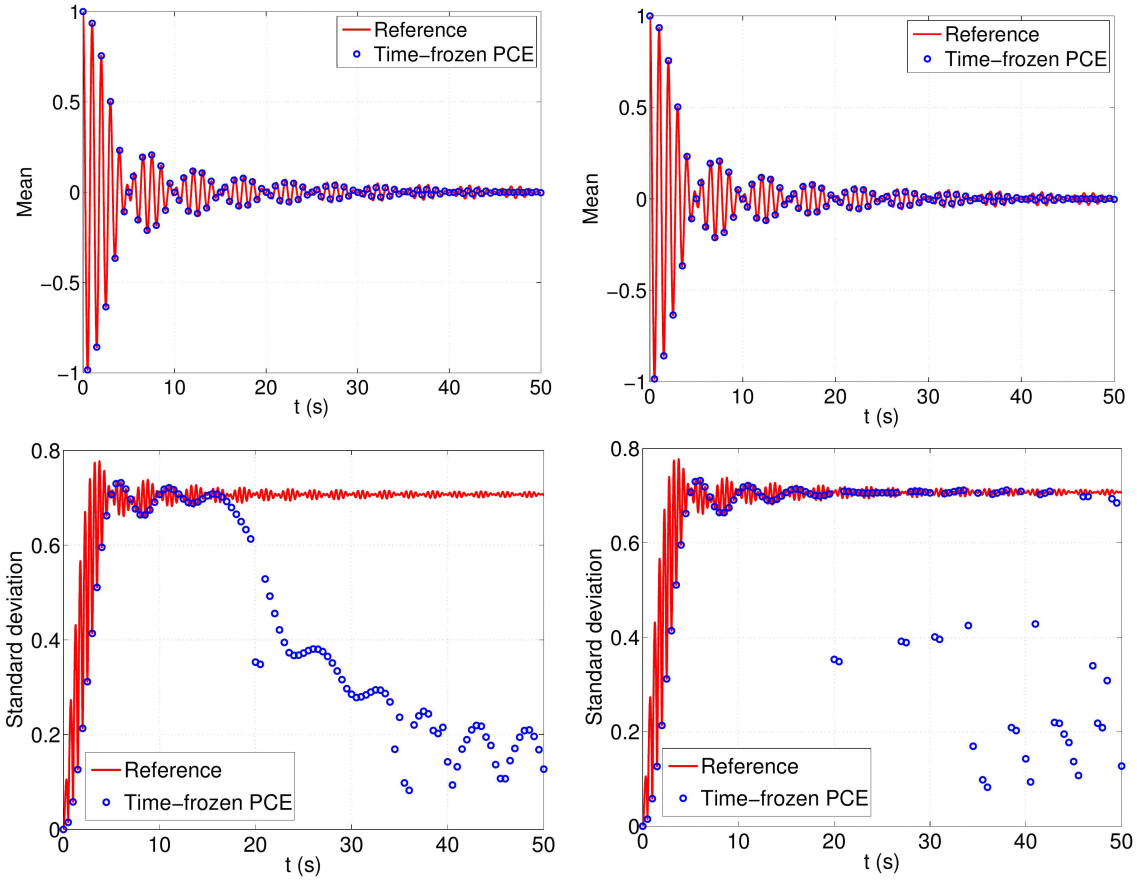

FIGURE 2.10 : Performance of the gPC expansion in capturing the variability of the oscillating function. First line present the approximation of the mean with gPC of order 10 (left) and 30 (right). The second line present the approximation of the standard deviation with gPC of order 10 (left) and 30 (right). *(Source : Chu Van Mai PhD Thesis, 2016.)*

2.4 Polynomial Chaos method for acoustic modes

Building a metamodel of the acoustic modes will allow us to restitute signals at any distance avoiding the long term integration problem. However, this method requires to calibrate a metamodel for each eigenvalue and eigenfunction of operator (2.10). Polynomial chaos expansions have proven to be efficient to capture the variability of spaced random eigenvalues in comparison to the Monte-Carlo approach [Ghosh et al., 2005, Ghanem and Ghosh, 2007]. The main difficulty arises in the case of a clustered spectrum where closely spaced modes can switch and eigenvalues associated with particular physical behaviors do not maintain the same ordering across statistical realizations of the random system [Ghosh and Ghanem, 2012]. This phenomenon occurs in the case of acoustic modes and needs to be investigated to restitute the real variability of the signals in a random medium.

The simplest solution to this problem would be to ignore the switching and only

rely on the ordering of eigenvalues but this would introduce non-smooth dependence on the parameters. As we will see in the next chapter, our strategy will be to use additional samples in order to track the eigenvalues even when the spectrum get clustered in order to maintain a smooth dependance of the eigenvalues on the inputs. This results in additional cost and penalizes the computational performance of the metamodel but enables to keep low order gPC. On the contrary introducing non-smooth dependence would necessitate polynomial of higher order to be captured and thereby additional samples to compute the expansion.

To summarize the approach that will be explained in detail in the next chapter, we are going to consider the acoustical modes as random variables depending on the realization of the celerity profile. Polynomial chaos expansions of those random variables can be computed and used as metamodel of the propagation in the random atmosphere we are considering. This approach requires a careful tracking of the modes under the perturbation but has the advantage to restitute signals at any distance and for any type of source.

Chapitre 3

Polynomial Chaos for Modeling Normal Modes in Atmospheric Boundary Layers

The use of polynomial chaos for incorporating atmospheric boundary layer variability into one-dimensional propagation models is investigated under the assumption that environmental variability exhibits a separation into wind-driven large structures and small-scale fluctuations, such as turbulence. The small-scale structures are described by a stochastic process and the representation of the wave field then consists of an expansion of normal modes that are represented in terms of orthogonal random polynomials. Issues concerning implementation of this multi-scale formalism, the accuracy of the approximation, and the computational burden necessary to evaluate the signal statistics are addressed. Very favourable agreements between the present approach and the usual Monte-Carlo approach are found. Furthermore, this approach offers an efficient mean for connection between the environmental variability and the structure of the propagated signals.

1 Introduction

Atmospheric Boundary Layers (ABLs) have been the subject of analytical, experimental and numerical research due to their relevance to a wide class of applications. Several field experiments [Wilson et al., 2003, Chunchuzov, 2004, Chunchuzov et al., 2005, Kosyakov et al., 2019] have been conducted to study the influence of mesoscale wind speed fluctuations on acoustic pulse propagation through the stable ABL and recent works have shown that even moderate atmospheric variability has an impact on acoustic propagation [Damiens et al., 2018]).

The ABL, which is the lowest part of Earth's atmosphere, has been numerically investigated by many researchers since the pioneering work of Deardorff in 1972 [Deardorff, 1972] related to unstable ABLs. The variety of ABL processes, e.g. wind, diurnal cycle, surface roughness, stratification, and Coriolis acceleration have led to numerous fluid mechanics contributions in recent decades on surface-atmosphere interactions. One key aspect of this problem concerns the behavior of the ABL at night. During

the day, solar radiation warms the surface of the Earth and, under normal conditions, the air temperature decreases with height. At night, however, this effect is reversed and surface cooling leads to the formation of a stable and stratified inversion layer with a potential for suppressing turbulent motions. The temperature inversion typically extends up to altitudes of a few hundreds of meters. Above the inversion there is generally a strong geostrophic Low Level Jet (LLJ), also known as the nocturnal jet [Whiteman et al., 1997, Garratt, 1994]. Similar stabilization of ABL turbulence is seen in winter-time conditions and polar regions.

The meteorology of the stable boundary layer includes states of continuous turbulence and intermittent turbulence [Businger, 1973]. Two prototypical states [Mahrt, 1998] are usually categorized : the weakly stable boundary layer and the very stable boundary layer. The weakly stable boundary layer is normally characterized by windy conditions such that the surface cooling is relatively low. The very stable boundary layer is characterized by weak winds and clear skies that lead to strong net radiative cooling at the surface. The weakly stable boundary layer is described by the Monin-Obukhov similarity theory [Monin, 1970], in which turbulence, although reduced, is continuous. On the contrary, the very stable boundary layer is characterized by global intermittency where turbulence is reduced for periods that are long compared with the time scale of individual eddies [Mahrt, 1989].

During the past few years a number of investigators [Waxler et al., 2008, Waxler, 2004, Waxler, 2002] studied the impact of a LLJ on ground to ground sound propagation. The downward refraction near the ground causes the propagation to be ducted, suggesting that the long range propagation is modal in nature. This duct is, however, leaky due to the upward refraction at high altitudes, and strongly affected by the LLJ. At long ranges (tens of kilometers) the structure of the signal consists of several distinct arrivals, built up by the superposition of many modes, followed by a low frequency tail made up of fewer modes. There are downwind convergence zones in which arrivals from the LLJ and from the nocturnal temperature inversion converge and then separate. The convergence leads to anomalously large signals, that cannot be predicted by the ray theory [Waxler et al., 2008]. At shorter ranges, however, the signals are qualitatively similar to that obtained without LLJ. Since LLJ are always observed in fair weather over flat land [Whiteman et al., 1997], the question arises of the sensitivity of the arrivals to both the meteorological model and the associated uncertainties. This motivates the present research that addresses how the LLJ and small-scale meteorological structures combine to produce qualitatively different waveforms at short and long ranges.

In classical normal mode theory [Jensen et al., 1994, Brekhovskikh and Godin, 2012, Brekhovskikh and Godin, 2013], the slowly-varying mean-state approximation is invariably used. Because of this, the full wave solution obtained for a given sound speed profile cannot be simply connected to the solution obtained for another (significantly different) sound speed profile. Except in the case of very-high-frequency waves, if the sound speed profile changes then some wave components would actually be moving with phase velocity larger than the sound speed which, as is well known, would lead immediately to leaky modes. Thus, for the problem under consideration here the classical approach must be modified to provide a smooth representation of

each mode as the sound speed profile is varied. The procedure adopted is similar to the method of [Damiens et al., 2018] and [Bertin et al., 2014], except that analytic continuation of modes is obtained through adding a Perfectly Matched Layer (PML). The LLJ is modelled using a Gaussian envelope whose amplitude and spread are randomly changed, so that the acoustic frequency cutoff is a random variable as well. The normal modes are expanded in terms of generalized Polynomial Chaos [Finette, 2006]. Small-scale structures are represented using a stochastic process with small variance, which allows use of perturbation theory to model the mode coupling. Further, unlike most of the works cited above, our approach enables the overall sensitivity of arrivals to be decomposed into Sobol indices (i.e. fractions which can be attributed to the amplitude and spread of the LLJ).

This chapter is organised as follows. In Section 2 a simplified model is introduced to mimic temperature and wind velocity fluctuations in nighttime ABLs. In section 3 the theoretical approach used to represent sound propagation and scattering in the ABL is described. In this chapter, the acoustic field is decomposed into normal modes. The details of the calculation of normal modes using polynomial chaos, including some technical aspects such as the 'mode switching' phenomenon, are given in Section 4. Section 5 discusses briefly applications and possible extensions of the approach.

2 The model environment

In this section, analytical forms are introduced to model the temperature and wind profiles. The models are extracted from the classical literature on sound propagation in the Atmospheric Boundary Layer (ABL) [Waxler, 2004, Waxler et al., 2008]. These models are simple but sufficient to capture gross features that are expected to occur in the ABL.

2.1 The meteorological model

The sound speed profile is obtained from a former study [Waxler, 2004, Waxler et al., 2008] in which the propagation in nocturnal conditions was studied. The temperature profile provides a natural duct that is accentuated by a nocturnal jet. The temperature profile is given by an exponential model [Waxler et al., 2008], the wind profile follows the Bussinger-Dyer model [Garratt, 1994] and the LLJ is modeled using a random jet. The wind then writes :

$$u(z) = \eta \left[u_{\mathrm{BD}}(z) + a\mathrm{e}^{-(z-z_J)^2/\sigma^2} \right], \tag{3.1}$$

where $\eta = 6.87$, $z_J = 250$ m and $u_{\mathrm{BD}}(z)$ is given by [Waxler, 2004]. Simplifying further by using the effective sound speed approximation [Godin, 2002, Waxler et al., 2015], the resulting effective sound speed is given by

$$c_0(z) = \sqrt{\gamma r_g T} + u(z), \tag{3.2}$$

where, for air, $\gamma = 1.4$ and $r_g = 287$ K.m^2.s^{-2}. Figure 3.1 shows the resulting profiles for the temperature, the wind and the effective sound speed profile.

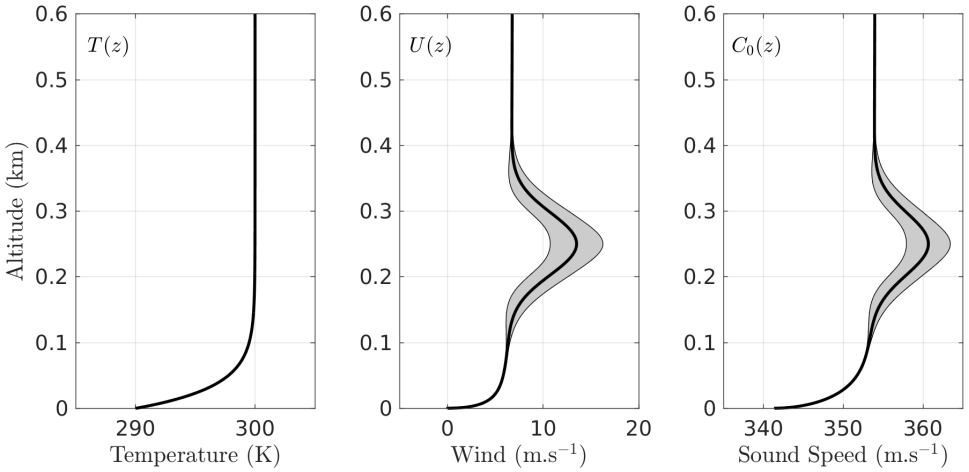

FIGURE 3.1 : From left to right, temperature profile, wind profile and effective sound speed profile. The shaded area presents the variability due to the jet. The envelope is delimited by $\overline{u} \pm 2\sigma_u$, where $\overline{u}$ and σ_u are the mean wind the standard deviation, respectively.

In equation (3.1), the random parameters a and σ are assumed gaussian, with $a = \mathcal{N}(1, 0.2)$ and $\sigma = \mathcal{N}(66.7, 10)$. These two parameters are the unique source of randomness in the sound speed profile. In figure 3.1 the thick solid line is associated with $a = 1$ and $\sigma = 66.7$. The first characteristic of this perturbation (which motivates the choice of standard deviations of a and σ) is its great amplitude. In fact, the variation of the sound speed profile is of around 5% of the mean profile which prevents us from using a perturbative approach.

For sake of simplicity the LLJ is the unique source of variability in the present study. While the ground impedance and the atmospheric absorption can also be considered as random, for low frequencies the impact of these phenomena on the signals can be neglected.

2.2 Small-scale structures

In solving the propagation problem, the significance of small-scale fluctuations has recently been underlined in various studies [Millet et al., 2007, Kulichkov et al., 2010, Chunchuzov et al., 2011, Green et al., 2011, Drob et al., 2013]. These small-scale fluctuations are typical of turbulent flow and should be considered. In order to superimpose small-scale fluctuations onto large-scale meteorological variability, the effective sound speed profile is assumed to be given by

$$c(z) = c_0(z)\left[1 + \epsilon\mu(z)\right], \tag{3.3}$$

where the mean state c_0 corresponds to the atmospheric specifications, ϵ is a small parameter and μ is a stationary random process with mean zero. c_0 contains the variability associated with the LLJ whereas the μ term describes the turbulent variability. This explains the presence of a small parameter ϵ that quantifies the amplitude

difference between the two terms. The random process μ is the Ornstein-Uhlenbeck process whose covariance function is given by

$$\mathbb{E}\left[\mu(z)\mu(z')\right] = \exp\left(-\lambda^{-1}|z-z'|\right), \tag{3.4}$$

where $\lambda > 0$ is the characteristic length-scale of the process (practically, "how far apart" two points z and z' have to be for μ to change significantly.)

The randomness introduced through equation (3.1), (3.2) and (3.3) implies a statistical treatment which can be numerically prohibitive. To circumvent this difficulty we have chosen to separate two scales : the large amplitude perturbation taken in a random $c_0(z) = c_0(z,\xi)$, where ξ is (a,σ), and the small amplitude turbulent flow modelled by the μ term. While the small-amplitude stochastic process can easily be incorporated using the perturbation theory [Waxler et al., 2017], the treatment of the random LLJ needs a more elaborate work, which is the subject of the following sections.

3 Modal propagation in a random waveguide

In this section we recall the modal decomposition of a random acoustic field. This decomposition is the building block used in the sequel to design a metamodel able to restitute random signals. We also give technical details which are essential for numerical application : the Perfectly Matched Layer allows seperating modes of different nature and the problem of mode switching needs to be understood to design an efficient mode tracking algorithm.

3.1 Basic formulation

We use as our starting point the wave equation in cylindrical coordinates (r,θ,z), where r and z are the distance from the source and altitude, respectively. The acoustic pressure is assumed to be initiated by a *deterministic* ground-based localized disturbance of frequency ω.

For a fixed frequency, the acoustic far-field can be expressed in terms of the propagating normal modes ϕ_{n0} of the mean background profile $c_0(z,\xi)$ using equation (2.36) of chapter 2 which writes in our case :

$$p(r,z,\xi;\omega) = \sum_{n=1}^{N} p_n(r,z,\xi;\omega) = \sum_{n=1}^{N} \frac{u_n(r,\xi;\omega)\phi_{n0}(z,\xi;\omega)}{\sqrt{r}}, \tag{3.5}$$

where the eigenmodes are solutions of

$$\left[\frac{\partial^2}{\partial z^2} + \frac{\omega^2}{c_0^2}\right]\phi_{n0} = k_n^2\phi_{n0}, \tag{3.6}$$

together with a Neumann boundary condition at $z = 0$. The subscript 0 refers to modes computed using the mean profile $c_0(z)$. For convenience, we drop the explicit

ω-dependence in the eigenvalues and eigenfunctions. Without loss of generality, the corresponding eigenfunction $\phi_{n0}(z,\xi;\omega)$ can be normalized so that

$$\int_0^\infty \phi_{n0}(z,\xi)\phi_{l0}(z,\xi)\,\mathrm{d}z = \delta_{nl}, \tag{3.7}$$

for all possible ξ, where δ_{nl} is the Kronecker delta. For unbounded atmospheres, the condition (3.7) requires a boundedness condition as $z \to \infty$. Here, and in the following, the continuous part of the spectrum of wave equation is neglected. This is a correct approximation provided we are sufficiently far from the source [Waxler, 2002, Waxler, 2003].

From this point, since every term depends on ξ we are going to drop out the dependance for reader's convenience.

We introduce the right-going and left-going mode amplitudes a_n^+ and a_n^- defined by

$$a_n^\pm(r) = \frac{ik_{n0}u_n(r) \pm u_n'(r)}{2i\sqrt{k_{n0}}}\mathrm{e}^{\mp ik_{n0}r}. \tag{3.8}$$

Substituting (3.3) in the wave equation, multiplying by $\phi_{l0}(z)$, and integrating over the vertical domain $z \geq 0$, we deduce from the orthogonality of the eigenmodes a system of coupled differential equations for the mode amplitudes

$$\frac{da_n^\pm}{dr} = \pm\frac{i\epsilon\omega^2}{2}\sum_{l=1}^{N}\frac{C_{nl}}{\sqrt{k_{n0}k_{l0}}}\left[a_l^+\mathrm{e}^{i(k_{l0}\mp k_{n0})r} + a_n^-\mathrm{e}^{-i(k_{l0}\pm k_{n0})r}\right], \tag{3.9}$$

where the coupling coefficients C_{nl} are defined by

$$C_{nl} = \int_0^\infty \frac{\mu(z)\phi_{n0}(z)\phi_{l0}(z)}{c_0^2(z)}\,\mathrm{d}z. \tag{3.10}$$

The weak uncertainties of the sound speed profile induce a coupling between the propagating modes which becomes of order one after a propagation distance of order ϵ^{-2}. It is possible to carry out a complete asymptotic analysis as $\epsilon \to 0$, including the coupling with the evanescent modes, in a general context or random differential equations, and specifically for random waveguide problems. Substituting the evanescent modes in terms of propagating modes in (3.9), we get an additional term of order ϵ^2. A detailed asymptotic analysis shows, however, that the coupling with evanescent modes does not remove energy from the propagating modes. The overall effect for the propagating modes is an additional dispersive term, which is given in terms of the two-point statistics of the random process μ.

In the forward-scattering approximation, we neglect the left-going (backward) propagating modes and consider the simplified coupled mode equation given by

$$\frac{da_n}{dr} = \frac{i\epsilon\omega^2}{2}\sum_{l=1}^{N}\frac{C_{nl}(r)\mathrm{e}^{i(k_{l0}-k_{n0})r}}{\sqrt{k_{l0}k_{n0}}}a_l, \tag{3.11}$$

where the superscript $+$ is dropped for convenience. Using the matrix form $\mathbf{a}' = \mathbf{H}\mathbf{a}$, this system can be solved using the propagator or transfer matrix, defined by

$\mathbf{T}' = \mathbf{HT}$, starting from $\mathbf{T}(r=0) = \mathbf{I}$. The (m,n) entry of the transfer matrix is the transmission coefficient $T_{mn}(R)$, *i.e.* the output amplitude of the mode m when the input wave is a pure n mode with amplitude one. Using $u_n\sqrt{k_n} = a_n \mathrm{e}^{ik_{n0}r}$, the transmitted field reads

$$p(r,z) = \sum_{n=1}^{N}\sum_{l=1}^{N} \frac{T_{nl}(r)a_l(r_0)\phi_{n0}(z)}{\sqrt{k_{n0}r}} \mathrm{e}^{ik_{n0}r}, \tag{3.12}$$

where $a_l(r_0)$ is the projection of the incident wave (at $r = r_0$) onto the l^{th} mode.

3.2 Perfectly matched layer

In order to simulate wave propagation in the atmosphere, we need a proper numerical wave handling due to the finite size of our computational domain. A common technique in the field of wave simulation consists in adding an absorbing layer at the upper boundary of the domain. One approach is the Perfectly Matched Layer (PML). From a mathematical point of view, introducing this layer becomes equivalent to a change of coordinates in the complex plane. This technique has been initially developed for finite difference time domain (FDTD) simulation ([Bérenger, 1994]) and then adapted to finite element simulation. However the literature is scarce about the use of this PML to modal expansion ([Olyslager, 2004]). An interesting impact of the PML on the spectrum of the propagating operator is nevertheless reported. When Bérenger introduced the PML in the context of electromagnetic simulation ([Bérenger, 1994]), he proposed a split field formalism. But the PML became more popular thanks to its reformulation as a complex stretching of the coordinates ([Chew et al., 1997]). In the case of the wave equation, the stretching had been studied to integrate system with singularities ([Boyd, 1985]).

Denoting H the original operator defined at (3.6) and $\tilde{H}$ the operator with the new coordinates. $\tilde{H}$ is obtained with the complex stretching of the coordinates $u = f(z)$:

$$H = \frac{d^2}{dz^2} + \frac{\omega^2}{c(z)^2}, \tag{3.13}$$

$$\tilde{H} = \frac{1}{f'(z)^2}\frac{d^2}{du^2} - \frac{f''(z)}{f'(z)^3}\frac{d}{du} + \frac{\omega^2}{c(u)^2}, \tag{3.14}$$

where $c(u)$ is the analytic continuation of the function c. Restrictions on f relate to its imaginary part which has to be zero in most of the domain $[0; z_{PML}]$ and to its smooth behaviour in the PML section $[z_{PML}; z_{max}]$ in order to avoid introducing discontinuities. In our case, the PML section starts above the nocturnal jet : $z_{PML} = 0.6$km. A possible function f was taken from [Zhu and Chen, 2013] :

$$\begin{cases} f(z) = z + \dfrac{15i}{2}\left[\tau^2 - \log(\tau^2+1)\right], \\ f'(z) = 1 + i\dfrac{15\tau^3}{1+\tau^2}, \\ f''(z) = i\dfrac{15\tau^2(\tau^2+3)}{(1+\tau^2)^2}, \end{cases} \tag{3.15}$$

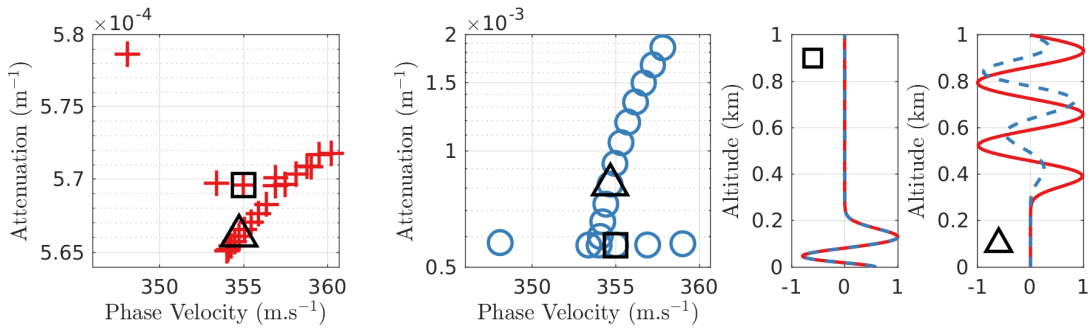

FIGURE 3.2 : Spectrum computed with a complex frequency parameter $\omega = 2\pi \times 20 + 0.2i$ but without PML (left) and the same spectrum computed with a PML (center). On the right, real parts of eigenfunctions corresponding to the discrete spectrum (square) and to the discretization of the continuous spectrum (triangle). Only modes of this second category are affected by the PML.

where $\tau = \left(\frac{z - z_{PML}}{z_{max} - z_{PML}}\right)_+$. From this definition, we can make two statements. First, the trapped modes are not affected by the PML : if (k_i, ϕ_i) is a mode of the original problem such that : $\forall z \geq z_{PML}, \phi(z) = 0$ then it is straightforward to see that (k_i, ϕ_i) satisfies the eigenvalue problem for $\tilde{H}$ and is also a mode of the new problem. Second, the modes that are impacted by the PML have an imaginary part on k_i which leads to an attenuation in the domain. Figure 3.2 shows the spectrum of the discretization of operators H (in red) and $\tilde{H}$ (in blue) in order to see the modification – both on eigenvalues and eigenfunctions – of the continuous part of the spectrum. Note that the two spectrum exhibit a continuous contribution which is here discretized and, as explained in the previous section, is not considered. The spectrum of the discretized operator contains a subset corresponding to the discretization of the continuous part of the original differential operator H. Thanks to the PML this subset is easy to recognize since only the continuous spectrum will be affected by the PML [Olyslager, 2004]. It is worthwhile to specify that in this work, the frequency parameter ω is treated as a Fourier-Laplace transform variable, $\omega = \omega_r + i\omega_i$, with a small positive imaginary part $\omega_i \ll 1$. The reason is related to the temporal response, which is obtained by applying the inverse Fourier transform along a path in the complex ω-plane [Bertin et al., 2014]. For this reason, the eigenvalues $k_n(\omega)$ lie in the upper half of the complex k-plane as shown on figure 3.2 even on the left where there is no PML but only a complex ω. This figure shows the impact of the PML on the continuous spectrum. This impact has been studied more extensively in [Goursaud, 2010] and corresponds to a rotation in the complex plane. On the eigenfunction it is worth noticing that the trapped modes (with a square) are not impacted by the PML whereas modes above the jet are impacted.

In this work, we use a boundary value (or implicit) method for solving the eigenvalue problem. The wave equation is reduced to a linear algebraic equation using a pseudo-spectral technique. The global eigenvalues are obtained by applying a spectral

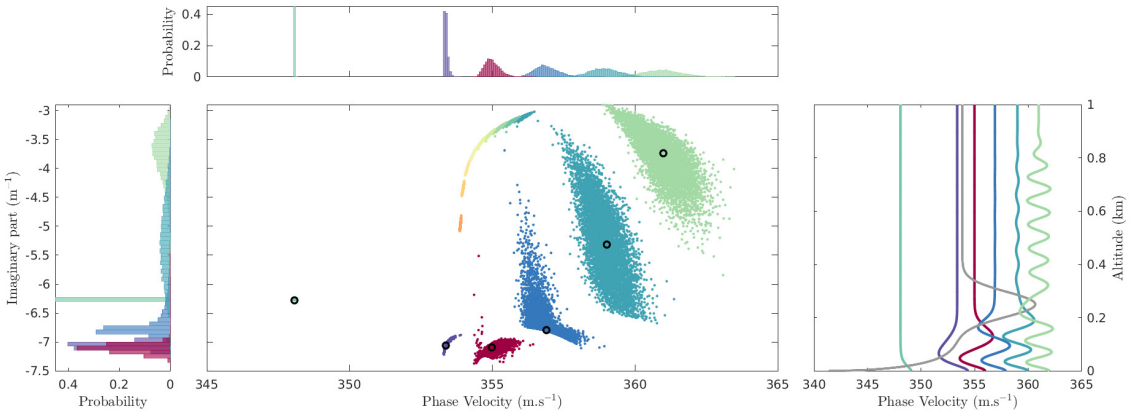

FIGURE 3.3 : Distribution of eigenvalues of operator 3.6 for 5000 realizations of the random celerity profile at 20Hz. Histograms show empirical distributions of phase velocities (top, in $\mathrm{m.s}^{-1}$) and imaginary part of the eigenvalues (left in logarithmic scale, in m^{-1}). Eigenfunctions corresponding to the eigenvalues encircled in black are plotted on the right with the mean celerity profile (up to 1km).

collocation algorithm to the generalized eigenvalue problem. While care must be taken to preclude spurious eigenvalues, it is worth mentioning that such an approach can be applied to problems in which the eigenvalue appears nonlinearly so that the absorption can be considered within the same numerical framework [Bridges and Morris, 1984].

3.3 Random modes

The eigenvalues can be identified as the zeroes of a dispersion relation $\mathcal{D}(k,\omega)$ that depends on $c_0(z)$. In our case, $c_0(z)$ depends on the random variable ξ. As a result, the eigenvalues k_n and eigenfunctions ϕ_n are random variables and random functions. Figure 3.3 shows five thousand spectra corresponding to five thousand realizations of the sound speed profile $c_0(z,\xi)$. The eigenfunctions[1], plotted on the right, show the interaction between the mode and the profile; each mode is refracted at a certain altitude. The modes trapped under the jet have much more variability than the one located under the boundary layer (with a phase velocity around $348\mathrm{m.s}^{-1}$), not impacted by the variations of ξ. The modes with little variability aligned around a phase velocity of $355\mathrm{m.s}^{-1}$ and with an imaginary part between 10^{-3} and 10^{-5} correspond to the discretization of the continuous part of the spectrum of (3.6).

Using $A_n = a_n \mathrm{e}^{ik_{n0}r}$ as our new variables, the system (3.11) reads

$$\frac{dA_n}{dr} = i\sum_{l=1}^{N}\left(k_{l0}\delta_{ln} + \epsilon\frac{\omega^2 C_{nl}}{2\sqrt{k_{n0}k_{l0}}}\right)A_l. \tag{3.16}$$

[1]The decreasing amplitude of the eigenfunctions for an altitude higher than 800m is due to the Perfectly Matched Layer (PML) detailed in the previous paragraph.

This system is readily integrated to yield

$$A_n(r) = A_n(0)\mathrm{e}^{ik_{n0}r}\left[1 + i\epsilon r\frac{\omega^2 C_{nn}}{2k_{n0}} + \epsilon\sum_{l\neq n}\alpha_{ln}\frac{A_l(0)}{A_n(0)} + O(\epsilon^2)\right], \tag{3.17}$$

where α_{ln} is bounded as $r \to \infty$. Thus, using the slow coordinate variable $R = \epsilon r$, and substituting (3.17) into (3.5), the perturbation expansion of the pressure far-field at ground level ($z = 0$) simply reads

$$p(\omega) = \sum_{n=1}^{N} p_n(\omega) = \frac{\mathrm{e}^{i\pi/4}}{\sqrt{8\pi}}\sqrt{\epsilon}\sum_{n=1}^{N}\frac{\phi_{n0}^2(0)\mathrm{e}^{ik_{n0}r}}{\sqrt{Rk_{n0}}}\left[1 + iR\frac{\omega^2 C_{nn}}{2k_{n0}} + O(\epsilon)\right], \tag{3.18}$$

where we have used the initial condition $A_n(0) = \phi_n(0)\mathrm{e}^{i\pi/4}/\sqrt{8\pi}$ that represents a point source at $r = 0$, $z = 0$. For $\epsilon = 0$, we recover the classical form of the impulse response [Jensen et al., 1994] for large r. Note that, to leading order in ϵ, we only get the diagonal elements of the matrix C_{nl}. Indeed, the more the diagonal terms dominate, the more the expansion (3.18) is valid for large R. On the other hand, given a typical propagation distance R, ϵ can be fixed for (3.18) to remain valid.

For small ϵ, we may use perturbation expansions to compute the eigenvalues k_n, assuming expansions of the form $k_n = k_{n0} + \epsilon k_{n1} + \cdots$ and $\phi_n = \phi_{n0} + \epsilon\phi_{n1} + \cdots$. Since the leading order k_{n0} depends only on the mean problem, it is determinate and can be computed by replacing the random coefficient in (3.6) by its respective mean value c_0 ; the remaining coefficients k_{nl} ($l > 1$) being random. Upon substituting the expansion of k_n into (3.6), expanding in powers of ϵ and identifying terms of equal orders of magnitude, one readily obtains $k_{n1} = \frac{1}{2}\omega^2 C_{nn}k_{n0}^{-1}$, which is the term being multiplied by iR in (3.18). We conclude, as expected intuitively, that the variance of p can be expressed in terms of variances and covariances of eigenvalue perturbations k_{n1}.

Figure 3.4 shows the link between the variability of $p(\omega)$ and the variance of the signal. On the first line (a-d) the probability distribution of $|p_n(\omega)|$ for the four first modes is plotted as a function of the frequency $f = \omega/2\pi$. The first mode (a) has very little variability, it is the one located on the left (with a phase velocity of 348m.s^{-1}) on figure 3.3, it is not impacted by the variablity of the LLJ. The black solid line is the same on those four plots and represents the spectrum of the source used to produce the signals presented on the second line (e-h). With light colors we have plotted the total signal and variance at a distance of 50 km and the dark colors plots are the signal and variance associated to only one mode. The variance has been evaluated using the Monte-Carlo estimator with a sample of 5000 signals obtained for 5000 profiles randomly drawn. Even though signals associated with the first mode (e) are very similar, there exists a small shift in time between one another, which gives the plotted variance. For the first arrival, the three other modes (f-h) give almost all the variance ; the covariance between the different modes seems to be negligible.

This variance decomposition shows why the acoustic modes are an interesting tool to isolate the different contribution depending on the localization of the perturbation.

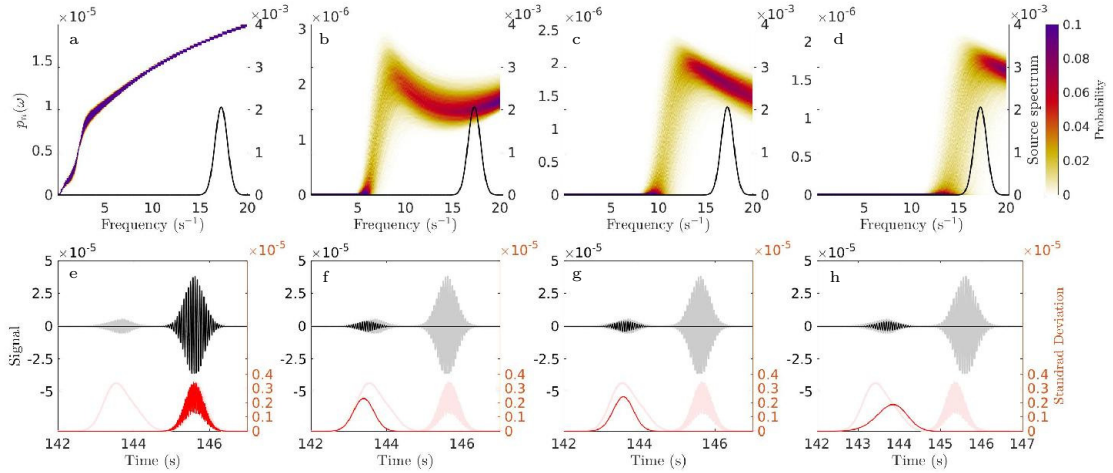

FIGURE 3.4 : Probability density functions of the residuals associated to the first four eigenvalues $|p_n(\omega)|_{n=1,..,4}$ for a frequency between 0 and 20Hz and at 50 km from the source are plotted in the first row. The spectrum of the source used to generate signals in the second row is represented in black. The signal in black corresponds to one realization at 50km of the whole signal (in gray) and for the corresponding mode (in black) and (in red) the partial standard deviation and (in light red) the total standard deviation.

In our case, we have located the perturbation on the nocturnal jet, only the three modes interacting with the jet are sensitive to this perturbation.

3.4 The mode-switching model

In the presence of clustered eigenvalues, switching of the ordering of eigenvectors can occur for a small change in the physical parameters, such as changing the frequency or the profile shape. This particularity consists in the ability of two different modes to suddenly switch when the parameters of the problem (either the frequency ω or the random parameters ξ) evolve. This behaviour arises in many different context and is sometimes called *mode veering*, it requires a special numerical treatment to be able to track the two modes [Pierre, 1988, Kalaba et al., 1981, Adhikari and Friswell, 2004, Ghanem and Ghosh, 2007, Georg et al., 2018].

On figure 3.5 the evolution of two modes with the frequency is plotted on the left. Each line corresponds to a different value of ξ (i.e. to a different profile). On the right, the eigenfunctions associated to those two eigenvalues are plotted. The two modes are associated with different acoustical situations : one is trapped under the jet and will contribute to the propagation in the case (considered here) of a source at the ground and the other case corresponds to an evanescent mode located mainly above the jet. However, on the second line, each mode evolves continuously towards the other situation. This phenomenon has practical implication especially when looking at the statistics of a mode.

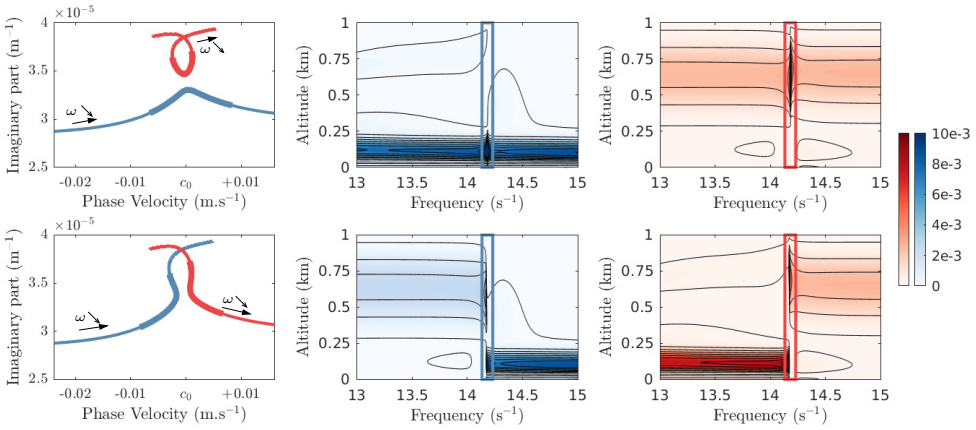

FIGURE 3.5 : Spatial branches of two eigenvalues between 13 and 15Hz for two different profiles for phase velocity around $c_0 = 353.8839$ m.s^{-1}. $\Re[\phi_i(z;\omega)^2]$ for the corresponding eigenfunctions are plotted on the right. The thick lines correspond to the switching frequencies, framed on the plot with the eigenfunctions.

In paragraph 3.3, we have presented the pdf of $|p_n(\omega)|_{n=1,..,4}$ computed with 5000 realizations. To be able to compute those pdf it has been necessary to identify in the set of eigenvalues of each realization, the considered mode n. The same problem arises when looking at the evolution of a mode as a function of the frequency to compute for instance the signal associated to one mode (like in figure 3.4, e-h).

This problem also appears for instance when one wants to take into account the horizontal variations of the atmosphere [Bertin et al., 2014, Damiens et al., 2018]. In this case, p_n depends on the integral of the mode alongside the propagation ; to compute this integral, it is thereby necessary to identify a mode at each point of the range discretization. A classical approach [Allemang, 2003] consists in looking at the scalar product between eigenfunctions and associate the eigenvalues whose eigenfunctions have the highest scalar product. Unfortunately, in our case the different eigenvalues have very similar shapes and the scalar product can be very similar between two different eigenfunctions. An option would be to discretize more finely but it would increase drastically the numerical cost. Our approach is detailed in annex 4.2 and relies on the continuous dependance of the eigenvalues in the parameters ω and ξ.

4 Polynomial chaos based metamodel

We have exposed in the previous chapter (section 2.2) the principle of generalized Polynomial Chaos (gPC) expansions. We have also exposed the limitation due to long term integration problem when it comes to oscillatory phenomena. We propose in this section a method to circumvent this limitation. The first paragraph gives the different steps of the construction of a metamodel based on the gPC expansion of the acoustic modes. Under a stochastic perturbation the modes evolve in the complex plane in a chaotic way, we explain in the second paragraph how we have proceeded

to track them. Some special behaviours appear under certain condition such as an exchange of two modes. This phenomenon can alter the convergence of the metamodel and we propose in the third paragraph an adaptation of the metamodel to adress this difficulty. The last paragraph explains how the perturbative treatment of the small-scale perturbations can be treated coherently with the metamodel.

4.1 gPC representation of the acoustic modes

If polynomial chaos-based stochastic approximation has shown very good results for many different applications, the method has issues in approximating a solution with unsteady dynamics over a long time interval [Gerritsma et al., 2010, Branicki and Majda, 2013]. One potential mitigation of the problem is to increase the approximation order [Lucor and Karniadakis, 2004, Hou et al., 2006, Ozen and Bal, 2017], but this is problematic since the amount of work does not scale linearly with the order of the polynomial basis functions and it becomes even more troublesome as the number of random variables increases. Early works [Wan and Karniadakis, 2006, Gottlieb and Xiu, 2008] have shown that even in the simple case of one-dimensional stochastic scalar advection in uniform media, a truncated gPC approximation of constant order produces an approximation error that increases linearly with time. In the case of wave propagation in a random medium, the same limitation arises for long range propagation. As a consequence, directly computing the gPC expansion of the stochastic pressure field is doomed to failure. To circumvent this limitation some studies in underwater acoustics have worked with an envelope function of the pressure field [Khine et al., 2010, Finette, 2006, Creamer, 2006].
A main contribution of this work is to propose to work instead in the Fourier space and to fine tune stochastic approximation of the normal modes decomposition of the Helmholtz operator. The modal decomposition has the advantage to exhibit the different waveguides of the considered medium which allows to study the impact of localized perturbation on the propagation. The gPC framework has been investigated for the study of random eigenvalue problems [Ghosh et al., 2005] and allows to take into account great deviation.

In our case, the random inputs are the variables describing the variation of the atmospheric boundary layer. It will be the shape parameters a and σ of the atmospheric boundary layer introduced in paragraph 2.1. Every realization of a and σ gives a particular celerity profile that will be an input for the acoustic modal decomposition. The quantity of interest Y is the acoustic output, it could be the signal itself or its characteristics : amplitude, time of arrival or duration. As explained in paragraph 4.1, our choice has been to decompose directly the acoustic modes and Y will be a vector containing a set of modes.

In this work we propose to build a surrogate model using gPC expansion of the normal modes. The calibration of the surrogate model is exposed in algorithm 1 and can be summarized as follows :

1. The first step consists in defining an appropriate design of experiment (DoE) $(\xi^{(k)})_{k\in\{1,..,K\}}$. This can be either a structured quadrature or a random sample

Algorithm 1 Polynomial Chaos metamodel to generate random signals

Input: A random celerity profile $c(z, \xi)$ with a model for uncertainties ξ.

1: Compute a design of experiment (DOE) $(\xi^{(k)})_{k\in\{1,..,K\}}$ and the corresponding celerity profiles $c(z, \xi^{(k)})_{k\in\{1,..,K\}}$.
2: Compute the spectra $(k_n(\omega, \xi^{(k)}), \phi_n(\omega, z, \xi^{(k)}))_{k\in\{1,..,K\}}$ for every point of the DOE and every frequency.
3: Densify the DOE and compute the spectra on those additional points for the last two frequencies.
4: **for** ω in $\{\omega_{max}, \omega_{max-1}\}$ **do**
5: Track each mode in the random realisations of the spectrum for every value of $\xi^{(k)}$ in the densified DOE.
6: **end for**
7: **for** k in $\{1, .., K\}$ **do**
8: **for** ω in $\{\omega_{max}, .., \omega_1\}$ **do**
9: Track each mode at every frequency.
10: **end for**
11: **end for**
12: **for** ω in $\{\omega_{max}, .., \omega_1\}$ **do**
13: Compute the gPC expansions of the modes for each frequency using their values on the DOE.
14: **end for**
15: Use the gPC expansions to generate an important sample of eigenvalues and eigenfunctions.
16: Set the distance and the Fourier spectrum of the source signal. Use them in addition to the generated eigenfunctions and eigenvalues to compute the corresponding sample of residual sums.
17: A Fourier transform will give the random signals.

Output: Set of random signals generated by the metamodel.

depending on the elected strategy. Then comes the computation of the spectra for the celerity profile associated to each configuration $c(z, \xi^{(k)})_{k\in\{1,..,K\}}$.

2. One then needs to gather all sample realizations pertaining to each eigenvalue for every frequency $(k_n(\omega, \xi^{(k)}), \phi_n(\omega, z, \xi^{(k)}))_{k\in\{1,..,K\}}$. This is not a trivial computational task due to potential clustered spectrum redistribution induced by the random medium. To tackle this difficult task we have developped a tracking algorithm detailled in the appendix 4.2.

3. Finally, the gPC approximation has to be built, using either a projection or a regression method. Cross-validation is used to select the optimal polynomial degree, given the sample.

Once the expansions of the eigenvalues and eigenfunctions are converged and accurate, they can be used at no cost in order to generate Green functions and temporal signals $\hat{p}(t, R, \xi)$. It has to be emphasized that the surrogate model can be evaluated at any distance and for any source without supplementary computational cost.

4.2 Tracking of clustered random eigenvalues

No matter the way gPC coefficients are computed (e.g. Galerkin projection or regression), a robust approach needs to make sense of the mapping between the quantity of interest and the different realizations of the random input ξ. In our case, the quantities of interest are the eigenvalues and eigenfunctions of the random propagating operator, and for every input realization of ξ and every frequency, the entire spectrum $\Lambda(\omega^{(i)}, \xi^{(j)}) \equiv \Lambda^{(i,j)}$ is computed. Overall, it takes the form of a collection of spectrum sets $\{\Lambda^{(i,j)}\}_{i\in J_\omega, j\in K}$.
In the following, the eigenspectrum mapping is determined based solely on the eigenvalues, the eigenfunctions mapping being straightforward thanks to their association to the eigenvalues. Monitoring the evolution of a particular eigenvalue as a function of the input parameter ξ (itself having multiple dimensions) and the frequency ω can be a difficult task as the dynamic of the eigenvalues is *a priori* unknown. Indeed, the closely spaced cluster of eigenvalues, due to the random perturbation, makes the eigenvalues tracking intricate [Georg et al., 2018, Rahman and Yadav, 2011] across the frequency range, with appearance of curve veering or mode switching [Ghosh and Ghanem, 2012].
We propose a strategy for the eigenmode tracking based on vicinity measures : – we first characterize the mapping of the eigenvalues for a given frequency, – then determine the mapping to the next frequency and – finally obtain the overall mapping for all frequencies. More specifically, the approach summarises in three step, step 1 and 2 are illustrated on figure 3.6 and step 3 on figure 3.7 :

1. for a given frequency, one must identify the sample of all parameter realizations pertaining to the same eigenvalue. This is facilitated by a densified version of the original DoE and with the help of a vicinity measure in the complex plane. The supplementary points added to the DoE will lengthen the computations

but, as long as we do those additional computations for only two frequencies, we will not slow down too much the all process.

2. for the next available frequency and a particular parameter realization, a modal sorting is initialized using the sorted eigenvalues at the previous frequency (cf. step 1.). The same tracking process across all parameter realizations is then repeated for the new frequency as in the first step.

3. Once this sorting has been done across these two frequencies, use it as a starting point to follow the eigenvalues over the entire frequency range. This sequential progression is done by linearly extrapolating a predicted state at the current frequency from the knowledge of the previous states at the two lower frequencies, and then replacing it according to the current closest available eigenvalue.

Our strategy is general as long as we assume some type of continuous dependence of the eigenvalues against both frequency $\omega^{(i)}$ and parametric $\xi^{(j)}$ changes. Moreover, our method relies on a measure of vicinity to follow an eigenvalue in the complex plane. In order to work with a reliable distance, we first rescale the imaginary part of the eigenvalues in order to have an aspect ratio close to one.

We now give a more detailed description of the method. We consider the set of spectrum corresponding to the last frequency : $\{\Lambda^{(N_\omega,j)}\}_{j\in K}$. For the first mode, we determine an optimal path covering all the samples of the dense DoE, based on a nearest-neighbor criterion. For ease of notation, we identify this path by the labels sequence $\{j_1, j_2, \ldots\}$. The elements of $\Lambda^{(N_\omega,j_1)}$ are then arbitrary labelled.
Moving to the next frequency dataset, every single eigenvalue from $\Lambda^{(N_\omega,j_2)}$ takes the label of its nearest point in $\Lambda^{(N_\omega,j_1)}$. This operation is sequentially repeated in order to label all the eigenvalues for each set from $\{\Lambda^{(N_\omega,j)}\}_{j\in K\setminus\{j_1,j_2\}}$. This explained why we have densified the DoE in order to recover small variations along two consecutive sets $\Lambda^{(N_\omega,j)}$ and $\Lambda^{(N_\omega,j+1)}$, and thereby avoiding wrong pairing.

At frequency $N_\omega - 1$, we proceed exactly the same way, but the tracking/labelling has to be coherent with the one of frequency N_ω. To this purpose, we first need to associate the closest neighbors of elements of $\Lambda^{(N_\omega-1,j_1)}$ in $\Lambda^{(N_\omega,j_1)}$.

Once the numbering has been done for the last two frequencies, we select a particular realization $j = j_0$ and label the spectra for all frequencies. To this end we compute for each set $\Lambda^{(i,j_0)}, i < N_\omega - 1$ the linear extrapolations of the eigenvalues, computed from the two preceding sets $\Lambda^{(i-1,j_0)}$ and $\Lambda^{(i-2,j_0)}$. This provides an initial guess that is corrected by searching for the nearest point in $\Lambda^{(i,j_0)}$. To initialize this process,we use the identification conducted for $i = N_\omega$ and $N_\omega - 1$. This approach across frequencies is repeated for all $j \in K$.
This technique is efficient if the evolution of the eigenvalues remains smooth over the different parametric ranges. It is usually the case for the dependence in frequency but it can be more complicated for the dependence in terms of the distribution of ξ. This motivates the choice for a denser DoE at step 1. Because we only compute the spectra on those points for the two last frequencies, the additional cost is not prohibitive.
On figure 3.7, we show the case of two switching modes. This phenomenon is quite common and we will explain in the next paragraph how to deal with those switching

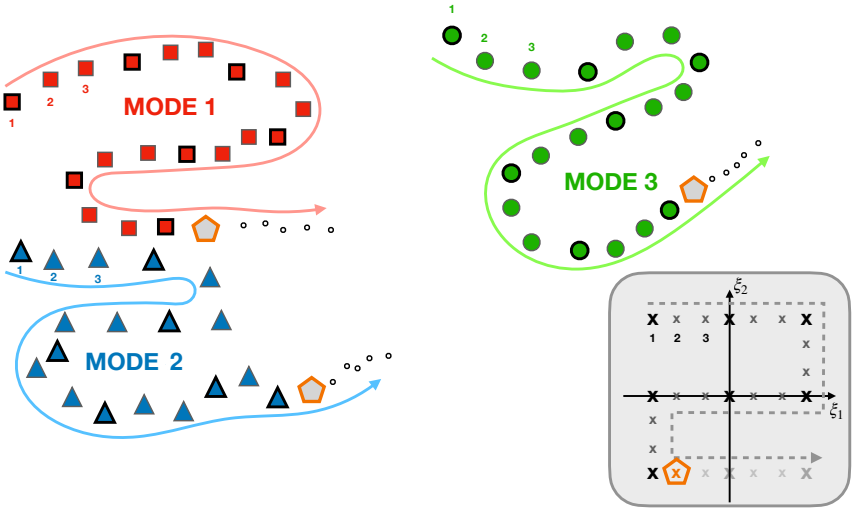

FIGURE 3.6 : Tracking the eigenvalues for a fixed frequency by crossing the quadrature points. The smaller markers indicates the supplementary points added to densify the quadrature to make tracking easier.

when computing the gPC expansions. Concerning the tracking of switching eigenvalue, our method revealed quite efficient thanks to the linear extrapolation (as we see on figure 3.7 a simple vicinity measure would not be conclusive). The efficacy of the method depends on the discretization in frequency but the Shanon theorem already impose a tight sampling to generate correct continuous-time signals.

To conclude this paragraph on the tracking strategies let us explain why some more classical methods have not revealed efficient in our case. A very classical way to track the eigenmodes consists in using the normalized scalar product between eigenfunctions :

$$\frac{\langle \phi_n, \phi_l \rangle}{||\phi_n|| ||\phi_l||}. \tag{3.19}$$

A given eigenfunction is associated with the eigenfunction of the next frequency with higher normalized scalar product. This method works very well when eigenfunctions are sufficiently different but in the case of two switching modes the two eigenfunctions look very similar and the technique fails. We have tried to adapt this method by densifying the frequency sampling when this situation occurs (the value of the normalized scalar product drops significatively which helps identifying this situation). Unfortunately this led to an explosion of the computations around the frequencies where some modes switch.

Another technique we have tried was to see the eigenproblem as an optimization problem. For a given value k we solve numerically two differential problems consisting

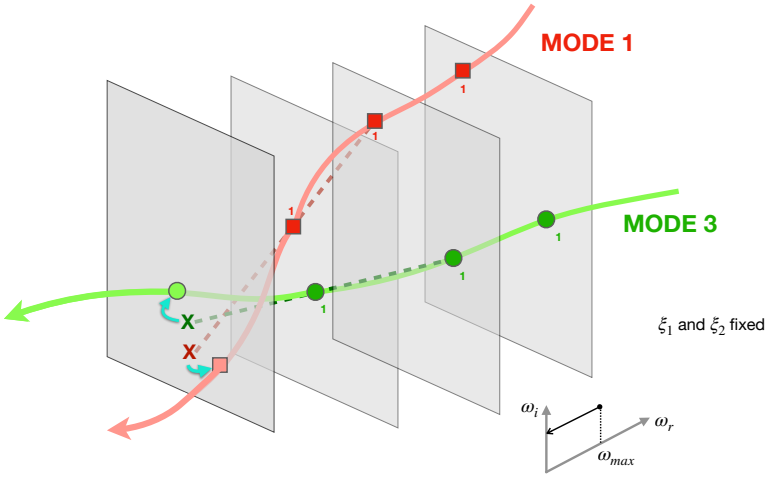

FIGURE 3.7 : Following eigenvalues in the frequency domain once the identification has been done on Λ_{i_1} and Λ_{i_2}. The dashed lines represent the linear extrapolation used to identify the modes in the last plane looking for the closest eigenvalue of each cross. As illustrated on this figure our method remains reliable in the case of switching eigenvalues.

in the Helmoltz equation starting from the ground or from the upper bound of the domain. This results in two solutions ϕ_u and ϕ_d satisfying equation (3.6) replacing k_n by k and the boundary conditions at the ground and at the top. The function to optimize is in this case :

$$\begin{aligned} D : \mathbb{C} &\to \mathbb{R} \\ k &\mapsto \begin{vmatrix} \phi_u(z_m,\omega,\xi) & \phi_d(z_m,\omega,\xi) \\ \partial_z\phi_u(z_m,\omega,\xi) & \partial_z\phi_d(z_m,\omega,\xi) \end{vmatrix}. \end{aligned} \tag{3.20}$$

where z_m is a point around the middle of the domain. If $D(k) = 0$ it means that the two solutions are proportionnal and k is an eigenvalue. This optimization formulation can be used to track eigenvalues using the eigenvalue at the preceding point as a starting point of the optimization at the current frequency. This method has the advantage of computing only the eigenvalues we are interested in which could save computational cost. Unfortunately the solution of the optimization algorithm is not always the nearest eigenvalue of the starting point. It depends on the form of the function and on the algorithm. We have observed the case of two close eigenvalues for which the optimization ends on the same eigenvalue starting from both eigenvalue at the preceding frequency. We have not tried to densifying the frequency sampling by fear of an explosion of the computational time.

4.3 gPC mixture for switching modes

The objective here is to build a continuous metamodel of the propagation despite the switching phenomenon. In fact, the switching leads to the clustering of the response depending on the values taken by the random perturbation (ξ_1, ξ_2). As a consequence, the surface response (QoI function of (ξ_1, ξ_2)) will present a discontinuity separating the two clusters (see 3.A for an illustration). If one computes a global gPC approximation of the response of the modal imaginary part for instance, high-order approximation will be needed in order to capture the jump in the response. This does not only demand a large sample size in order to accurately build the approximations, but it may also entails the quality of the approximation due to Gibbs-like phenomena inducing large spurious oscillations of the response surface.
In fact, this different behaviour leads to a bimodal distribution which deteriorates the convergence of a classical gPC [Nouy, 2010, Soize, 2015]. We propose to adopt a more local approach by partitioning the parametric domain and construct multiple metamodels based on different subsamples. The determination of this partition is key and cannot be performed *a priori* based on the design of experiments (DoE). It has to be determined based on the response sample. To this purpose, we use an indicator function $I(\xi_1, \xi_2)$ which will take a value of 1 or 2 depending where the mode is localized. The location of the indicator discontinuity in the parametric space may be somewhat complex to determine from the sample. A numerical stratagem we use is to build a metamodel of this indicator function itself. Here, we use a kriging approach which is more efficient for modeling discontinuous functions. The idea is to model the unknown function as a realization of a gaussian process conditionally to the observed points, its approximation being provided by the process mean $M(\xi_1, \xi_2)$, with some level of confidence given by its standard deviation $\Sigma(\xi_1, \xi_2)$.
This gaussian regression provides us a metamodel of the indicator function :

$$\hat{I}(\xi_1, \xi_2) = \begin{cases} 1 \text{ if } M(\xi_1, \xi_2) \leq 0.5, \\ 2 \text{ if } M(\xi_1, \xi_2) > 0.5, \end{cases} \tag{3.21}$$

but also a validity domain $\mathcal{D}$, here empirically taken as :

$$\mathcal{D} = \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 \,|\, \Sigma(\xi_1, \xi_2) < 1/3 \right\}. \tag{3.22}$$

In fact, we consider that when the standard deviation is greater than $\frac{1}{3}$, the metamodel is not good enough to seperate the two regions : $\mathcal{R}_1 = \{(\xi_1, \xi_2) \in \mathbb{R}^2 | I(\xi_1, \xi_2) = 1\}$ and $\mathcal{R}_2 = \{(\xi_1, \xi_2) \in \mathbb{R}^2 | I(\xi_1, \xi_2) = 2\}$. Adaptive strategies exist to improve the approximation by proposing new optimal samples depending on the chosen optimality criteria [Wan and Karniadakis, 2005]. Here, we keep the approach simple and do not enrich the DoE.

The indicator function can then be used to approximate the eigenvalue in each region using a different gPC expansion. Optimal polynomial order for each expansion is obtained from cross-validation (Leave-one-Out technique) ; the classical gPC expansion has been truncated at order 6 whereas green and yellow surfaces have been truncated at order 2 and 4. With a regression on each part of the surface response, we

can obtain the coefficients of the two expansions $(\boldsymbol{a}_{\boldsymbol{\gamma}}^{(1)})_{\boldsymbol{\gamma}\in\boldsymbol{\Lambda_p}}$ on $\mathcal{R}_1$ and $(\boldsymbol{a}_{\boldsymbol{\gamma}}^{(2)})_{\boldsymbol{\gamma}\in\boldsymbol{\Lambda_p}}$ on $\mathcal{R}_2$. Then, we can lump the information into a single expression for the metamodel :

$$\widehat{\Im[k]}(\xi_1,\xi_2) = (2-\hat{I}(\xi_1,\xi_2))\sum_{\boldsymbol{\gamma}\in\Lambda_p} a_{\boldsymbol{\gamma}}^{(1)} H_{\boldsymbol{\gamma}}(\xi_1,\xi_2) + (\hat{I}(\xi_1,\xi_2)-1)\sum_{\boldsymbol{\gamma}\in\Lambda_p} a_{\boldsymbol{\gamma}}^{(2)} H_{\boldsymbol{\gamma}}(\xi_1,\xi_2). \tag{3.23}$$

Exactly the same technique can be used for the real part and the eigenfunction, the indicator function remaining the same, as it is set using global consideration on the mode behaviour (based on the altitude of refraction).

This technique is essential on the case of long range propagation for which modes frequently switch due to the variability of the stratospheric jet. Taking into account the effect of this variability is essential to restitute the statistics of the different arrivals of signal propagating on large distances (see 3.A).

4.4 gPC representation of the coupling coefficients

We have presented in the previous paragraphs of this section the polynomial decomposition of the acoustic modes. Thanks to this decomposition we are able to simulate efficiently the acoustic propagation in a ABL with a perturbation of great amplitude. On the top of this simulation we want in this paragraph to take into account the perturbation of the propagation due to the small-scale structures of the turbulent flows modelled by the process $\mu(z)$. As explained in section 3.1, thanks to the small amplitude of μ, the contribution can be added as a coupling effect through a coupling coefficient C_{nl}.

The coupling coefficients C_{nl} depend on the realization of the profile $c_0(z,\xi)$ and thereby on the random parameter ξ. It also depends on the realization of $\mu(z)$, the small scale perturbation. We explain here how gPC decomposition of the eigenfunctions can be used to compute the coupling coefficient. This means that there is not any expensive computation to do to take into account the small scale structures on top of a large scale perturbation.

In order to compute those coefficients, we introduce the Karhunen-Loève decomposition of $\mu(z)$:

$$\mu(z) = \sum_{i=1}^{M} X_i \sqrt{\lambda_i} e_i(z), \tag{3.24}$$

where $(X_i)_{i=1,..,M}$ are gaussian centered independent random variables, $(\lambda_i, e_i)_{i=1,..,M}$ are the eigenvalues and eigenfunctions of the covariance kernel. We suppose here that $\mu(z)$ is a centered process that is why there is no mean term. In the case of Ornstein-Uhlenbeck process (whose covariance is given by (3.4) the Karhunen-Loève decomposition can be computed analytically.

By definition (3.10), the coupling coefficients C_{nl} depends on μ and on the eigenfunctions $\phi_{n0}(z,\xi)$ and $\phi_{l0}(z,\xi)$. Then, the gPC expansion of the eigenfunctions $\phi_{n0} = \sum_{k=0}^{+\infty} \alpha_k^{(n)}(z) H_k(\xi)$ can be use to compute C_{nl} (we drop the dependence in ξ and

z from third line for convenience) :

$$C_{nl} = \int_0^\infty \frac{\mu(z)\phi_{n0}(z,\xi)\phi_{l0}(z,\xi)}{c_0^2(z,\xi)}\,\mathrm{d}z, \tag{3.25}$$

$$= \int_0^\infty \frac{1}{c_0^2(z)} \sum_i X_i \sqrt{\lambda_i} e_i(z) \sum_j \alpha_j^{(n)}(z) H_j(\xi) \sum_k \alpha_k^{(l)}(z) H_k(\xi)\,\mathrm{d}z, \tag{3.26}$$

$$= \int_0^\infty \frac{1}{c_0^2} \sum_{i,j,k} \sqrt{\lambda_i} e_i \alpha_j^{(n)} \alpha_k^{(l)} X_i H_j H_k\,\mathrm{d}z, \tag{3.27}$$

$$= \sum_{i,j,k} \beta_{i,j,k} X_i H_j H_k, \tag{3.28}$$

where

$$\beta_{i,j,k} = \sqrt{\lambda_i} \int_0^\infty \frac{e_i \alpha_j^{(n)} \alpha_k^{(l)}}{c_0^2}\,\mathrm{d}z. \tag{3.29}$$

This expression gives a way of computing the coupling coefficients without any supplementary cost, using the gPC coefficients of the eigenfunctions.

5 Applications of gPC metamodelling

In this section, we will use the case of sound propagation in an atmospheric boundary layer to illustrate the use of gPC expansion in the context of atmospheric infrasound propagation. We will also provide some comparison with statistics obtained by Monte-Carlo simulations to numerically validate the convergence of our surrogate model.

5.1 Sensitivity based eigenvalues selection

For the acoustic modes, we compute the gPC expansions of the eigenvalues and of the eigenfunctions at the ground. A first application of the metamodel concerns the selections of acoustic modes depending on their acoustic contribution. In [Bertin et al., 2014], it has been shown that the modes contributing the most are those with higher sobol' index. This property opens the way to model reduction by considering only the most important modes as explained in the article. For this reason, the computation of Sobol' indices of the modes are a key quantity we can easily compute thanks to the metamodel.

The acoustic modes depend on the random input parameters ($\xi = (a, \sigma)$ amplitude and spread) but also on the frequency ω. For low frequencies a cut-off appears naturally, defined by $\omega_{co} = \Re[k] \times \max c(z)$; the analytic continuation of the mode can be defined under this frequency but has no physical interpretation. We have used this analytic continuation to follow the modes on the entire range of frequency and avoid thereby any discontinuity which would have limited the performance of gPC. Before using those expansions to generate acoustical quantities such as Green functions and signals, we first use the expansions to compute variances of each mode and compute their sensitivity to the two inputs. To do so we have used the Sobol' indices

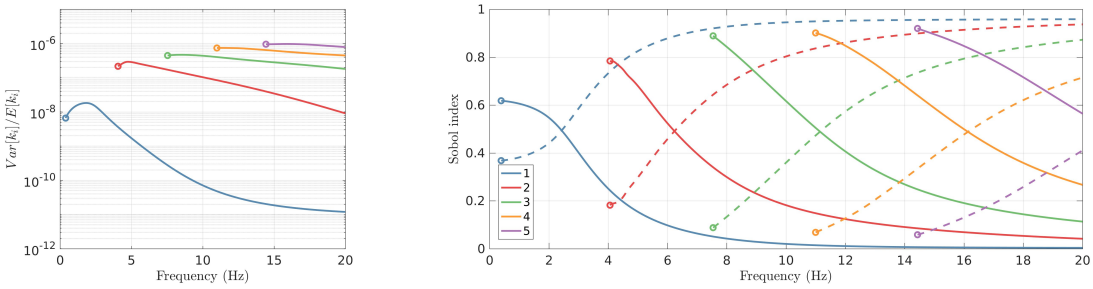

FIGURE 3.8 : Variance (left) and Sobol' indices (right) associated to the real part of the eigenvalues. Those quantities are represented for frequencies higher than the cut-off frequency (i.e. when the phase velocity is higher than 365m.s^{-1}). Solid (dashed) lines refer to the amplitude (spread) parameter.

which have already been used in [Bertin et al., 2014] in a slightly different context. The Sobol' indices represent the sensitivity of the quantity of interest (in terms of variance contribution) to the different inputs. It consists in decomposing the variance of the quantity of interest (the acoustic modes in our case) into the sum of the variances caused either by a single input or by the coupling between several inputs. Mathematically, the first-order Sobol' indices (assessing the impact of a single input on the variance of the output) is defined as the conditional expectation of the output conditionnaly to a given input (3.32). A key property of the gPC is their ability to give a direct evaluation of those Sobol' indices which are not straightforward to assess otherwise. In fact, the variance of the output is given by the l^2-norm of its gPC coefficients (3.31) and the Sobol'indices are obtained by summing only the coefficients before a polynomial function of the considered input. In our case for each acoustic mode $k_n, n = 1, .., 5$ whose expansion writes (3.30), the mean, variance and Sobol' indices are given by :

$$k_n = \sum_{\boldsymbol{\gamma} \in \Lambda_p} a_{\boldsymbol{\gamma}} H_{\boldsymbol{\gamma}}(\xi), \tag{3.30}$$

$$\mathbb{E}[k_n] = a_0,\ Var(k_n) = \sum_{\boldsymbol{\gamma} \in \Lambda_p \setminus \{0\}} a_{\boldsymbol{\gamma}}^2, \tag{3.31}$$

$$S_i = \frac{Var(k_n|\xi_i)}{Var(k_n)} = \sum_{\boldsymbol{\gamma} \in \mathcal{S}_i} a_{\gamma}^2, \tag{3.32}$$

where $\mathcal{S}_i \subset \Lambda_p$ is the set of indices such that such that $H_\gamma(\xi)$ depends on ${\xi_i}^2$.
Because the Sobol' indices are normalized by the variance of the output it is important to look at the variance of each mode in order understand the range of frequency for which the impact is more important. The variance is again easy to assess using the gPC expansion of the given mode (3.31). Figure 3.8 shows in a logarithmic scale the variance of each mode. The first thing to see is the respective arrangement of the modes : the first mode has a smaller variance than the other modes for all frequencies

[2]For a detailed description of the Sobol' indices of higher order, see Chapter 1, section 4.1.

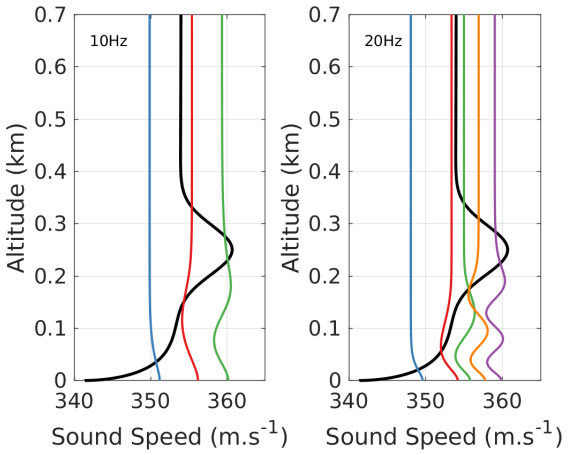

FIGURE 3.9 : Eigenfunctions centered on the phase velocity of the mode at two frequencies : 10Hz on the left and 20 Hz on the right. There exist only three modes at 10Hz because we are under the cutoff frequency of the two others.

where they are defined. This observation is coherent with the location of the maximum of variance at a frequency of 2.5Hz that corresponds to a range of frequencies for which the mode interact with the jet whereas it is confined in the boundary layer for higher frequencies (see figure 3.9).

An important factor when looking at the Sobol' indices is the variation coefficient of the inputs. In fact, it has to be of the same order for each input otherwise it would artificially enlarge the sensitivity due to the parameter. In our case the standard deviation of our normal laws has been chosen at 20% of the mean for the amplitude and 15% for the spread. After this verification, we can look at figure 3.8 where the evolution of the Sobol' is presented. Each mode is plotted in a given color and the dashed line represents the Sobol index relative to the spread whereas the continuous line represents the Sobol index relative to the amplitude. It appears clearly that all the modes have the same behaviour with a shift in frequency : for low frequencies (here low frequencies means just above the cut-off) the amplitude of the jet is the most influential factor whereas the spread gets more important for higher frequencies. Again, this behaviour can be justified by looking at the interaction with the profile : the amplitude of the jet has almost no impact on the boundary layer but a huge one on the jet. On the contrary, the spread has an influence on the form of the sound speed profile around the jet but also at lower altitudes. As a consequence, when the mode is trapped under the boundary layer (for high frequencies), it is more sensitive to the spread whereas the amplitude of the jet gets more influential at low frequencies when the mode is located around the jet.
At last, it is interesting to note that the sum of the two Sobol' indices is close to one. This means that there is almost no interaction effect between the two variables. In fact, interaction effects are measured by the second order Sobol' indices that are equal to $1 - S_1 - S_2$.

The key element to understand the evolution of the variance and Sobol' indices

shown on figure 3.8 is the interaction of each mode with the different parts of the profile. In fact, we have used a very localized variability (it affects only the jet) and the different modes account for contribution to the propagation of the different part of the profile. To know the part of the profile interacting with the mode the key element is the phase velocity (defined as $v_\varphi = \omega/\Re[k]$). Figure 3.9 shows this phenomenon : the eigenfunctions are plotted centered at the phase velocity of each mode at two frequencies : 10Hz and 20Hz. The phase velocity decreases with the frequency, which means that modes are interacting with the boundary layer for high frequencies and to the jet for smaller frequencies before the cut-off. Then, the two random parameters a and σ do not have the same impact on the profile. While a modifies essentially the extremity of the jet, σ deforms the profile around the junction between the jet and the boundary layer. Those two facts explain the shape of the Sobol' indices on figure 3.8. For instance, the third mode (in green) is almost only sensitive to the spread at 20Hz whereas both amplitude and spread have an impact on this mode at 10Hz and the amplitude is dominating for lower frequencies.

Concerning model reduction, the conclusion of this numerical example is the dependance of the Sobol' indices on the frequency range. In practice, the frequency range is determined by the support of the Fourier transform of the source ; the multiplication with the sum of residuals will indeed remove all the effects occuring out of this range. For example, if we look at figure 3.8, someone studying the impact of the amplitude of the jet around 15Hz can focus on modes 4 and 5 whereas mode n^{o}2 will be the most important for someone looking at 5Hz.

5.2 Computing signal statistics

We have explained in chapter 2 the problem of long term integration for reproducing random signals with a gPC metamodel. To circumvent this issue we have decided to build directly a metamodel on the modes and, then, use the metamodel of the modes to generate random signals. We give in this paragraph a numerical illustration of this strategy, showing random signals whose quality does not depend on the distance of propagation. To demonstrate the efficiency of our method we give a comparison of statistics of signals generated using the metamodel with statistics from direct Monte-Carlo simulations. For Monte-Carlo simulations the modes have been computed for every simulation which led to much longer computations, affordable thanks to the reduced height of the domain.

Before computing the signals one must deal with the complex part of the spectrum induced by the imaginary part of ω. The spatial eigenvalues $k_j(\omega)$ have been introduced as solutions of the dispersion relation when the contour $\mathcal{F}$ in the complex ω-plane differs from the real axis. Although a straightforward interpretation is missing, the temporal branches $k_j(\omega)$ are objects that are naturally involved whenever the initial contour is gradually displaced upward from the real axis. According to the frequency shifting property of the Fourier-Laplace transform, the ground-based signal can be obtained from the classical FFT algorithm (for $\omega = \omega_r$) together with the mapping $p(t) \mapsto e^{\omega_i t}$.

If the computation of the complete signal is based on the Fourier transform of

$p(\omega)$ given at eq. (3.18) multiplied by the Fourier transform of the source. Here, we use a simplified source model defined by :

$$s(t) = \frac{1}{2}\sin\left(\omega_s(t-3\sigma_s)+\pi/2\right)\exp\left(-\left(-\frac{t-3\sigma_s}{\sigma_s}\right)^2\right), \tag{3.33}$$

with $\omega_s = 2\pi f_s$ where $f_s = 17Hz$ is the central frequency of the source and $\sigma_s = 0.35s$ sets the length of this source. The Fourier transform of this source is given by the black line on the first row of figure 3.4, its support is included between 15 and 20 Hz. We have chosen this type of source to be able to select a desired range of frequency and decided to set $f_s = 17Hz$ to have five modes contributing to the signal (the different contributions vanish under the cut-off frequency of each mode). We can also compute the signal associated with only one mode by limiting the sum to $p_n(\omega)$, the resulting signal $p_n(t)$ is called the n^{th} wavepacket.

Once reliable gPC expansions of eigenvalues/eigenfunctions at the ground have been computed, they can be directly used to generate pressure signals. In fact, for any random scenario ξ within the variability range, rather than solving the eigenvalue problem for this celerity profile, we only need to inquire our surrogate model in order to predict the appropriate spectrum utilized to produce the signals at the ground for a given source and at a given distance.

Figure 3.10 shows the signals and their variances at three distances computed either with the gPC expansions (in orange) and by computing the acoustic modes (in blue). The source model for this simulation is given by equation (3.33) and for the metamodel the gPC expansions has been used with an order 5. The very good agreement validate the ability of the metamodel to produce signals with the statistics implied by a certain perturbation of the medium. The very small number of runs used to calibrate the metamodel (49 computations of acoustic modes) compared to Monte Carlo simulations (5000 computations) advocates for the use of the metamodel for statistical studies. In order to assess the convergence of our surrogate model we have made the effort of generating with the full solver a validation dataset of solutions for 5000 randomly selected inputs. It is then possible, for comparison purposes, to generate from the surrogate model the same number of samples. A key property of our metamodel is its ability to reproduce signal statistics for any source and at any distance and figure 3.10 presents comparison of standard deviation for different distances. Chosen measure of discrepancy is a normalized mean quadratic error :

$$\text{err} \equiv \frac{1}{\max_{t,i}|p_i^{MC}(t)|^2}\max_t\left(\frac{1}{N}\sum_{i=1}^{N}|p_i^{gPC}(t)-p_i^{MC}(t)|^2\right), \tag{3.34}$$

where $p_i^{(\cdot)}(t)$ is the i^{th} wavepacket realization, evaluated either with the gPC expansion or with the direct solver. We have computed this measure for each mode and for different polynomial degrees. gPC results are accurate for degree ≥ 2, and as expected the surrogate model gives better results for higher degree. Spectral convergence is recovered for most cases as figure 3.11 shows. We have represented the convergence for the different wavepackets (i.e. the signal associated with only one mode). This

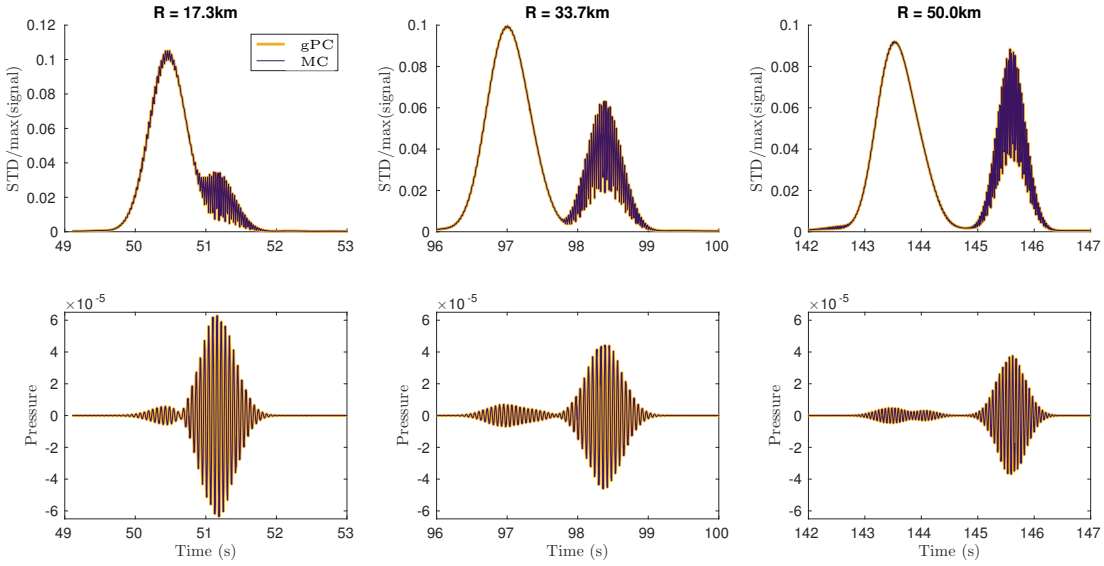

FIGURE 3.10 : Signals produced either by the metamodel (in orange) and the direct simulation (in black) at different distances.The first line gives an estimation of the standard deviation at different distances while the second line gives an example of signal with the same time scale.

figure 3.11 confirms the spectral convergence of the gPC expansions and justifies the very good agreement shown on figure 3.10 concerning the signals generated using those expansions.

6 Conclusion

To simulate numerically the acoustic propagation through a turbulent boundary layer taking into account the inhomogeneities of different amplitudes, we have presented an approach separating the different scales. The modal decomposition of the acoustic field allows to circumvent the problem of long term integration that arises when studying random oscillatory phenomena. It also provides an easy way to take into account perturbations of small amplitude on top of the large ones using perturbative approach.

We have illustrated our method with the case of an atmospheric boundary layer with a random jet and shown the convergence of the metamodel on the signals and their variance. Our method enables reducing drastically the numerical cost in comparison with Monte Carlo estimation : instead of 5000 simulations, we have been able to compute the variance of the temporal signal with only 49 simulations. The main limitation of this method comes from the necessity to follow the modes as functions of the frequency but also as functions of the random parameters. In the case of complex perturbations or in high dimension, this task can reveal very challenging.

To illustrate the usefulness of this kind of metamodels, we show in the next chapter how it can be used to localize a source. To do so we have used Quantities of Interest

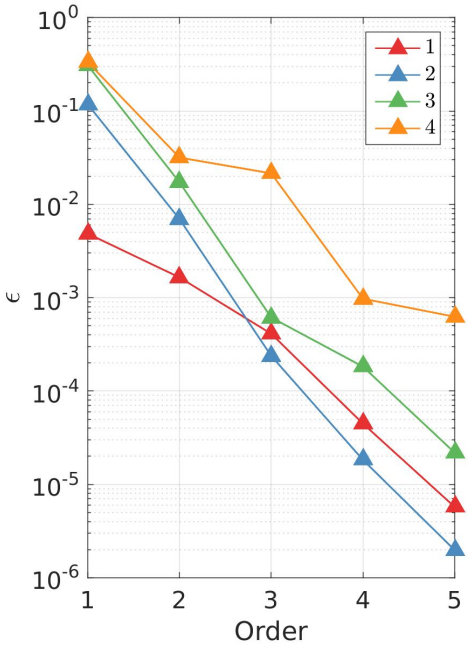

FIGURE 3.11 : Mean square convergence error of the different wavepackets, cf. Eq. (3.34) with $N = 5000$. The results are displayed for different gPC degrees (horizontal axis). Spectral convergence is verified.

extracted from the signals and do not need to restitute entire signals. For this reason, we have considered a slightly different approach and computed directly the gPC expansion of the extracted quantites ; this avoids us a complex tracking.

However, in a high dimension problem where one would want to restitute the entire signal our technique would be limited by the numerical tracking of the acoustic modes. In this case, machine learning techniques provide an interesting framework to improve the tracking and perhaps adapt our algorithm to more real-world cases.

Appendix:

3.A gPC mixture for switching modes in the case of long range propagation

In this chapter we have shown that modes can switch as perturbation varies. This causes a change of regime which results in a discontinuity on the surface response (i.e. the surface described by the mode as (ξ_1, ξ_2) vary). We have explained our method in paragraph 4.3. We show in this appendix a numerical application in a slightly different case : the impact of the variability of the stratospheric jet on the propagation of infrasound over large distances. This is another framework where large deviations of the wind flow occur and have to be simulated numerically to reproduce statistics of the propagated signals.

The stratospheric jet has a decisive impact on the long distance propagation. Due to its randomness, small stratospheric jet changes can lead to very different regimes that are difficult to reproduce by a global approximation. Here, we propose a method to adapt the metamodeling in order to accurately capture this sensitive change of behaviour. We will illustrate our approach with a theoretical atmosphere, where the nominal effective sound speed profile is described by the sum of four gaussian functions :

$$c(z) = \sum_{i=1}^{4} A_i \exp\left(-\frac{(z-\mu_i)^2}{2\sigma_i^2}\right), \tag{3.35}$$

with $(A_i)_{i=1..4} = (-60, 70, -70, 300)$, $(\mu_i)_{i=1..4} = (25, 40, 60, 200)$km and $(\sigma_i)_{i=1..4} = (20, 15, 20, 65)$km. A random perturbation, also defined in terms of a gaussian function centered in $\mu_0 = 40$km, with amplitude $0.03 \times \xi_1$ and $\sigma_0 = 10 + \xi_2$km with (ξ_1, ξ_2) two independant gaussian random variables, is added to the profile.

In order to build a continuous metamodel of the propagation through this medium, we compute the acoustic modes at a frequency of 1Hz for a random sample of 200 realizations of (ξ_1, ξ_2). On figure 3.12 we show the mean and standard deviation of the gaussian process conditionally to the 200 points computed using the original model. We also show the changes in the mode behaviour which can be either stratospheric or thermospheric : cf. corresponding eigenfunction on figure 3.12, and this is confirmed by the clustering of the response depending on the values taken by the random perturbation (ξ_1, ξ_2). Figure 3.12 shows a global gPC approximation in blue and the two element gPC in green and yellow. The blue surface exhibits the Gibbs-like oscillation leading to a poor quality of the approximation. Regarding the performance, we can compare the statistics of the original sample with those given by the two metamodels evaluated on a sample of 20 000 points (see Figure 3.13). If the main statistics are globally the same, the number of extreme values is greatly reduced with our new approach.

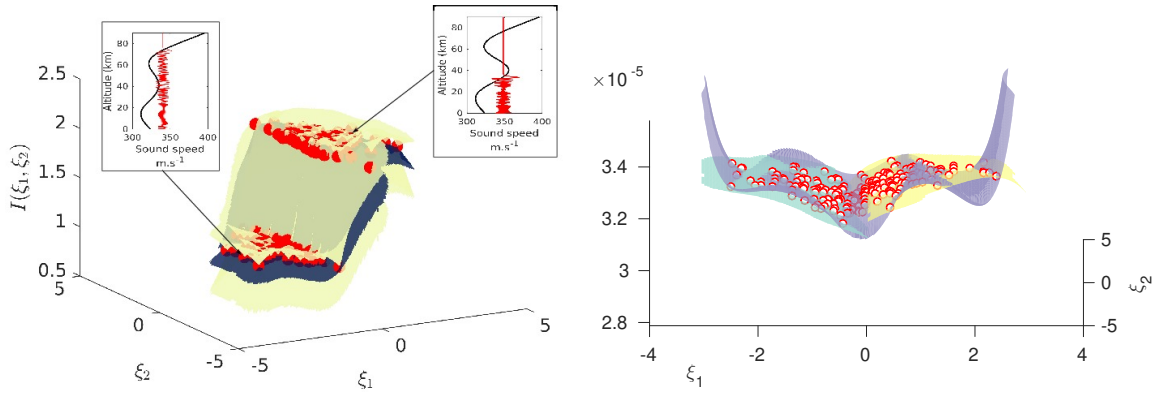

FIGURE 3.12 : Metamodel of the indicator function (on the left) and the response surface on the right with a classical gPC (in blue) and a two elements gPC (in green and yellow) for the imaginary part of an eigenvalue.

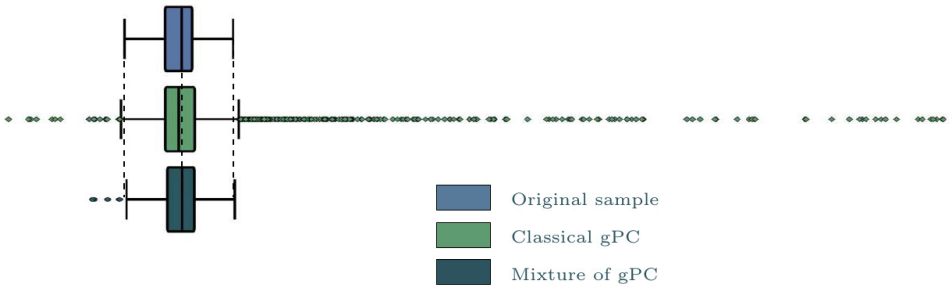

FIGURE 3.13 : Comparison of the classical gPC regression with optimal order selection and mixture of gPC

Chapitre 4

Metamodeling for Propagation Based Source Localization

Throughout this manuscript we have been concerned with constructing and examining gPC models of infrasound propagation, with the aim of computing waveforms at the lowest possible CPU cost. Such models are valuable for processing data from the International Monitoring System (IMS), which includes waveform sensor stations. According to the classification proposed by [Mialle et al., 2019], the problem is generally divided in four main steps : detection (analysis of raw sensor data to detect and classify signal arrivals at each station), association (of signals from different stations that have presumably detected the same event), localization and characterization. The first two steps are far from trivial and recently, they have been formulated using Bayesian inference, using discrete detections as observations [Mialle et al., 2019, Mialle and Arora, 2018, Arora et al., 2013] or waveforms [Moore and Russell, 2017]. In this chapter we focus on the localization problem, using an approach that bridges the gap between the Bayesian framework [Evans and Stark, 2002, Kaipio and Somersalo, 2006, Mohammad-Djafari, 1996, Tarantola, 2005] for solving the inverse problem and metamodeling [Marzouk et al., 2007, Marzouk and Najm, 2009, Marzouk and Xiu, 2009] for updating the waveforms from the atmospheric state at the time of the event.

1 Introduction

In the last ten years, a number of methods have been developed to combine infrasonic observations with the aim of estimating a location and time for infrasonic source events [Ceranna et al., 2009, Modrak et al., 2010, Blom et al., 2015]. In most source location methodologies, arrival time and backazimuth (i.e. the azimuth from which the observed signal appears to be generated) are combined in a manner which produces an estimate of the source location and time. The Bayesian Infrasonic Source Localization (BISL) framework proposed by [Modrak et al., 2010] and applied by [Park et al., 2014] uses Bayes' theorem to combine detections across a spatially separated infrasound network in order to estimate a spatial and temporal distribution describing

the source event. The BISL method has been updated by [Blom et al., 2015] using tools employed in directional statistics as well as a stochastic treatment of infrasonic propagation. Bayesian inference has also been shown to be useful for solving the characterization problem [Blom et al., 2018]. Since theses methods primarily rely, at their core, on minimization of the difference between observed and predicted quantities of interest (QoI), the common thread through all these inverse problems is the need for propagation models.

A propagation model can be described as a map $\mathbf{F} : \mathbf{X} \mapsto \mathbf{Y}$ between a multi-dimensional input vector $\mathbf{X}$ describing the source and at least one QoI, which is the output vector $\mathbf{Y}$ of the model. Except the classification tasks, the output $\mathbf{Y}$ takes continuous values as $\mathbf{X}$ is varied. The usual statistical approaches for estimating $\mathbf{Y}$ are based on either generative models [Moore and Russell, 2017, Arora et al., 2013] or stochastic propagation models [Blom et al., 2018, Blom et al., 2015, Marcillo et al., 2014]. These models can be constructed by numerically propagating infrasound through a large suite of atmosphere specifications and defining a distribution for arrival characteristics from the combined results. Using a pre-defined function from which arrivals are expected to be realized [Marcillo et al., 2014, Morton and Arrowsmith, 2014] increases the efficiency of the procedure without repeatedly updating the atmospheric state on a case-by-case basis. But such stochastic models do not include current atmospheric specifications knowledge and additional analysis is generally necessary to better refine the source location estimate. Alternately, atmospheric specifications at the time of an event of interest can be perturbed to improve agreement between infrasonic observations and propagation modeling predictions [Lalande et al., 2012, Arrowsmith et al., 2013, Assink et al., 2013, Fricke et al., 2014]. Such methods typically use empirical orthogonal functions (EOFs) constructed from a sequence of atmospheric specifications at a given location and compute the coefficients of the EOFs which correspond to the atmosphere state most consistent with observations. However, estimating the impact of a wide range of EOFs on the statistics of $\mathbf{Y}$ is computationally demanding. Since many thousands of detections per day are recorded by national data centers, this approach is not suited for continuous, real-time monitoring.

The approach set forth in this chapter for solving the localization problem is based on a combination of Bayesian inference and metamodeling, which allows to replace an original propagation model with an inexpensive-to-run but accurate metamodel. Examples of approaches combining metamodeling and Bayesian inference have been published by [Marzouk et al., 2007, Marzouk and Najm, 2009, Marzouk and Xiu, 2009] and a few others. In the present study we use polynomial chaos expansions (PCE), which can be viewed as an interpretable machine learning regression algorithm [Vapnik, 1995], to predict $\mathbf{Y}$ given the location $\mathbf{x}_0$ and time t_0 of the infrasound event. PCE can be used to derive estimates of various statistics of $\mathbf{Y}$ such as its moments and its sensitivities to different components of $\mathbf{X} = (\mathbf{x}_0, t_0)$ [Saltelli et al., 2000]. The simulation data (i.e. the set $(\mathbf{X}, \mathbf{Y})$ of input values and corresponding responses) are produced by a normal-mode-based model which belongs to the class of the so-called full-wave models. Unlike high-frequency propagation models in which expertise plays an important role in the interpretation of results, full-wave models produce waveforms from which a large variety of QoI can be extracted using signal

processing. It is worthwhile to point out that, under fixed computing budget, normal mode expansions can be truncated using model reduction algorithms [Bertin et al., 2014] so as to reduce the size of involved matrices in the numerical solving method.

The purpose of this chapter is to show that (i) a metamodel purely trained on few simulated data yields reliable probability distributions of location and time of real-world infrasound events, and (ii) the Bayesian inference problem can be reformulated to incorporate both simulated data and uncertainties associated with extracting the QoI (signal processing, arrival time picking, etc.). While for simplicity the QoI are restricted to a single arrival time for each infrasound station, the generalisation to multivariate outputs is straightforward. Another motivation for the present work is to show that the event location can be updated using all the available knowledge on the atmospheric state, including subgrid-scale stochastic processes (e.g., gravity waves, turbulence, etc.). This is especially important since stochastic parameterizations [Berner et al., 2017] are an essential component of the models currently in use in the operational weather centers, and neglecting this random component is known to produce biases.

The chapter is organized as follows. Section 2 is dedicated to an overview of the underlying theory. In section 3, this technique is applied to the localization of a bolide that caused a huge explosion over the Bering sea, near Russia's Kamchatka Peninsula, on December 18th, 2018. The concluding section discusses applications and possible extensions of the approach.

2 PCE-based Bayesian localization

2.1 The Bayesian framework

Localizing a point-like source occurring in the atmosphere consists in finding the latitude x_0, longitude y_0 and altitude z_0 characterizing the source position from detections $(\mathbf{D}_j)_{j\leq n}$ recorded at n stations. For impulsive sources, it is often of interest to represent the origin time t_0 as an independent unknown variable. In the Bayesian setting $\mathbf{X} = (x_0, y_0, z_0, t_0)$ and $\mathbf{D} = (\mathbf{D}_1, \mathbf{D}_2, \ldots)$ are random vectors. Bayesian inference derives the posterior probability density function $\mathrm{P}(\mathbf{X}|\mathbf{D})$ as a consequence of two antecedents : a prior probability $\mathrm{P}(\mathbf{X})$ and a "likelihood function" describing the probability of observing $\mathbf{D}$ under the assumption that $\mathbf{X}$ is valid. Bayesian inference computes the posterior probability according to Bayes' theorem, which reads as

$$\pi\left(\mathbf{X}\right) = \mathrm{P}\left(\mathbf{X}|\mathbf{D}\right) = \frac{\mathrm{P}\left(\mathbf{D}|\mathbf{X}\right)\mathrm{P}\left(\mathbf{X}\right)}{\mathrm{P}\left(\mathbf{D}\right)}. \tag{4.1}$$

In equation (4.1), the term $\mathrm{P}\left(\mathbf{D}\right)$ is often considered as a normalization constant, $\mathrm{P}(\mathbf{X})$ is the prior probability (i.e. the estimate of the probability of $\mathbf{X}$ before the data $\mathbf{D}$) and $L\left(\mathbf{X}\right) = \mathrm{P}\left(\mathbf{D}|\mathbf{X}\right)$ is the likelihood function.

The likelihood function relating the detections with the source hypothesis can

then be specified by the product $\mathrm{P}(\mathbf{D}_1|\mathbf{X})\mathrm{P}(\mathbf{D}_3|\mathbf{X})\ldots\mathrm{P}(\mathbf{D}_n|\mathbf{X})$, with

$$\mathrm{P}(\mathbf{D}_j|\mathbf{X}) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left[-\frac{(D_j - F_j(\mathbf{X}))^2}{2\sigma_j^2}\right] \tag{4.2}$$

for $j = 1, \ldots, n$, where $F_j(\mathbf{X}) \in \mathbb{R}$ is considered as a single-valued function. Hence, the simulated data F_j can be concatenated in a single vector $\mathbf{F} \in \mathbb{R}^n$ containing the infrasonic propagation characteristics and the standard deviations σ_j can be adjusted for the different stations. It should be noted that the normal distributions in (4.2) can be replaced by other distributions, depending on the model used to compute $\mathbf{F}(\mathbf{X})$. Further, when several quantities of interest (e.g., arrival time, duration, amplitude, etc.) are incorporated into $\mathbf{D}_j$ or when multiple infrasound arrivals are detected, equation (4.2) should be adapted to a multidimensional framework. Here, and in the following, the hypothesis parameter space is defined as the spatial source description and each detection consists of a single observed arrival time.

The posterior probability density function (PDF) produced via the Bayesian approach can be analysed by computing a number of characteristics of the distribution. The maximum *a posteriori* (MAP) solution is often used as an estimate of the event location [Blom et al., 2015]. This location, which is defined by those parameter values that maximize the posterior PDF, corresponds to the best guess for the source localization given the set of detections. It is referred to as $\mathbf{X}_{\mathrm{MAP}}$ in the following. Other characteristics can be computed from the posterior PDF, such as confidence ellipses or likelihood ratios. The way these ratios are used to assess the goodness of fit of two different competing localizations is detailed in section 3.3.

2.2 Simulation-driven likelihood

For each station, computing the probability $\mathrm{P}(\mathbf{D}_j|\mathbf{X})$ reduces to computing infrasonic propagation characteristics under the assumption that the source location is specified by $\mathbf{X}$. The predicted arrival time for station $\#j$ can be decomposed into

$$F_j(\mathbf{X}) = t_j(x_0, y_0) + t_0, \tag{4.3}$$

where t_j is the travel time, t_0 is the origin time, and the altitude z_0 of the source is dropped for simplicity. The travel time depends on the location of the station, which is fixed and denoted by $\mathbf{x}_j = (x_j, y_j)$. The most commonly used estimate of t_j is obtained from

$$t_j = \frac{||\mathbf{x}_0 - \mathbf{x}_j||}{c_j}, \tag{4.4}$$

where c_j is the celerity and $\mathbf{x}_0 = (x_0, y_0)$ is obtained from sampling the prior distribution $\mathrm{P}(\mathbf{x}_0, t_0) = \mathrm{P}(\mathbf{x}_0)\mathrm{P}(t_0)$, for a fixed t_0. Provided a suitable model for c_j, and using m source locations, the corresponding arrival times can be readily calculated from (4.4) so as to obtain a vector $\mathbf{t}_j$ of size $m \times 1$. The celerity c_j can be specified by a statistical distribution (i.e. the probability of observing specific celerities) via tapered box-car distributions [Blom et al., 2015]. Propagation modeling methods can also be used to construct catalogues of propagation predictions at specific geographic

locations using historical atmospheric models [Marcillo et al., 2014, Morton and Arrowsmith, 2014, Nippress et al., 2014, Blom et al., 2015]. However, the validity of these models is restricted to the training data and thus, these models are not suited to the study of a new particular atmospheric state, at a particular time.

The limitation inherent to statistical propagation models can be avoided by constructing a polynomial approximation of (4.3) over the support of the prior distribution [Marzouk and Xiu, 2009]. This approximation then defines a surrogate posterior probability density $\widetilde{\pi}(\mathbf{X})$ that can be estimated from a set $\{(\mathbf{x}_{0i}, t_{ji}), i = 1, \ldots, m\}$ of simulated data. If we denote $\{H_{\mathbf{k}}(.), \mathbf{k} \in \mathbb{N}^d\}$ the polynomial chaos basis, the arrival time can be represented as

$$\widetilde{F}_j(\mathbf{X}) = \sum_{\mathbf{k} \in \mathbb{N}^2} a_{j\mathbf{k}} H_{\mathbf{k}}(x_0, y_0) + t_0, \tag{4.5}$$

or, equivalently $\widetilde{F}_j = \widetilde{t}_j + t_0$, with

$$\widetilde{t}_j(\mathbf{x}_0) = \sum_{k=1}^{|\Lambda|} a_{jk} H_{\mathbf{k}_k}(\mathbf{x}_0), \tag{4.6}$$

where the coefficients a_{jk} can be computed either by projection or regression, using the orthogonality property[1]. Without any loss of generality, we use approximation spaces of total degree and introduce, for this purpose, the index set $\Lambda = \{\mathbf{k} \in \mathbb{N}^2 : ||\mathbf{k}||_1 \leq d\}$. Using the regression method, the coefficients of the expansion are determined from the m simulated arrival times t_{ji} by solving the ordinary least square problem

$$\mathbf{a}_j = \arg \min_{\mathbf{a}_j \in \mathbb{R}^p} ||\mathbf{t}_j - \mathbf{H}\mathbf{a}_j||_2, \tag{4.7}$$

where the k-th component of $\mathbf{a}_j$ is given by a_{jk} and

$$||\mathbf{t}_j - \mathbf{H}\mathbf{a}_j||_2 = \frac{1}{m} \sum_{i=1}^{m} \left[t_{ji} - \sum_{k=1}^{|\Lambda|} a_{jk} H_{\mathbf{k}_k}(\mathbf{x}_{0i}) \right]^2, \tag{4.8}$$

with $|\Lambda| = p = (2 + d)!/(2!d!)$, where d is the chosen maximum polynomial degree. When the random vector $\mathbf{x}_0$ has a bivariate normal distribution with a diagonal covariance matrix, the polynomial chaos basis can be obtained from the tensor product of univariate Hermite polynomials. If the input vector $\mathbf{x}_0$ has statistically dependent components, constructing a basis of orthogonal polynomials is still possible, but computationally demanding. For this reason, and because there is no reason why x_0 and y_0 should be correlated, we use the assumption $\mathrm{Cov}(x_0, y_0) = 0$ in the following.

Once the coefficients a_{jk} have been computed, the travel time can be evaluated at arbitrary values of $\mathbf{x}_0$ without resorting to additional simulations. It turns out that solving the least square problem requires $m > p$. While the regression method tends

[1]The projection method consists in computing directly the scalar product whereas the regression method uses the fact that orthogonal projection realizes the distance.

to overfit the simulated data for large $p = |\Lambda|$, simpler models with fewer coefficients can be constructed by sparse regression. Solving (4.7), the solution reads as $\mathbf{a}_j = \mathbf{G}\mathbf{t}_j$, where the matrix G is defined by $\mathbf{G} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}$, where $H_{ik} = H_{\mathbf{k}_k}(\mathbf{x}_{0i})$. Then, the posterior PDF can be split into

$$\widetilde{\pi}(\mathbf{x}_0, t_0) = \prod_{j=1}^{n} \widetilde{\pi}_j(\mathbf{x}_0, t_0), \tag{4.9}$$

where the posterior for station $\#j$ is

$$\widetilde{\pi}_j(\mathbf{x}_0, t_0) = \frac{\mathrm{P}(\mathbf{x}_0)\mathrm{P}(t_0)}{\mathrm{P}(\mathbf{D})}\mathrm{P}(D_j|\mathbf{x}_0, t_0), \tag{4.10}$$

with

$$\mathrm{P}(D_j|\mathbf{x}_0, t_0) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_j^2}\left(\mathbf{G}\mathbf{t}_j\mathbf{H}(\mathbf{x}_0) + t_0 - D_j\right)^2\right], \tag{4.11}$$

where the components of vector $\mathbf{H}(\mathbf{x}_0)$ are defined by $H_{\mathbf{k}_k}(\mathbf{x}_0)$ for $k = 1, \ldots, p$.

Given a set of simulated arrival times, the present approach capitalises on polynomial chaos expansion to produce a new version of the likelihoood, where the computational model $F_j(\mathbf{X})$ is replaced with an inexpensive metamodel that retains the statistics of the output. For a given station and fixed t_0, it can be shown that the rate of convergence of $\widetilde{\pi}_j$ to π_j is greater than the L^2 convergence rate of $\widetilde{t}_j$ to t_j, as noticed by [Marzouk and Xiu, 2009], and is twice faster in the case of a uniform prior [Birolleau et al., 2014]. More precisely, following [Marzouk and Xiu, 2009] and using the Pinsker's inequality, a straightforward calculation leads to

$$D_{KL}\left(\pi_j \mid \widetilde{\pi}_j\right) \leq C\|\mathbf{G}\mathbf{t}_j\mathbf{H}(\mathbf{x}_0) - t_j\left(\mathbf{x}_0\right)\|_{L^2} \tag{4.12}$$

where D_{KL} is the Kullback-Leibler (KL) divergence and C is a positive constant. It is worthwhile to note that although the KL divergence is not a distance (because it is not symmetric and does not respect the Minkowski's inequality), an analogous result [Lie et al., 2018] can be given in terms ot the total variation between two PDF.

3 Application to the Bering sea bolide

3.1 Recorded signals and detections

To validate our approach, we consider an event that was recorded by several IMS infrasound stations on December 18, 2018. This event is known to correspond to a small asteroid that entered the atmosphere at approximately 30 km.s^{-1} velocity, 69° elevation and 349° azimuth, and exploded at roughly 25 kilometers altitude over the Bering Sea[2]. The event, the second largest of its kind observed in the past twenty years, after the 2013 Chelyabinsk bolide, released an estimated 173 kilotons of energy.

[2]Center for Near Earth Object Studies."Fireballs." Jet Propulsion Laboratory. Retrieved July 23, 2019. https ://cneos.jpl.nasa.gov/fireballs/

The meteor-generated infrasound was recorded by 19 infrasound stations of the 49 certified IMS infrasound stations, ranging from 1030 km up to 15000 km away from the location of the event, as reported in the Reviewed Event Bulletin (REB) of the International Data Center (IDC). Further, 62 stations of the USArray Transportable Array [de Groot-Hedlin and Hedlin, 2015] also monitored the event. On December 18, NASA's Terra satellite passed over the Bering Sea a few minutes after the large bolide event's peak brightness at 23 :48 :20 UTC. The Multi-angle Imaging SpectroRadiometer (MISR), one of the scientific instruments on Terra, consists of nine cameras pointed at different along-track view angles, enabling multiangular views of Earth in four spectral bands. Between 23 :51 :40 and 23 :58 :04 UTC, MISR's nine pushbroom cameras each observed, in succession, the debris cloud and its shadow. This set of observations, apparently unique in MISR's nearly 20-year record, offer exceptional opportunities for the study of such an event. In the following, the localization given by MISR is considered as the reference and noted $\mathbf{X}_{\text{NASA}}$. Similarly, the localization provided by the IDC is noted $\mathbf{X}_{\text{IDC}}$

Due to the good low frequency detection capability of IMS infrasound stations at the time of event, very-long-period acoustic signals, with dominant periods spanning the 20 to 25 s interval, were clearly observed. While this event was theoretically detectable at each of the existing IMS infrasound stations [Pilger et al., 2020], all detections are not equal. In fact three main factors influence the infrasound signal detectability : (1) the ducting behavior of the acoustic waves travelling from the source to any of the globally distributed receivers ; (2) the daytime, during which the signal arrival at the infrasound array takes place ; (3) the directivity of the station towards the source. Ideal conditions correspond to quiet nighttime conditions, a 90° perpendicular direction towards the line source trajectory, and the availability of a stratospheric ducting. Here, to avoid complications due to post-processing of noisy signals, we restrict the detections used to localize the event to that provided by six infrasound stations : IS53, IS30, IS18, IS56, IS57 and IS59. These stations are shown in figure 4.1 as well as ten recent fireball events that occurred between summer 2018 and summer 2019. These events were selected and analyzed in the context of the NEar real-time MOnitoring system (NEMO) project [Drolshagen et al., 2019]. This selection is in good agreement with the fact that these events yielded the most energy during the period of observation.

The usual approach to locate fireball events is based on the triangulation method, which uses the detected backazimuths provided by the Progressive Multi-Channel Correlation (PMCC) method [Cansi, 1995]. The backazimuths are projected along the great-circle propagation path for three stations at a time so that if three intersection points are found, their geographical center gives the triangulation result. This procedure can be subsequently repeated for all conceivable combinations of three stations that detected the event. The observed backazimuths can also be incorporated in the Bayesian framework (via the detection vector $\mathbf{D}$) together with a ray tracing technique to compute infrasonic propagation characteristics [Assink et al., 2013, Blom et al., 2015, Vanderbecken et al., 2020, among others]. For the event considered in this chapter, however, the frequency is typically smaller than 0.5 Hz, a frequency for which ray tracing is likely to introduce discontinuities in the predicted arrival times as

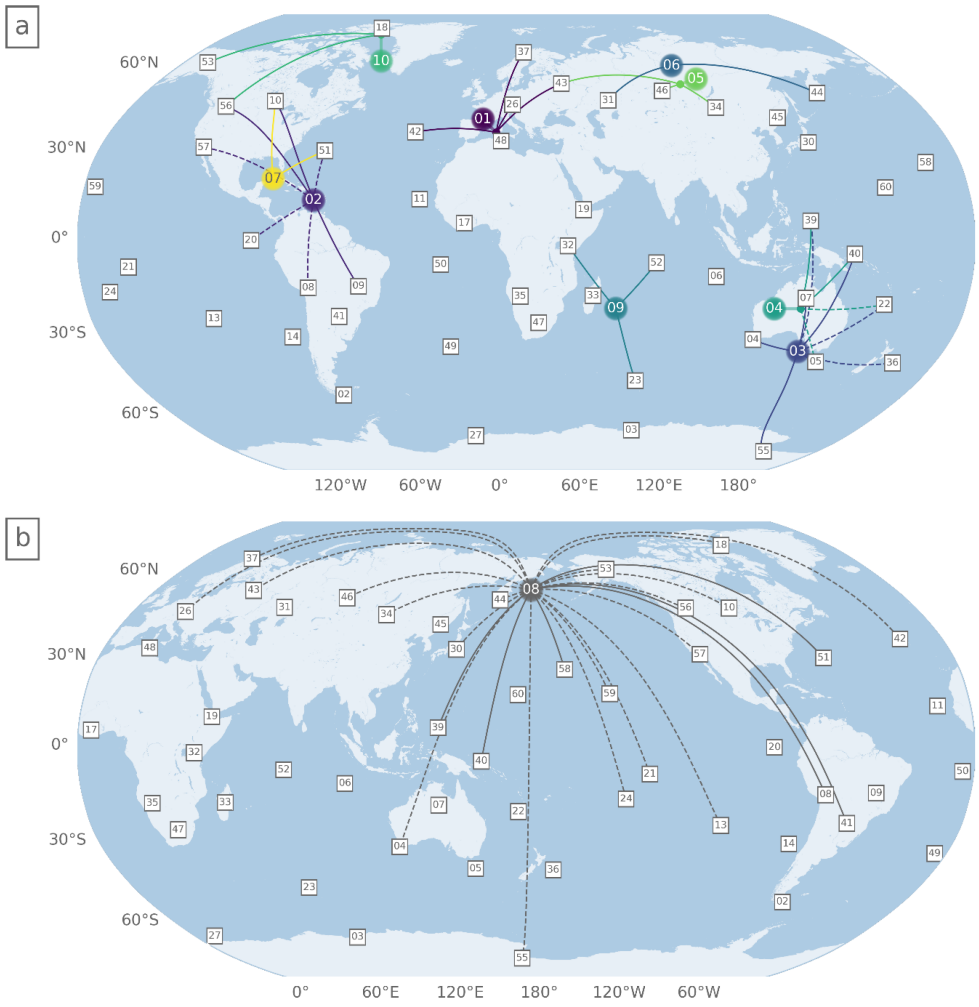

FIGURE 4.1 : (a,b) Location of fireball events (circles) and positions of certified IMS infrasound stations (rectangles). Stations and events are connected by lines if infrasound from an event was recorded at the respective stations ; the line is dashed, if the detection is part of the REB. (b) Bering Sea event. From [Pilger et al., 2020].

the event location $\mathbf{x}_0 = (x_0, y_0)$ is varied. Bearing this limitation in mind, we resort to computing the pressure field using normal mode expansions at the locations $\mathbf{x}_j$ of the infrasound stations. Following [Bertin et al., 2014] and using cylindrical coordinates, the Green's functions are computed for each possible source location, using a Fast Fourier Transform algorithm. The resulting signals can easily be computed using a Ricker-like wavelet for representing the source function and applying the convolution property. While the frequency of the source is fixed to $\omega_0 = 0.2 \times 2\pi$ rad.s^{-1} in this work, this parameter can also be incorporated in $\mathbf{X}$ as well as any parameter that may have an effect on the quantities of interest (QoI).

Although full-wave modeling is better suited for propagating low-frequency waves, there is the question of how QoI can be extracted from complex, noisy signals. A simple approach for determining an arrival time from a signal consists in computing

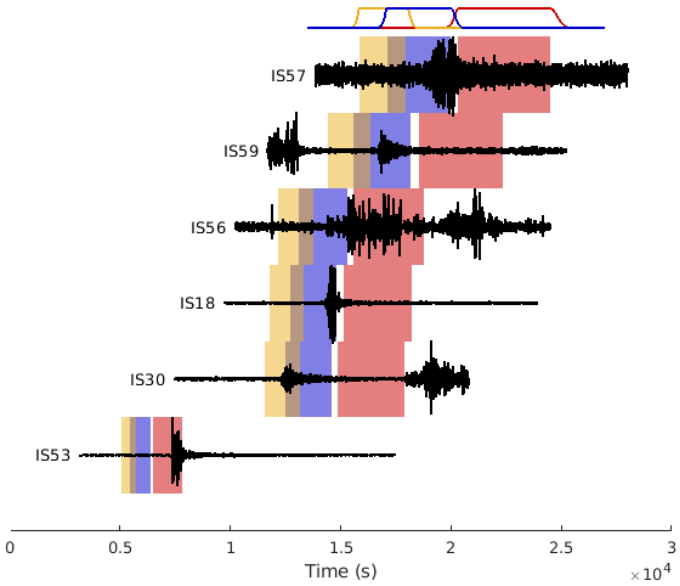

FIGURE 4.2 : Recorded signals and regions of plausible arrival times for tropospheric (yellow), stratospheric (blue) and thermospheric (red) infrasonic arrivals, obtained using the tapered box-car celerity distributions of [Blom et al., 2015] and the localization provided by the International Data Center (IDC).

the moving root-mean-square (rms) over a sliding window of length $\tau = T/N$, where T is the duration of the signal and N is the number of samples. The arrival time is then defined as the time for which the amplitude of the rms first exceeds a given threshold, which is here specified by the product βA_j, where A_j is the maximum amplitude of the rms and β is fixed between 0 and 1. Figure 4.3 shows the arrival time as a function of β for the recorded signals at IS53, IS30, IS18, IS56, IS57 and IS59. The sample size is fixed to $N = 1000$ to reduce the sensitivity of the arrival time to β. As might be expected, for sufficiently large values of β, the arrival time slowly increases but, as β becomes smaller, smaller values of t_j and larger sensitivity occur. While each jump of $t_j(\beta)$ is related to a pattern in the signal, figure 4.3 shows that the range of possible arrival times increases with increasing distance from the source, except for the signal recorded at IS56. This effect is not purely due to distance, it is also related to properties of the atmosphere along the source receiver path, as section 3.2 shows. Nevertheless, for fixed τ, the uncertainty associated with arrival time picking is incorporated in the probabilistic approach using

$$F_j(\mathbf{X}; \beta) = t_j(\mathbf{x}_0; \beta) + t_0, \tag{4.13}$$

where β is sampled from a normal distribution $N(\mu_\beta, \sigma_\beta)$ with a mean μ_β and a standard deviation σ_β. It turns out that knowledge of the arrival time as β changes randomly clearly allows a much deeper appreciation of uncertainty than is possible by mere point estimates. This implies that σ_j in equation (4.2) depends on β, as well

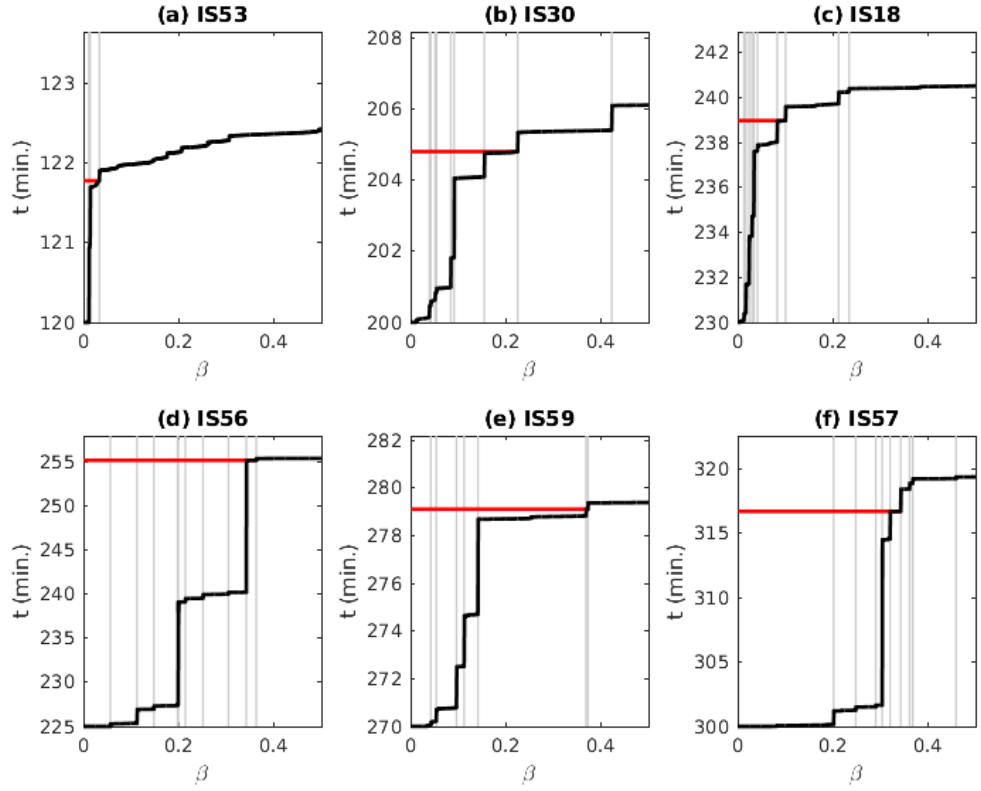

FIGURE 4.3 : Arrival times as function of β for each recorded signal and possible values (red lines) for constructing the detection vector in (4.2).

as the posterior distribution $\mathrm{P}(\mathbf{X}|\mathbf{D}, \beta)$ and various other aspects of the distribution (marginal distributions, confidence regions, etc.).

3.2 The propagation conditions

Following the detectability estimations provided by [Pilger et al., 2020], the event is expected to be best observed at nearby stations in the Pacific Ocean and along the eastern coast of Asia as well as the western coast of North America. In fact, the predominant direction for stratospheric ducting (here, to the east on the northern hemisphere) is less important for very strong events. Such events have a very low dominant frequency, which leads to low atmospheric attenuation along propagation paths of even many thousands of kilometers. Further, for strong events it is possible that the ducted propagation takes place in both directions, upstream and downstream the predominant stratospheric wind. It may take the shape of elevated ducting, when propagating against the dominant wind direction.

For each source-to-receiver path, along-path wind speeds can be calculated using the projection of the horizontal wind vector $\mathbf{U}$ into the direction from source to receiver $\mathbf{e}_0$, where $\mathbf{e}_0 = \mathbf{x} - \mathbf{x}_0$. The components of $\mathbf{x}$ are the longitude x and latitude

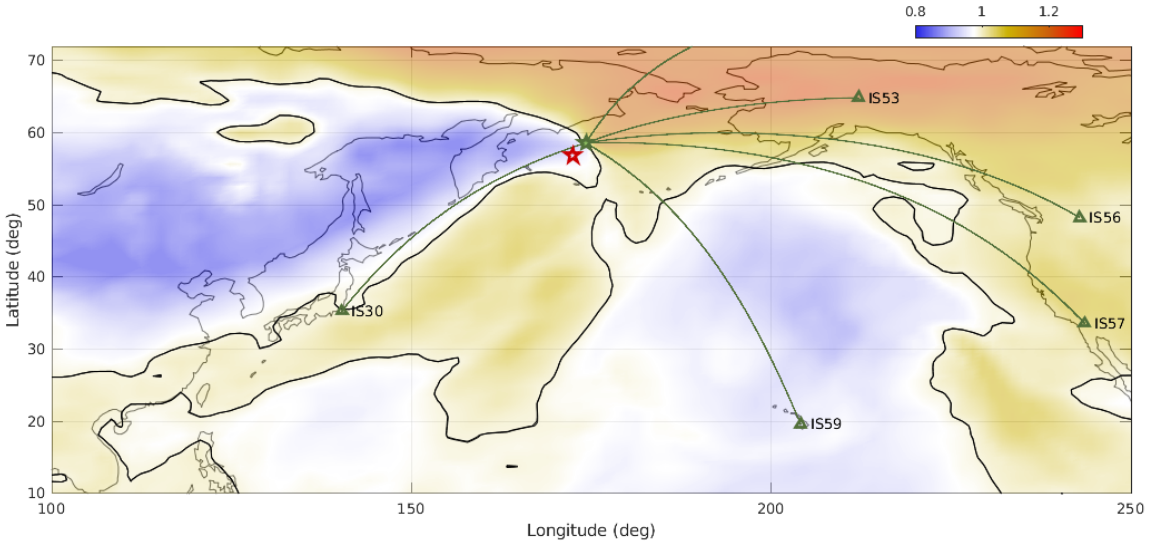

FIGURE 4.4 : Maximum effective sound speed ratio $R_z(\mathbf{x})$ over the region of interest on december 19, 2018 (0h UT). The source location for calculating R is that provided by the IDC and the contour $R_z = 1$ is represented by a solid black line. The green and blue stars give the locations provided by the IDC and the NASA, respectively. Each infrasound station is represented by a green triangle.

y, and $\mathbf{x}_0$ corresponds to the source location, which is given by the NASA (172.4 deg and 56.9 deg, respectively) or the IDC (58.61 deg and 174.23 deg, respectively). Propagation conditions can be quantified by the maximum value of the effective sound speed ratio, which is defined by

$$R_z(\mathbf{x}) = \max_{z_0 \leq z \leq z_1} R(\mathbf{x}, z) \quad \text{with} \quad R(\mathbf{x}, z) = \frac{c_e(\mathbf{x}, z)}{c_e(\mathbf{x}, 0)}, \tag{4.14}$$

where the effective sound speed is given by

$$c_e(\mathbf{x}, z) = \sqrt{\gamma r_g T(\mathbf{x})} + \mathbf{U}(\mathbf{x}, z).\mathbf{e}_0(\mathbf{x}), \tag{4.15}$$

where γ is the ratio of specific heats, r_g is the gas constant for air and T is the ambient temperature. Figure 4.4 gives the ratio R_z as a function of longitude and latitude, using $z_0 = 40$ km and $z_1 = 80$ km in equation (4.14) for the source location estimated by the IDC. The atmospheric specifications are obtained from the analysis provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). This analysis provides estimates, among other parameters, of atmospheric temperatures and windspeeds with a horizontal resolution of 1.5 deg, and a vertical structure composed of 91 points between the surface and a maximum altitude of approximately 80 km.

According to the classical ray theory, strong arrivals (totally ducted) are refracted from altitudes for which the effective sound speed is larger than that at ground level, i.e. $R > 1$. The contour $R_z = 1$ (solid black line in figure 4.4) splits the domain into regions where sound waves propagate efficiently within the stratosphere ($R_z > 1$) and

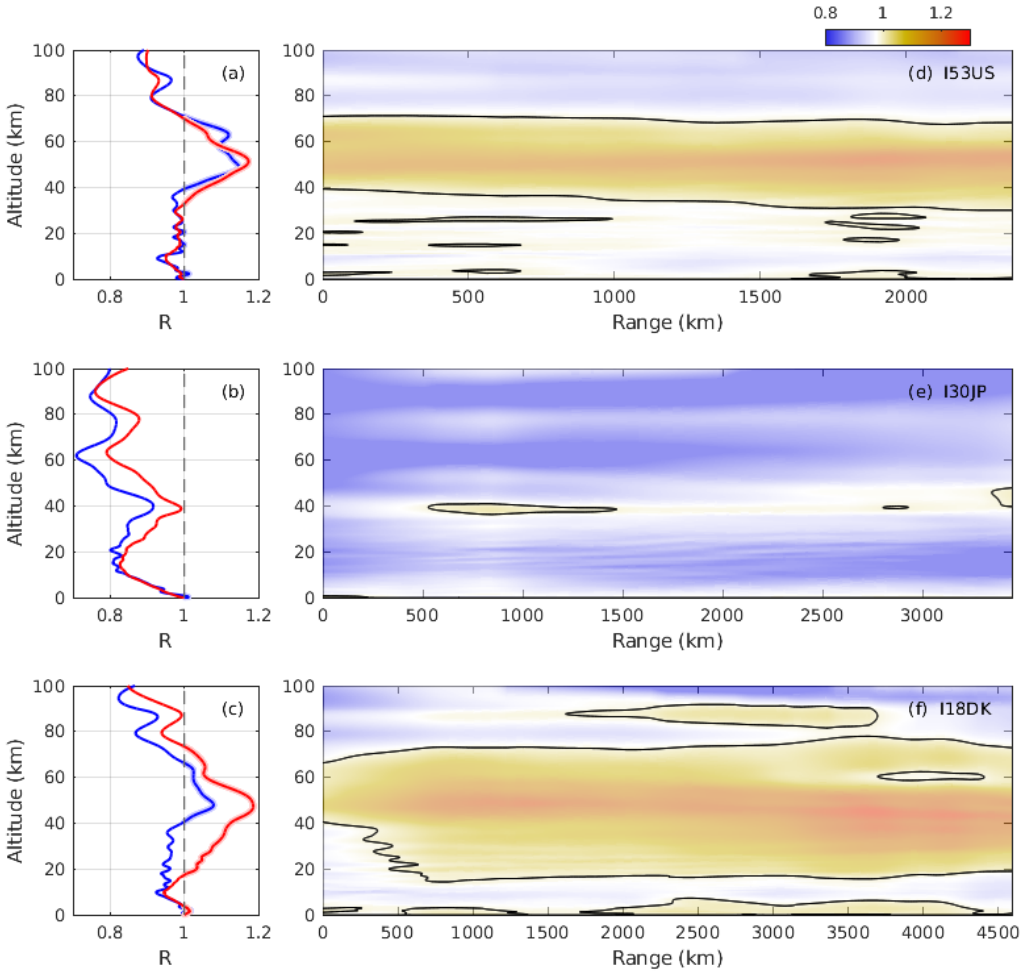

FIGURE 4.5 : (a,b,c) Effective sound speed-ratio profiles at the location estimated by the IDC (blue) and averaged profiles (red) along the source-receiver path. (c,e,f) Effective sound speed-ratio as a function of altitude and range for infrasound stations IS53 (d), IS30 (e) and IS18 (f).

regions where the stratospheric waveguide weakens or disappear ($R_z < 1$). The ratio R can be viewed as a simple parameterization of the stratospheric waveguide structure. The along-path variability of the propagation conditions can thus be quantified by $R(r, z)$, where r is the distance from the source location, as estimated by the IDC. Figures 4.5 and 4.6 show this ratio for altitudes below 100 km, using a combination of cubic splines and empirical atmospheric models (HWM and MSISE) in the upper layer, for $z > 80$ km. For northeastward propagation (IS53 and IS18), the ducting layer is formed at altitudes above $z = 40$ km at source location and gradually spreads downward as the range increases, except for IS30 for which we have $R(0) < 1$. The along-path variability, in terms of the length of the interval mapped by R_z, is obtained for the closest station (IS53) and given by $\Delta R_z \simeq 0.07$. For eastward and southeastward propagations the waveguides weaken along the path and may literally disappear, as figure 4.6 shows. The along-path variability ΔR is given by 0.16, 0.17 and 0.16

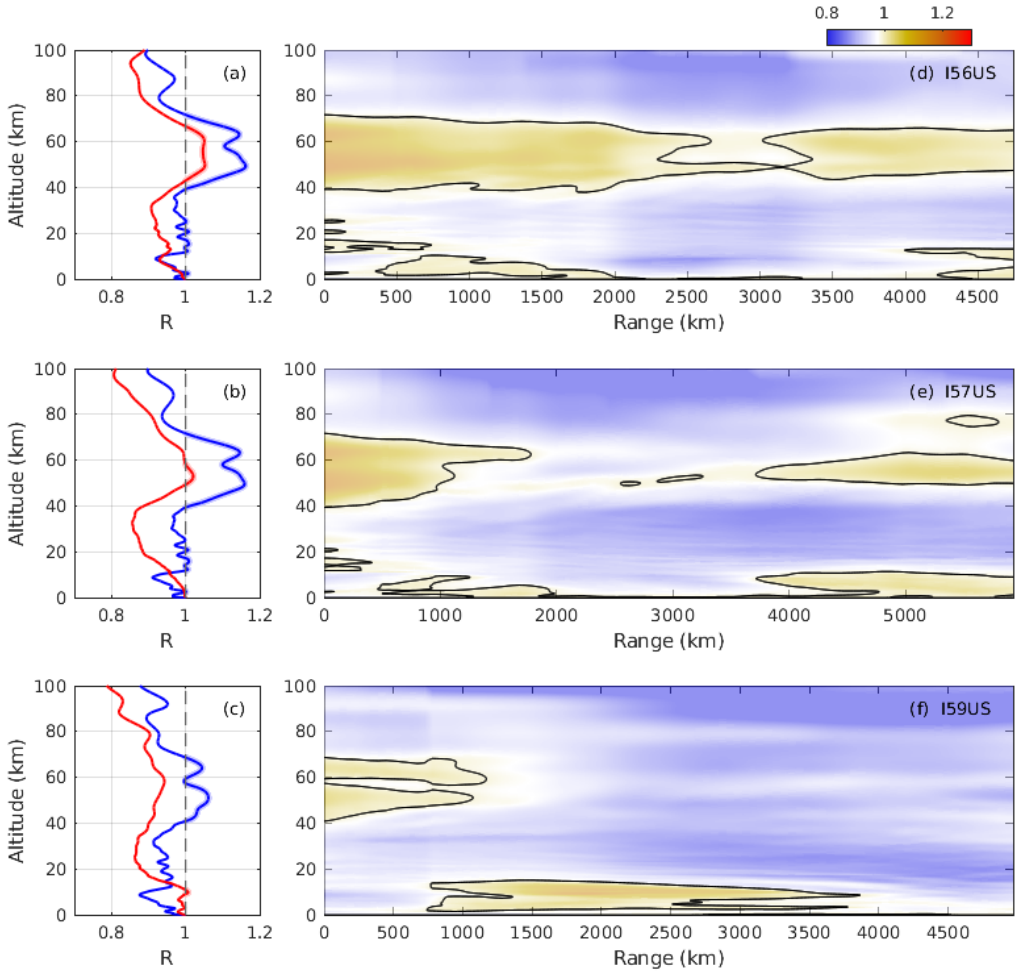

FIGURE 4.6 : (a,b,c) Effective sound speed-ratio profiles at the location estimated by the IDC (blue) and averaged profiles (red) along the source-receiver path. (c,e,f) Effective sound speed-ratio as a function of altitude and range for infrasound stations IS56 (d), IS57 (e) and IS59 (f).

for IS56, IS57 and IS59, respectively. This strong along-path waveguide variability restricts the signal celerities that can be supported by stratospheric waveguide propagation, thereby leading to shorter signals as the geometrical optics approximation suggests [Green and Nippress, 2019]. Further, the stronger the waveguide, in terms of R, the steeper the maximum launch angle at which acoustic energy can be turned and returned to Earth's surface. These steeper launch angle arrivals, which propagate further and faster before returning to Earth's surface, belong to the fast stratospheric branch described by [Waxler et al., 2015].

As mentioned above, the profiles shown in figures 4.5 and 4.6 are obtained using the IDC estimate for $\mathbf{x}_0$. For other source locations, the waveguide characteristics may be inferred by suitably projecting the windspeed onto the plausible source-receiver plane. When the source location is randomly selected, the effective sound speed profile should be viewed as a random function and R_z, as a random variable. Figure 4.7 shows

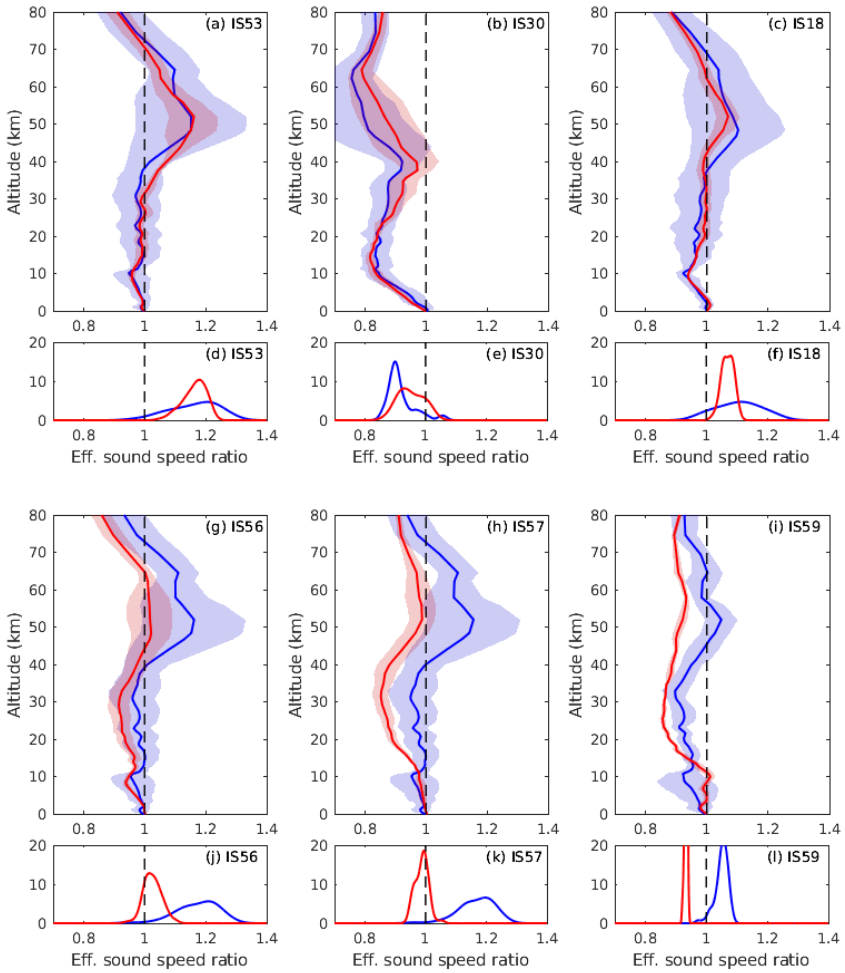

FIGURE 4.7 : (a,b,c,g,h,i) Mean ratio $\bar{R}$ when the source location is picked randomly around the IDC estimate. Blue and red solid lines are the profile at the source and the averaged profile along the source-receiver path, respectively. The associated shaded area gives the region between $\pm 2\sigma_R$. (d,e,f,j,k,l) Probability density functions of R_z.

the dispersion of the resulting profiles using 70 plausible locations picked randomly around the IDC estimate. In this example the longitude and latitude are considered as being uncorrelated and sampled using a bivariate normal distribution $N(\mathbf{x}_0, \sigma\mathbf{I})$, with $\sigma = 3$ deg. The mean effective sound speed-ratio $\bar{R}$ and the standard deviation σ_R are computed with the profiles at the source (blue line) and with the averaged profiles along the source-receiver path (red line). The corresponding locations are given in figure 4.8. For each station, the kernel probability density estimate of the random variable R_z is obtained using a normal kernel function. While sound waves remain trapped in the stratospheric waveguide for eastward propagations, as figure 4.7 shows for the paths to IS53, IS18 and IS56, some azimuths are more problematic when the source location varies. For IS30 and IS59 (figures 4.7(b,h)), it is evident that a change in $\mathbf{x}_0$ can affect acoustic ducting at source location, in terms of R_z (blue line). The practical implication is that small uncertainties in the source location may result in

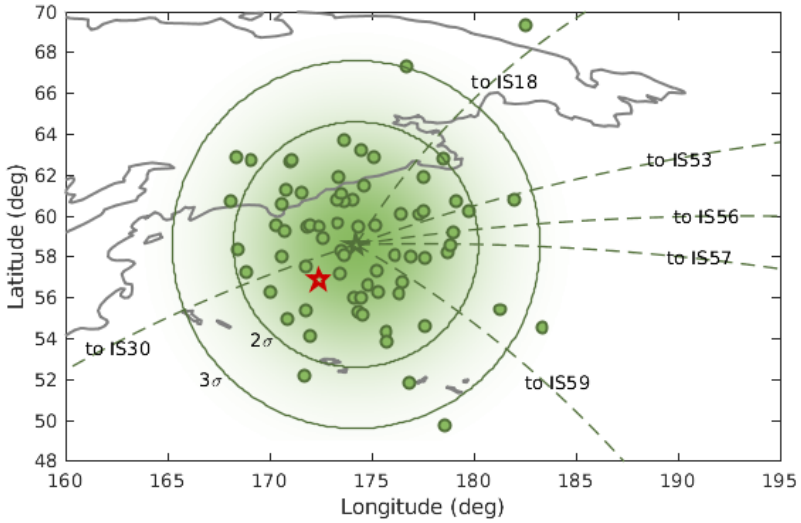

FIGURE 4.8 : Source locations (full circles) used to calibrate the gPC metamodel. The locations are sampled in the vicinity of the IDC estimate ($\star$, green). Dashed lines give the source-receiver paths for the six stations considered. The radius of the two large circles correspond to two and three standard deviations from the mean *a priori* location. The NASA estimate is identified by a red star.

large changes in the probability density functions of the QoI so that perturbation theory is not applicable to compute the distributions. It is worthwhile to point out that we should be cautious in practical inverse problems of relying solely on range-independent propagation modeling, for if $\mathrm{P}(R_z|\mathbf{x}_0, \mathbf{x})$ were changing drastically over the range in which $\mathrm{P}(\mathbf{x}|\mathbf{x}_0)$ was appreciable, it would be important to know about it. Thus, in principle, we should always make a study of $\mathrm{P}(R_z|\mathbf{x}_0, \mathbf{x})$ for a number of profiles extracted along the source-receiver path and indeed, in the present case, figure 4.7(k,l) show that two opposite conclusions, in terms of R_z, can be drawn from the same set of source locations.

3.3 Localization estimates and confidence regions

For each of the 70 plausible locations given on figure 4.8 a complete simulation is done which gives 70 times of arrival for each station. From this sample we have computed the coefficients (a_{jk}), polynomial chaos decomposition of the time of arrival, at each station. With this operation we have a metamodel given the time of arrival function of the source location for each station. This metamodel is very cheap from a computational point of view and can be used to generate the surface response on a thiner grid to get a precise localization.
To take into account the variability due to the extraction from the registered signals we have taken a random β (defined in paragraph 3.1) : $\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta)$. This parameter is

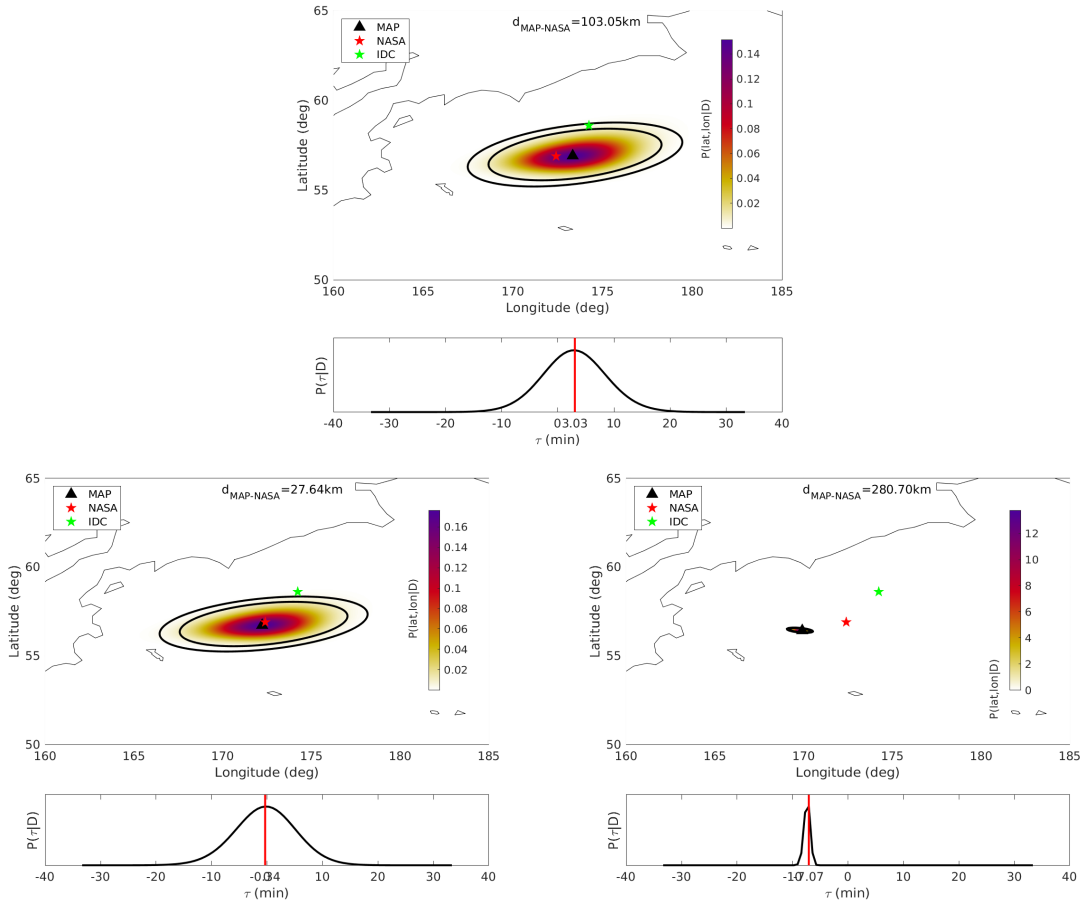

FIGURE 4.9 : Posterior distributions for different definitions of the threshold β. The only difference between those three plots comes from the values of the parameter $(\mu_\beta, \sigma_\beta)$: $(0.3, 0.2)$ on the top, $(0.45, 0.3)$ on the left and $(0.8, 0.22)$ on the right.

used to ajust the *noise* at each station : a sample of 1000 β gives a standard deviation σ_j for the 1000 corresponding D_j, this σ_j is then used to compute the likelihood with formula (4.2). Figure 4.9 give the marginal distributions for the location (x_0, y_0) and the time of the event t_0 for different values of the parameters $(\mu_\beta, \sigma_\beta)$. The ellipses give the confidence region : according to the posterior distribution the event took place in the smaller ellipse with a probability 0.95 and in the wider one with probability 0.99. In the three configurations the time of the event seems close to the estimation given by the NASA (in red), only the standard deviation of the estimation changes depending on the value of σ_β. The localization is more sensitive to the extraction process : the size of the ellipses seems to depend on the ratio σ_β/μ_β and the maximum a posteriori which would be the location proposed by the bayesian process can change of several thousand of kilometers. To look at the impact of the different stations, we have plotted on figure 4.10 the station whose partial likelihood $\tilde{\pi}_j(\mathbf{x}_0, t_0, \beta)$ is maximal. The directions of the different areas are globally orthogonal to the corresponding source-to-receiver paths. The other map on figure 4.10 gives the number $N(\mathbf{x})$ of

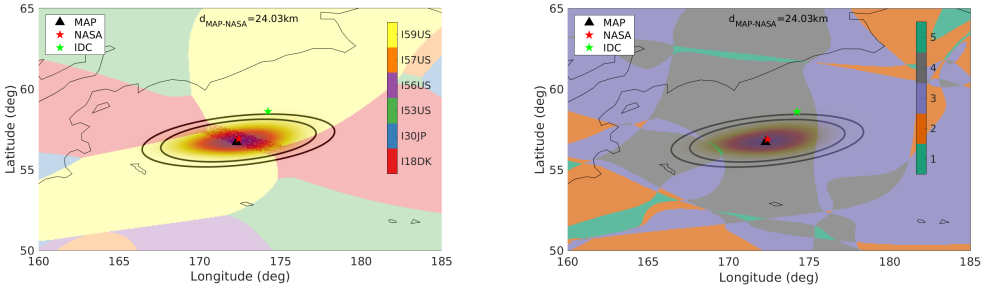

FIGURE 4.10 : Station with a predominant contribution to the likelihood (on the left) and number of station with significative contribution $N(\mathbf{x})$ (on the right). Here, $(\mu_\beta, \sigma_\beta) = (0.45, 0.3)$ for the two figures.

station with a significative contribution to the likelihood defined by :

$$N(\mathbf{x}) = \sum_{j=1}^{n} \zeta_j(\mathbf{x}) \tag{4.16}$$

where ζ_j is given by :

$$\zeta_j(\mathbf{x}_0, t_0, \beta) = 1 \quad \text{if} \quad \widetilde{\pi}_j(\mathbf{x}_0, t_0, \beta) > \frac{1}{n}\tilde{\pi}(\mathbf{x}_0, t_0, \beta), \tag{4.17}$$

and $\zeta_j(\mathbf{x}_0, t_0, \beta) = 0$ otherwise. This second plot indicates that around the maximum a posteriori, the likelihood takes information from four different stations which guarantees a more robust localization. We show on figure 4.10 the map only for $(\mu_\beta, \sigma_\beta) = (0.45, 0.3)$ but the two other maps for $(\mu_\beta, \sigma_\beta) = (0.3, 0.2)$ and $(0.8, 0.22)$ give similar results. Consequently, this cannot be a criterium to choose the value of $(\mu_\beta, \sigma_\beta)$.

4 Conclusion

We have shown in this chapter how polynomial chaos based metamodels can be used in a bayesian inversion context to lacalize the bolide. Bayes' theorem describes, in a fundamental way, the process of learning from experience, and shows how knowledge about the state of nature represented by X is continually modified as new data becomes available. Mathematically, Bayes' formula is merely a statement of conditional probability, and as such its validity is not in question. In practice, our inversion is based on the signal of six stations of the IMS network and uses the time of arrival as a quantity of interest. With this strategy we have been able to reduce the error of localization from 220 km (for the IDC point) to less than 50 km for a certain extraction of the times of arrival from the registered signals – and assuming that the NASA point is exact. An important feature of the metamodel based approach relies

in its low computational cost wich allows a fine grid computation of the posterior probability distribution and thereby a better precision of the location estimation.
We have also emphasized the impact of the extraction of the quantity of interest. In fact, signals registered at the different stations are affected by ambient noise and by the measurement noise. Then the extraction of the quantity of interest (time of arrival in our case) depends on arbitrary choices : where does the signal start ? Is its maximum a relevant point ? How many arrivals do we consider ? We have shown in this chapter that the uncertainty due to the extraction procedure is of similar amplitude to the impact of the propagation effects. We have modelled this uncertainty using a random threshold for extracting the time of arrival however fine-tuning the parameters of this random threshold remains an open question.

Possible other improvments of our method inculde the use of range dependant simulations to take into account the atmospheric variability alongside the perturbation [Damiens et al., 2018, Waxler et al., 2017]. Then, the metamodel only takes into account the uncertainty associated with the localization although there exists many other sources of uncertainties (in particular atmospheric variability [Drob, 2019, Waxler and Assink, 2019, Chunchuzov and Kulichkov, 2019]). Adding more uncertainties will increase the dimension of the perturbation and limit the benefit of using a metamodel. To avoid increasing too much the dimension, a preliminary study (for instance the Morris screening technique, like in Rojano *et al.* [Rojano et al., 2019]) can be conducted to select the most influencing variables. Our inversion is only based on the time of arrival although other quantities can be extracted from the signal. The benefit of using more quantities to improve the localization needs to be investigated.

Chapitre 5

Conclusion

We have explained in chapter 3 our method to design a metamodel able to return signals propagating in a random atmosphere. Our approach is based on a spectral decomposition of the random acoustic field :

$$p(r,z,\xi;\omega)=\sum_{n=1}^{N}p_n(r,z,\xi;\omega)=\sum_{n=1}^{N}\frac{u_n(r,\xi;\omega)}{\sqrt{r}}\phi_{n0}(z,\xi;\omega), \tag{5.1}$$

where the basis functions $(\phi_{n0})_{n=1,..,N}$ are the eigenfunctions of the Helmholtz operator $\mathcal{H}$:

$$\mathcal{H}\phi_{n0}=\frac{d^2\phi_{n0}}{dz^2}+\frac{\omega^2}{c(z,\xi)^2}\phi_{n0}=k_n^2\phi_{n0}. \tag{5.2}$$

The decomposition on this basis characterizes the propagation, we have chosen to build a metamodel of this basis by decomposing $(\phi_{n0})_{n=1,..,N}$ and $(u_n)_{n=1,..,N}$ on a polynomial chaos basis. A great advantage of this method is the ability to restitute random signals even for large distances despite the long term integration problem (see chapter 2). In addition, our metamodel is valid for every source and every distance of propagation without any supplementary cost. The main limitation comes from the necessity of tracking the modes impacted by the perturbation. Once the metamodel is computed, the impact of the small scale part of the perturbation is integrated by means of the coupling coefficients. The numerical impact of those coefficients and their dependence on the different components of the perturbation has not been evaluated numerically in this thesis, their computation would enlighten the role of those small scale perturbations. We have always considered in this work a propagation model invariant in the propagation direction and a natural perspective would be to consider range dependant simulations. However, range dependant modal decompositions allow to take into account horizontal variations of the atmospheric state as long as the horizontal inhomogeneities of the medium are small over a typical acoustic wavelength [Pierce, 1965, Waxler et al., 2008]. This cannot be guaranteed in the case of atmospheric variability, random perturbations can be of great amplitude as illustrated by the gravity waves breaking in chapter 1.

In decomposition (5.1), the stochasticity is brought both by the basis functions and the coefficients of the decomposition. The dynamically orthogonal decomposition (DO) introduced by Sapsis and Lermusiaux [Sapsis and Lermusiaux, 2009] have proposed a separation : a deterministic basis adapted to the physical phenomenon and the stochastic coefficients characterizing the variability due to the randomness of the parameters. In our case, this would give a single basis adapted to the acoustic propagation and to the uncertainty at the same time instead of decomposing the acoustic basis on another stochastic basis. The dynamically orthogonal decomposition has been introduced in 2009 just before Cheng *et al.* introduce the bi-orthogonal (BO) decomposition [Cheng et al., 2013a, Cheng et al., 2013b]. In both work the objective is to construct the sparsest representation of a solution of a stochastic partial differential equation exploiting its dynamics. Choi *et al.* have shown in 2014 the equivalence of those two decompositions [Choi et al., 2014] and we are going to focus here on the work of Sapsis and Lermusiaux. In our case, let us consider the evolution of $p(r, z, \xi)$, the random acoustic field, which depends on the altitude z, the range r and the random parameters ξ, we drop the frequency ω that is supposed to be fixed here :

$$\frac{\partial \Psi(r, z, \xi)}{\partial r} = \mathcal{L}\left[\Psi(r, z, \xi); \xi\right], \tag{5.3}$$

where $\Psi(r, z, \xi) = \begin{bmatrix} p(r, z, \xi) \\ \dfrac{\partial p(r, z, \xi)}{\partial r} \end{bmatrix}$ and $\mathcal{L}$ is a differential operator defined by :

$$\mathcal{L} = \begin{bmatrix} 0 & 1 \\ -\frac{\partial^2}{\partial z^2} - \frac{\omega^2}{c(z,\xi)^2} & -\frac{1}{r} \end{bmatrix}. \tag{5.4}$$

This operator models the acoustic propagation for cylindrical coordinates (r, z) for a medium invariant in the propagation direction. This DO-decomposition is based on a decomposition of the solution Ψ into a mean $\overline{\Psi}$ and a stochastic dynamical component. The stochastic part is decomposed on an orthonormal basis $\left(\Psi_i\left(r, z\right)\right)_{i=1,..,s}$ of the subspace $V_s = \text{Span}\left\{\Psi_i, i = 1, .., s\right\}$ where the stochasticity 'lives' and the coefficients $Y_i\left(r; \xi\right)$ describing how the stochasticity evolves as the subspace changes alongside the propagation,

$$\Psi(r, z; \xi) = \overline{\Psi}\left(r, z\right) + \sum_{i=1}^{s} Y_i\left(r; \xi\right) \Psi_i\left(r, z\right). \tag{5.5}$$

This decomposition is more general than a classical gPC approach in the sense that the basis is not known a priori but adapts to the dynamics. In fact, the evolution of Ψ_i and Y_i is governed by $s + 1$ deterministic PDEs based on the operator $\mathcal{L}$. More precisely the evolution of the basis Ψ_i is driven by an orthogonality condition :

$$\frac{\partial V_s}{\partial r} \perp V_s \Leftrightarrow \left\langle \frac{\partial \Psi_i}{\partial r}; \Psi_j \right\rangle = 0, \, \forall (i, j) \in \{1, .., N\}^2. \tag{5.6}$$

The idea of this orthogonality relation is to avoid redundancy in the representation. In fact, the variation of the stochastic coefficients $Y_i\left(r; \xi\right)$ can express exclusively the

evolution of uncertainty within the space V_s. On the other hand, the variation of the basis $(\Psi_i(r,z))_{i=1,..,s}$ expresses both the evolution of uncertainty within V_s and also normal to V_s. Therefore, we see that the source of redundancy comes from the evolution of uncertainty that can be described by both the variation of the stochastic coefficients and the basis. This explains the restriction on the evolution of the basis $(\Psi_i(r,z))_{i=1,..,s}$: it has to be normal to the space V_s since the evolution within V_s can be described completely by the stochastic coefficients.

In chapter 4, we have shown how a metamodel technique can be used to take advantage of propagation models in a localization process. The originality of this chapter stands in the combination of bayesian inference method and polynomial chaos based metamodel for the propagation part. The use of a metamodel is essential here due to the numerical cost of a simulation on such distances. A great strength of our method is to use full-wave signals, which carry much more information than ray tracing or generative models used in other studies based on the bayesian approach [Moore and Russell, 2017, Arora et al., 2013, Blom et al., 2018]. However, we have only worked here with the time of arrival extracted from the signal and not the signal itself. Taking into account other quantities of interest such as the amplitudes and the durations of the signals would complete this study. Another advantage of the full-wave simulation is to restitute multi-arrival signals when atmospheric conditions create different waveguide. It was not the case for the Bering bolide but could be used to increase the number of QoI (time between two arrival, amplitude ratio,...). Then, only the uncertainty due to the localization has been treated here : the uncertainty on the source model as well as the impact of the atmospheric variability has not been assessed, this could also improve the localisation.

A more complex question lies in the a priori part. We have been considering an a priori based on informations given by the IDC seen as the expert judgment. The importance given to this a priori compared to the simulation and the translation of this expert judgment into a probability measure can be questionned. A possibility to circumvent this issue could be to recursively consider the a posteriori as a new prior and start again the computations with a new sampling adapted to this new prior. This would lead to an increase of the numerical cost which could be affordable or not depending on the convergence rate of such a method. But the convergence itself needs to be studied first, depending on the structure of the likelihood.

Bibliographie

[Abramowitz and Stegun, 1965] Abramowitz, M. and Stegun, I. A. (1965). Handbook of mathematical functions with formulas, graphs, and mathematical table. In *US Department of Commerce*. National Bureau of Standards Applied Mathematics series 55. [*Cited on pages 45 and 46.*]

[Adhikari and Friswell, 2004] Adhikari, S. and Friswell, M. (2004). Random eigenvalue problems in structural dynamics. In *45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference*, page 1747. [*Cited on page 71.*]

[Alexanderian, 2015] Alexanderian, A. (2015). A brief note on the karhunen-loève expansion. *arXiv preprint arXiv :1509.07526*. [*Cited on page 19.*]

[Allemang, 2003] Allemang, R. J. (2003). The modal assurance criterion–twenty years of use and abuse. *Sound and vibration*, 37(8) :14–23. [*Cited on page 72.*]

[Arora et al., 2010] Arora, N., Russell, S. J., Kidwell, P., and Sudderth, E. B. (2010). Global seismic monitoring as probabilistic inference. In *Advances in neural information processing systems*, pages 73–81. [*Cited on page 2.*]

[Arora et al., 2013] Arora, N. S., Russell, S., and Sudderth, E. (2013). Net-visa : Network processing vertically integrated seismic analysis. *Bulletin of the Seismological Society of America*, 103(2A) :709–729. [*Cited on pages 11, 16, 91, 92, and 111.*]

[Arrowsmith et al., 2013] Arrowsmith, S. J., Marcillo, O., and Drob, D. P. (2013). A framework for estimating stratospheric wind speeds from unknown sources and application to the 2010 december 25 bolide. *Geophys. J. Int.*, 195 :491–503. [*Cited on page 92.*]

[Assink et al., 2013] Assink, J., Waxler, R., Frazier, W., and Lonzaga, J. (2013). The estimation of upper atmospheric wind model updates from infrasound data. *Journal of Geophysical Research : Atmospheres*, 118(19) :10–707. [*Cited on pages 8, 92, and 97.*]

[Assink et al., 2017] Assink, J., Waxler, R., and Velea, D. (2017). A wide-angle high mach number modal expansion for infrasound propagation. *The Journal of the Acoustical Society of America*, 141(3) :1781–1792. [*Cited on pages 4, 22, and 33.*]

[Attenborough, 1988] Attenborough, K. (1988). Review of ground effects on outdoor sound propagation from continuous broadband sources. *Applied acoustics*, 24(4) :289–319. [*Cited on page 34.*]

[Bartlett and Mendelson, 2002] Bartlett, P. L. and Mendelson, S. (2002). Rademacher and gaussian complexities : Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov) :463–482. [*Cited on page 24.*]

[Beck et al., 2005] Beck, V., Malick, J., and Peyré, G. (2005). *Objectif agrégation : mathématiques*. H & K. OCLC : 469495218. [*Cited on page 50.*]

[Berner et al., 2017] Berner, J., Achatz, U., Batté, L., Bengtsson, L., de la Cámara, A. Christensen, H. M., Colangeli, M., Coleman, D. R. B., Daan Crommelin, D., Dolaptchiev, S. I., Franzke, C. L. E., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajan, S., Palmer, T. N., Penland, C. andSakradzija, M., von Storch, J.-S., Weisheimer, A., Weniger, M., Williams, P. D., and Yano, J.-I. (2017). Stochastic parameterization : Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, 98 :565–588. [*Cited on page 93.*]

[Bertin, 2014] Bertin, M. (2014). *Propagation des incertitudes dans un modèle réduit de propagation des infrasons*. PhD thesis, École normale supérieure de Cachan. [*Cited on page 37.*]

[Bertin et al., 2014] Bertin, M., Millet, C., and Bouche, D. (2014). A low-order reduced model for the long range propagation of infrasounds in the atmosphere. *The Journal of the Acoustical Society of America*, 136(1) :37–52. [*Cited on pages 4, 5, 22, 63, 68, 72, 81, 82, 93, and 98.*]

[Birolleau et al., 2014] Birolleau, A., Poëtte, G., and Lucor, D. (2014). Adaptive bayesian inference for discontinuous inverse problems, application to hyperbolic conservation laws. *Communications in Computational Physics*, 16(1) :1–34. [*Cited on page 96.*]

[Blanc-Benon et al., 2001] Blanc-Benon, P., Dallois, L., and Juvé, D. (2001). Long range sound propagation in a turbulent atmosphere within the parabolic approximation. *Acta Acustica united with Acustica*, 87(6) :659–669. [*Cited on pages 4 and 22.*]

[Blatman and Sudret, 2010] Blatman, G. and Sudret, B. (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*, 25(2) :183–197. [*Cited on page 57.*]

[Blom et al., 2018] Blom, P. S., Dannemann, F. K., and Marcillo, O. E. (2018). Bayesian characterization of explosive sources using infrasonic signals. *Geophysical Journal International*, 215(1) :240–251. [*Cited on pages 92 and 111.*]

[Blom et al., 2015] Blom, P. S., Marcillo, O. E., and Arrowsmith, S. J. (2015). Improved bayesian infrasonic source localization for regional infrasound. *Geophysical Journal International*, 203(3) :1682–1693. [*Cited on pages 91, 92, 94, 95, 97, and 99.*]

[Blumer et al., 1989] Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1989). Learnability and the vapnik-chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4) :929–965. [*Cited on page 24.*]

[Boyd, 1985] Boyd, J. P. (1985). Complex coordinate methods for hydrodynamic instabilities and sturm-liouville eigenproblems with an interior singularity. *Journal of Computational Physics*, 57(3) :454–471. [*Cited on page 67.*]

[Branicki and Majda, 2013] Branicki, M. and Majda, A. J. (2013). Fundamental limitations of Polynomial Chaos for uncertainty quantification in systems with intermittent instabilities. *Commun. Math. Sci.*, 11(1) :55–103. [*Cited on page 73.*]

[Brekhovskikh and Godin, 1999] Brekhovskikh, L. M. and Godin, O. (1999). *Acoustics of Layered Media II : Point Sources and Bounded Beams*, volume 10. Springer Science & Business Media. [*Cited on page 33.*]

[Brekhovskikh and Godin, 2012] Brekhovskikh, L. M. and Godin, O. A. (2012). *Acoustics of layered media I : Plane and quasi-plane waves*, volume 5. Springer Science & Business Media. [*Cited on page 62.*]

[Brekhovskikh and Godin, 2013] Brekhovskikh, L. M. and Godin, O. A. (2013). *Acoustics of layered media II : Point sources and bounded beams*, volume 10. Springer Science & Business Media. [*Cited on page 62.*]

[Bridges and Morris, 1984] Bridges, T. and Morris, P. J. (1984). Differential eigenvalue problems in which the parameter appears nonlinearly. *Journal of Computational Physics*, 55(3) :437–460. [*Cited on page 69.*]

[Broto et al., 2018] Broto, B., Bachoc, F., and Depecker, M. (2018). Variance reduction for estimation of shapley effects and adaptation to unknown input distribution. *arXiv preprint arXiv :1812.09168*. [*Cited on page 29.*]

[Businger, 1973] Businger, J. A. (1973). Turbulence transfer in the atmospheric surface layer. In *Workshop on micrometeorology*, pages 67–100. Amer. Meteor. Soc. [*Cited on page 62.*]

[Bérenger, 1994] Bérenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of computational physics*, 114(2) :185–200. [*Cited on pages 11 and 67.*]

[Cameron and Martin, 1947] Cameron, R. H. and Martin, W. T. (1947). The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, pages 385–392. [*Cited on pages 50 and 53.*]

[Cansi, 1995] Cansi, Y. (1995). An automatic seismic event processing for detection and location : The pmcc method. *Geophysical research letters*, 22(9) :1021–1024. [*Cited on pages 2 and 97.*]

[Canuto et al., 2006] Canuto, C., Hussaini, M. Y., Quarteroni, A., and Zang, T. A. (2006). *Spectral methods*. Springer. [*Cited on pages 41 and 44.*]

[Ceranna et al., 2009] Ceranna, L., Le Pichon, A., Green, D. N., and Mialle, P. (2009). The buncefield explosion : a benchmark for infrasound analysis across central europe. *Geophys. J. Int.*, 177(2) :491–508. [*Cited on page 91.*]

[Chastaing et al., 2015] Chastaing, G., Gamboa, F., and Prieur, C. (2015). Generalized sobol sensitivity indices for dependent variables : numerical methods. *Journal of Statistical Computation and Simulation*, 85(7) :1306–1333. [*Cited on page 26.*]

[Cheng et al., 2013a] Cheng, M., Hou, T. Y., and Zhang, Z. (2013a). A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations i : Derivation and algorithms. *Journal of Computational Physics*, 242 :843–868. [*Cited on page 110.*]

[Cheng et al., 2013b] Cheng, M., Hou, T. Y., and Zhang, Z. (2013b). A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations ii : Adaptivity and generalizations. *Journal of Computational Physics*, 242 :753–776. [*Cited on page 110.*]

[Chew et al., 1997] Chew, W., Jin, J., and Michielssen, E. (1997). Complex coordinate system as a generalized absorbing boundary condition. In *Antennas and Propagation Society International Symposium, 1997. IEEE., 1997 Digest*, volume 3, pages 2060–2063. IEEE. [*Cited on page 67.*]

[Choi et al., 2014] Choi, M., Sapsis, T. P., and Karniadakis, G. E. (2014). On the equivalence of dynamically orthogonal and bi-orthogonal methods : Theory and numerical simulations. *Journal of Computational Physics*, 270 :1–20. [*Cited on page 110.*]

[Choi et al., 2004] Choi, S.-K., Grandhi, R. V., Canfield, R. A., and Pettit, C. L. (2004). Polynomial Chaos expansion with Latin Hypercube Sampling for estimating response variability. *AIAA journal*, 42(6) :1191–1198. [*Cited on page 55.*]

[Chunchuzov, 2004] Chunchuzov, I. (2004). Influence of internal gravity waves on sound propagation in the lower atmosphere. *Meteorology and Atmospheric Physics*, 85(1) :61–76. [*Cited on page 61.*]

[Chunchuzov and Kulichkov, 2019] Chunchuzov, I. and Kulichkov, S. (2019). Internal gravity wave perturbations and their impacts on infrasound propagation in the atmosphere. In *Infrasound Monitoring for Atmospheric Studies*, pages 551–590. Springer. [*Cited on pages 3 and 108.*]

[Chunchuzov et al., 2005] Chunchuzov, I., Kulichkov, S., Otrezov, A., and Perepelkin, V. (2005). Acoustic pulse propagation through a fluctuating stably stratified atmospheric boundary layer. *The Journal of the Acoustical Society of America*, 117(4) :1868–1879. [*Cited on page 61.*]

[Chunchuzov et al., 2011] Chunchuzov, I., Kulichkov, S., Popov, O., Waxler, R., and Assink, J. (2011). Infrasound scattering from atmospheric anisotropic inhomogeneities. *Izvestiya, Atmospheric and Oceanic Physics*, 47(5) :540. [*Cited on page 64.*]

[Cohen et al., 2013] Cohen, A., Davenport, M. A., and Leviatan, D. (2013). On the stability and accuracy of least squares approximations. *Foundations of computational mathematics*, 13(5) :819–834. [*Cited on page 56.*]

[Creamer, 2006] Creamer, D. B. (2006). On using polynomial chaos for modeling uncertainty in acoustic propagation. *The Journal of the Acoustical Society of America*, 119(4) :1979–1994. [*Cited on page 73.*]

[Cugnet et al., 2019] Cugnet, D., de la Camara, A., Lott, F., Millet, C., and Ribstein, B. (2019). Non-orographic gravity waves : representation in climate models and effects on infrasound. In *Infrasound Monitoring for Atmospheric Studies*, pages 827–844. Springer. [*Cited on page 8.*]

[Damiens et al., 2018] Damiens, F., Millet, C., and Lott, F. (2018). An investigation of infrasound propagation over mountain ranges. *The Journal of the Acoustical Society of America*, 143(1) :563–574. [*Cited on pages 61, 63, 72, and 108.*]

[de Groot-Hedlin and Hedlin, 2014] de Groot-Hedlin, C. D. and Hedlin, M. A. (2014). Infrasound detection of the chelyabinsk meteor at the usarray. *Earth and Planetary Science Letters*, 402 :337–345. [*Cited on page 16.*]

[de Groot-Hedlin and Hedlin, 2015] de Groot-Hedlin, C. D. and Hedlin, M. A. H. (2015). A method for detecting and locating geophysical events using groups of arrays. *Geophys. J. Int.*, 203(2) :960–971. [*Cited on page 97.*]

[de la Cámara et al., 2014] de la Cámara, A., Lott, F., and Hertzog, A. (2014). Intermittency in a stochastic parameterization of nonorographic gravity waves. *Journal of Geophysical Research : Atmospheres*, 119(21) :11–905. [*Cited on pages 18 and 20.*]

[Deardorff, 1972] Deardorff, J. W. (1972). Numerical investigation of neutral and unstable planetary boundary layers. *Journal of the Atmospheric Sciences*, 29(1) :91–115. [*Cited on page 61.*]

[Demeure, 2021] Demeure, N. (2021). *Gestion du compromis entre la performance et la précision de code de calcul*. Theses, Université Paris-Saclay. [*Cited on page 48.*]

[Demeure et al., 2021] Demeure, N., Chevalier, C., Denis, C., and Dossantos-Uzarralde, P. J. (2021). Encapsulated error : A direct approach to evaluate floating-point accuracy. *Submitted to ACM Transactions on Mathematical Software*. [*Cited on page 47.*]

[Der Kiureghian and Ditlevsen, 2009] Der Kiureghian, A. and Ditlevsen, O. (2009). Aleatory or epistemic ? does it matter ? *Structural Safety*, 31(2) :105–112. [*Cited on page 22.*]

[Drob, 2019] Drob, D. (2019). Meteorology, climatology, and upper atmospheric composition for infrasound propagation modeling. In *Infrasound Monitoring for Atmospheric Studies*, pages 485–508. Springer. [*Cited on page 108.*]

[Drob et al., 2013] Drob, D., Broutman, D., Hedlin, M., Winslow, N., and Gibson, R. (2013). A method for specifying atmospheric gravity wavefields for long-range infrasound propagation calculations. *Journal of Geophysical Research : Atmospheres*, 118(10) :3933–3943. [*Cited on pages 4, 18, 22, and 64.*]

[Drolshagen et al., 2019] Drolshagen, E., Ott, T., Koschny, D., Drolshagen, G., Mialle, P., Pilger, C., Vaubaillon, J., and Poppe, B. (2019). Nemo—a global near real-time fireball monitoring system. In *1st NEO and Debris Detection Conference*. [*Cited on page 97.*]

[Evans and Stark, 2002] Evans, S. N. and Stark, P. B. (2002). Inverse problems as statistics. *Inverse problems*, 18(4) :R55. [*Cited on page 91.*]

[Finette, 2006] Finette, S. (2006). A stochastic representation of environmental uncertainty and its coupling to acoustic wave propagation in ocean waveguides. *The Journal of the Acoustical Society of America*, 120(5) :2567–2579. [*Cited on pages 63 and 73.*]

[Fouque et al., 2007] Fouque, J.-P., Garnier, J., Papanicolaou, G., and Solna, K. (2007). *Wave propagation and time reversal in randomly layered media*, volume 56. Springer Science & Business Media. [*Cited on page 8.*]

[Fricke et al., 2014] Fricke, J. T., Evers, L. G., Smets, P. S. M., Wapenaar, K., and Simons, D. G. (2014). Infrasonic interferometry applied to microbaroms observed at the large aperture infrasound array in the netherlands. *J. Geophys. Res.*, 119(16) :9654–9665. [*Cited on page 92.*]

[Garratt, 1994] Garratt, J. R. (1994). The atmospheric boundary layer. *Earth-Science Reviews*, 37(1-2) :89–134. [*Cited on pages 62 and 63.*]

[Geisser, 1975] Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American statistical Association*, 70(350) :320–328. [*Cited on page 24.*]

[Georg et al., 2018] Georg, N., Ackermann, W., Corno, J., and Schöps, S. (2018). Uncertainty quantification for geometry deformations of superconducting cavities using eigenvalue tracking. *CoRR*. [*Cited on pages 10, 71, and 75.*]

[Gerritsma et al., 2010] Gerritsma, M., Van der Steen, J.-B., Vos, P., and Karniadakis, G. (2010). Time-dependent generalized polynomial chaos. *Journal of Computational Physics*, 229(22) :8333–8363. [*Cited on pages 57 and 73.*]

[Ghanem and Ghosh, 2007] Ghanem, R. and Ghosh, D. (2007). Efficient characterization of the random eigenvalue problem in a polynomial chaos decomposition. *International Journal for Numerical Methods in Engineering*, 72(4) :486–504. [*Cited on pages 58 and 71.*]

[Ghanem and Spanos, 1991] Ghanem, R. G. and Spanos, P. D. (1991). *Stochastic Finite Elements : A Spectral Approach*. Springer-Verlag, Berlin, Heidelberg. [*Cited on pages 7, 50, and 52.*]

[Ghosh and Ghanem, 2008] Ghosh, D. and Ghanem, R. (2008). Stochastic convergence acceleration through basis enrichment of polynomial chaos expansions. *International journal for numerical methods in engineering*, 73(2) :162–184. [*Cited on page 57.*]

[Ghosh and Ghanem, 2012] Ghosh, D. and Ghanem, R. (2012). An invariant subspace-based approach to the random eigenvalue problem of systems with clustered spectrum. *International Journal for Numerical Methods in Engineering*, 91(4) :378–396. [*Cited on pages 10, 58, and 75.*]

[Ghosh et al., 2005] Ghosh, D., Ghanem, R. G., and Red-Horse, J. (2005). Analysis of eigenvalues and modal interaction of stochastic systems. *AIAA journal*, 43(10) :2196–2201. [*Cited on pages 58 and 73.*]

[Ghosh and Iaccarino, 2007] Ghosh, D. and Iaccarino, G. (2007). Applicability of the spectral stochastic finite element method in time-dependent uncertain problems. *Annual Research Briefs of Center for Turbulence Research*, pages 133–141. [*Cited on page 57.*]

[Godin, 2002] Godin, O. A. (2002). An effective quiescent medium for sound propagating through an inhomogeneous, moving fluid. *The Journal of the Acoustical Society of America*, 112(4) :1269–1275. [*Cited on pages 20 and 63.*]

[Gottlieb and Xiu, 2008] Gottlieb, D. and Xiu, D. (2008). Galerkin method for wave equations with uncertain coefficients. *Communications in Computational Physics*, 3(2) :505–518. [*Cited on page 73.*]

[Goursaud, 2010] Goursaud, B. (2010). *Etude mathématique et numérique de guides d'ondes ouverts non uniformes, par approche modale*. PhD thesis, Ecole Polytechnique X. [*Cited on page 68.*]

[Green and Nippress, 2019] Green, D. N. and Nippress, A. (2019). The effects of propagation distance and waveguide structure. *Geophys. J. Int.*, 216 :1974–1988. [*Cited on page 103.*]

[Green et al., 2011] Green, D. N., Vergoz, J., Gibson, R., Le Pichon, A., and Ceranna, L. (2011). Infrasound radiated by the gerdec and chelopechene explosions : Propagation along unexpected paths. *Geophysical Journal International*, 185(2) :890–910. [*Cited on page 64.*]

[Hampton and Doostan, 2015a] Hampton, J. and Doostan, A. (2015a). Coherence motivated sampling and convergence analysis of least squares Polynomial Chaos regression. *Computer Methods in Applied Mechanics and Engineering*, 290 :73–97. [*Cited on page 56.*]

[Hampton and Doostan, 2015b] Hampton, J. and Doostan, A. (2015b). Compressive sampling of polynomial chaos expansions : Convergence analysis and sampling strategies. *Journal of Computational Physics*, 280 :363–386. [*Cited on page 57.*]

[Hazard, 2009] Hazard, C. (2009). A distribution framework for the generalized fourier transform associated with a sturm-liouville operator. Technical report, Rapport de Recherche INRIA 6885. [*Cited on page 35.*]

[Hedlin and Drob, 2014] Hedlin, M. A. and Drob, D. P. (2014). Statistical characterization of atmospheric gravity waves by seismoacoustic observations. *Journal of Geophysical Research : Atmospheres*, 119(9) :5345–5363. [*Cited on pages 4 and 22.*]

[Heuveline and Schick, 2014] Heuveline, V. and Schick, M. (2014). A hybrid generalized polynomial chaos method for stochastic dynamical systems. *International Journal for Uncertainty Quantification*, 4(1). [*Cited on page 57.*]

[Hoeffding, 1948] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, pages 293–325. [*Cited on page 26.*]

[Homma and Saltelli, 1996] Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1) :1–17. [*Cited on page 27.*]

[Hou et al., 2006] Hou, T. Y., Luo, W., Rozovskii, B., and Zhou, H.-M. (2006). Wiener chaos expansions and numerical solutions of randomly forced equations of fluid mechanics. *Journal of Computational Physics*, 216 :687–706. [*Cited on page 73.*]

[Iooss and Prieur, 2017] Iooss, B. and Prieur, C. (2017). Shapley effects for sensitivity analysis with correlated inputs : comparisons with sobol'indices, numerical estimation and applications. *arXiv preprint arXiv :1707.01334*. [*Cited on page 28.*]

[Jensen et al., 1994] Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (1994). *Computational Ocean Acoustics*. Springer New York. [*Cited on pages 62 and 70.*]

[Kac and Siegert, 1947] Kac, M. and Siegert, A. (1947). An explicit representation of a stationary gaussian process. *The Annals of Mathematical Statistics*, 18(3) :438–442. [*Cited on pages 8 and 19.*]

[Kaipio and Somersalo, 2006] Kaipio, J. and Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media. [*Cited on page 91.*]

[Kalaba et al., 1981] Kalaba, R., Spingarn, K., and Tesfatsion, L. (1981). Individual tracking of an eigenvalue and eigenvector of a parameterized matrix. *Nonlinear Analysis : Theory, Methods & Applications*, 5(4) :337–340. [*Cited on pages 10 and 71.*]

[Karhunen, 1947] Karhunen, K. (1947). *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, volume 37. Sana. [*Cited on pages 8 and 19.*]

[Kato, 2013] Kato, T. (2013). *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media. [*Cited on page 8.*]

[Khine et al., 2010] Khine, Y. Y., Creamer, D. B., and Finette, S. (2010). Acoustic propagation in an uncertain waveguide environment using stochastic basis expansions. *Journal of Computational Acoustics*, 18(04) :397–441. [*Cited on page 73.*]

[Klyatskin and Tatarskii, 1970] Klyatskin, V. and Tatarskii, V. (1970). The parabolic equation approximation for propagation of waves in a medium with random inhomogeneities. *Sov. Phys. JETP*, 31(2) :335–339. [*Cited on pages 4 and 22.*]

[Koltchinskii, 2001] Koltchinskii, V. (2001). Rademacher penalties and structural risk minimization. *IEEE Transactions on Information Theory*, 47(5) :1902–1914. [*Cited on page 24.*]

[Kosyakov et al., 2019] Kosyakov, S., Kulichkov, S., and Chunchuzov, I. (2019). Influence of the stability of the atmospheric boundary layer on the parameters of propagating acoustic waves. *Acoustical Physics*, 65(4) :407–417. [*Cited on page 61.*]

[Kulichkov et al., 2010] Kulichkov, S., Chunchuzov, I., and Popov, O. (2010). Simulating the influence of an atmospheric fine inhomogeneous structure on long-range propagation of pulsed acoustic signals. *Izvestiya, Atmospheric and Oceanic Physics*, 46(1) :60–68. [*Cited on page 64.*]

[Lalande et al., 2012] Lalande, J. M., Sebe, O., Landes, M., Blanc-Benon, P., Matoza, R. S., Le Pichon, A., and Blanc, E. (2012). Infrasound data inversion for atmospheric sounding. *Geophys. J. Int.*, 190(1) :687–701. [*Cited on page 92.*]

[Lalande and Waxler, 2016] Lalande, J.-M. and Waxler, R. (2016). The interaction between infrasonic waves and gravity wave perturbations : Application to observations using uttr rocket motor fuel elimination events. *Journal of Geophysical Research : Atmospheres*, 121(10) :5585–5600. [*Cited on pages 4, 22, and 33.*]

[Langenhove et al., 2016] Langenhove, J. V., Lucor, D., and Belme, A. (2016). Robust uncertainty quantification using preconditioned least-squares polynomial approximations with ℓ_1-regularization. *International Journal for Uncertainty Quantification*, 6(1) :57–77. [*Cited on page 56.*]

[Le Maître et al., 2010] Le Maître, O., Mathelin, L., Knio, O., and Hussaini, M. (2010). Asynchronous time integration for polynomial chaos expansion of uncertain periodic dynamics. *Discrete Contin. Dyn. Syst*, 28 :199–226. [*Cited on pages 9, 56, and 57.*]

[Le Maître et al., 2007] Le Maître, O. P., Najm, H. N., Pébay, P. P., Ghanem, R. G., and Knio, O. M. (2007). Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM Journal on Scientific Computing*, 29(2) :864–889. [*Cited on page 57.*]

[Le Pichon et al., 2010] Le Pichon, A., Blanc, E., and Hauchecorne, A. (2010). *Infrasound monitoring for atmospheric studies*. Springer Science & Business Media. [*Cited on page 18.*]

[Lie et al., 2018] Lie, H. C., Sullivan, T., and Teckentrup, A. L. (2018). Random forward models and log-likelihoods in bayesian inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4) :1600–1629. [*Cited on page 96.*]

[Loeve, 1948] Loeve, M. (1948). Fonctions aléatoires du second ordre, supplement to p. *Lévy, Processus Stochastiques et Mouvement Brownien, Gauthier-Villars, Paris*. [*Cited on pages 8 and 19.*]

[Lott et al., 2012] Lott, F., Guez, L., and Maury, P. (2012). A stochastic parameterization of non-orographic gravity waves : Formalism and impact on the equatorial stratosphere. *Geophysical Research Letters*, 39(6). [*Cited on page 18.*]

[Lott and Millet, 2010] Lott, F. and Millet, C. (2010). The representation of gravity waves in atmospheric general circulation models (gcms). In *Infrasound monitoring for atmospheric studies*, pages 685–699. Springer. [*Cited on page 18.*]

[Luchtenburg et al., 2014] Luchtenburg, D. M., Brunton, S. L., and Rowley, C. W. (2014). Long-time uncertainty propagation using generalized polynomial chaos and flow map composition. *Journal of Computational Physics*, 274 :783–802. [*Cited on page 57.*]

[Lucor and Karniadakis, 2004] Lucor, D. and Karniadakis, G. E. (2004). Adaptive generalized polynomial chaos for nonlinear random oscillators. *SIAM Journal on Scientific Computing*, 26(2) :720–735. [*Cited on pages 57 and 73.*]

[Mahrt, 1989] Mahrt, L. (1989). Intermittency of atmospheric turbulence. *Journal of the Atmospheric Sciences*, 46(1) :79–95. [*Cited on page 62.*]

[Mahrt, 1998] Mahrt, L. (1998). Stratified atmospheric boundary layers and breakdown of models. *Theoretical and computational fluid dynamics*, 11(3-4) :263–279. [*Cited on page 62.*]

[Mai et al., 2016] Mai, C., Spiridonakos, M., Chatzi, E., and Sudret, B. (2016). Surrogate modelling for stochastic dynamical systems by combining narx models and polynomial chaos expansions. *arXiv preprint arXiv :1604.07627*. [*Cited on page 57.*]

[Mai, 2016] Mai, C. V. (2016). *Polynomial chaos expansions for uncertain dynamical systems-applications in earthquake engineering*. PhD thesis, ETH Zurich. [*Cited on page 56.*]

[Mai and Sudret, 2017] Mai, C. V. and Sudret, B. (2017). Surrogate models for oscillatory systems using sparse polynomial chaos expansions and stochastic time warping. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1) :540–571. [*Cited on pages 9 and 57.*]

[Marcillo et al., 2014] Marcillo, O., Arrowsmith, S., Whitaker, R., Anderson, D., Nippress, A., Green, D. N., and Drob, D. (2014). Using physics-based priors in a bayesian algorithm to enhance infrasound source location. *Geophys. J. Int.*, 196(1) :375–358. [*Cited on pages 92 and 95.*]

[Marrel et al., 2009] 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. [*Cited on page 27.*]

[Marzouk and Xiu, 2009] Marzouk, Y. and Xiu, D. (2009). A stochastic collocation approach to bayesian inference in inverse problems. [*Cited on pages 11, 91, 92, 95, and 96.*]

[Marzouk and Najm, 2009] Marzouk, Y. M. and Najm, H. N. (2009). Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6) :1862–1902. [*Cited on pages 11, 91, and 92.*]

[Marzouk et al., 2007] Marzouk, Y. M., Najm, H. N., and Rahn, L. A. (2007). Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2) :560–586. [*Cited on pages 11, 91, and 92.*]

[Matthies, 2007] Matthies, H. G. (2007). Quantifying uncertainty : modern computational representation of probability and applications. In *Extreme man-made and natural hazards in dynamics of structures*, pages 105–135. Springer. [*Cited on page 22.*]

[McKay et al., 1979] McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2) :239–245. [*Cited on page 23.*]

[Mialle and Arora, 2018] Mialle, P. and Arora, N. S. (2018). Progresses with the international data centre infrasound system. *Journal of the Acoustical Society of America*, 143(3) :1782–1782. [*Cited on page 91.*]

[Mialle et al., 2019] Mialle, P., Brown, D., Arora, N., et al. (2019). Advances in operational processing at the international data centre. In *Infrasound Monitoring for Atmospheric Studies*, pages 209–248. Springer. [*Cited on pages 2, 11, and 91.*]

[Migliorati et al., 2014] Migliorati, G., Nobile, F., von Schwerin, E., and Tempone, R. (2014). Analysis of discrete ℓ_2 projection on polynomial spaces with random evaluations. *Foundations of Computational Mathematics*, 14(3) :419–456. [*Cited on page 56.*]

[Millet et al., 2007] Millet, C., Robinet, J.-C., and Roblin, C. (2007). On using computational aeroacoustics for long-range propagation of infrasounds in realistic atmospheres. *Geophysical Research Letters*, 34(14). [*Cited on pages 4, 22, and 64.*]

[Modrak et al., 2010] Modrak, R. T., Arrowsmith, S. J., and Anderson, D. N. (2010). A bayesian framework for infrasound location. *Geophys. J. Int.*, 181(1) :399–405. [*Cited on page 91.*]

[Mohammad-Djafari, 1996] Mohammad-Djafari, A. (1996). A full bayesian approach for inverse problems. In *Maximum entropy and Bayesian methods*, pages 135–144. Springer. [*Cited on page 91.*]

[Monin, 1970] Monin, A. (1970). The atmospheric boundary layer. *Annual Review of Fluid Mechanics*, 2(1) :225–250. [*Cited on page 62.*]

[Moore and Russell, 2017] Moore, D. A. and Russell, S. J. (2017). Signal-based bayesian seismic monitoring. In W&CP, J., editor, *20th International Conference on Artificial Inbtelligence and Statistics (AISTATS)*, volume 54. [*Cited on pages 91, 92, and 111.*]

[Morton and Arrowsmith, 2014] Morton, E. A. and Arrowsmith, S. J. (2014). The development of global probabilistic propagation look-up tables for infrasound celerity and back-azimuth deviation. *Seismol. Res. Lett.*, 85(6) :1223–1233. [*Cited on pages 92 and 95.*]

[Nippress et al., 2014] Nippress, A., Green, D. N., Marcillo, O. E., and Arrowsmith, S. J. (2014). Generating regional infrasound celerity-range models using ground-truth information and the implications for event location. *Geophys. J. Int.*, 197(2) :1154–1165. [*Cited on page 95.*]

[Nouy, 2010] Nouy, A. (2010). Identification of multi-modal random variables through mixtures of polynomial chaos expansions. *Comptes Rendus Mécanique*, 338(12) :698–703. [*Cited on pages 11 and 79.*]

[Olver, 1974] Olver, F. (1974). 5 - differential equations with regular singularities ; hypergeometric and legendre functions. In Olver, F., editor, *Asymptotics and Special Functions*, pages 139 – 189. Academic Press. [*Cited on page 45.*]

[Olyslager, 2004] Olyslager, F. (2004). Discretization of continuous spectra based on perfectly matched layers. *SIAM Journal on Applied Mathematics*, 64(4) :1408–1433. [*Cited on pages 11, 67, and 68.*]

[Ostashev et al., 2019] Ostashev, V. E., Muhlestein, M. B., and Wilson, D. K. (2019). Extra-wide-angle parabolic equations in motionless and moving media. *The Journal of the Acoustical Society of America*, 145(2) :1031–1047. [*Cited on pages 4 and 22.*]

[Owen, 2014] Owen, A. B. (2014). Sobol'indices and shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1) :245–251. [*Cited on page 28.*]

[Ozen and Bal, 2017] Ozen, H. C. and Bal, G. (2017). A dynamical polynomial chaos approach for long-time evolution of SPDEs. *Journal of Computational Physics*, 343 :300 – 323. [*Cited on page 73.*]

[Park et al., 2014] Park, J., Arrowsmith, S. J., Hayward, C., Stump, B. W., and Blom, P. (2014). Automatic infrasound detection and location of sources in the western united states. *J. Geophys. Res.*, 119(13) :7773–7798. [*Cited on page 91.*]

[Pettit and Beran, 2006] Pettit, C. and Beran, P. (2006). Spectral and multiresolution wiener expansions of oscillatory stochastic processes. *Journal of Sound and Vibration*, 294(4-5) :752–779. [*Cited on pages 9, 56, and 57.*]

[Pierce, 1965] Pierce, A. D. (1965). Extension of the method of normal modes to sound propagation in an almost-stratified medium. *The Journal of the Acoustical Society of America*, 37(1) :19–27. [*Cited on page 109.*]

[Pierre, 1988] Pierre, C. (1988). Mode localization and eigenvalue loci veering phenomena in disordered structures. [*Cited on page 71.*]

[Pilger et al., 2020] Pilger, C., Gaebler, P., Hupe, P., Ott, T., and Drolshagen, E. (2020). Global monitoring and characterization of infrasound signatures by large fireballs. *Atmosphere*, 11(1) :83. [*Cited on pages 97, 98, and 100.*]

[Rahman and Yadav, 2011] Rahman, S. and Yadav, V. (2011). Orthogonal polynomial expansions for solving random eigenvalue problems. *International Journal for Uncertainty Quantification*, 1(2). [*Cited on page 75.*]

[Resmini et al., 2016] Resmini, A., Peter, J., and Lucor, D. (2016). Sparse grids-based stochastic approximations with applications to aerodynamics sensitivity analysis. *International Journal for Numerical Methods in Engineering*, 106(1) :32–57. [*Cited on page 56.*]

[Rojano et al., 2019] Rojano, R. M., Mendez, S., Lucor, D., Ranc, A., Giansily-Blaizot, M., Schved, J.-F., and Nicoud, F. (2019). Kinetics of the coagulation cascade including the contact activation system : sensitivity analysis and model reduction. *Biomechanics and Modeling in Mechanobiology*, 18 :1139–1153. [*Cited on page 108.*]

[Sabatini et al., 2016] Sabatini, R., Marsden, O., Bailly, C., and Bogey, C. (2016). A numerical study of nonlinear infrasound propagation in a windy atmosphere. *The Journal of the Acoustical Society of America*, 140(1) :641–656. [*Cited on page 22.*]

[Sabatini et al., 2019a] Sabatini, R., Marsden, O., Bailly, C., and Gainville, O. (2019a). Three-dimensional direct numerical simulation of infrasound propagation in the earth's atmosphere. *Journal of Fluid Mechanics*, 859 :754–789. [*Cited on page 22.*]

[Sabatini et al., 2019b] Sabatini, R., Snively, J. B., and Hickey, M. P. (2019b). An investigation of the scattering of infrasonic acoustic waves by turbulent fluctuations generated by the breaking of gravity waves. *The Journal of the Acoustical Society of America*, 145(3) :1901–1901. [*Cited on page 22.*]

[Saltelli, 2002] Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer physics communications*, 145(2) :280–297. [*Cited on page 27.*]

[Saltelli et al., 2000] Saltelli, A., Chan, K., and Scott, E. (2000). *Sensitivity analysis*. J. Wiley & Sons. [*Cited on page 92.*]

[Saltelli et al., 2008] 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. [*Cited on page 26.*]

[Sapsis and Lermusiaux, 2009] Sapsis, T. P. and Lermusiaux, P. F. (2009). Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D : Nonlinear Phenomena*, 238(23-24) :2347–2360. [*Cited on page 110.*]

[Sobol', 1967] Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4) :784–802. [*Cited on page 23.*]

[Sobol', 1993] Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical modelling and computational experiments*, 1(4) :407–414. [*Cited on pages 26 and 27.*]

[Soize, 2015] Soize, C. (2015). Polynomial chaos expansion of a multimodal random vector. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1) :34–60. [*Cited on pages 11 and 79.*]

[Soize and Ghanem, 2004] Soize, C. and Ghanem, R. (2004). Physical systems with random uncertainties : chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing*, 26(2) :395–410. [*Cited on page 50.*]

[Song et al., 2016] Song, E., Nelson, B. L., and Staum, J. (2016). Shapley effects for global sensitivity analysis : Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1) :1060–1083. [*Cited on pages 28 and 29.*]

[Spiridonakos and Chatzi, 2015] Spiridonakos, M. D. and Chatzi, E. N. (2015). Metamodeling of dynamic nonlinear structural systems through polynomial chaos narx models. *Computers & Structures*, 157 :99–113. [*Cited on page 57.*]

[Stone, 1974] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society : Series B (Methodological)*, 36(2) :111–133. [*Cited on page 24.*]

[Sudret, 2008] Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7) :964–979. [*Cited on page 27.*]

[Tarantola, 2005] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*, volume 89. siam. [*Cited on page 91.*]

[Trefethen, 2000] Trefethen, L. N. (2000). *Spectral methods in MATLAB*, volume 10. Siam. [*Cited on page 41.*]

[Trefethen and Embree, 2005] Trefethen, L. N. and Embree, M. (2005). *Spectra and pseudospectra : the behavior of nonnormal matrices and operators*. Princeton University Press. [*Cited on page 39.*]

[Valiant, 1984] Valiant, L. G. (1984). A theory of the learnable. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 436–445. ACM. [*Cited on page 24.*]

[Vanderbecken et al., 2020] Vanderbecken, P. J., Mahfouf, J.-F., and Millet, C. (2020). Bayesian selection of atmospheric profiles from an ensemble data assimilation system using infrasonic observations of may 2016 mount etna eruptions. *J. Geophys. Res. Atmos.*, 125(2) :e2019JD031168. [*Cited on page 97.*]

[Vapnik, 1995] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York. [*Cited on pages 23 and 92.*]

[Wan and Karniadakis, 2005] Wan, X. and Karniadakis, G. E. (2005). An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics*, 209(2) :617–642. [*Cited on page 79.*]

[Wan and Karniadakis, 2006] Wan, X. and Karniadakis, G. E. (2006). Long-term behavior of polynomial chaos in stochastic flow simulations. *Computer methods in applied mechanics and engineering*, 195(41-43) :5582–5596. [*Cited on pages 56 and 73.*]

[Waxler, 2002] Waxler, R. (2002). A vertical eigenfunction expansion for the propagation of sound in a downward-refracting atmosphere over a complex impedance plane. *The Journal of the Acoustical Society of America*, 112(6) :2540–2552. [*Cited on pages 11, 34, 37, 62, and 66.*]

[Waxler, 2003] Waxler, R. (2003). On the use of modal expansions to model broadband propagation in the nighttime boundary layer and other downward refracting atmospheres over lossy ground planes. *The Journal of the Acoustical Society of America*, 113(4) :2313–2313. [*Cited on pages 11, 36, and 66.*]

[Waxler, 2004] Waxler, R. (2004). Modal expansions for sound propagation in the nocturnal boundary layer. *The Journal of the Acoustical Society of America*, 115(4) :1437–1448. [*Cited on pages 62 and 63.*]

[Waxler and Assink, 2019] Waxler, R. and Assink, J. (2019). Propagation modeling through realistic atmosphere and benchmarking. In *Infrasound Monitoring for Atmospheric Studies*, pages 509–549. Springer. [*Cited on pages 3 and 108.*]

[Waxler et al., 2017] Waxler, R., Assink, J., and Velea, D. (2017). Modal expansions for infrasound propagation and their implications for ground-to-ground propagation. *The Journal of the Acoustical Society of America*, 141(2) :1290–1307. [*Cited on pages 22, 33, 65, and 108.*]

[Waxler et al., 2015] Waxler, R., Evers, L. G., Assink, J., and Blom, P. (2015). The stratospheric arrival pair in infrasound propagation. *The Journal of the Acoustical Society of America*, 137(4) :1846–1856. [*Cited on pages 63 and 103.*]

[Waxler et al., 2008] Waxler, R., Gilbert, K. E., and Talmadge, C. (2008). A theoretical treatment of the long range propagation of impulsive signals under strongly ducted nocturnal conditions. *The Journal of the Acoustical Society of America*, 124(5) :2742–2754. [*Cited on pages 62, 63, and 109.*]

[Wegman, 1972] Wegman, E. J. (1972). Nonparametric probability density estimation : I. a summary of available methods. *Technometrics*, 14(3) :533–546. [*Cited on page 17.*]

[Whiteman et al., 1997] Whiteman, C. D., Bian, X., and Zhong, S. (1997). Low-level jet climatology from enhanced rawinsonde observations at a site in the southern great plains. *Journal of Applied Meteorology*, 36(10) :1363–1376. [*Cited on page 62.*]

[Wiener, 1938] Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, 60(4) :897–936. [*Cited on pages 7, 50, and 52.*]

[Wilson et al., 2003] Wilson, D. K., Noble, J. M., and Coleman, M. A. (2003). Sound propagation in the nocturnal boundary layer. *Journal of the atmospheric sciences*, 60(20) :2473–2486. [*Cited on page 61.*]

[Witteveen and Bijl, 2008] Witteveen, J. A. and Bijl, H. (2008). An alternative unsteady adaptive stochastic finite elements formulation based on interpolation at constant phase. *Computer Methods in Applied Mechanics and Engineering*, 198(3-4) :578–591. [*Cited on page 57.*]

[Xiu and Karniadakis, 2002] Xiu, D. and Karniadakis, G. E. (2002). The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2) :619–644. [*Cited on pages 50 and 53.*]

[Zhu and Chen, 2013] Zhu, J. and Chen, Z. (2013). On the validity of modal expansion in pekeris waveguide with pml. *Journal of Applied & Computational Mathematics*, 2(1) :–. [*Cited on page 67.*]

université PARIS-SACLAY

ÉCOLE DOCTORALE
de mathématiques
Hadamard (EDMH)

Titre : Une approche spectrale de la métamodélisation multi-échelle appliquée à la propagation acoustique

Mots clés : infrasons - modes normaux - polynôme de chaos - propagation en milieu aléatoire.

Résumé : De nombreuses méthodes permettent de simuler numériquement la propagation d'une onde dans un milieu complexe avec une excellente précision. Cependant, la prise en compte des fluctuations du milieu de propagation requiert un traitement statistique nécessitant un grand nombre d'appel à des codes de calcul souvent coûteux. Afin de rendre accessible ces études nous proposons la construction d'un métamodèle basé sur une décomposition en polynômes de chaos des modes normaux. Cette approche permet de restituer les statistiques des signaux se propageant dans un milieu aléatoire avec un coût de calcul moindre.
Les applications proposées dans cette thèse concernent la propagation d'ondes acoustiques dans l'atmosphère terrestre. En effet, les fluctuations météorologiques modifiant considérablement les conditions de propagation, leur prise en compte est indispensable. Le coût numérique de la simulation sur un domaine de plusieurs centaines de milliers de kilomètres carrés justifie pleinement l'utilisation d'un métamodèle. Une application à la localisation de source couplant ces techniques de métamodèlisation avec une approche bayésienne est aussi proposée. En effet, le cadre bayésien permet une résolution du problème inverse dans un cadre probabiliste capable de prendre en compte les fluctuations du milieu et l'incertitude sur la localisation de la source.

Title : A Spectral Approach of Multi-Scale Metamodelling Applied to Acoustic Propagation

Keywords : infrasound - normal modes - polynomial chaos - propagation in random media.

Abstract : There exists many numerical methods to simulate wave propagation through complex media with a very good precision. However, taking into account the fluctuations of the propagation medium necessitates a statistical approach implying a prohibitive numerical cost. To have those studies affordable, we propose the construction of a metamodel based on a polynomial chaos decomposition of normal modes. This approach presents the great advantage to give statistics of signals propagating in a random medium at an affordable numerical cost.
Those results are illustrated with acoustic propagation in the atmosphere. In fact, meteorological fluctuations have a critical impact on the propagation, it is therefore essential to take them into account. The numerical cost of a simulation over thousands of kilometers fully justifies the use of a metamodel. An application to source localization is proposed to illustrate the joint use of a metamodel and a bayesian inversion. The bayesian framework allows a resolution of the inverse problem in a probabilistic context able to take into account the fluctuations of the medium and uncertainties due to unknown source localization.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France