A surrogate model of elastic wave propagation to quantify uncertainties in seismic hazard analysis

Fanny Lehmann

université PARIS-SACLAY

A surrogate model of elastic wave propagation to quantify uncertainties in seismic hazard analysis

Un méta-modèle de la propagation des ondes élastiques pour quantifier les incertitudes liées à l'aléa sismique

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 579 : sciences mécaniques et énergétiques, matériaux et géosciences (SMEMaG)
Spécialité de doctorat : Mécanique des solides et des structures
Graduate School : Sciences de l'ingénierie et des systèmes
Référent : CentraleSupélec

Thèse préparée au **LMPS - Laboratoire de Mécanique Paris-Saclay (Université Paris-Saclay, CentraleSupélec, ENS Paris-Saclay, CNRS)**,
sous la direction de **Didier CLOUTEAU**, Professeur des universités,
la co-direction de **Filippo GATTI**, Maître de conférences
et la co-supervision de **Michaël BERTIN**, Ingénieur-chercheur

Thèse soutenue à Paris-Saclay, le 28 juin 2024, par

Fanny LEHMANN

Composition du jury

Membres du jury avec voix délibérative

Mathilde MOUGEOT Professeure des universités, ENS Paris-Saclay	Présidente
Yann CAPDEVILLE Directeur de recherche, Nantes Université	Rapporteur & Examinateur
Roger GHANEM Professor, University of South California	Rapporteur & Examinateur
Sylvie LE HEGARAT Professeure des universités, Université Paris-Sud	Examinatrice
Siddhartha MISHRA Professor, ETH Zürich	Examinateur

THESE DE DOCTORAT

NNT : 2024UPAST074

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences mécaniques et énergétiques, matériaux et géosciences (SMEMAG)

Titre: Un méta-modèle de la propagation des ondes élastiques pour quantifier les incertitudes liées à l'aléa sismique

Mots clés: équation d'ondes élastique; aléa sismique; intelligence artificielle; méta-modèle; transfert d'apprentissage; quantification d'incertitudes

Résumé: La propagation des ondes sismiques dans le sol est soumise à de nombreuses sources d'incertitudes, allant de l'activité incertaine des failles géologiques à notre connaissance imparfaite des propriétés mécaniques à l'intérieur de la croûte terrestre. Pour évaluer correctement l'aléa sismique, il est donc essentiel de quantifier l'influence des incertitudes sur l'intensité des mouvements du sol générés par les séismes.

Dans les zones à sismicité faible à modérée, comme la plupart des régions de France métropolitaine, les enregistrements sismiques ne sont pas suffisants pour évaluer les incertitudes des mouvements du sol. Dans cette situation, les simulations numériques sont la seule option pour estimer l'intensité du mouvement du sol, mais les coûts de calcul élevés rendent impossible la plupart des analyses d'incertitude. Dans cette thèse, on propose un méta-modèle qui permet de remplacer le solveur numérique en réduisant considérablement les coûts de calcul tout en conservant sa flexibilité et une précision satisfaisante.

On illustre tout d'abord l'influence des hétérogénéités géologiques sur l'intensité du mouvement du sol dans le contexte du séisme du Teil (Mw4.9, Ardèche, France, 2019). Des hétérogénéités sont ajoutées à un modèle géologique régional sous la forme de champs aléatoires et on montre que le mouvement sismique ainsi généré est plus réaliste. Cependant, les hétérogénéités induisent également une grande variabilité entre les réalisations.

Pour étudier cette variabilité de manière systématique, on élabore une base de données de 30 000 modèles géologiques hétérogènes en 3D et à l'intérieur de chaque géologie, des ondes sismiques sont propagées à partir d'une source aléatoire en utilisant le code d'éléments spectraux SEM3D. La base de données est ensuite utilisée pour entraîner le méta-modèle.

Notre méta-modèle est une extension de l'opérateur neuronal de Fourier appelée opérateur neuronal de Fourier à entrées multiples (MIFNO). Le MIFNO prend en entrée une géologie 3D et un vecteur de paramètres de source pour prédire le mouvement du sol en 3D. Le mouvement du sol est un champ d'ondes enregistré à la surface du domaine en fonction du temps. Il est prédit en une seule itération grâce à une conversion profondeur-temps. On caractérise ensuite l'erreur de prédiction du MIFNO et on explore sa capacité de généralisation aux données hors de la distribution d'entraînement.

Enfin, on utilise l'apprentissage par transfert pour améliorer la précision du MIFNO dans le contexte du séisme du Teil. Avec ce méta-modèle spécifique, on obtient des distributions statistiques de plusieurs quantités d'intérêt pour l'aléa sismique. Ces distributions sont cohérentes avec les simulations numériques et fournissent des intervalles de confiance qui étaient hors de portée avec les méthodes précédentes.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences mécaniques et énergétiques, matériaux et géosciences (SMEMAG)

Title: A surrogate model of elastic wave propagation to quantify uncertainties in seismic hazard analysis
Keywords: elastic wave equation; seismic hazard assessment; artificial intelligence; surrogate model; transfer learning; uncertainty quantification

Abstract: The propagation of seismic waves in the ground is subject to many sources of uncertainties, ranging from the uncertain activity of geological faults to the incomplete knowledge of mechanical properties inside the Earth's crust. To properly assess seismic hazard, it then becomes essential to quantify how uncertainties influence the intensity of ground motion generated by earthquakes.

In areas with low-to-moderate seismicity, like most regions in metropolitan France, seismic records are too sparse to evaluate ground motion uncertainties. In this situation, numerical simulations are the only option to estimate ground motion intensity, but their high computational costs prevent most uncertainty analyses. In this thesis, we design a surrogate model that can replace the numerical solver by drastically reducing the computational costs while preserving its flexibility and a satisfying accuracy.

We first illustrate the influence of geological heterogeneities on ground motion intensity in the context of the Mw4.9 Le Teil earthquake (Ardèche, France, 2019). Heterogeneities are added to a regional geological model in the form of random fields, and we show that it generates more realistic ground motion. However, heterogeneities also lead to a large variability between samples.

To study this variability systematically, we build a database of 30 000 heterogeneous 3D geological models, and inside each geology, seismic waves are propagated from a random source using the spectral element code SEM3D. The database is then used to train a surrogate model in a purely data-driven framework.

To design the surrogate model, we propose an extension of the Fourier Neural Operator called the Multiple Input Fourier Neural Operator (MIFNO). The MIFNO takes as inputs a 3D geology and a vector of source parameters to predict 3D ground motion. Ground motion is a time-dependent surface wavefield, but we do not need any time iteration thanks to a depth-to-time conversion. We characterize the MIFNO prediction error and explore the MIFNO generalization ability to out-of-distribution data.

We finally take advantage of transfer learning to further improve the MIFNO accuracy in the context of the Le Teil earthquake. With this fine-tuned surrogate model, we obtain statistical distributions of several quantities of interest in seismic hazard assessment. They are coherent with numerical simulations and provide confidence intervals that were out of reach with existing methods.

A Alba,
Rêve ta vie en grandes lettres colorées et ose vivre tes rêves

Contents

Acknowledgements

I would like to start by expressing my deepest gratitude to the members of the jury who gave me the honour of assessing my work; Prof. Mathilde Mougeot, for presiding the jury and for your advice from the MASCOT-NUM workshop where we first met; Prof. Roger Ghanem and Prof. Yann Capdeville, thank you for your time reading my manuscript and your encouraging report; Prof. Sylvie Le Hegarat and Prof. Mishra, for sharing your expertise on machine learning. Your questions and suggestions are immensely valuable.

I am deeply indebted to my supervisors, Didier Clouteau, Filippo Gatti, and Michaël Bertin, for their continuous support and precious guidance during those three years. Working under your supervision was an absolute pleasure that made me learn and grow, as a woman and as the researcher I have become. You challenged me to open new scientific doors and encouraged me to embrace opportunities. I am privileged to have shared enlightening moments with you.

I would also like to extend my sincere thanks to Fernando Lopez-Caballero and Pierre-Alain Boucard for welcoming me to the Laboratoire de Mécanique Paris-Saclay. I am grateful to Nicolas Lardjane and Amaury Vallage for the support and resources of the CEA without which this research would not have been possible.

My sincere appreciation goes to Divya Madhavan, Melissa Ann-Thomas, Andrew Crawford, and Calvin Peck at the Academic Writing Center for the immensely rich resources you offer us and the amazing 3-minute thesis experience we shared. Your workshops and coaching sessions have been truly transformative. I am also grateful to the mentorship program of the Université Paris-Saclay, and especially my mentor, Anaïs Pitto-Barry, for helping me navigate the academic world and anticipate the post-PhD path. Our discussions were helpful breathes along this PhD journey.

I would also like to thank the many colleagues and friends I met. Sharing office, meals, and activities with you made my PhD even more enjoyable. Valeria, Hugo, Exneyder, Juan Camillo, Julian, Yilun, Briac, Valentin, Ali, Gianluca, Mouhammed, Kuralay, Min, Eloi, Wilfried, Jean-Baptiste, Gottfried, Gerardo, Camille, Héloïse, and all the members of the OMEIR team. Benjamin, Pierre L., Pierre A., Alice, Aurélie, Marine, Jérôme, Clara, Elodie, Nadège, and all my colleagues at the CEA. Special thanks are due to Christophe Millet for your collaboration.

And lastly, I am infinitely grateful to my family and friends, whose unlimited love and support have always been essential. Your achievements are a daily motivation, and I am proud to share a little bit of my world with you.

Résumé étendu

La propagation des ondes sismiques dans le sol est soumise à de nombreuses sources d'incertitudes, allant de l'activité incertaine des failles géologiques à notre connaissance imparfaite des propriétés mécaniques à l'intérieur de la croûte terrestre. Pour évaluer correctement l'aléa sismique, il est donc essentiel de quantifier l'influence des incertitudes sur l'intensité des mouvements du sol générés par les séismes. Dans les zones à sismicité faible à modérée, comme la plupart des régions de France métropolitaine, les enregistrements sismiques ne sont pas suffisants pour évaluer les incertitudes des mouvements du sol. Dans cette situation, les simulations numériques sont la seule option pour estimer l'intensité du mouvement du sol, mais les coûts de calcul élevés rendent impossible la plupart des analyses d'incertitude.

On illustre tout d'abord l'influence des hétérogénéités géologiques sur l'intensité du mouvement du sol dans le contexte du séisme du Teil (Mw4.9, Ardèche, France, 2019). Des simulations numériques sont réalisées dans une région de 50 km $\times$ 115 km avec deux modèles géologiques nationaux (un modèle unidimensionnel, 1D, et un modèle tridimensionnel, 3D) et deux modèles de source sismique (un point source et un modèle de rupture dynamique du plan de faille). Le code SEM3D basé sur la méthode des éléments spectraux est employé pour les simulations. Les deux modèles géologiques permettent de reproduire les caractéristiques principales du mouvement du sol enregistré. Cependant, le modèle unidimensionnel crée un guide d'ondes absent des sismogrammes et le modèle tridimensionnel induit des amplitudes plus élevées que les données enregistrées du fait d'un manque de fluctuations géologiques. Des hétérogénéités sont alors ajoutées au modèle géologique unidimensionnel sous la forme de champs aléatoires et on montre que le mouvement sismique ainsi généré est plus réaliste. En particulier, les effets du guide d'ondes sont atténués par la présence des hétérogénéités. Cependant, les hétérogénéités induisent également une grande variabilité entre les trois réalisations de champs aléatoires, ce qui montre la nécessité d'une analyse quantitative de la variabilité ainsi obtenue.

Pour étudier cette variabilité de manière systématique, on élabore une base de données de 30 000 modèles géologiques hétérogènes en 3D. Les modèles géologiques sont construits comme la superposition de couches horizontales choisies aléatoirement, ce qui représente la structure 1D de la géologie. Ensuite, des champs aléatoires sont ajoutés indépendamment dans chaque couche. A l'intérieur de chaque géologie, une source sismique ponctuelle est placée de manière aléatoire avec une orientation aléatoire. Les ondes sismiques sont alors propagées avec le code SEM3D, depuis la source jusqu'à la surface où elles sont synthétisées par une grille de capteurs virtuels. On constitue ainsi la base de données HEMEW^S-3D de 30 000 triplets (géologie 3D, propriétés de source, mouvements du sol en surface). Pour caractériser la base de données, plusieurs définitions de la dimension intrinsèque des données sont explorées : l'analyse en composantes principales, une estimation à partir du maximum de vraisemblance, la dimension de corrélation et un indice de similarité structurelle. Si les valeurs de la dimension intrinsèque fluctuent grandement selon les méthodes, il apparaît néanmoins clairement que les données sont de grande dimension. La base de données HEMEW^S-3D est ensuite utilisée pour entraîner le méta-modèle, ce qui constitue le cœur de cette thèse.

Notre méta-modèle est une extension de l'opérateur neuronal de Fourier appelée opérateur neuronal de Fourier à entrées multiples (*Multiple Input Fourier Neural Operator*, MIFNO). Le MIFNO prend en entrée une géologie 3D et un vecteur de paramètres de source pour prédire le mouvement du sol en 3D. Le premier point important du MIFNO est d'utiliser une architecture factorisée dans laquelle les transformées de Fourier sont effectuées séparément sur les trois axes. Cela permet de traiter des données 3D avec un nombre de paramètres raisonnable. Deuxièmement, le mouvement du sol est un champ d'ondes enregistré à la surface du domaine en fonction du temps. Il est prédit en une seule itération grâce à une conversion profondeur-temps. Ainsi, le MIFNO prédit la solution d'une équation aux dérivées partielles en 3D dépendant du temps sans nécessiter de variables en quatre dimensions. La conversion profondeur-temps évite également le recours à des méthodes auto-régressives qui ont tendance à diminuer la précision en temps long. La troisième caractéristique du MIFNO est de contenir une branche dédiée pour la source. On peut ainsi considérer un vecteur quelconque de paramètres de source sans être contraint par la structure 3D

des géologies, ce qui n'était pas possible avec les opérateurs neuronaux existant auparavant.

Après avoir entraîné le MIFNO sur la base de données HEMEW^S-3D , on illustre sa capacité de prédiction sur des données de test. L'erreur de prédiction est quantifiée avec plusieurs métriques pour analyser différentes propriétés physiques. En particulier, des métriques de biais fréquentiel montrent que les erreurs sont plus importantes sur les composantes hautes fréquences du mouvement du sol, ce qui est cohérent avec le biais spectral des réseaux de neurones. De plus, des critères de *Goodness-Of-Fit* habituels en sismologie montrent que les prédictions sont généralement excellentes pour la phase, ce qui traduit des temps d'arrivée précis, et elles sont bonnes à excellentes pour l'enveloppe, ce qui indique la précision sur les amplitudes prédites. Le MIFNO est également capable de généraliser les prédictions à des géologies hors de la distribution d'entraînement, comme illustré avec le modèle géologique de Marmousi, ainsi qu'à des sources placées hors du domaine d'entraînement.

Enfin, on utilise l'apprentissage par transfert pour améliorer la précision du MIFNO dans le contexte du séisme du Teil. Dans ce cadre, une base de données de 4000 simulations est réalisée avec des géologies dérivées du modèle géologique 1D de référence pour la région du Teil et des caractéristiques de source cohérentes avec la faille ayant généré le séisme du Teil. La base de données contient moins de réalisations que la base de données générique HEMEW^S-3D car sa variabilité est plus faible. Le MIFNO pré-entraîné est alors spécialisé avec les réalisations de la base de données du Teil, ce qui améliore ses précisions. On montre qu'environ 500 simulations spécifiques sont nécessaires pour obtenir d'excellentes prédictions. Par application du principe de superposition des sources, le MIFNO fournit d'excellentes prédictions pour la rupture dynamique de la faille, bien que les caractéristiques spatiales et temporelles du mouvement du sol soient plus complexes que dans le cas d'un point source. Avec ce méta-modèle spécifique, on obtient également des distributions statistiques de plusieurs quantités d'intérêt pour l'aléa sismique, en particulier l'accélération pseudo-spectrale. Ces distributions sont cohérentes avec les simulations numériques, ce qui montre la fiabilité du MIFNO dans un cadre statistique. Grâce à la rapidité du MIFNO, il est possible d'obtenir rapidement plusieurs centaines de milliers de prédictions. On en déduit alors des intervalles de confiance qui étaient hors de portée avec les méthodes précédentes. On propose également une première application pour conditionner les paramètres de source et de géologie à partir d'observations fictives.

En conclusion, cette thèse a proposé un méta-modèle pour la propagation des ondes sismiques dans des domaines tridimensionnels. Le méta-modèle générique est précis, applicable à un grand nombre de configurations et capable de généraliser à des données hors de la distribution d'entraînement. Il peut être spécialisé pour des contextes spécifiques avec des coûts de calcul réduits grâce au transfert d'apprentissage. Cela permet ainsi des études de quantification d'incertitudes avec des coûts énergétiques divisés d'un facteur 100 par rapport à l'utilisation exclusive des simulations numériques.

Introduction

Devastating earthquakes regularly remind us of the vulnerability of infrastructures to extreme natural hazards. In 2011, the Tohoku earthquake (Japan) generated a tsunami that led to a major nuclear accident at the Fukushima-Daïchi power plant. In response, nuclear authorities worldwide requested new safety analyses of the existing structures. In France, this gave rise to the SINAPS@ project (2012-2017), which was devoted to the full evaluation of seismic safety from the earthquake source to the infrastructure, and the characterization of uncertainties (Berge-Thierry et al. 2020). In parallel, the SIGMA (2012-2017) and SIGMA-2 (2017-2022) projects involved 30 European partners with the aim to improve the modelling of seismic ground motion and, similarly to SINAPS@, reduce uncertainties (Pecker et al. 2017). Despite numerous achievements in the evaluation of ground motion, the development of numerical tools to simulate earthquakes, and propositions of new regulations, both projects showed that a proper quantification of uncertainties was beyond reach with the existing framework (Berge-Thierry et al. 2020).

On November 11, 2019, an earthquake of magnitude M_W 4.9 hit the Le Teil village in South-Eastern France, damaging more than 900 houses and buildings. This is the most destructive earthquake felt in metropolitan France since the Arlette earthquake (Pyrenees) in 1967 (Ritz et al. 2020), and it raised several questions about the adequacy of regional ground motion estimates. Indeed, the Le Teil surroundings were considered a low-seismicity region, and few earthquakes had been recorded since the deployment of seismic instruments (from 1962 in France, Duverger et al. 2021). In addition, the Le Teil earthquake occurred within the Rhône valley, in the vicinity of several industrial installations, including the Cruas and Tricastin nuclear power plants, respectively 14 km and 20 km away from the epicenter. Installations of the French Alternative Energies and Atomic Energy Commission (CEA) are also located within the Rhône valley and the CEA is responsible for the assessment of seismic hazard around its installations (*Règle fondamentale de sûreté n°2001-01*). For these reasons, the Le Teil earthquake called for new seismic hazard investigations in low-to-moderate seismicity regions and is one of the motivations of this doctoral thesis.

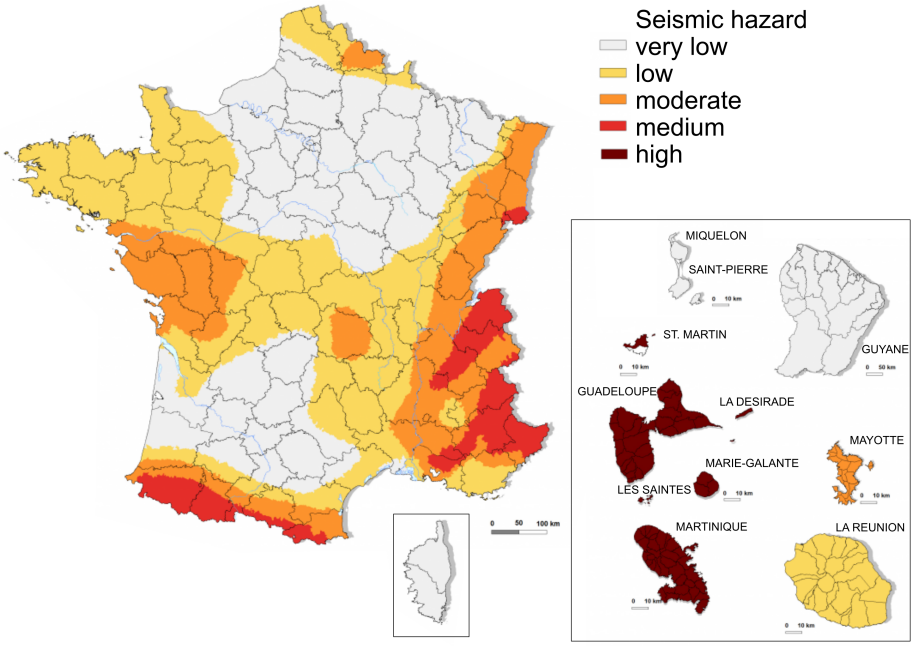


Figure 0.1: Map of seismic hazard in France, colored by five intensity zones (credit: Ministry of ecological transition)

From seismic hazard to physics-based simulations

Assessing seismic hazard amounts to determining the ground motion intensity that could be reached at a given location with a defined return period. A typical question can be formulated as "What is the maximum ground motion acceleration that has a 10 % probability of being observed in the next 50 years in this region?" Seismic hazard assessment is, therefore, strongly related to probabilistic evaluations, and, as such, it requires large amounts of data to derive meaningful distributions of ground motion intensities. Thanks to a massive deployment of seismic sensors worldwide, large datasets of recorded earthquakes are now available (the ESM database in Europe (Luzi et al. 2016), the NGA-West 2 database of crustal events in active tectonic regions (Ancheta et al. 2012), the K-Net and KikNet databases in Japan (C. Zhu et al. 2021), etc.). These datasets have led to a wide variety of Ground Motion Models, also called Ground Motion Prediction Equations (GMPEs), that estimate ground motion intensity from the earthquake parameters and geological characteristics at the site of interest. GMPEs are the most common method to estimate seismic hazard but have several limitations.

In low-to-moderate seismicity regions such as most of metropolitan France, datasets of recorded ground motion are too sparse to inform GMPEs. Then, the only option is to use numerical simulations to complement recorded earthquakes. Thanks to a continuous increase in computational resources and improvements in numerical schemes over the last decades, physics-based earthquake simulations have reached a high level of fidelity. They now enable the simulation of past earthquakes with complex physical phenomena (dynamic fault rupture, realistic geological properties, three-dimensional wave propagation, etc.). As an illustrative example, high-fidelity simulations have been conducted for the 1976 M_w 7.6 Tangshan earthquake (China, H. Fu et al. 2017; W. Zhang et al. 2019), the 1992 M_w 7.3 Landers earthquake (United States of America, Heinecke et al. 2014), the 2005 M_w 7.6 Kashmir earthquake (Pakistan, Khan et al. 2020). The Le Teil earthquake has also been simulated with different numerical settings (El Haber et al. 2021; F. Lehmann et al. 2022; Smerzini et al. 2023).

However, large-scale three-dimensional (3D) simulations require huge computational costs, on the order of tens of thousands of scalar CPU hours (Central Processing Unit hours, total time of all cores in the parallel simulation). In seismic hazard analyses, ground motion must be assessed for a large number of potential seismic sources and a wide variety of geological properties. Conducting these analyses with 3D physics-based simulations would require multiple calls to the earthquake simulator. But due to their prohibitive computational costs, high-fidelity simulations cannot be used directly to assess seismic hazard.

Scientific machine learning to design surrogate models

Thanks to the large datasets of ground motion records, machine learning has a long tradition in seismology (Mousavi and Beroza 2023). In particular, analytical GMPEs can be replaced by deep neural networks to offer more flexibility to ground motion predictions. However, these methods still depend on the quantity of recorded data in the region under study and are of limited use in low-to-moderate seismicity regions.

Recent developments in deep learning offer new perspectives to replace physics-based simulations with surrogate models that preserve the accuracy of simulations while drastically reducing computational costs (Z. Li et al. 2020; Lu et al. 2021; Brandstetter et al. 2022). In this work, we define a surrogate as a numerical model that has the same inputs as the parameters of the physics-based simulations and that estimates ground motion at any location and time without fully resolving the wave propagation equation. It is important to emphasize that the surrogate model should predict time series and not only one measure of ground motion intensity. The need for the full ground motion history in non-linear structural dynamic analysis and earthquake engineering justifies this major difference with GMPEs.

The intersection between physics-based simulations and deep learning has given rise to the field of *Scientific Machine Learning* (SciML). One prominent method in SciML is the Fourier Neural Operator (FNO, Z. Li et al. 2021) that - in a concise and oversimplified summary - predicts the solution of a Partial Differential Equation (PDE) by learning its representation in the frequency space of Fourier coefficients. Soon after the introduction of the FNO, Y. Yang et al. 2021 showed that it could predict the solution of the two-dimensional (2D) acoustic wave equation. These works and many others inspired us to direct our research towards a FNO-based surrogate model of seismic wave propagation.

In F. Lehmann et al. 2023, we proposed the first FNO application to predict the solution of the 3D elastic wave equation. This work and the followings build on large datasets of physics-based simulations (F. Lehmann et al. 2024a) that enable the surrogate model to learn the simulation outputs when provided with the simulation parameters. Then, a faster FNO was proposed by Tran et al. 2023, and we showed that this surrogate model significantly

improves the prediction accuracy (F. Lehmann et al. 2024b). However, none of the existing works allowed to vary the earthquake source, which severely restricts the applicability of the surrogate model. To address this limit, we proposed a Multiple Input Fourier Neural Operator (MIFNO) that considers the dependency on both geological properties and source parameters when predicting ground motion.

Quantifying and propagating uncertainties

A common objective of the SINAPS@ and SIGMA projects was to characterize, quantify, and ultimately reduce the numerous uncertainties in seismic hazard estimates. The earthquake source is the first uncertain parameter. In intraplate regions, earthquakes mainly occur on geological faults that are well identified. However, the Le Teil earthquake occurred on the La Rouvière fault, whose seismogenic activity was ill-defined (Marconato et al. 2022). After identifying the fault, the location of the rupture and its dynamic along the fault are also difficult to characterize. Nevertheless, the influence of the earthquake source uncertainties on ground motion is often studied, for instance by simulating different fault rupture scenarios of the Le Teil earthquake (Causse et al. 2021) or by running sensitivity analyses on the source location (Laporte et al. 2024).

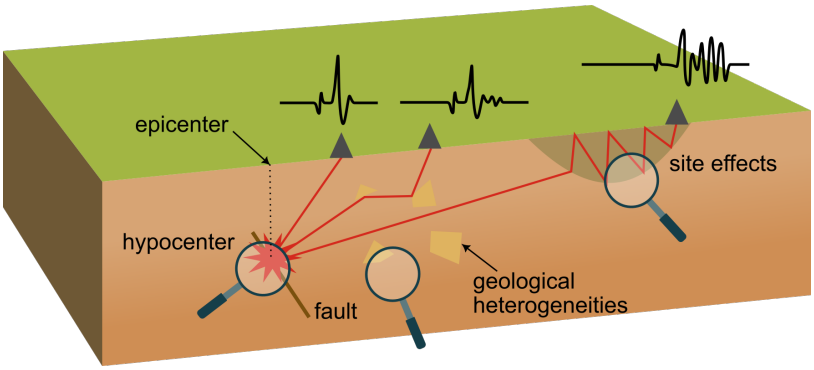


Figure 0.2: Sources of uncertainties in seismic hazard assessment: position and characteristics of the earthquake source, mechanical properties of the Earth's crust (including geological heterogeneities), geological singularities that create site effects.

Another source of uncertainties concerns the properties of the propagation domain. Seismic waves propagate in three dimensions, over long distances, and in deep heterogeneous materials where measurements are challenging or even impossible. Although geological models provide essential information on the material properties, they are only an approximation of the Earth's mechanical characteristics. Geological fluctuations impact the propagation of seismic waves between the source and the surface, resulting in different ground motion time series at neighbouring locations (R.-S. Wu and Aki 1988). Yet localized peculiarities, such as the presence of a sedimentary basin, can also greatly amplify seismic waves and lead to severe damage (Kawase 2003). These *site effects* constitute a significant concern in seismic hazard analyses, and a 3D physics-based surrogate model is one approach to quantifying their influence.

Finally, model uncertainties must be quantified. High-fidelity simulations rely on assumptions that differ from the propagation of waves in reality, for instance, through the implementation of boundary conditions to mimic an infinite propagation domain. The accuracy of numerical solvers is also limited by the available computational resources, implying for instance that high frequencies cannot be resolved numerically. Numerical errors are also a potential source of uncertainties. Then, when using a surrogate model to replace simulations, the approximation error should be added to the other sources of uncertainties.

If previous works (e.g. within the SINAPS@ and SIGMA projects) have greatly improved the identification of uncertainties, propagating them from their different sources to the resulting ground motion requires an efficient and accurate model that includes a detailed description of the earthquake source and the 3D geological properties. Using the Le Teil earthquake as a case study, we show that our MIFNO surrogate model is a promising tool towards this objective.

Outline of the thesis

Chapter 1 presents the computational framework of this thesis. All earthquake simulations are conducted with SEM3D, an open-source code developed in collaboration between the CEA, CentraleSupélec, and IPGP (Touhami et al. 2022). SEM3D solves the elastic wave equation in 3D heterogeneous media, thanks to the Spectral Element Method. The main simulation parameters are introduced from a seismological perspective, and their implementation is detailed. We focus on the description of the earthquake source, which is at the origin of seismic waves, and the geological model since it characterizes the properties of the propagation domain and controls the wave propagation. In this chapter, we also present our simulations of the Le Teil earthquake, emphasizing the importance of geological heterogeneities to reproduce realistic ground motion.

Chapter 2 presents our database of seismic simulations that was used to train the surrogate model. Its ground motion characteristics are analyzed and we provide insights on the *intrinsic dimension* of the data. The high dimensions justify that traditional surrogate models are not suitable to predict space- and time-dependent ground motion from 3D highly heterogeneous geological parameters and varying earthquake sources. In an attempt to reduce the data dimensionality, two common approaches are explored and compared, namely the Principal Component Analysis (PCA) and a deep neural network called UNet auto-encoder.

Chapter 3 reviews state-of-the-art surrogate models to predict the solution of parametric PDEs. It introduces the FNO methodology and exposes our MIFNO surrogate model. From a 3D geological model, a description of the earthquake source position and orientation, the MIFNO predicts the ground motion wavefield at the surface of the propagation domain. The MIFNO predictions are assessed with different metrics and several architectures are compared to exhibit optimal hyperparameters. We also highlight relationships between input parameters and the prediction accuracy that help to understand situations where the surrogate model can give the best predictions. Generalization to corrupted and out-of-distribution data is also investigated.

In **Chapter 4**, we focus on the application of our surrogate model to the Le Teil earthquake. We illustrate the benefits of transfer learning to improve the predictions in a target region. In particular, ground motion generated by a extended fault source is predicted with our surrogate model. Then, we quantify the influence of geological uncertainties on ground motion and show that our surrogate model provides meaningful distributions of ground motion intensity measures. Thanks to the speed of the MIFNO predictions, we also propose a first application of parameter conditioning based on fictive observations and provide confidence intervals of ground motion intensities. Throughout this thesis, we estimate the energy consumption of the various methods and highlight the benefits of our surrogate model in this respect.

Chapter 1

Physics-based earthquake simulations

In low-to-moderate seismicity regions, numerical simulations are necessary to complement recorded ground motion data. They rely on numerical schemes to solve the wave propagation equation, namely the elastic wave equation. This equation describes the motion of particles in an elastic medium defined by its mechanical properties when particles are subject to an external force such as an earthquake source. Among the existing numerical schemes, the Spectral Element Method (SEM, Faccioli et al. 1997; Komatitsch and Tromp 1999) has several advantages in computational seismology and constitutes the backbone of our work (Section 1.1).

Several factors determine the accuracy of earthquake simulations. One crucial parameter is the 3D description of the mechanical properties of the Earth, which we refer to as a geological model (Section 1.2). Due to the difficulty of conducting geophysical measurements, the geological model is often poorly known and too simplistic. When seismic waves propagate through the geological model, this can result in significant discrepancies between the synthetic and recorded seismograms. One standard method to account for geological uncertainties consists in adding random fluctuations to the reference geological model. Depending on the random fields characterization, seismic waves encounter different perturbations and can be rendered more realistically by numerical simulations.

Second, the design of the simulation mesh also influences the accuracy of ground motion. One aims at high-frequency simulations to approach the needs of structural engineering, but this leads to many degrees of freedom. In addition, earthquake simulations are run on large spatial scales to encompass the seismic source and distant areas that can encounter high ground motion intensities. These two factors lead to high computational costs that make large-scale simulations very demanding. Table 1.1 references some large-scale high-frequency earthquake simulations, and their limited number indicates the difficulty of conducting these simulations.

These challenges are illustrated with our simulations of the Le Teil earthquake. We use two geological models, namely a 1D national and a 3D regional geological model. Two parametrizations of the earthquake source are also compared: a kinematic fault rupture model and an equivalent point-wise source. Although the overly simplistic geological models are able to capture only the main ground motion features, we show that adding geological heterogeneities can produce more realistic synthetic seismograms (Section 1.4). This work was published in F. Lehmann et al. (2022). "Machine Learning Opportunities to Conduct High-Fidelity Earthquake Simulations in Multi-Scale Heterogeneous Geology". In: *Frontiers in Earth Science* 10. ISSN: 2296-6463. DOI: 10.3389/feart.2022.1029160.

Table 1.1: Non-exhaustive summary of large-scale high-frequency earthquake simulations. f_{max} is the maximal frequency accurately resolved. See Poursartip et al. 2020 for additional references

Reference	Case study	Domain size	f_{max}
Heinecke et al. 2014	1992 M_W 7.3 Landers earthquake (USA)	-	10 Hz
Maufroy et al. 2016	19 earthquakes in the Mygdonian basin (Greece)	69 km × 69 km × 30 km	4 Hz
H. Fu et al. 2017	1976 M_W 7.6 Tangshan earthquake (China)	320 km × 312 km × 40 km	18 Hz
W. Zhang et al. 2019	1976 M_W 7.6 Tangshan earthquake (China)	320 km × 320 km × 40 km	7.3 Hz
Khan et al. 2020	2005 M_W 7.6 Kashmir earthquake (Pakistan)	40 km × 40 km × 40 km	5.5 Hz
Touhami et al. 2022	2014 M_W 6.0 Cephalonia earthquake (Greece)	44 km × 44 km × 63 km	10 Hz
Smerzini et al. 2023	2019 M_W 4.9 Le Teil earthquake (France)	45 km × 70 km × 8.5 km	8 Hz
this work	2019 M_W 4.9 Le Teil earthquake (France)	50 km × 115 km × 38 km	5 Hz

1.1 Numerical simulations in seismology

Seismic waves are a type of mechanical waves that are generated by a disturbance in the solid Earth. Although earthquakes are the most obvious source of seismic waves, they can also originate from other sources such as volcanic tremors, mining, explosions, noise from railroads, etc. The propagation of seismic waves inside the Earth obeys the elastic wave equation, which is introduced below.

1.1.1 Wave propagation equations

In a solid $\Omega \subset \mathbb{R}^d$, the propagation of disturbances generates particle motion in several directions, implying that the displacement $\boldsymbol{u}$ is a vector : $\boldsymbol{u} : \Omega \times [0,T] \to \mathbb{R}^d$, with $[0,T]$ being the temporal interval. The displacement is governed by the elastic wave equation, for which two equivalent formulations are (the derivation can be found, for instance, in Shearer 2019, p.40-44)

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} - \nabla \cdot \boldsymbol{\sigma}(\boldsymbol{u}) = \boldsymbol{f} \tag{1.1}$$

where $\boldsymbol{\sigma}(\boldsymbol{u}) = C : \nabla \boldsymbol{u}$ is the stress tensor depending on the elasticity tensor C, and

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} - \nabla \lambda \left(\nabla \cdot \boldsymbol{u}\right) - \nabla \mu \left[\nabla \boldsymbol{u} + \left(\nabla \boldsymbol{u}\right)^T\right] - \left(\lambda + 2\mu\right) \nabla \left(\nabla \cdot \boldsymbol{u}\right) + \mu \nabla \times \nabla \times \boldsymbol{u} = \boldsymbol{f} \tag{1.2}$$

In equation 1.2, the material is described by three parameters: its density ρ and the Lamé parameters λ, μ. In heterogeneous media, they all depend on the spatial point $\boldsymbol{x} \in \Omega$. The right-hand side $f : (\boldsymbol{x}, t) \to \mathbb{R}$ is an external body force (detailed in Section 1.1.3) and $\boldsymbol{u} : (\boldsymbol{x}, t) \to \mathbb{R}^d$ is the solution of the elastic wave equation, written as a displacement. From the parameters (ρ, λ, μ), one can express the velocity of compression waves (P waves, V_P) and shear waves (S waves, V_S) through

$$V_P = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad V_S = \sqrt{\frac{\mu}{\rho}} \tag{1.3}$$

The analysis of P and S waves is of practical importance in seismology and our results will often be discussed in terms of P and S waves components. Equation 1.3 especially implies that the velocity of P-waves is always higher than the S-wave velocity (see Fig. 1.1 where P waves arrive before S waves).

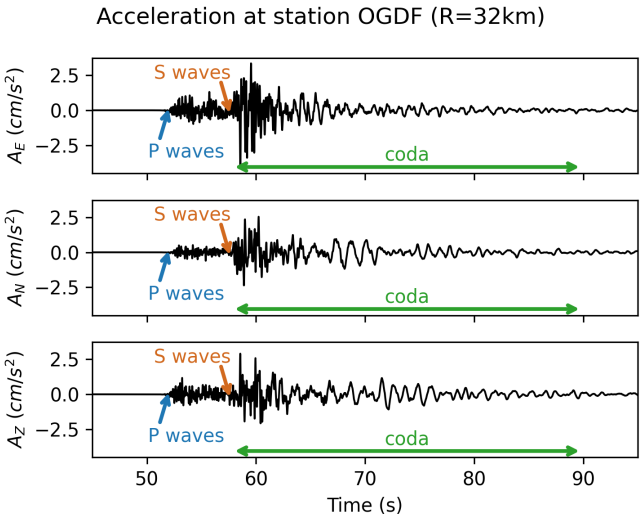


Figure 1.1: Acceleration recorded at station OGDF during the Le Teil earthquake, located 32 km away from the epicenter (more details in Section 1.4). Sub-panels show the East-West (E), North-South (N), and vertical (Z) components. Arrows indicate the approximate arrival time of P and S waves, as well as the coda.

Because it is the most general and accurate, the elastic wave equation 1.2 is the main focus of this work. However, beyond the case of an infinite homogeneous medium or some forms of approximations, analytical solutions

do not exist and numerical approximations are sought after. It is also important to note that the wave equation is not linear with respect to the material parameters, although it is linear with respect to the source.

Due to the hyperbolicity of the elastic wave equation, it is challenging to obtain accurate numerical solutions. The wave equation is a prototypical example of hyperbolic equations. Other examples include the linear advection equation $\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = f$ and the Burger's equation $\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}$. In hyperbolic equations, information travels at a finite speed, referred to as the wavespeed. This means that points where the wave has not yet arrived have received no information. Hyperbolic equations are also characterized by the propagation or appearance of discontinuities, controlled by the initial and boundary conditions. Hyperbolicity therefore often results in solutions with high complexity (Thomas 1995).

Before explaining the numerical scheme used to solve the elastic wave equation, one should mention other forms of wave equations that will be encountered in the deep learning literature review (Section 3.1). First, the Helmholtz equation is obtained by taking the Fourier transform in time (denoted $\mathcal{F}$ or $\hat{}$) of the acoustic equation. When propagation involves a scalar quantity $u_P : (\boldsymbol{x}, t) \to \mathbb{R}$ (e.g. pressure field or P waves only) instead of a vector, the wave equation is restricted to the acoustic equation

$$\rho(\boldsymbol{x})\frac{\partial^2 u_P}{\partial t^2}(\boldsymbol{x}, t) - \nabla \cdot (\mu(\boldsymbol{x})\nabla u_P(\boldsymbol{x}, t)) = f(\boldsymbol{x}, t) \tag{1.4}$$

Let us define $\hat{u}(\boldsymbol{x}, \omega)$ as the Fourier transform of u at point $\boldsymbol{x}$ and circular frequency ω, and recall that $\mathcal{F}\left(\frac{\partial^2 u}{\partial t^2}\right)(\omega) = -\omega^2\mathcal{F}(u)(\omega)$. Then, the Helmholtz equation reads

$$\rho(\boldsymbol{x})\omega^2\hat{u}(\boldsymbol{x}, \omega) + \nabla \cdot (\mu(\boldsymbol{x})\nabla\hat{u}(\boldsymbol{x}, \omega)) = -\hat{f}(\boldsymbol{x}, \omega) \tag{1.5}$$

Equation 1.5 is an elliptic equation, implying that its solutions are smooth, even if the initial and boundary conditions are rough (although there may be singularities at sharp corners). Therefore, solving the elastic wave equation or the Helmholtz equation involves different complexities.

Another form of the wave equation is given by the eikonal equation. It is a high-frequency approximation of the elastic wave equation, leading to a time-independent first-order non linear hyperbolic PDE. It is often used to describe the traveltime of waves from the source to any point in the domain.

1.1.2 Physics-based simulations with SEM3D

Numerically solving the elastic wave equation requires a discrete approximation of the problem, both spatially and temporally. Several discretization schemes exist and have been used for the wave equation, such as the finite difference method, pseudospectral methods, boundary element methods, finite element method, etc. A methodological review can be found in Seriani and Oliveira 2020 and references of early work are given in Komatitsch and Tromp 1999. The Spectral Element Method (SEM, Faccioli et al. 1997; Komatitsch and Tromp 1999) is a special type of finite element method, well-used for wave propagation problems. The computational code SEM3D, originally developed by the CEA, CentraleSupélec, and IPGP relies on the SEM and constitutes the backbone of this work (Touhami et al. 2022). In the following, we describe the main ideas of the SEM and its implementation in SEM3D. More details can be found in the works referenced below.

The SEM does not solve directly the elastic wave equation 1.1. Instead, this equation is integrated against any test function $\boldsymbol{\phi} : \Omega \to \mathbb{R}^d$ to give the weak formulation of the wave equation (Komatitsch and Tromp 1999, eq. 12)

$$\int_\Omega \rho\boldsymbol{\phi} \cdot \partial_t^2\boldsymbol{u}\,\mathrm{d}^3\boldsymbol{x} = -\int_\Omega \nabla\boldsymbol{\phi} : \boldsymbol{\sigma}(\boldsymbol{u})\,\mathrm{d}^3\boldsymbol{x} - \int_\Omega \boldsymbol{\phi} \cdot \boldsymbol{f}\,\mathrm{d}^3\boldsymbol{x} + \int_\Gamma \rho\left(\boldsymbol{\sigma}(\boldsymbol{u}) \cdot \vec{\boldsymbol{n}}\right) \cdot \boldsymbol{\phi}\,\mathrm{d}^2\boldsymbol{x} \tag{1.6}$$

The boundary of the spatial domain Ω is separated into a stress-free surface $\partial\Omega_{top}$ (usually the upper surface) and an absorbing boundary Γ which mimics a semi-infinite domain (waves going through Γ are supposed to escape the domain, Fig. 1.2). In equation 1.6, $\vec{\boldsymbol{n}}$ is the outward unitary vector normal to the boundary surface Γ. The upper surface $\partial\Omega_{top}$ does not appear in the integration since this stress-free surface must satisfy $\boldsymbol{\sigma}(\boldsymbol{u}) \cdot \vec{\boldsymbol{n}} = 0$. Therefore, this boundary condition is necessarily satisfied when satisfying the weak formulation of the wave equation 1.6.

In the SEM, the spatial domain Ω is discretized in a mesh of hexahedral elements $\{\Omega_a\}_a$ and similarly, the boundary Γ is discretized with the corresponding quadrilateral elements $\{\Gamma_b\}_b$. Then, the wave equation 1.6 can

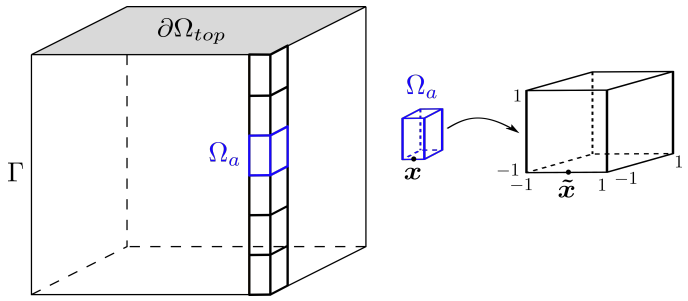


Figure 1.2: Schematic illustration of the spatial domain Ω divided into mesh elements $\{\Omega_a\}_a$. $\partial\Omega$ is the upper surface boundary and Γ contains all the other boundaries.

be solved in parallel in each mesh element (with points common to several elements managed carefully), which considerably reduces the computational time.

Computing integrals as in equation 1.6 is always a computational challenge. One solution is to use quadrature points and compute the integral as the sum over the function value at the quadrature points. First, notice that each mesh element Ω_a is isomorphous to the reference cube $[-1,1]^3$. This implies that any function ϕ evaluated at point $\boldsymbol{x} = (x,y,z) \in \Omega_a$ can be expressed in terms of points $(\tilde{x},\tilde{y},\tilde{z}) \in [-1,1]^3$ and the element-wise function ϕ_a. Integrals can then be rewritten as

$$\int_{\Omega_a} \phi(\boldsymbol{x})\, \mathrm{d}^3\boldsymbol{x} = \int_{[-1,1]^3} \phi_a(\tilde{x},\tilde{y},\tilde{z}) J_a(\tilde{x},\tilde{y},\tilde{z})\, \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{y}\, \mathrm{d}\tilde{z} \tag{1.7}$$

where J_a is the Jacobian expressing the change of coordinates between $[-1,1]^3$ and Ω_a. One specific feature of the SEM is that the quadrature points are chosen as the Gauss-Lobato-Legendre (GLL) points, defined as the roots of $(1-\tilde{x}^2)P_n'(\tilde{x}) = 0$ where P_n is the Lagrange polynomial of degree n. If the GLL points are denoted $\tilde{x}_0,\cdots,\tilde{x}_n$ along the $\tilde{x}$ axis and similarly $\tilde{y}_0,\cdots,\tilde{y}_n$ and $\tilde{z}_0,\cdots,\tilde{z}_n$ along the $\tilde{y}$ and $\tilde{z}$ axes, the integral in equation 1.7 can be approximated by

$$\int_{[-1,1]^3} \phi_a(\tilde{x},\tilde{y},\tilde{z}) J_a(\tilde{x},\tilde{y},\tilde{z})\, \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{y}\, \mathrm{d}\tilde{z} \simeq \sum_{i,j,k=0}^{n} w_i w_j w_k \phi_a(\tilde{x}_i,\tilde{y}_j,\tilde{z}_k) J_a(\tilde{x}_i,\tilde{y}_j,\tilde{z}_k) \tag{1.8}$$

where w_i, w_j, w_k are the weights associated with the GLL points and can be pre-computed.

From these developments, the displacement is expressed at each grid point $\boldsymbol{x}_m$ as a global vector $U = (u(x_m))_m$ and the global system is written in the form

$$M\ddot{U} + F^{int}(U;\dot{U}) = F^{ext} \tag{1.9}$$

where M is the mass matrix, F^{int} the matrix of internal forces, and F^{ext} the external force matrix. The coefficients of those matrices are found by inserting the integral expression 1.8 into the weak formulation 1.6. One advantage of the SEM on the basis of Lagrange polynomials is that the mass matrix is diagonal, meaning that its inversion is computationally cheap.

Equation 1.9 involves time derivatives that also need to be discretized temporally. The temporal discretization is a velocity leap-frog scheme, where the velocity $\dot{U}$ at time t_n is approximated by $\dot{U}_n \simeq \dot{U}(t_n)$ and is computed from

$$M\frac{\dot{U}_{n+1}-\dot{U}_n}{\Delta t} + F^{int}\left(U_{n+1/2};\dot{U}_{n+1/2}\right) = F^{ext}_{n+1/2} \tag{1.10}$$

The elastic wave equation is solved with initial conditions $\boldsymbol{u}(\boldsymbol{x},t=0)=0$ and $\frac{\partial \boldsymbol{u}}{\partial t}(\boldsymbol{x},t=0)=0$ to reflect that the medium is at rest before the seismic event. One should also note that the leap-frog scheme is conditionally stable with a Courant–Friedrichs–Lewy (CFL) stability condition determined by the size of the mesh elements and the maximum wave velocity (i.e. the P-wave velocity). This means that the time step becomes very small when small mesh elements are needed and the P-wave velocity is large. This leads to a large number of time iterations and therefore a high computational time.

Design of the mesh

Although the computation of the time step is managed automatically in SEM3D from the CFL value imposed by the user (CFL=0.2 in this work), the design of the mesh is left to the user and has a crucial importance. The mesh elements should be small enough to accurately represent the material properties and the propagation of waves. In the meantime, the computational time increases with the number of elements, meaning that we aim for elements as large as numerically possible.

The most constraining factor is the accurate representation of the wave propagation. Following (De Martin 2011), at least 5 GLL points should be used to discretize the minimum wavelength propagating inside the domain. The minimum wavelength λ_{min} depends on the minimum wave velocity, i.e. $V_{S,min}$, and the maximum frequency of the waves one aims at representing, f_{max}, following

$$\lambda_{min} = \frac{V_{S,min}}{f_{max}} \tag{1.11}$$

The element size can be increased if the number of GLL points N_{GLL} is larger than 5 since the distance between quadrature points will remain of the same order. Therefore, the mesh element size Δx is computed with

$$\Delta x = \frac{N_{GLL}}{5}\lambda_{min} = \frac{N_{GLL}}{5}\frac{V_{S,min}}{f_{max}} \tag{1.12}$$

The material properties detailed in Section 1.2 will be designed in accordance with the mesh element size.

It was mentioned earlier that the domain Ω is delineated by an absorbing boundary Γ. This allows to approximate the propagation of waves in a semi-infinite domain where waves hitting the absorbing boundary should escape the domain and no longer influence the propagation inside the domain. SEM3D employs the Perfectly Matched Layers (PML) as absorbing boundary conditions (Berenger 1994; Festa and Vilotte 2005). This means that layers of mesh elements are added on the boundary of the domain and the amplitude of the incident wave is exponentially reduced inside the PML. When hitting the external PML boundary, the wave amplitude is negligible, so that no reflecting wave can be seen.

SEM3D has been validated for different benchmarks (Touhami 2020), widely employed to simulate past earthquakes (Touhami et al. 2021), and to assess the seismic response of nuclear sites and urban areas (Gatti et al. 2018; Korres et al. 2022). Therefore, under the assumption that the input parameters are accurate, numerical results of SEM3D reflect the reality. In the following, results of SEM3D simulations are sometimes referred to as *ground truth*, in opposition with the prediction of the surrogate model. This term implies *simulated ground truth*.

1.1.3 Modelling earthquake sources

The elastic wave equation 1.2 involves a source term $\boldsymbol{f}(\boldsymbol{x}, t)$ that describes the distribution of force density acting on a point $\boldsymbol{x}$ at time t. The simplest description of the source is a point-source located at $\boldsymbol{x}_s \in \Omega$, with a time evolution described by the seismic moment function $\boldsymbol{M}(t)$. The seismic moment function is a time-dependent 3×3 symmetric tensor describing the forces acting at the source location. The point source expression can then be written

$$\boldsymbol{f}(\boldsymbol{x}, t) = \nabla \cdot (\boldsymbol{M}(t)\delta(\boldsymbol{x} - \boldsymbol{x}_s)) \tag{1.13}$$

It is well-known that the far-field components of ground motion are proportional to the time-derivative of the seismic moment function (see e.g. Madariaga 2015, eq. 21). Knowing that the geometry of the point source is generally independent from its temporal evolution, it is then convenient to describe the moment rate function $\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t}(t)$ as

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t}(t) = \boldsymbol{M}s(t) \tag{1.14}$$

where $s(t)$ is the source time function and $\boldsymbol{M}$ is the seismic moment

$$\boldsymbol{M} = \begin{pmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{xy} & M_{yy} & M_{yz} \\ M_{xz} & M_{xy} & M_{zz} \end{pmatrix} \tag{1.15}$$

The component M_{ij} corresponds to the action of two point forces parallel to the i axis (with opposite signs) and separated by a small distance in the j direction. In particular, the diagonal elements represent force dipoles while the off-diagonal elements correspond to force couples.

When an earthquake is generated by a fault rupture, the seismic moment function describes the slip history along the fault. Then, the moment tensor can be parametrized by four variables characterizing the rupture process (Aki and Richards 2002):

- the scalar seismic moment M_0 (unit $N.m$),
- the dip angle of the fault (δ)
- the strike angle of the fault with respect to the north (ϕ)
- the rake angle of the fault (λ), which is the angle of the slip vector with respect to the horizontal

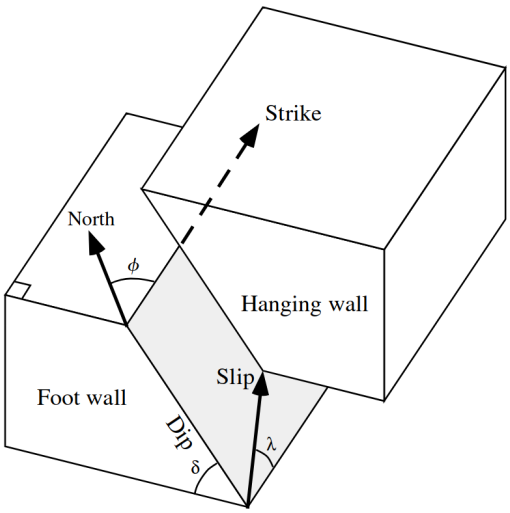


Figure 1.3: Description of a fault from its dip (δ), strike (ϕ), and rake (λ) angles. Figure extracted from Shearer 2019

In this work, faults are parametrized either with their three descriptive angles or with the moment tensor components. Equivalences between the two representations are obtained through (Aki and Richards 1980)

$$\begin{aligned}
M_{xx} &= -M_0(\sin\delta\cos\lambda\sin(2\phi) + \sin(2\delta)\sin\lambda\sin^2\phi) \\
M_{xy} &= M_0(\sin\delta\cos\lambda\cos(2\phi) + 0.5\sin(2\delta)\sin\lambda\sin(2\phi)) \\
M_{xz} &= -M_0(\cos\delta\cos\lambda\cos\phi + \cos(2\delta)\sin\lambda\sin\phi) \\
M_{yy} &= -M_0(\sin\delta\cos\lambda\sin(2\phi) - \sin(2\delta)\sin\lambda\cos^2\phi) \\
M_{yz} &= -M_0(\cos\delta\cos\lambda\sin\phi - \cos(2\delta)\sin\lambda\cos\phi) \\
M_{zz} &= M_0(\sin(2\delta)\sin\lambda)
\end{aligned} \tag{1.16}$$

Source time function

The source time function describes the temporal evolution of the earthquake process. As such, it provides rich information to seismologists about the fault rupture duration and its frequency content. Assuming a simple linear slip of the fault leads to the ramp seismic moment function depicted in Fig. 1.4a whose source time function is a boxcar function. The boxcar length is equal to the rise time τ of the fault, i.e. the time to evolve from the initial state to the final state. It is known that small faults lead to shorter rise times, hence, source time functions that are closer to a Dirac impulse (Fig. 1.4b).

Fortunately, the numerical analysis of source time functions is made easier by the introduction of the Green function $\boldsymbol{G}_{\boldsymbol{M},\boldsymbol{x}_s,t_0}(\boldsymbol{x},t)$, which is the fundamental solution of the elastic wave equation 1.2 when the source is an impulse point force located at $\boldsymbol{x}_s$ and occurring at $t = t_0$,

$$\boldsymbol{f}(\boldsymbol{x},t) = \nabla\cdot(\boldsymbol{M}\delta(t-t_0)\delta(x-x_s)) \tag{1.17}$$

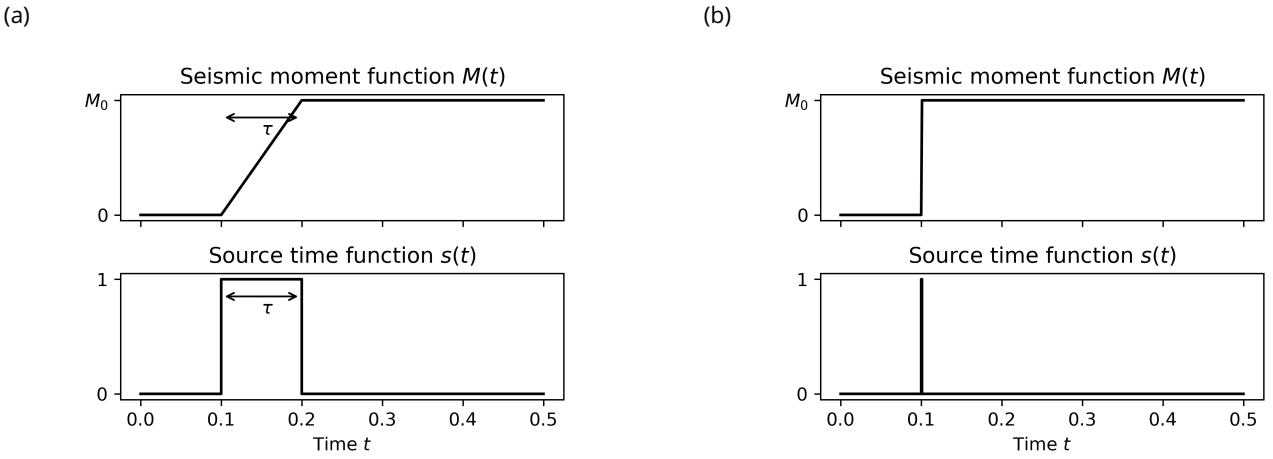


Figure 1.4: Seismic moment functions $M(t)$ of a simple ramp (a) and a Heaviside step (b). The associated source time functions are proportional to the moment rate function $\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t}(t)$.

For any source time function $s(t)$, the general solution of the elastic wave equation 1.2 is then obtained from the convolution of the Green function with the source time function

$$\boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{G}_{\boldsymbol{M},\boldsymbol{x}_s,t_0}(\boldsymbol{x},t) * s(t) \tag{1.18}$$

Assuming that the solution u_1 of the elastic wave equation is known for a given source time function s_1 and provided that the moment tensor $\boldsymbol{M}$ remains the same, it is easy to obtain the solution u_2 for any new source time function s_2. The steps are as follows:

1. compute the Fourier transform of the original source time function $\hat{s}_1 := \mathcal{F}(s_1)$ and the solution $\hat{u}_1 := \mathcal{F}(u_1)$
2. derive the Green function in the frequency domain $\hat{G}_1 = \dfrac{\hat{u}_1}{\hat{s}_1}$
3. compute the Fourier transform of the new source time function $\hat{s}_2$
4. compute the new solution in the frequency domain $\hat{u}_2 = \hat{G}_1 * \hat{s}_2$
5. deduce the new solution in the temporal domain $u_2 = \mathcal{F}^{-1}(\hat{u}_2)$

This procedure is illustrated in Fig. 1.5, starting from a *spice bench* seismic moment function M_1. It is often used to describe the temporal evolution of a typical point source and is used as the reference source time function in this work. The solution u_1 is simulated with SEM3D (Fig. 1.5a). The associated Green function G_1 allows to reconstruct the solution for a new seismic moment function M_2. Figure 1.5b shows that the reconstructed solution $G_1 * s_2$ perfectly matches the true solution u_2.

Although the temporal evolution of the solution is crucial, the Green function approach explains that the temporal evolution of the source can be fixed when designing the metamodel (Chap. 3). In Chapter 4, convolutions are employed to obtain predictions for different source time functions.

1.2 Geological models and their uncertainties

Numerical simulations require a description of the material properties: density ρ, S-wave velocity V_S and P-wave velocity V_P , that together define a unique elastic wave equation 1.2. These properties correspond to the mechanical behaviour of rocks inside the Earth crust, from the surface to several dozens of kilometers depth, on all the spatial extend of the region under study (on the order of 10^2 to 10^4 km^2). Hence, it is extremely difficult to measure the material properties in the whole study domain and one needs to rely on models with different levels of fidelity.

In this section, we give a brief overview of some types of geological models commonly used in seismic numerical simulations to highlight their strengths and limitations. In particular, we focus on the modelling of geological heterogeneities to ensure that our data are as realistic as possible.

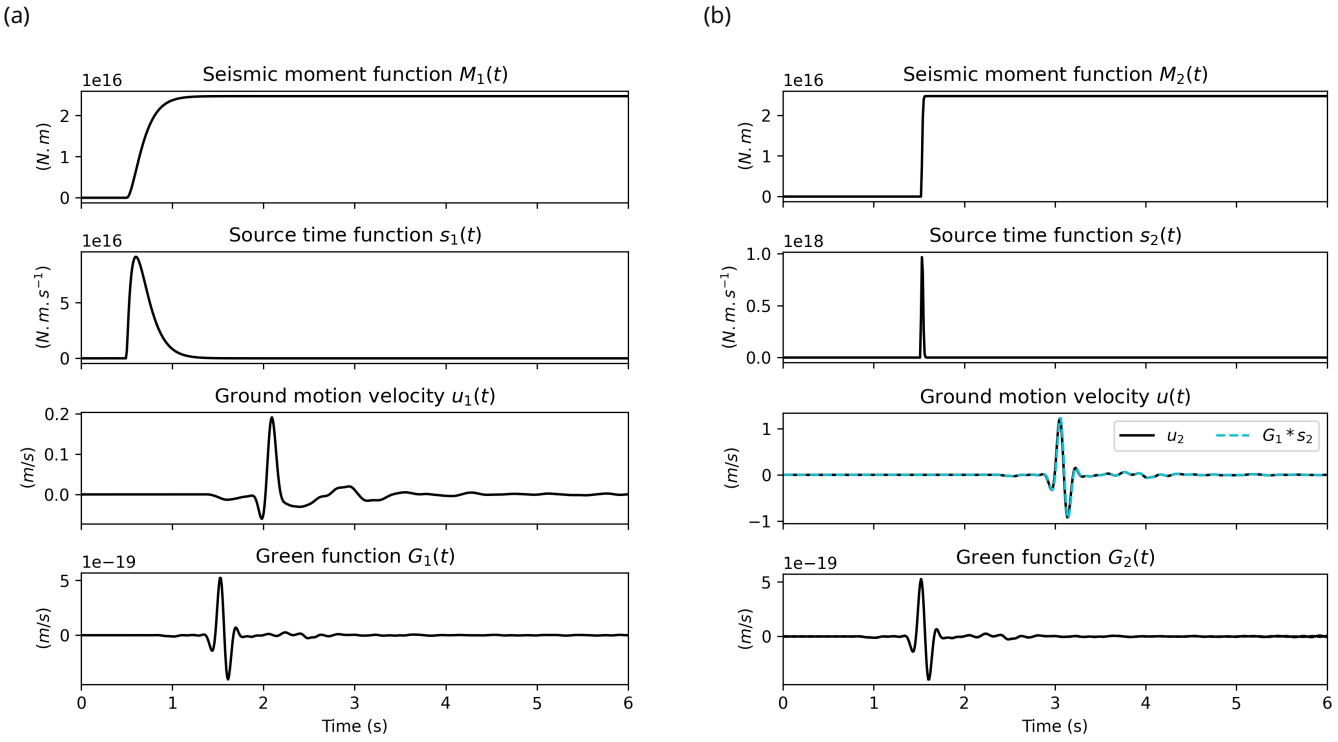


Figure 1.5: For the original source (a) and the new source (b), derivation of the Green functions G_1 and G_2. The Green function G_1 is then used to obtain the solution $G_1 * s_2$ for a new source time function s_2.

1.2.1 Geological models

Let us define a geological model a as the description of one or more material properties (with g being the number of properties) at each point of the domain

$$a : \begin{array}{ccc} \Omega \subset \mathbb{R}^d & \rightarrow & \mathbb{R}^g \\ \boldsymbol{x} & \mapsto & a(\boldsymbol{x}) \end{array} \tag{1.19}$$

The coarsest level of fidelity is a one-dimensional (1D) geological model that describes a as a function of depth. It assumes a 1D structure of the Earth with parallel horizontal layers. This type of model can be obtained from borehole measurements to give accurate but very localized information. The depth is also limited to a few hundred (rarely thousand) meters (Vergniault and Mari 2020). Alternatively, a 1D model can be obtained as the result of an optimization problem based on the observation of seismic wave arrivals. Starting from an initial hypothetical model, one tries to minimize the error between the seismic waves simulated with the hypothetical model and the observed seismic waves. The CEA LDG provides a 1D model for metropolitan France (Duverger et al. 2021, detailed in Section 1.4). In the vicinity of the Le Teil earthquake, Causse et al. 2021 obtained a 1D model of V_P and V_S with this optimization process (Fig. 1.6a).

One main limitation of 1D geological models is their ignorance of horizontal variations in material properties that can lead to site effects. Site effects refer to the amplification (or deamplification) of ground motion due to geological or topological characteristics in the vicinity of the site under study. Following this definition, they are considered as a local effect and should be distinguished from the path effect, i.e. the modification of ground motion due to the propagation of seismic waves between the source and the surface. Well-known features leading to site effects are the presence of sedimentary basins where seismic waves get trapped inside the sediments due to their lower velocity. Other characteristics such as the presence of a hill are also known to create site effects but topography is not considered in this work.

3D geological models are required to assess the site effects (Makra and Chávez-García 2016; Smerzini et al. 2011) but they are expensive to acquire. When devoting much effort into the modelling of a region of interest, a local model describes the interfaces between geological layers. It is common to describe at least the interface

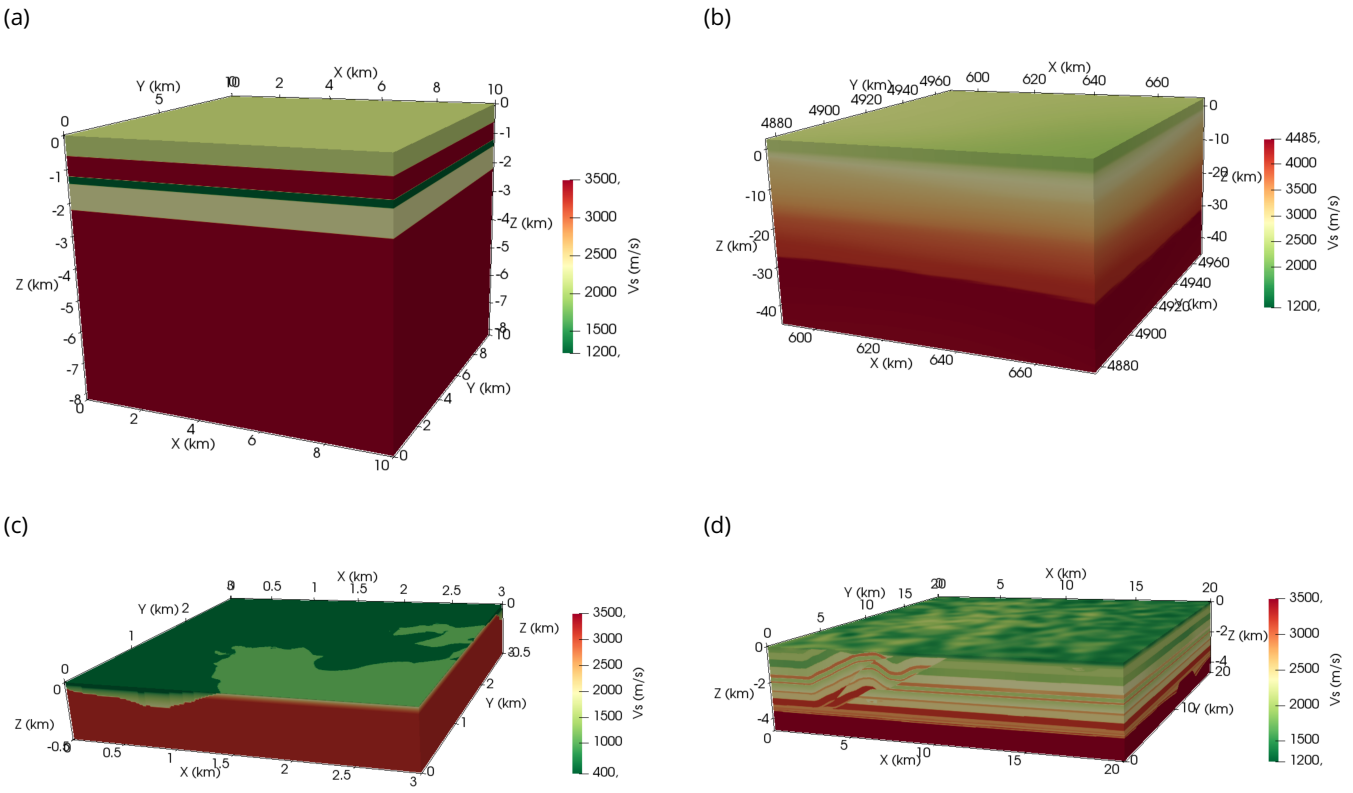


Figure 1.6: Example of (a) a *layer cake* 1D geological model (Smerzini et al. 2023) adapted from (Causse et al. 2021), (b) a 3D regional model (Arroucau 2020), (c) a local 3D model with a sedimentary basin, (d) a local 3D model with an overthrust of thin sedimentary layers (Aminzadeh et al. 1997). Note the different spatial scales between the models

between the sedimentary basin and the bedrock (Fig. 1.6c) and more detailed models describe thinner sedimentary layers (1.6d). Geological parameters are then inferred in each layer, which is also subject to the extrapolation of point-wise measurements.

Alternatively, 3D regional models can be obtained on a large scale (several hundreds of kilometers wide and tens of kilometers deep to represent the Moho, the crust-mantle boundary) from seismic tomography. For instance, Arroucau 2020 provides a geological model for metropolitan France with a 10 km horizontal and 0.5 km vertical resolution (a small area of this model is shown in Fig. 1.6b). However, this resolution is too coarse to include any site effect.

All the above-mentioned geological models are subject to uncertainties, related to the depth and shape of the main structures, and the parameter values in-between (Wellmann and Caumon 2018). Beyond these uncertainties, it is also crucial to acknowledge that the Earth crust is not a homogeneous material. Indeed, it is made of composite minerals and fractures of various sizes that induce a deviation of the mechanical properties from their mean values.

1.2.2 Geological heterogeneities

As stated by R.-S. Wu and Aki 1988, "the scale length of heterogeneities revealed by seismic waves, not including the laboratory measurements of rock samples, spans 8 orders of magnitude". Thus, before considering the effects of geological heterogeneities, we firstly need to define the appropriate scales for our problem. We aim at conducting local seismic simulations on areas with a typical length $L \sim 10\,\text{km}$. The minimum wavelength defined by equation 1.11 is $\lambda \sim 100\,\text{m}$ while the largest wavelengths associated with low frequencies extend up to $\lambda \sim 50\,\text{km}$. Geological heterogeneities are parametrized by their correlation length ℓ_c, which remains to be characterized. The correlation length reflects the typical size of heterogeneities and indicates the distance above which two spatial points are assumed uncorrelated.

When heterogeneities are small compared to the wavelength ($\ell_c \ll \lambda$), homogenization can be used to convert

the heterogeneous medium into an *effective* smooth medium where the *effective* wave equation can be solved efficiently (Capdeville and Marigo 2007; Cupillard and Capdeville 2018). In our work, we focus on larger heterogeneities inducing scattering, i.e. a transfer of the coherent energy in the incoming signal to an incoherent signal, further leading to a modification of seismic waves. Scattering is responsible for the coda of seismic signals (i.e. the seemingly random wave train visible after the arrival of the coherent pulses, Fig. 1.1), the envelope broadening (i.e. S waves having a longer duration when the travel distance is longer), and the differences between ground motion observed at neighbouring sensors (R.-S. Wu and Aki 1988). Two main scattering regimes exist.

Large-angle scattering is characterized by $\ell_c \simeq \lambda$. In this situation, heterogeneities mainly cause scattering attenuation, i.e. the wave amplitude globally decreases with increasing distance. Hence, including this type of heterogeneities in the geological models is important to make the simulations more realistic. When $\ell_c \simeq \lambda$, one can introduce an additional characteristic length, the transport mean free path, that describes the mean distance between two scatterers. Although it is useful to obtain numerical solutions with the radiative transfer equations, materials are rarely characterized with the transport mean free path.

In the small-angle scattering regime defined by $\ell_c \gg \lambda$, the seismic energy is scattered into the forward direction. This mainly causes travel time and amplitude fluctuations since waves travelling through a heterogeneity are subject to a different velocity than the background model. While seismic tomography revealed crustal heterogeneities between 1 km- and 10 km-wide (Cottereau 2016 and references therein), several studies used 3D numerical simulations to reproduce seismic recordings by altering the reference geological models with random heterogeneities. The following studies were conducted on scales similar to ours and therefore, serve as a reference to define the correlation lengths in our work.

In the San Francisco bay area, correlation lengths of 4 km were found by Thompson et al. 2007 in the first 10 m of the basin and Hartzell et al. 2010 confirmed correlations between 5 km and 10 km. In southern Nevada, Scalise et al. 2021 experimented horizontal correlation lengths between 0.6 km and 2.4 km and vertical correlation lengths from 0.15 km to 0.6 km. Finally, based on generic geological models, Imperatori and Mai 2013 found that correlation lengths between few hundreds meters and few kilometers produce good agreement between simulations and recordings.

Random fields

Random fields are a convenient approach to model geological heterogeneities since they cannot be located at a precise location in the domain. A random field A can be thought of as a stochastic process defined on a probabilistic space and the physical domain $\Omega \subset \mathbb{R}^d$. For a given realization η in the probabilistic space, A_η is a deterministic function defining the value of the random field at any point, $A_\eta(\boldsymbol{x})$. Alternatively, when fixing the spatial point $\boldsymbol{x} \in \Omega$, $A(\boldsymbol{x})$ is a random variable describing the material properties. Under some conditions on the individual random components of the elasticity tensor, the random field describing the material is a second-order random field (Soize 2008). Log-normal random fields are conveniently used to preserve the positivity of material properties by assuming that any finite-dimensional variable $\{\log(A)(\boldsymbol{x}_1), \cdots, \log(A)(\boldsymbol{x}_n)\}$ is a multivariate Gaussian random variable. Gaussian random fields are characterized by their mean $m(\boldsymbol{x})$ and covariance $Cov(\boldsymbol{x}, \boldsymbol{x}')$ functions

$$\begin{aligned} m(\boldsymbol{x}) &:= \mathbb{E}\left[A(\boldsymbol{x})\right] \\ Cov(\boldsymbol{x}, \boldsymbol{x}') &:= \mathbb{E}\left[\left(A(\boldsymbol{x}) - m(\boldsymbol{x})\right)\left(A(\boldsymbol{x}') - m(\boldsymbol{x}')\right)\right] \end{aligned} \tag{1.20}$$

The mean function is obtained from the reference geological model. The covariance function is conveniently defined from a correlation model R that assumes some relationships between the material properties at different points. In particular, the stationarity and isotropy assumption allows to write the correlation model as a function of distance $r = \|\boldsymbol{x} - \boldsymbol{x}'\|$ instead of the exact location of points $\boldsymbol{x}$ and $\boldsymbol{x}'$. The covariance function is related to the correlation model through the standard deviation of $A(\boldsymbol{x})$, assumed to be constant for all $\boldsymbol{x} \in \Omega$, and hence denoted σ

$$Cov(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 R(\|\boldsymbol{x} - \boldsymbol{x}'\|) \tag{1.21}$$

It is commonly accepted that geological parameters can be appropriately defined with a von Karman correlation model (e.g. Hartzell et al. 2010 and references therein). The von Karman correlation model reads

$$R(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\ell_c}\right)^\nu K_\nu \left(\frac{r}{\ell_c}\right) \tag{1.22}$$

where ν is the Hurst exponent characterizing the roughness of the material (ν is fixed to 0.1 in this work), Γ is the Gamma function, and K_ν is a second-order Bessel function. One can note that the correlation length can be made

anisotropic by defining correlation lengths ℓ_x, ℓ_y, ℓ_z along each direction and scaling r independently along each axis

$$\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|}{\ell_c} = \sqrt{\left(\frac{x-x'}{\ell_x}\right)^2 + \left(\frac{y-y'}{\ell_y}\right)^2 + \left(\frac{z-z'}{\ell_z}\right)^2} \tag{1.23}$$

To compute efficiently the random fields, the spectral representation technique (Shinozuka and Deodatis 1991; de Carvalho Paludo et al. 2019) is adopted within the SEM3D framework. With this formulation, a centered Gaussian random field A can be decomposed as a sum of independent identically distributed random variables $(\phi_n)_{-N\leq n\leq N}$, with uniform distribution over $[0, 2\pi]$

$$A(\boldsymbol{x}) = \sum_{n=-N}^{N} \sqrt{2\hat{\mathcal{R}}(n\boldsymbol{\Delta k})}\cos(n\boldsymbol{\Delta k}\cdot\boldsymbol{x}+\phi_n)$$

where $\hat{\mathcal{R}}$ is the Fourier transform of the correlation kernel $\mathcal{R}$ and $\boldsymbol{\Delta k}$ the unit volume in $\mathbb{R}^3$. Figure 1.7 illustrates two heterogeneous geological models obtained with this method, different correlation lengths and coefficients of variation.

(a) $\ell_x = \ell_y = 6$km, $\ell_z = 3$km, $\sigma = 40\%$

(b) $\ell_x = \ell_y = \ell_z = 1.5$km, $\sigma = 20\%$

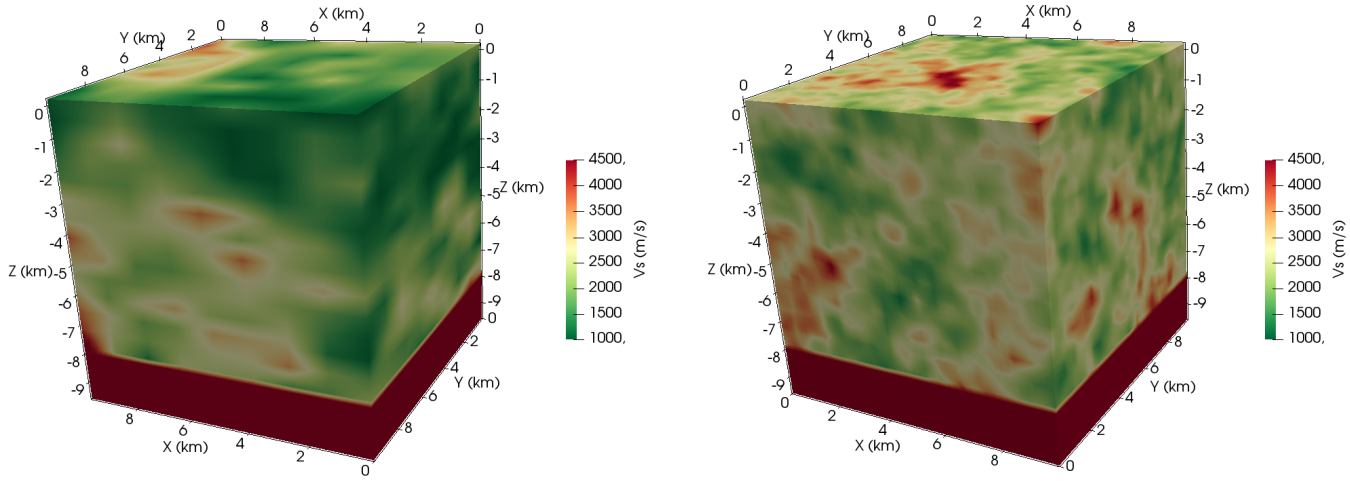


Figure 1.7: (a) Geology with large-scale heterogeneities: correlation lengths of 6 km in directions x and y, 3 km in direction z, coefficient of variation of 40 %. (b) Geology with small-scale heterogeneities: correlation lengths of 1.5 km in directions x, y and z, coefficient of variation of 20 %.

With these definitions, we are left with the choice of the coefficient of variation σ. The above-mentioned works (Hartzell et al. 2010; Imperatori and Mai 2013; Scalise et al. 2021) use coefficients of variation between 5 % and 13 %. However, in local studies focusing on borehole measurements, coefficients of variation as large as 40 % have been found (El Haber et al. 2019; Chaljub et al. 2021). Although those values are associated with smaller correlation lengths, it indicates that it is worth exploring large coefficients of variation.

1.3 Evaluation metrics

It is often required to compare ground motion time series, either between numerical simulations and recorded seismograms, or between numerical simulations and the predictions of a surrogate model. The comparison should account for the three components of ground motion, the temporal evolution, as well as all spatial points where ground motion is available. No single metric can provide an exhaustive measure of the agreements between two sets of time series and we need to evaluate several metrics to get a proper understanding of the agreement between the time series.

Let us consider a reference velocity time series $\boldsymbol{v}(\boldsymbol{x},t)$ and an approximation $\widetilde{\boldsymbol{v}}(\boldsymbol{x},t)$, both depending on space $\boldsymbol{x}$ and time t. Their numerical representation is given at N_x spatial points $\boldsymbol{x_1},\cdots,\boldsymbol{x_{N_x}}$ and N_t time instants $t_1,\cdots,t_{N_t}$. Our typical variable of interest is the ground motion velocity but the metrics can be defined for other variables. In the usual framework, the space is three-dimensional ($d=3$), which means that ground motion has three components (u_E, u_N, u_Z, see Fig. 1.1).

1.3.1 Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)

The Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are common metrics to compare data in machine learning. They are often used in their relative formulation

$$rMAE(\boldsymbol{x}) = \frac{1}{N_t}\sum_{t=1}^{N_t}\frac{\|\widetilde{\boldsymbol{v}}(\boldsymbol{x},t)-\boldsymbol{v}(\boldsymbol{x},t)\|_1}{\|\boldsymbol{v}(\boldsymbol{x},t)\|_1+\varepsilon} \tag{1.24}$$

$$rRMSE(\boldsymbol{x}) = \sqrt{\frac{1}{N_t}\sum_{t=1}^{N_t}\frac{\|\widetilde{\boldsymbol{v}}(\boldsymbol{x},t)-\boldsymbol{v}(\boldsymbol{x},t)\|_2^2}{\|\boldsymbol{v}(\boldsymbol{x},t)\|_2^2+\varepsilon^2}} \tag{1.25}$$

where ε (typically ε=0.01 m/s in this work) prevents numerical artefacts due to small values.

1.3.2 Frequency biases

The MAE and RMSE are point-wise metrics that do not give information on the accuracy of the approximated physical phenomena. We propose to use frequency biases to represent the agreement between the Fourier coefficients of the approximation $\widetilde{\boldsymbol{v}}$ and the *ground truth* $\boldsymbol{v}$. To do so, we compute the mean of Fourier coefficients on several frequency intervals: low frequency (0-1Hz), medium frequency (1-2Hz), and high frequency (2-5Hz)

$$\overline{\mathcal{F}\left(v(\boldsymbol{x})\right)}_{low} = \frac{1}{N_f}\sum_{0\leq f<1}|\mathcal{F}(v(\boldsymbol{x}))(f)| \tag{1.26}$$

In equation (1.26), N_f denotes the number of frequencies f between 0 and 1Hz and $\mathcal{F}(v(\boldsymbol{x}))(f)$ is the Fourier coefficient of the time series $v(\boldsymbol{x})$ at frequency f. Definitions naturally extend to medium and high frequencies. Frequency biases are then defined by

$$rFFT_{low}(\boldsymbol{x}) = \frac{\overline{\mathcal{F}(\widetilde{v}(\boldsymbol{x}))}_{low}-\overline{\mathcal{F}(v(\boldsymbol{x}))}_{low}}{\overline{\mathcal{F}(v(\boldsymbol{x}))}_{low}} \tag{1.27}$$

It should be noted that positive frequency biases indicate an overestimation of the frequency content compared to the reference ($\overline{\mathcal{F}\left(\widetilde{v}(\boldsymbol{x})\right)} \geq \overline{\mathcal{F}\left(v(\boldsymbol{x})\right)}$) while negative values indicate an underestimation. Noting that the mean of Fourier coefficients amplitudes $\overline{\mathcal{F}\left(v(\boldsymbol{x})\right)}$ is always positive, this implies that the frequency biases cannot be lower than -1. Also, the frequency biases are computed individually for each of the three components of ground motion and then, averaged.

Alternatively, Anderson 2004 defines the frequency bias as the logarithm ratio

$$\ln\left(\frac{\overline{\mathcal{F}(\widetilde{v}(\boldsymbol{x}))}_{low}}{\overline{\mathcal{F}(v(\boldsymbol{x}))}_{low}}\right) \tag{1.28}$$

The two definitions are not very different when the error between the prediction and the *ground truth* is small, which is our main focus of interest. Indeed, writing $\overline{\mathcal{F}(\widetilde{v}(\boldsymbol{x}))}_{low} = (1+\alpha)\overline{\mathcal{F}(v(\boldsymbol{x}))}_{low}$, our definition of the frequency bias (equation 1.27) becomes α while the logarithm ratio (equation 1.28) becomes $\ln(1+\alpha)$. When α is relatively small (typically $|\alpha|\leq 0.5$), both values are close.

1.3.3 Goodness-Of-Fit (GOF)

In seismology, time series are often compared in terms of time-frequency envelope and phase misfits Kristeková et al. 2009. The time-frequency representation W of the signal v is computed from the continuous wavelet transform

$$W_{\boldsymbol{x}}(t,f) = \sqrt{\frac{2\pi|f|}{\omega_0}} \int_{-\infty}^{\infty} v(\boldsymbol{x},\tau)\psi^* \left(2\pi f \frac{\tau - t}{\omega_0}\right) d\tau \tag{1.29}$$

Misfits are evaluated on the envelope of the signals $|W_{\boldsymbol{x}}(t,f)|$ and on the phase $Arg(W_{\boldsymbol{x}}(t,f))$. They can be summarized by two scalar values: the envelope Goodness-Of-Fit (GOF) and the phase GOF. The envelope GOF expresses the error of the predicted amplitude compared to the reference time series while the phase GOF describes the error of the phase arrivals. GOFs are given on a scale between 0 and 10 where 10 means a perfect agreement and the following score assessment is well accepted: 0-4 is a poor score, 4-6 is fair, 6-8 is good, 8-10 is excellent (Kristeková et al. 2009). Figure 1.8 illustrates the time-frequency GOF between a reference simulation and a surrogate model prediction. Lower GOF values are found at the time instants where signals are the most different.

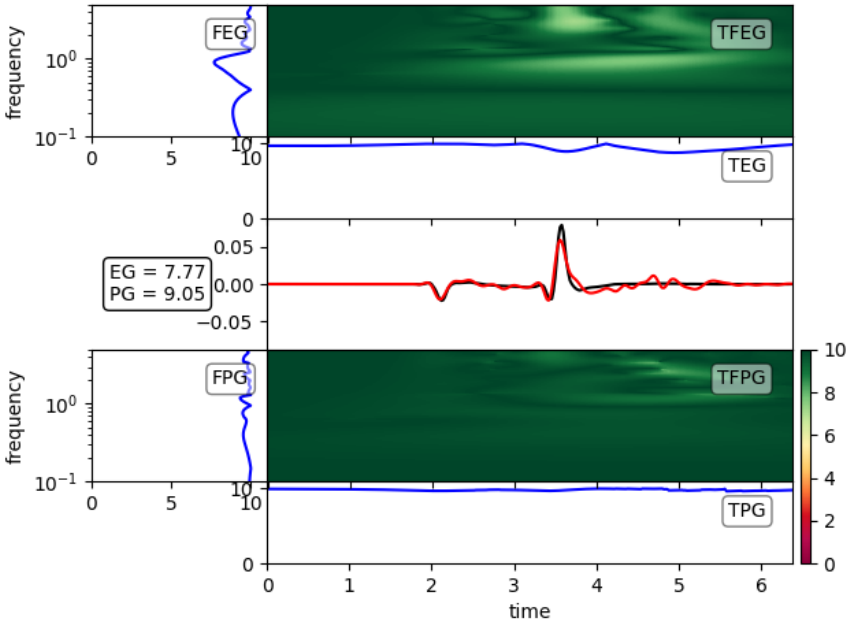


Figure 1.8: Illustration of the time-frequency representation and GOF values. In the central panel, the black line is the reference simulated ground motion and the red line is the ground motion predicted by our surrogate model (same configuration as Fig. 3.8 but for a different sensor). At the top, the envelope GOF is shown with respect to time and frequency. On the bottom, the phase GOF is shown. The final GOF scores are the envelope GOF (EG) and phase GOF (PG).

Olsen and Mayhew 2010 states that GOFs are well adapted for low-frequency comparisons, which is the framework of this work since the maximum frequency solved in the simulations is 5 Hz. Also, one should note that the GOFs are computed simultaneously on the three components and the normalization accounts for the three components.

1.3.4 Intensity measures

Several intensity measures are used in seismology to evaluate specific features of ground motion. They are often related with structural response.

The Peak Ground Velocity (PGV) is computed as the maximum absolute value of the velocity time series

$$PGV(v)(\boldsymbol{x}) = \max_t |v(\boldsymbol{x},t)| \tag{1.30}$$

Similarly, the Peak Ground Displacement and Peak Ground Acceleration correspond to the maximum of the displacement and acceleration time series. The PGV can be computed component-wise, on the norm of the three components or more specifically, on the geometrical mean of the two horizontal components.

The Relative Significant Duration (RSD) indicates the duration of ground motion. It corresponds to the duration of the signal between 5% and 95% of the Arias intensity (I_A) (Arias 1970)

$$I_A = \frac{\pi}{2g} \int_0^T a^2(t) dt \tag{1.31}$$

where $a(t)$ is the acceleration field and T is the total duration of the signal.

The Cumulative Absolute Velocity (CAV) is often computed to provide a proxy of the velocity field's energy content since it is found to be related to structural damage (M.-H. Wu et al. 2022)

$$CAV = \int_0^T |a(t)| dt \tag{1.32}$$

1.3.5 Pseudo-Spectral Acceleration (PSA)

The Pseudo-Spectral Acceleration (PSA) is the most commonly used metric to estimate structural response. If a structure is approximated by a linear oscillator with a given natural period T_0 and a damping level (usually 5 % damping), one can compute the displacement d_{osc} and acceleration a_{osc} of the oscillator when the structure is subject to ground acceleration $\boldsymbol{a}(t)$. Then, the spectral displacement is $SD(T_0) = \max_t |d_{osc}(t)|$ and the spectral acceleration is $SA(T_0) = \max_t |a_{osc}(t)|$. Despite being close to the spectral acceleration, the "true mesure" (J. W. Baker and Cornell 2006) of the structural response is the pseudo-spectral acceleration (PSA), defined as

$$PSA(T_0) = \left(\frac{2\pi}{T_0}\right)^2 SD(T_0) \tag{1.33}$$

1.4 Simulations of the Le Teil earthquake (France, 2019)

This section presents the work done in 2021-2022 to simulate the Le Teil earthquake with SEM3D and the data available at this time. In particular, this study was conducted before the publication of Smerzini et al. 2023 and El Haber et al. 2021 within the framework of the SIGMA-2 project. They investigated several simulation methods to reproduce the recordings of the Le Teil earthquake, including large-scale 3D physics-based simulations. We do not aim at comparing our results with theirs (geological models and source parameters are different) but instead, we intend to show the limitations of regional models and how random heterogeneities can complement these models. This work was published in F. Lehmann et al. 2022 and is mostly reproduced below.

1.4.1 Context of the work

The Rhône valley was hit by a M_W4.9 earthquake on November, 11th, 2019 known as the Le Teil earthquake. The Le Teil earthquake is the most damaging earthquake of the last decade in metropolitan France. While the seismic source can be constrained by InSAR data (De Novellis et al. 2020; Vallage et al. 2021), determining local geological models is more challenging. In this region with a low instrumental seismicity (Larroque et al. 2021), poor geological measurements are available at present. Although geophysical campaigns were conducted after the earthquake, data remain too sparse to design a geological model at the scale of interest (Marconato et al. 2022). Therefore, the most specific geological model is the 3D regional geological model (P- and S-waves) recently elaborated within the framework of the SIGMA-2 international project (Arroucau 2020). Thus, in the absence of a validated local model, it is legitimate to explore the abilities of regional-scale models in regional numerical simulations.

However, due to its insufficient resolution, this 3D geological model shows only gentle horizontal variations representing the main geological patterns of the region (i.e. a 50 km $\times$ 115 km domain across the Rhône Valley characterized by a crystalline basement with higher velocities in the Massif Central in the Northwest and sediments in the plain (Ritz et al. 2020)). The 3D model especially lacks the description of the sedimentary basin extending along the Rhône valley (Bravard and Gaydou 2015). This prevents the possibility to constrain numerical simulations in a broad-band spectrum (0-10 Hz at least). One possible remedy consists in modelling the sedimentary basin from

a Digital Elevation Model. Then, velocity fields can be inferred from 1D velocity profiles acquired by seismic inversion inside and outside the basin, as done by Smerzini et al. 2023; El Haber et al. 2021. This method has already been adopted in other low-to-moderate seismicity regions such as the Grenoble region (Chaljub 2006; Stupazzini et al. 2009).

Another modelling challenge is introduced by small-scale heterogeneity, which plays a crucial role in seismic wave propagation, especially at high-frequency (Vyas et al. 2018; Scalise et al. 2021). Heterogeneity is rarely included in numerical simulations due to the increased computational demand. It can be represented by random fields that add small scale fluctuations to the geophysical models (Gatti et al. 2017; Svay et al. 2017).

1.4.2 Data

Source models

The Le Teil earthquake occurred on the La Rouvière fault, a fault which was not considered as active despite being part of the larger Cevennes fault system which was potentially active (Ritz et al. 2020). Remarkably, this shallow earthquake (1.3 km depth ± 0.5 km) induced a surface rupture. Figure 1.9 shows the map of the region of interest, including the available recording stations and the trace of the La Rouvière fault.

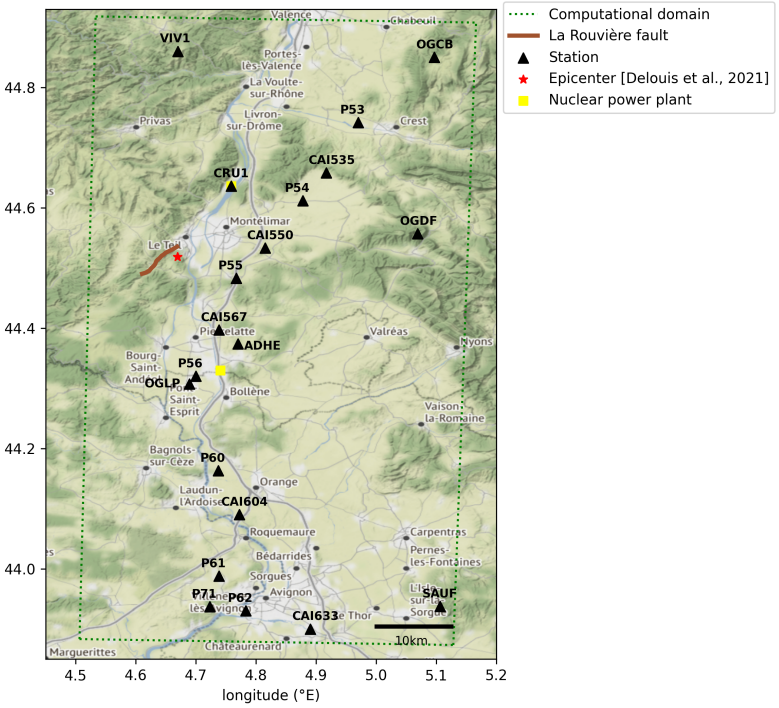


Figure 1.9: Map of the region affected by the 2019 Le Teil earthquake, in South-Eastern France. The computational domain is indicated with the dotted box. Velocimeters and accelerometers are shown with black triangles (details in Tab. A.1).

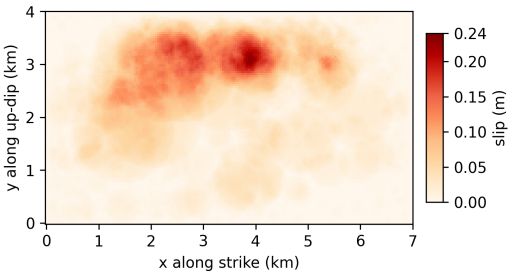


Figure 1.10: Final slip given by the kinematic fault model (Vallage et al. 2021).

The Le Teil earthquake was recorded by 22 stations within 70 km from the fault (Table A.1). Some velocimeters (e.g. PAUL and BOLL in Figure 1.9) saturated while recording and they could not be considered for further analyses. This study focuses on the stations OGDF, OGCB, CRU1 located out of the sedimentary basin since its absence from our geological models is likely to impact synthetic ground motions inside the basin.

The hypocenter was obtained from the results of the waveform inversion (44.5188°N, 4.6694°E, depth −1.3 km, Delouis et al. 2021). In this study, two types of seismic sources were compared, with a target seismic moment $M_0 = 2.47 \times 10^{16}$ N m. Namely, a double-couple point source was assumed with strike = 48°, dip = 45°, and rake = 88° (Delouis et al. 2021). Its temporal evolution is a spice bench

$$t \mapsto 1 - \left(1 + \frac{t}{\tau}\right) e^{-\frac{t}{\tau}}, \tag{1.34}$$

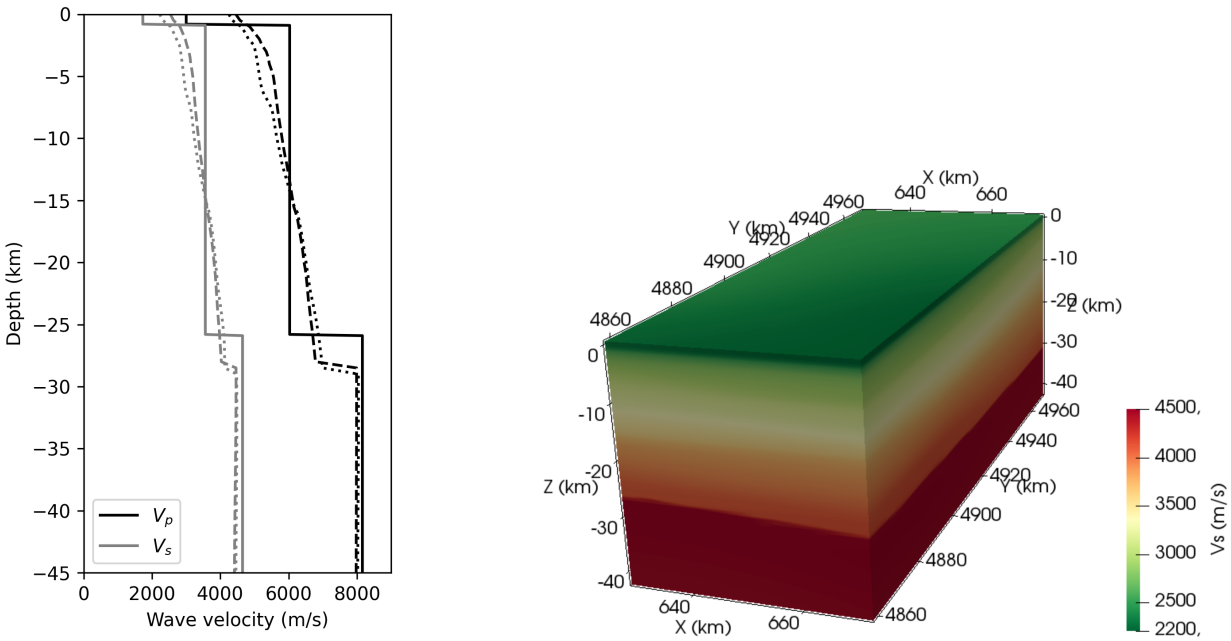
with $\tau = 10^{0.5(M_W - 6.69)}$ = 0.127 s (Dreger et al. 2007; Gatti 2017). In addition, a kinematic fault model was designed from the inversion of InSAR satellite observations performed by Vallage et al. 2021 (Fig. 1.10). The fault model was computed following the Ruiz's Integral Kinematic (RIK) numerical scheme (Ruiz et al. 2011) implemented by Gallovič 2016. The fault was represented as a 7 km-long and 4 km-deep plane with a 60° dip and reaching the surface at its highest point. It was further discretized in triangular patches. A bidirectional rupture front starts from the nucleation point located at the hypocenter. Each patch activates when it is reached by the rupture front, under the constraint that its final slip corresponds to the one obtained from the InSAR inversion.

Geological models

In the absence of a local model, two geological models for metropolitan France were adopted in this study and compared. The first one is a 1D model used by the Geophysical and Detection Laboratory (LDG) of the French Alternative Energies and Atomic Energy Commission (CEA) to locate seismic events (Table 1.2, Figure 1.11a and Duverger et al. 2021). It was obtained using the Pg, Sg, Pn, and Sn phases of a series of 50+ well identified earthquakes (Veinante-Delhaye and Santoire 1980). It presents a thin sedimentary subsurface layer with low velocity (V_S = 1730 m/s), overlying a 25 km thick crustal layer (V_S = 3560 m/s). The average ratio between P- and S- wave velocities is 1.69. The bedrock is described by S-wave velocities of 4650 m/s.

layer	thickness (m)	V_P (m/s)	V_S (m/s)	ρ (kg/m^3)	Q_P	Q_S
1	900	3000	1730	2700	200	61
2	25,000	6030	3560	2700	300	128
3	15,000	8160	4650	3300	1000	500

Table 1.2: 1D geological model used by the CEA-LDG to locate seismic events. Described in more details in (Duverger et al. 2021).



(a) V_P (black) and V_S (grey) velocity profiles for the 1D geological model (solid line), 3D geological model in station VIV1 (dashed line), and station SAUF (dotted line).

(b) 3D geological model for S-wave velocity from Arroucau 2020

Figure 1.11: Geological models used in the simulations

Alternatively, Arroucau 2020 proposed in the framework of the SIGMA2 project a 3D geological model for S- and P-waves in metropolitan France. This model includes topography/bathymetry and was built as an improved

and homogeneized version of partial previous models (mainly *EPcrust* (Molinari and Morelli 2011), combined with ambient noise and teleseismic surface wave tomography models, a CSS-derived model, and a local earthquake tomography study, see references in Arroucau 2020). This model has a 10 km × 10 km × 0.5 km resolution and shows horizontal variability even at the regional scale of interest. Figure 1.11b for example shows higher surface velocities in the North-Western mountainous part of the region compared to the plains in the South-East. The thickness of the Earth crust is between 26 km and 31 km in the region of interest, consistently with geophysical knowledge (Larroque et al. 2021). This model also leads to a mean V_P /V_S ratio of 1.72 (between 1.68 and 1.8), which is slightly lower than the 1.9 ratio recommended by Delouis et al. 2021 to recover S-waveforms of the Le Teil earthquake.

Heterogeneous geological models

Random fields were added to the 1D geological model to create small scale heterogeneities. They were drawn independently in each of the three layers, thus possibly creating sharp interfaces between layers. Following previous works described in Section 1.2, we adopted correlation lengths of 10 km in the horizontal direction and 1 km in the vertical direction, associated with a 10 % coefficient of variation. This parametrization is coherent with the results obtained for the Le Teil earthquake via a Monte-Carlo approach of particles diffusion in a heterogeneous Earth crust (Heller 2021). Thanks to the spectral representation (Equation 1.2.2), random fields were generated in a few seconds for the whole domain with a 2 km × 2 km × 0.2 km resolution. They were later interpolated on the Gauss-Lobato-Legendre (GLL) points as part of the computation process.

Simulation framework

The 50 km × 115 km × 38 km computational domain was discretized on a hexahedral mesh with 6.7 million elements. With a minimum S-wave velocity of 1730 m/s (in the 1D geological model) and 5 GLL points per element, this mesh allowed wave propagation up to 5 Hz. Simulations were run on 2048 cores AMD Milan @2.45GHz (AVX2) operated by the *Très Grand Centre de Calcul* (TGCC, France). Thanks to this computational power, simulations were obtained in 5800 h CPUs for 60 s of simulated signal.

Figure 1.12 summarizes the configurations investigated in this study, with different combinations of sources and geological models. When running the simulations with the heterogeneous models, we focus on the point-source description to limit the interactions between the source and the medium which may alter the results independently from path effects.

1.4.3 Comparison of simulations with records

Numerical results were compared with seismograms records in several stations to evaluate the parameters choices. First, the point-source was used in conjunction with the 1D geological model. Figures 1.13 and A.1 show that despite the simplicity of the models, the level of agreement is surprisingly good. Indeed, the frequency response spectra show a correct corner frequency around 0.8 Hz, as well as similar slopes and amplitudes.

In addition, the wave arrival times are well reproduced by the simulation, proving that the mean velocity on the source-to-site path is correctly modelled (Fig. 1.14). The very first oscillations are also coherent with the records, especially in station OGDF (Fig. 1.14a).

However, one can notice in Figures 1.14a and 1.14b that the numerical simulation produces late oscillations with higher amplitudes than the recorded ones. They can be seen from 27 s in station OGDF and from 36 s in station OGCB. These oscillations may come from the thin upper layer with a high velocity contrast defined in the 1D geological model that creates a wave guide. In fact, when the 3D geological model is used instead of the 1D geological model, those late oscillations are no more present, as can be seen in Figures A.2a and A.2b.

Yet, the use of the 3D geological model leads to other issues. Figures A.2a and A.2b indeed show high velocity peaks at the beginning of the signal. The peaks's amplitude is higher than the maximal velocities recorded during the earthquake. As an example, the horizontal Peak Ground Velocity (PGV) was 5 times higher than records in station OGDF and 7 times higher in station OGCB. The velocity peaks indicate that the energy content of the signal is concentrated in the first wave arrivals. This results from the smoothness of the 3D geological model: the absence of inter-layer discontinuity prevents the multiple wave refractions and reflections that tend to spread the energy distribution over time.

To better represent the time duration of the signal, the point-source was replaced by an extended fault model and used in conjonction with the 3D model. Figures 1.15 and A.3 show a satisfactory agreement between the

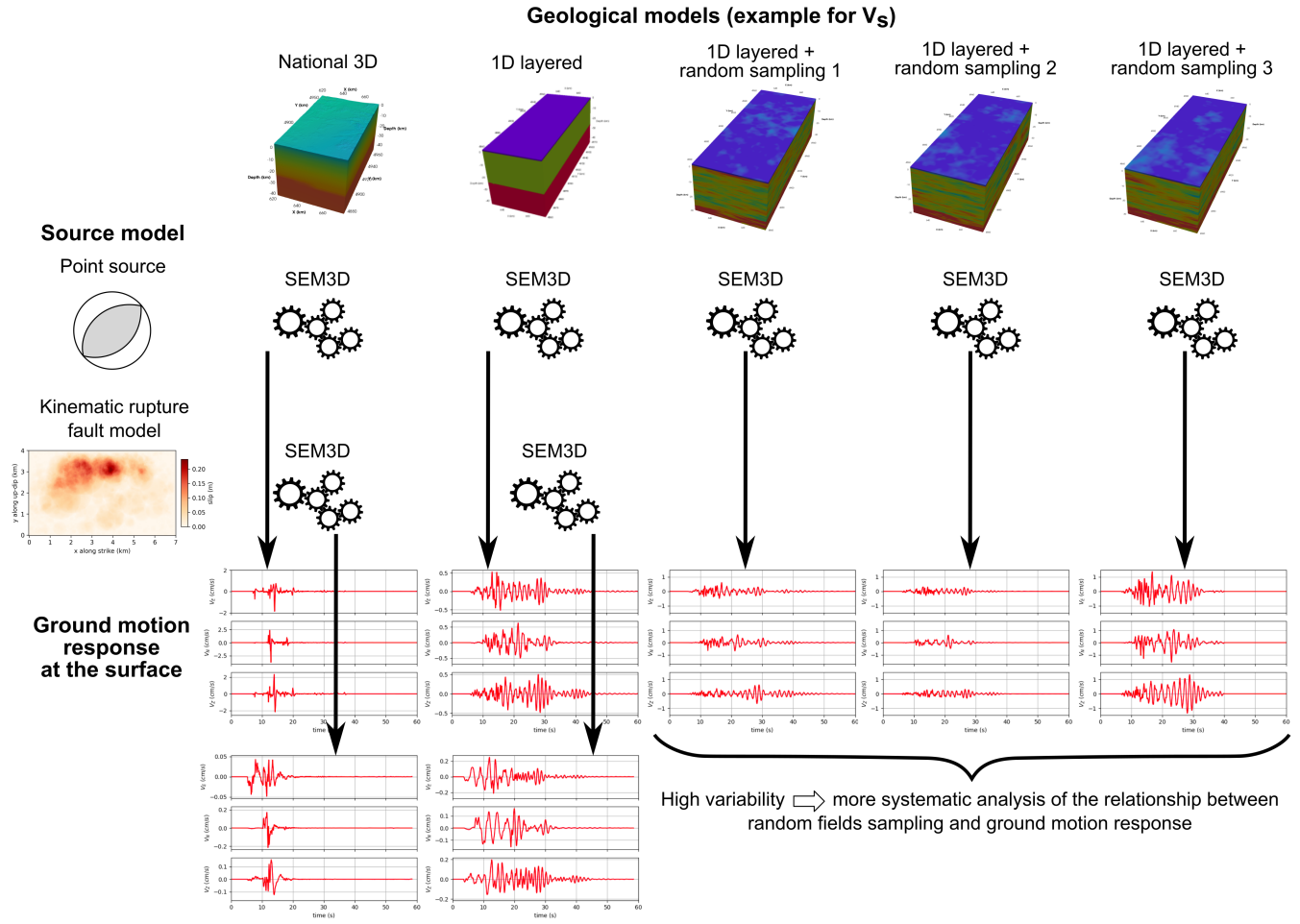


Figure 1.12: The simulation workflow involves a 3D geological model and a 1D geological model that was perturbed by three random field samplings. Inside each model, seismic waves from a point source were simulated. Ground motion generated by the kinematic fault model was also computed in both homogeneous models.

recorded and synthetic frequency response spectra in this setting. As expected, the successive nucleation of points on the fault plane creates an energy distribution that avoids the large peaks observed with the point source (Figure A.4). The horizontal PGV was hence reduced to 1.15 and 5 times the recorded one in stations OGDF and OGCB respectively.

1.4.4 Effects of geological heterogeneities on simulated ground motion

Figure 1.16 shows that introducing heterogeneities reduces the early peaks' amplitude compared to the signals generated with the homogeneous 1D geological model. Heterogeneities also limit the duration and scale of the surface wave oscillations. This behaviour is interpreted as a consequence of the diffraction induced by heterogeneities that spread the energy content and limit the wave guide effect. Therefore, the signal obtained with a heterogeneous medium seems more realistic than those originating from the homogeneous model.

However, one cannot claim that one sampling of a random field can represent the variability of all possible heterogeneities. To assess the possible impacts of heterogeneities while maintaining reasonable computational costs, two other different random fields were drawn and added to the 1D geological model. Figures A.5a and A.5b show the ground motion response in station OGDF with those random fields. One can notice that those samplings tend to increase the amplitude of the late oscillations compared to those obtained with the homogeneous medium. Therefore, the influence of random fields on the surface wave oscillations was not consistent between samplings.

Given the large variability in ground motion responses arising from the three random fields samplings, it is cru-

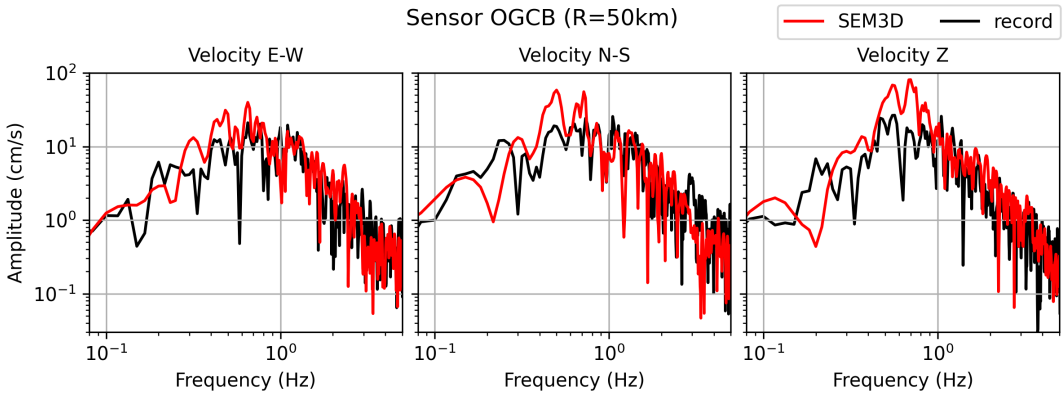


Figure 1.13: Frequency response spectra of the numerical simulation (in red) obtained with the 1D geological model and a point source. Comparison with seismograms records (in black). Velocities are given in the East-West (E-W), North-South (N-S), and vertical (Z) directions for the station OGCB. R denotes the epicentral distance.

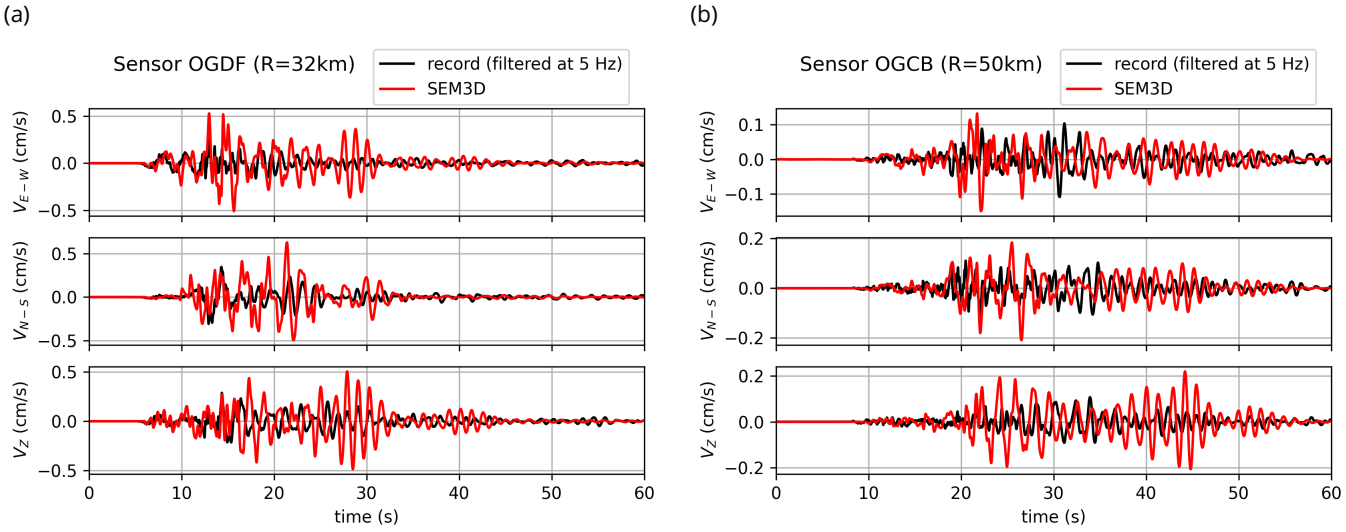


Figure 1.14: Results of the numerical simulation (in red) obtained with the 1D geological model and a point source. Comparison with seismograms records (in black) filtered at 5 Hz. Velocities are given in the East-West (E-W), North-South (N-S), and vertical (Z) directions for stations OGDF (a) and OGCB (b)

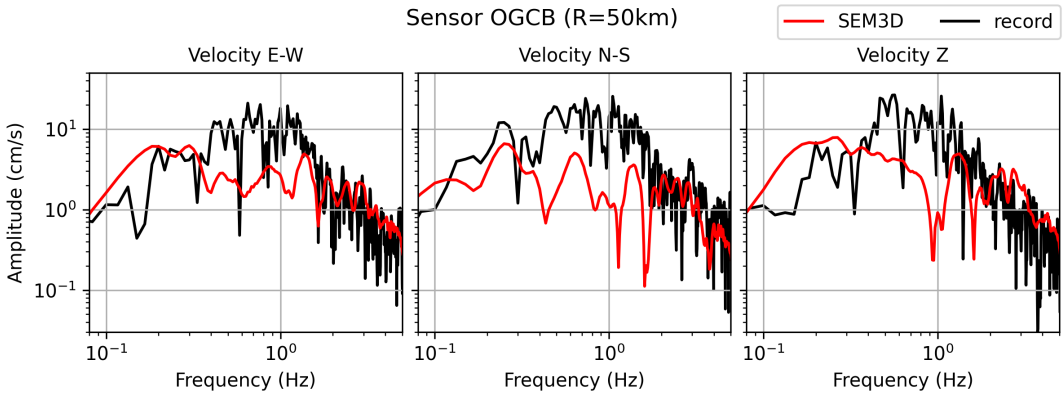


Figure 1.15: Frequency response spectra of the numerical simulation (in red) obtained with the 3D geological model and a kinematic fault model. Comparison with seismograms records (in black). Velocities are given in the East-West (E-W), North-South (N-S), and vertical (Z) directions for the station OGCB.

cial to ensure that those responses remain physically plausible. This was done by computing the mean horizontal Pseudo Spectral Acceleration (PSA) in three stations for the 1D geological model and the three heterogeneous models. Then, we compared the synthetic PSA with the one given by a regional Ground Motion Prediction Equation (GMPE, Berge-Thierry et al. 2003). Figure 1.17 shows that the synthetic PSAs were within the confidence bounds of the GMPE, thereby ensuring that our heterogeneous models were realistic. More precisely, the first and second random field samplings were close to the GMPE. The third sampling led to PSA slightly higher than the upper bound of the GMPE for stations CRU1 and OGDF. However, it is noteworthy that the record in station OGCB was also out of the confidence interval. Therefore, we had no reason to reject the third sampling based on the sole analysis of the PSA.

Moreover, Figure 1.17 shows that both the homogeneous 1D geology and the second sampling of the heterogeneous geology were close to the records. Although comparing the PSA did not lead to a preference for one of these two models, we showed above that the heterogeneous model was able to reduce the surface wave oscillations.

1.4.5 Partial conclusion

Considering the sparsity of available geological data in the Le Teil region, a 1D layered geological model was not rich enough to accurately simulate the Le Teil earthquake. Geological models can be improved by the addition of random fields that however yield a large ground motion variability. We showed that the main ground motion characteristics of the Le Teil earthquake can be reproduced using regional geological models. The frequency response spectra were in satisfactory agreement with the recorded seismograms. However, the 1D geological model, with its peculiar subsurface layer, induced surface waves with high-amplitude oscillations. These oscillations were significantly reduced when adding random fields to the geological model, therefore leading to more realistic signals.

With the 3D geological model, surface wave oscillations disappeared. However, this model needed to be used in conjunction with the kinematic fault model to ensure that the signal energy was correctly spread over time. Otherwise, the point source model led to peak ground velocities much larger than the recorded ones. We also found some differences between ground motion generated from a point source and from an extended kinematic fault model. We interpret these differences as plausibly coming from the low depth of the Le Teil fault.

Although the addition of random fields on the 1D geological model could reduce the unrealistic surface waves oscillations, this effect was not necessarily consistent between stations and between different random fields samplings. Therefore, a larger diversity of random fields is necessary to better understand the impacts of heterogeneities on ground motion.

1.5 Conclusion

We have presented the elastic wave equation that models the propagation of seismic waves in the Earth's crust. This equation is solved numerically with the Spectral Element code SEM3D that provides an efficient High-Performance Computing (HPC) framework. Simulations are parametrized by a 3D geological model that often relies on random

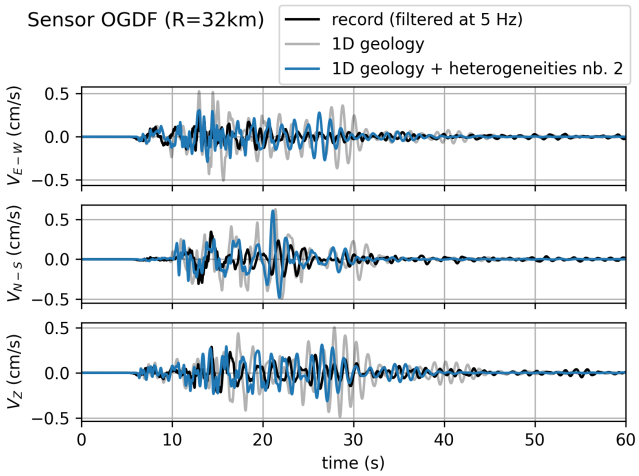


Figure 1.16: Numerical results in station OGDF obtained with a point source and the 1D geological model enhanced with random fields (blue). Comparison with the results of the homogeneous 1D geological model (grey) and the records filtered at 5 Hz (black).

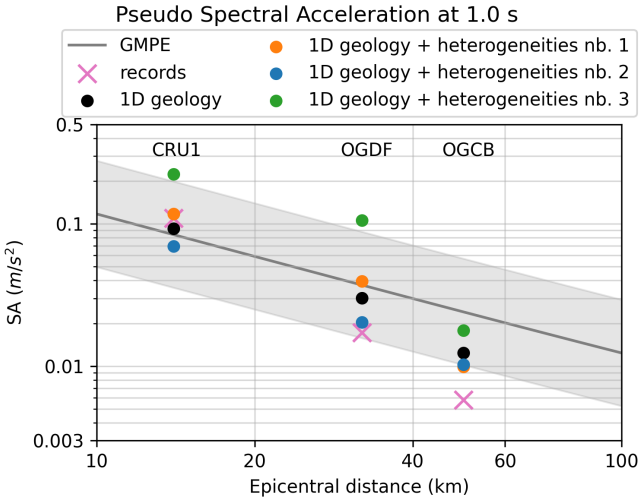


Figure 1.17: Pseudo Spectral Acceleration (PSA) at 1 s in 3 stations (CRU1, OGDF, OGCB). Mean horizontal PSA for records (pink cross), 1D model (black dots), 1D model with random fields (colored dots), and a GMPE (grey line) with the associated confidence interval (Berge-Thierry et al. 2003).

fields to represent geological heterogeneities and account for our lack of geophysical knowledge in the region of interest. Although the seismic source can be defined as an extended fault plane, it is often approximated by the equivalent point-wise source with the moment tensor description.

Simulations of the Le Teil earthquake highlighted that numerical simulations can reproduce recorded seismograms, provided that geological models and source parameters are accurate enough. However, quantitative analyses of the variability induced by heterogeneous models would require hundreds of thousands of simulations in a typical Monte Carlo framework, which is computationally intractable. Based on the sensitivity analysis performed in Chapter 4, let us assume $N_{MC} = 10^5$ Monte-Carlo evaluations of ground motion with different geological models. Knowing that one SEM3D simulation in the whole domain lasts for 5800 h CPUs, this would amount to an energy consumption of 2.8 GWh (Tab. 1.3).

Task	Number of realizations	Energy consumption (MWh)
SEM3D simulations 50 km × 115 km × 38 km, T_f=60 s	$N_{MC} = 10^5$	2800

Table 1.3: Estimation of the energy consumption that would be necessary for uncertainty quantification analyses with large-scale SEM3D simulations. The energy consumption of 1 CPU is estimated to 4.88 Wh.

This study illustrates the need for a surrogate model able to predict the ground motion generated by a realistic earthquake source when seismic waves propagate through a complex 3D geological medium. The surrogate model should be fast enough to allow repeated runs while preserving a good accuracy.

Chapter 2

First attempts to build a reduced-order model

In the perspective of seismic hazard analyses, we focus on the forward problem of wave propagation and we investigate the design of a surrogate model to replace SEM3D physics-based simulations. The surrogate should offer a flexibility similar to simulations by

1. taking as input a 3D geological model that enables a detailed description of the Earth's crust properties, and especially geological heterogeneities
2. considering different earthquake source properties
3. predicting ground motion time series for each spatial location

The surrogate model should be accurate enough so that predicted ground motion is reliable and physically consistent, and it should be significantly faster than the physics-based simulations.

Machine learning provides several well-used surrogate models, such as Gaussian processes, polynomial chaos expansions, and projection-based reduced-order models. Despite many developments and several applications in seismology, none of these surrogate models offers a framework that meets all our requirements (Section 2.1). Surrogate models that depend on a 3D heterogeneous geology are especially rare, although we showed in Chapter 1 that the influence of geological heterogeneities on ground motion should be accurately quantified.

To fill in this gap, we design a database of 30 000 heterogeneous geological models and compute the ground motion generated by an earthquake source in each geology (Section 2.2). Existing databases in the literature either compile recorded ground motion (e.g. STanford EArthquake Dataset (Mousavi et al. 2019), MLAAPDE dataset (Cole et al. 2023)), in which case information on the geological properties between the earthquake source and the recording sensor are almost nonexistent, or provide synthetic 3D geological models (e.g. Noddyverse dataset Jessell et al. 2022) with no associated ground motion. We provide the first database of physics-based earthquake simulations in complex 3D domains with enough samples to allow machine learning applications. Our HEMEW-3D database is publicly available (https://doi.org/10.57745/LAI6YU) and its analysis is under review in F. Lehmann et al. (2024a). "Synthetic Ground Motions in Heterogeneous Geologies: The HEMEW-3D Dataset for Scientific Machine Learning". In: *Earth System Science Data Discussions* 2024, pp. 1–26. DOI: 10.5194/essd-2023-470.

Due to their dependence on three spatial coordinates, our geological models are described by a large number of variables. In many surrogate models, one aims at reducing the dimension of variables to capture only the meaningful relationships in a latent space of smaller dimension than the initial data space. To this end, we compare a linear (Principal Component Analysis, PCA) and a non-linear (UNet auto-encoder) dimensionality reduction method (Section 2.3). After evaluating the quality of the reduced geological dimension, we show its influence on ground motion and highlight several shortcomings of this approach. This work was published as the second part of F. Lehmann et al. (2022). "Machine Learning Opportunities to Conduct High-Fidelity Earthquake Simulations in Multi-Scale Heterogeneous Geology". In: *Frontiers in Earth Science* 10. ISSN: 2296-6463. DOI: 10.3389/feart.2022.1029160.

2.1 Machine learning-based surrogate models

In the following sections, we highlight some commonly used surrogate models like Gaussian processes, polynomial chaos expansion and projection-based reduced order models. While acknowledging the numerous theoretical advancements that extend those methods beyond the reference framework presented here, we focus on studies that

have direct applications in the prediction of ground motion. Several of these works also inspired further developments of our surrogate model and their significance should not be understated.

2.1.1 Gaussian processes

A Gaussian Process (GP) is "a collection of random variables $(Y_{\mathbf{x}})_{\mathbf{x}\in\mathcal{X}}$, any finite number of which have a Gaussian distribution" (Rasmussen and Williams 2006). The index set $\mathcal{X}$ is the set of possible inputs, for instance $\mathcal{X} \subset \mathbb{R}^d$ may represent the spatial coordinates of the Gaussian process. As examples, one can think of Y as describing geological properties or ground motion displacement as a function of space. A Gaussian Process is entirely characterized by its mean function m and covariance function k

$$m(\mathbf{x}) = \mathbb{E}[Y_{\mathbf{x}}] \tag{2.1}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(Y_{\mathbf{x}} - m(\mathbf{x}))\left(Y_{\mathbf{x}'} - m(\mathbf{x}')\right)\right] \tag{2.2}$$

One suitable property of Gaussian processes lies in the ease of obtaining new data from a set of training data. More precisely, let us assume zero mean, a given expression of the covariance function (e.g. the squared exponential covariance function $k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2\ell}\|\mathbf{x} - \mathbf{x}'\|^2\right)$) and a set of training data of the form $(\mathbf{x}_i, Y_{\mathbf{x}_i})_{1\leq i\leq n}$. Training data may come from (assumed noise-free) observations and can be written in the short form $\mathbf{x}_{1:n} := (\mathbf{x}_1, \cdots, \mathbf{x}_n)$, $Y_{1:n} := (Y_{\mathbf{x}_1}, \cdots, Y_{\mathbf{x}_n})$. Then, for a new observation point $\mathbf{x}_* \in \mathcal{X}$, it is well-known that the output Y_* can be computed from the joint prior distribution conditioned on the observations (Rasmussen and Williams 2006, eq. 2.19)

$$\begin{aligned} Y_*|\mathbf{x}_*, \mathbf{x}_{1:n}, Y_{1:n} \sim \mathcal{N}\big(&k(\mathbf{x}_*, \mathbf{x}_{1:n})k(\mathbf{x}_{1:n}, \mathbf{x}_{1:n})^{-1}Y_{1:n}, \\ &k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{x}_{1:n})k(\mathbf{x}_{1:n}, \mathbf{x}_{1:n})^{-1}k(\mathbf{x}_{1:n}, \mathbf{x}_*)\big) \end{aligned} \tag{2.3}$$

One major difficulty is to define a suitable covariance function and adjust its hyperparameters (σ and ℓ in the example of the squared covariance function) to match the training data.

Illustrative examples of Gaussian processes in our seismological objective can be divided in two broad classes. In the first class, the geological model is prescribed and one aims at finding an interpolator of ground motion displacement between a set of observation points. When considering the 3D homogeneous acoustic wave equation, Henderson et al. 2023 showed that it is possible to build a Gaussian process that approximates the numerical solution from a small number of observations (30 sensors inside a cubic domain). However, the geological parameter is constant and fixed, meaning that the obtained Gaussian process is specific to a given geological model. A similar application was conducted by Tamhidi et al. 2022 from numerical simulations of the 1906 M_w 7.9 San Francisco earthquake. They were able to estimate the acceleration time series at any point from a set of synthetic seismograms but, once again, the geological model used in the simulation is fixed. In this setting, Gaussian Processes are similar to Physics-Informed Neural Networks (PINNs, Chapter 3) in the sense that they output (an approximation of) the solution of the wave equation in a given geological model.

In the second class of Gaussian processes applications, one considers varying geological configurations and one wants to estimate some ground motion features corresponding to each configuration. In this context, ground motion is not predicted as an entire time series but with some scalar features such as the spectral acceleration at a given period (corresponding to the random variable Y). And the input $\mathbf{x}$ is a set of parameters related to the earthquake (e.g. magnitude, epicentral distance) and the ground characteristics at the location of interest (e.g. $V_{S,30}$). This is the typical framework of ground motion models where ground motion and inputs can be modelled with Gaussian processes (Alimoradi and Beck 2015; Lacour 2023). However, this framework does not satisfy our requirement that the surrogate model should provide ground motion time series.

2.1.2 Polynomial chaos expansion

Polynomial chaos expansion (PCE) is another well-used surrogate model, especially thanks to its strong relationship with sensitivity analysis (Sudret 2008). This method relies on the decomposition of a random variable Y as a sum of orthogonal polynomials ϕ (Wiener 1938; Ghanem and Spanos 1991)

$$Y(\boldsymbol{\xi}) = \sum_{\boldsymbol{k}\in\mathbb{N}^p} \alpha_{\boldsymbol{k}}\phi_{\boldsymbol{k}}(\boldsymbol{\xi}) \tag{2.4}$$

Inputs are random variables $\xi_1, \cdots, \xi_p$ that parametrize, for instance, the earthquake source or the geological model. After deciding the form of the polynomial basis functions $\phi_{\boldsymbol{k}}$ and determining the coefficients $\alpha_{\boldsymbol{k}}$ from a training database, the surrogate model can infer outputs corresponding to any new values of the random inputs.

Cruz-Jiménez et al. 2018 studied the influence of the fault plane on the Peak Ground Velocity (PGV) thanks to 8000 simulations of fault rupture. Each rupture is parametrized by 7 random variables and a 9-th order PCE models the PGV at a given station. To study the spatial variability of ground motion, one new surrogate model is built for each station, which increases the required computation. To benefit from the spatial correlation in ground motion, Sochala et al. 2020 applied the PCE to the dominant spatial modes of the Peak Ground Displacement (PGD). Their model also differs by considering parameters of the geological model as input random variables instead of the source parameters. For a given region near Thessaloniki (Greece), the 3D geological model was parametrized by 7 random variables that describe the interfaces between layers and the velocity values per layer.

The first two studies predict the maximum displacement, which does not give information on the ground motion dynamics. To obtain ground motion time series, F. Wang and Sett 2016 rely on an intrusive approach that intricates the PCE inside the numerical solver of the wave equation. By doing so, both the source time function and the geological model can be parametrized by random variables. This method was extended to a 3D domain in F. Wang and Sett 2019. However, there are numerous situations where the numerical solver cannot be accessed by the end-user, or even there exists no such solver. Non-intrusive approaches are therefore the only approach in those situations.

2.1.3 Reduced-order models

The above mentioned works parametrize the inputs with a small number of random variables that correspond to tangible physical parameters. However, when acquiring geological models from observations for instance, inputs are high-dimensional images and there exists no close-form relationship to express the geological model as a function of parameters. In this case, one generally assumes that there exists a representation of the physical processes in a lower-dimensional subspace and one looks for a reduced representation. Proper Orthogonal Decomposition, Principal Component Analysis (PCA), Empirical Orthogonal Functions, and Karhunen-Loeve expansion are all different terms to define the decomposition of a function as a sum of components which correspond to a projection on a reduced basis. As an illustrative example, let us consider the ground motion displacement u generated by parameters $\boldsymbol{\theta}$ as a function of space $\boldsymbol{x}$ and time t. It can be decomposed as

$$u(\boldsymbol{x}, t; \boldsymbol{\theta}) = \sum_{m=1}^{\infty} v_m(t; \boldsymbol{\theta}) \phi_m(\boldsymbol{x}; \boldsymbol{\theta}) \tag{2.5}$$

where the space and time dependencies are split between two sets of basis functions, respectively temporal eigenmodes $v_m(t)$ and spatial eigenmodes $\phi_m(\boldsymbol{x})$. Eigenfunctions are classically obtained by computing the eigenvectors of the empirical covariance matrix. However, this matrix becomes very large when the number of snapshots $(u(\boldsymbol{x}_i, t_j))_{i,j}$ is large, which can make the computation intractable.

Ignoring the spatial dependency, a reduced basis can be obtained from a collection of seismograms and further used to fit a Ground Motion Model (Alimoradi 2011; Hu et al. 2023). However, the generalizability of this approach is strongly dependent on the region where seismograms were recorded. To account for different earthquake scenarios, Imai et al. 2021 expressed earthquake shaking maps (spatial description of PGV) as the PCA of space- and scenario-dependent functions. The reduced-order model was built from a database of 12 scenarios and new scenarios were generated while preserving the probability distribution of the scenario eigenmodes. A similar approach was followed by Rekoske et al. 2023 who predicted PGV maps for a region in south California. The input parameters θ correspond to the description of the earthquake source (depth, strike, dip, and rake) and the spatial eigenmodes account for the spatial correlation of ground motion. The training database was obtained from 5000 3D earthquake simulations with a reference Californian geological model. However, the POD formulation in Equation 2.5 highlights one challenge of this method since the basis functions v_m and ϕ_m have been obtained only for a set of parameters $\boldsymbol{\theta}$ that belong to the training database. If one wants to approximate $u(\boldsymbol{x}, t; \boldsymbol{\theta}^*)$ for a new $\boldsymbol{\theta}^*$, more advanced methods like Empirical Interpolation Methods need to be employed (a summary can be found in Ghattas and Willcox 2021). Rekoske et al. 2023 resolved this issue by an explicit parametrization of the source eigenmodes with radial basis functions which allows them to predict the PGV for new source parameters that do not belong to the training database. However, the surrogate model cannot be applied with a different geological model than the one used in the training database.

Similarly to the intrusive approach mentioned with the PCE, the POD is also appropriate to obtain a reduced basis for the wave equation. This approach was adopted by Hawkins et al. 2023 to express time-dependent ground motion with less variables and therefore enable the decomposition of the empirical covariance matrix. Although their reduced-order model generates reasonable predictions when the geological model is slightly perturbed from the reference, it is still closely tied to the geological configuration used during training.

2.2 HEMEW-3D database

Our literature review has highlighted several works that relate the source parameters with ground motion intensity, either in the form of intensity measures or as time series. However, much less investigations have been conducted on the relationship between geological models and ground motion, especially when considering heterogeneous geological models. Our study of the Le Teil earthquake showed that a better understanding of the influence of geological heterogeneities is crucial to include site effects in 3D seismic hazard analyses. Therefore, it appears that we need a database of various heterogeneous geological models with their associated ground motion to study the relationship between geology and ground motion.

This section introduces the HEMEW-3D and HEMEW^S-3D databases that have been computed for this objective. Both databases contain 30 000 geological models whose description is provided in Section 2.2.2. The main difference between the two databases lies in the fact that the earthquake source has random properties and location in HEMEW^S-3D while it is fixed in HEMEW-3D (Section 2.2.3). Since most of the results presented in Chapter 3 have been obtained with a random source, the emphasis is put on the HEMEW^S-3D database in this section and notable differences with HEMEW-3D are pointed out when necessary. For each geological model and each earthquake source, the solution of the elastic wave equation was solved with SEM3D and acquired at the earth surface (Section 2.2.4).

2.2.1 Related work

There exist many datasets of worldwide seismograms that provide ground motion time series at monitoring stations for local as well as distant earthquakes (e.g. Bahrampouri et al. 2021; Michelini et al. 2021; Mousavi et al. 2019). Although the datasets of recorded ground motion are incredibly important for machine learning applications in seismology, they show several limitations. In particular, recordings are sparse in regions with low-to-moderate seismicity or poor instrumental coverage. In addition, the only geological data associated with the recordings are, at best, local properties at the recording station which are not sufficient to inform 3D physics-based simulations. The BB-SPEEDset dataset provides a compilation of simulation results for several past earthquakes (Paolucci et al. 2021) but the number of simulations is not appropriate for machine learning approaches.

Due to the high computational costs of solving 3D Partial Differential Equations (PDEs), only very few 3D datasets are publicly available for general physical problems. In geophysics, CO2 underground storage has been explored with machine learning based on 3D numerical simulations (Grady et al. 2023; Wen et al. 2023; Witte et al. 2023). To support the study of Witte et al. 2023, Annon 2022 provided 4000 simulation results for 3D CO2 flow through geological models based on the Sleipner dataset complemented by random fields (Equinor 2020). The Kimberlina dataset also contains 6000 CO2 leakage rates simulations (Mansoor et al. 2020). However, the geological models in both databases are all variants of the geological model carefully estimated for a given region, thereby limiting the reproducibility in other areas.

A few datasets of realistic geological units have been developed, such as the Noddyverse dataset (Jessell et al. 2022). In this dataset, geological models result from the deformation of horizontal layers by successive geological events (folds, faults, unconformities, dykes, plugs, shear zones, and tilts) but no associated ground motion is provided. Along the same line of geological deformation, some databases combine geological models with associated waveforms, and target 2D geophysical inversion as the main application (Deng et al. 2022; Liu et al. 2021). The OpenFWI database also includes real geological models from field survey areas and models of CO2 geological storage (Deng et al. 2022). The $\mathbb{E}^{FWI}$ database is an extension of this work to the elastic wave equation, thereby providing two-dimensional ground motion time series (Feng et al. 2023).

As a summary, no database provides solutions of the elastic wave equation in 3D domains. Our HEMEW-3D and HEMEW^S-3D databases intend to fill this gap.

Table 2.1: Summary of datasets providing geological models and seismic wavefields. Dimension of geological models: number of grid points in (width, depth) for 2D datasets, in (width, length, depth) for 3D datasets. Domain: size of the physical domain. Dimension of seismic wavefields: (receivers along width, time steps) for 2D datasets, (receivers along width, receivers along length, time steps) for 3D datasets. Components (abbrv. Comp.): number of velocity components for each sensor (1 means that the acoustic wave equation is solved, 2 or 3 means that the elastic wave equation is solved).

		Geological models				Seismic wavefields		
Dataset	train/test	Dimensions	Domain	Values V_S	Construction	Dimensions	Comp.	Source
Noddyverse (Jessell et al. 2022)	1M/-	$200 \times 200 \times 200$	$4 \times 4 \times 4\text{km}^3$	categorical	succession of geological events	N/A	N/A	N/A
OpenFWI (Deng et al. 2022)	408k/62k	70×70	$0.7 \times 0.7\text{km}^2$	1500; 4500 m/s	mathematical, from recorded images, and from geological faults	70×1000	1	5 fixed sources at the surface
OpenFWI Kimberlina CO2 (Deng et al. 2022)	15k/4k	401×141	$4 \times 1.4\text{km}^2$	1200; 3600 m/s	from real data	101×1251	1	9 fixed sources at the surface
OpenFWI Kimberlina 3D (Deng et al. 2022)	1.6k/163	$400 \times 400 \times 350$	$4 \times 4 \times 3.5\text{km}^3$	?	from real data	$40 \times 40 \times 5001$	1	25 fixed sources at the surface
$\mathbb{E}^{\mathbf{FWI}}$ (Feng et al. 2023)	144k/24k	70×70	$0.35 \times 0.35\text{km}^2$	V_S : 612; 3000 m/s V_P : 1500; 4500 m/s	mathematical and from geological faults	70×1000	2	5 fixed sources at the surface
HEMEW-3D (this work)	27k/3k	$32 \times 32 \times 32$	$9.6 \times 9.6 \times 9.6\text{km}^3$	1071; 4500 m/s	horizontal layers + random fields	$16 \times 16 \times 2000$	3	1 fixed source
HEMEW^S-3D (this work)	27k/3k	$32 \times 32 \times 32$	$9.6 \times 9.6 \times 9.6\text{km}^3$	1071; 4500 m/s	horizontal layers + random fields	$32 \times 32 \times 800$	3	1 source with random location and random orientation

2.2.2 Heterogeneous geological models

The 3D geological models are designed in terms of S-wave velocities since the average velocity of S-waves in the top 30 m (denoted $V_{S,30}$) is a well-used proxy to characterize the geological properties of a given site. Geological models $V_S(\boldsymbol{x})$ are non-stationary random fields defined as a 1D model (horizontal homogeneous layers) to which fluctuations are added, as illustrated in Figure 2.1. The domain is a cube of size 9.6 km $\times$ 9.6 km $\times$ 9.6 km, discretized in 32 $\times$ 32 $\times$ 32 voxels of size 300 m $\times$ 300 m $\times$ 300 m.

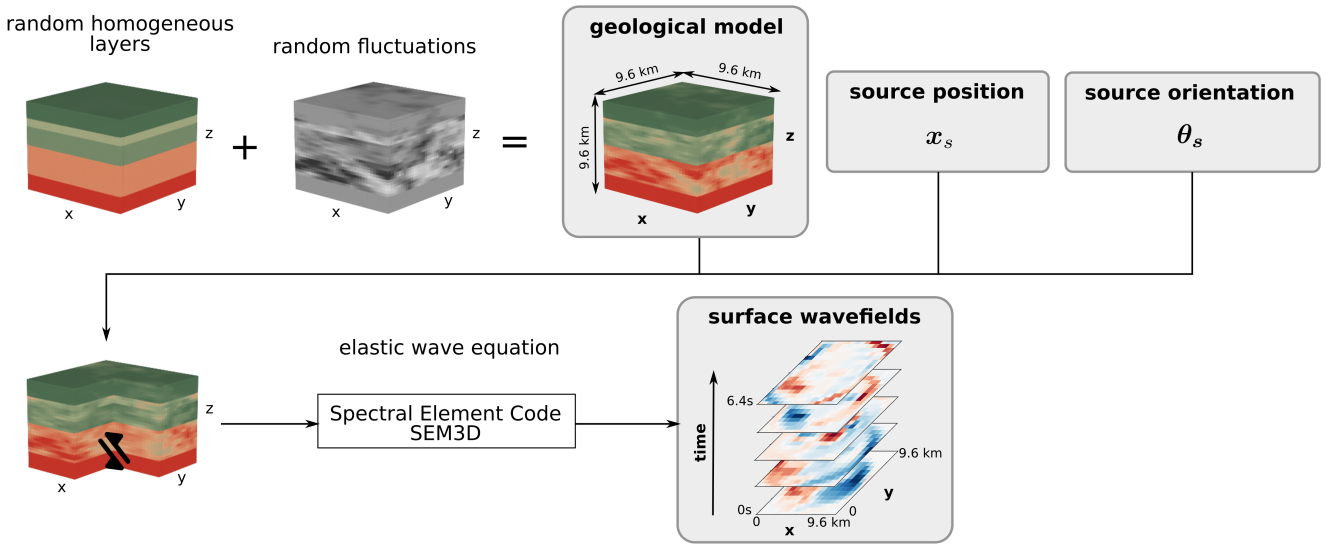


Figure 2.1: Geological models are built by adding heterogeneities to randomly chosen horizontal layers. Then, elastic waves are propagated from a source located randomly inside the domain with random orientation up to the surface, where velocity wavefields are acquired.

Homogeneous models

A 1.8 km-thick homogeneous layer is imposed at the bottom of each geological model, with a constant V_S value of 4500 m/s. This layer was originally conceived to contain the source in the HEMEW-3D database and ensure that the seismic waves would always originate from a region with the same properties. This assumption is relaxed in the HEMEW^S-3D database but the same geological models are used in both databases for the sake of consistency. Above the bottom layer, a random arrangement of horizontal layers is designed. The number of horizontal layers N_ℓ and their thickness are randomly chosen for each geological model, with the sole constraint to fill the remaining depth of $9.6 - 1.8 = 7.8$km with 1 to 6 layers. Then, the mean layer-wise V_S value is drawn from the uniform distribution $\mathcal{U}([1785; 3214 \text{ m/s}])$. The bounds of the uniform distribution were determined to ensure that most values remain bounded between $V_{S,min}$=1071 m/s and $V_{S,max}$=4500 m/s after the addition of random fields ($1785 = \frac{1071}{1-0.4}, 3214 = \frac{4500}{1+0.4}$ and 0.4 corresponds to an interval of two 0.2 standard deviation around the mean, interval which contains 95 % of values following a normal distribution). Table 2.2 gives a summary of the geological parameters.

We can already make a few observations. First, the minimum S-wave velocity of 1071 m/s is rather high when compared to S-wave velocities in soft sediments (typically of few hundreds of m/s) but coherent for hard sediments (Molinari and Morelli 2011). One should also note that the vertical resolution of the geological models is 300 m while very low V_S values are more commonly encountered in the first tens of meters and they would be averaged in our models. Reducing the minimum velocity poses no theoretical limitation but would increase the computational cost of the subsequent numerical simulations since it increases the number of mesh elements (equation 1.12). Second, the maximum S-wave velocity of 4500 m/s corresponds to typical V_S values at the bottom of the Earth's crust (Molinari and Morelli 2011). Third, we do not constrain the ordering of layer-wise V_S values, meaning that some geological models may contain layers with V_S values decreasing with depth, which is unphysical. However, the physics of wave propagation is still satisfied in those situations and these geological models increase the variability of our database.

Parameter	Statistical distribution
Number of heterogeneous layers N_ℓ	$\mathcal{U}(\{1,2,3,4,5,6\})$
Layers' thickness $h_1, \cdots, h_{N_\ell}$	$\mathcal{U}\left(\{(h_1, \cdots, h_{N_\ell}) > 0 \vert h_1 + \cdots + h_{N_\ell} = 7.8\text{km}\}\right)$
Mean V_S value per layer	$\mathcal{U}([1785, 3214])$
Layer-wise coefficient of variation	$\vert\mathcal{N}(0.2, 0.1)\vert$
Layer-wise correlation length along x	$\mathcal{U}(\{1.5, 3, 4.5, 6\text{ km}\})$
Layer-wise correlation length along y	$\mathcal{U}(\{1.5, 3, 4.5, 6\text{ km}\})$
Layer-wise correlation length along z	$\mathcal{U}(\{1.5, 3, 4.5, 6\text{ km}\})$

Table 2.2: Statistical distribution of each parameter describing the geological models. Mean V_S values, coefficients of variation, and correlation lengths are chosen independently in each layer. Since the bottom layer has a constant thickness of 1.8 km, it is not included in these parameters.

To solve the elastic wave equation with SEM3D, the velocity of P waves V_P and the density ρ are required in addition to V_S. To reduce the complexity of the problem we aim at solving, we assume that V_P and ρ can be fully determined from V_S. More precisely, the ratio of P- to S-wave velocity is fixed to $V_P/V_S = 1.7$ and the density ρ is computed as a function of the P-wave velocity (Molinari and Morelli 2011)

$$\rho = 1.6612V_P - 0.4721V_P^2 + 0.0671V_P^3 - 0.0043V_P^4 + 0.000106V_P^5 \tag{2.6}$$

Attenuation factors for P-waves (Q_P) and S-waves (Q_S) are computed as

$$Q_P = \max\left(\frac{V_P}{20}, \frac{V_S}{5}\right); Q_S = \frac{V_S}{10} \tag{2.7}$$

Addition of heterogeneities

The layers' thickness and mean value describe the general structure of the propagation domain and they correspond to the prior physical information usually available. However, materials in the Earth's crust contain much more variability, especially along the horizontal directions. Following the methodology presented in Section 1.2, heterogeneities are represented as Gaussian random fields with a von Karman correlation kernel and a Hurst exponent of 0.1 (marginal distributions are log-normal to preserve positive values). We choose correlation lengths randomly in $\{1.5\,\text{km}, 3\,\text{km}, 4.5\,\text{km}, 6\,\text{km}\}$, to mix samples with small- and large-scale heterogeneity. In addition, large coefficients of variation were chosen to provide high geological contrasts, following a folded normal distribution $\vert\mathcal{N}(0.2, 0.1)\vert$, with mean 0.2 and standard deviation 0.1. After the addition of random fields on the 1D geological model, V_S values are clipped between $V_{S,\text{min}}$ and $V_{S,\text{max}}$. Figure 1.7 illustrates two geological models from the HEMEW^S-3D database. By allowing large correlation lengths and large coefficients of variation, we create regions with singular properties that can mimic some forms of sedimentary basins for instance.

It should be noted that all layers have distinct coefficients of variation and correlation lengths, meaning that different random fields are drawn inside each layer. In particular, this implies that there is no continuation of random fields across two layers. To facilitate further analyses, a small dataset of 1000 *consistent* geological models was also designed. The only difference with the HEMEW^S-3D database is that all layers in the *consistent* model have the same coefficient of variation σ and correlation length ℓ_c. Random fields are still drawn independently inside each layer but it allows to identify one geological model with a pair (σ, ℓ_c).

2.2.3 Source parameters

The earthquake source is represented as a point-wise source, defined by its position $\boldsymbol{x_s}$ and its orientation $\boldsymbol{\theta_s}$. In the HEMEW^S-3D database[1], the source is located randomly inside the propagation domain, not too close from the boundaries to avoid numerical issues due to absorbing boundary conditions. The source position $\boldsymbol{x_s} = (x_s, y_s, z_s)$

[1] In the HEMEW-3D database, the source position is fixed at (4.8, 4.8, −8.4 km).

is chosen from a Latin Hypercube sampling with

$$x_s \in [1.2; 8.4\ km]$$
$$y_s \in [1.2; 8.4\ km]$$
$$z_s \in [-9.0; -0.6\ km]$$

The source orientation can be described alternatively by the three angles (strike, dip, rake) or by the 6 components of the moment tensor[2]. In the first representation, $\boldsymbol{\theta_s} \in \mathbb{R}^3$, the angles are sampled from a Latin Hypercube with a strike between 0° and 360°, dip between 0° and 90°, and rake between 0° and 360°. In the second representation, the equivalent moment tensor $\boldsymbol{\theta_s} \in \mathbb{R}^6$ is computed for each set of angles (Section 1.1.3).

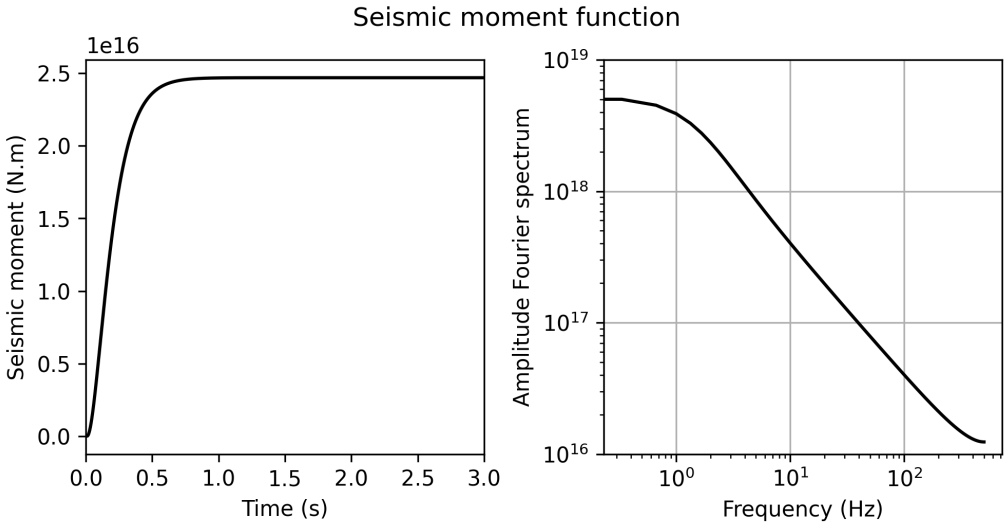


Figure 2.2: The seismic moment function in the HEMEW-3D and HEMEW^S-3D databases

The source amplitude corresponds to a seismic moment M_0 = 2.47 $\times$ 10^{16} N m and the source time evolution is a spice bench given by $s(t) = 1 - \left(1 + \frac{t}{\tau}\right) e^{-\frac{t}{\tau}}$ with τ = 0.1 s (Fig. 2.2).

2.2.4 Solutions of the wave equation

The elastic wave equation is solved in each domain with SEM3D. The maximum frequency f_{max} one aims at exactly resolving is chosen at 5 Hz, which is relatively high for this type of simulations. Indeed, many simulations have been conducted so far with an accuracy up to 1 or 2 Hz (Rekoske et al. 2023; Rosti et al. 2023), while high-fidelity simulations for local realistic earthquake scenarios extend up to 10 Hz, and exceptionally up to 18 Hz (Table 1.1). With 7 Gauss-Lobatto-Legendre quadrature points per side of mesh element and a minimum velocity of 1071 m/s, this leads to elements of size 300 m. This choice corresponds to the discretization of the geological models in voxels of size 300 m $\times$ 300 m $\times$ 300 m.

Since the numerical scheme of SEM3D solves for velocity values, we consider only ground motion velocity; displacement and acceleration can be obtained afterwards via integration or derivation of the velocity time series. To maintain reasonable computational loads and reflect real-life situations, velocity wavefields are recorded only at the surface of the propagation domain. A regular grid of 32 $\times$ 32 sensors[3] is placed between 150 m and 9450 m in both horizontal directions. At each monitoring sensor, the three-component velocity field is synthetized with a 100 Hz sampling frequency between 0 and 8 s[4]. With the surrogate model described in Chapter 3, the velocity wavefields were low-pass filtered at 5 Hz to match the maximum frequency of the simulation, restricted to the time window [0 s; 6.4 s] and downsampled from 100 Hz to 50 Hz.

[2] In the HEMEW-3D database, the source orientation is fixed and corresponds to the source of the Le Teil earthquake, strike = 48°, dip = 45°, and rake = 88° (Delouis et al. 2021)

[3] In the HEMEW-3D database, the grid is coarser and has only 16 $\times$ 16 sensors

[4] In the HEMEW-3D database, velocity fields are synthetized up to 20 s but it appeared that most of the ground motion occurs before 8 s. Therefore, the time duration was shortened in the following HEMEW^S-3D database

2.2.5 Descriptive statistics

Since most of the geological parameters are chosen uniformly randomly (Table 2.2), the geological dataset is well-balanced: geological models with 1 to 6 layers are equipartitioned and all random fields parameters have approximately the same frequency. Mean V_S values range from 1756 m/s to 3145 m/s.

Since our surrogate model needs to predict time series, it is interesting to explore the temporal evolution of the target signals. The first wave arrival time is a crucial parameter for earthquake early warning. It depends on the distance between the earthquake source and the monitoring sensor, as well as the geological properties on the propagation path. Wave arrival times are usually determined from recordings, either manually by experts, or with machine learning methods. However, it is possible to compute approximated arrival times from synthetic velocity fields since ground motion is almost zero before the first wave arrival. Therefore, we obtained the wave arrival times for the P-waves as the first time where the amplitude exceeds 0.1 % of the maximum amplitude. Due to the variability in the source depth and the different wave velocities in the geological models, first wave arrival times vary significantly between samples and between sensors. Figure 2.3a shows that 10 % of velocity time series are initiated before 0.66 s while 10 % of time series are still null after 2.18 s.

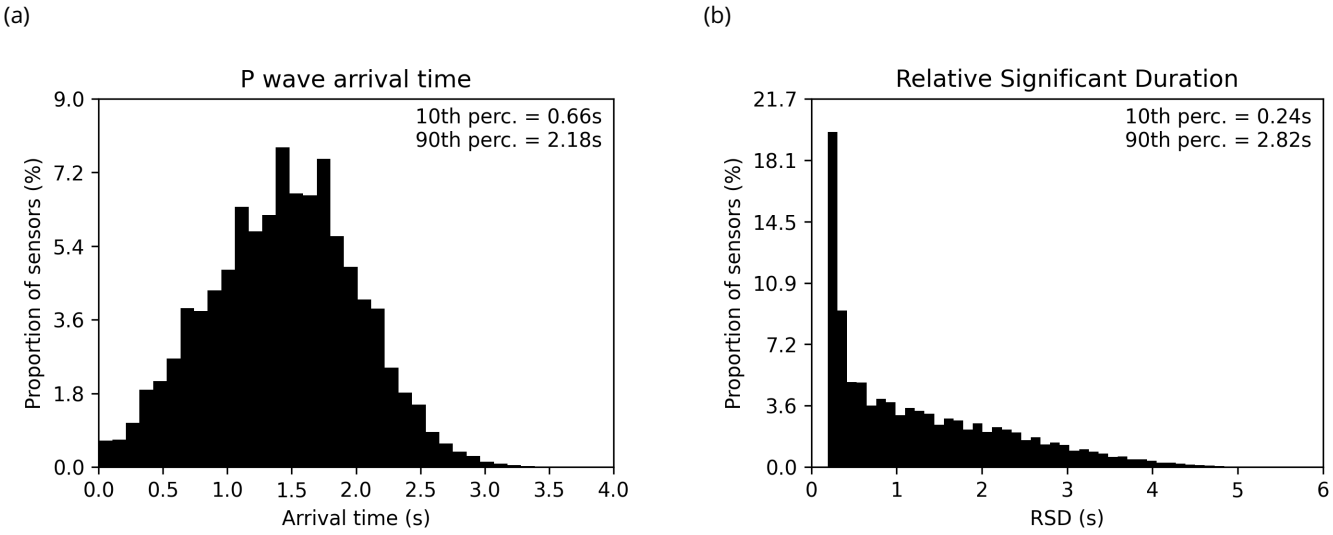


Figure 2.3: Distributions of the temporal features of velocity time series at each monitoring sensor and for 30 000 samples. (a) the first P-wave arrival time is computed on the vertical component (b) the Relative Significant Duration (RSD) is shown for the East-West component, results are very similar for the two other components

The temporal evolution of ground motion can also be characterized by its Relative Significant Duration (RSD, defined in Section 1.3.4). Figure 2.3b shows that many time series are relatively short since 25 % of them have a RSD smaller than 0.36 s. This is not surprising since synthetic time series do not contain the high-frequency components that form most of the seismic coda. However, there are still 10 % of ground motion time series that last more than 2.82 s. These observations on the first wave arrival time and on the RSD indicate that the surrogate model needs to predict time series with very different temporal features.

Figure 2.4a firstly shows that the Peak Ground Velocity (PGV) extends over three orders of magnitude, with the first percentile being equal to 0.89 cm/s while the 99th percentile equals 129.3 cm/s. When the propagation path is longer, seismic waves encounter more geological heterogeneities. They create a dispersion and diffraction of waves that spread the energy signal over time. For an earthquake source located at (x_s, y_s, z_s) and a monitoring sensor at (x, y, z), the hypocentral distance is defined as

$$R_{hypo} = \sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2} \tag{2.8}$$

Larger hypocentral distances are associated with longer propagation paths. Figure 2.4a then shows that the PGV is negatively correlated with the hypocentral distance.

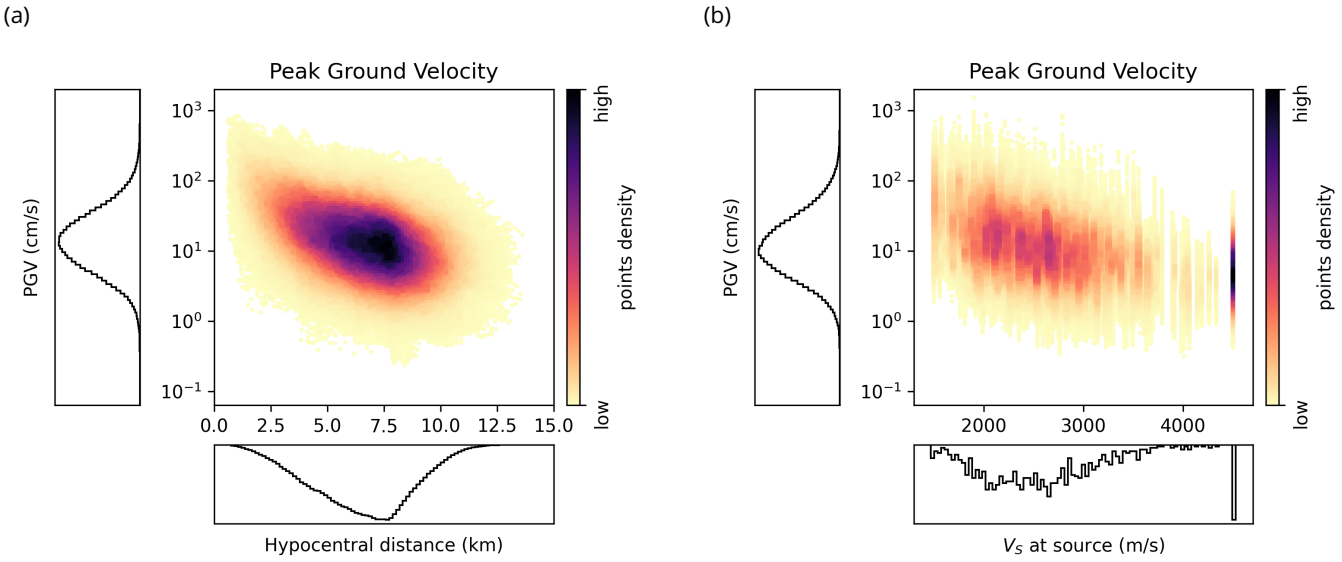


Figure 2.4: For each sample and each sensor, the PGV is shown against (a) the hypocentral distance, (b) the S-wave velocity at the source location. The PGV is computed on the East-West component, results are very similar for the two other components.

It is also known that the seismic energy E_s generated by a fault rupture is

$$E_s = \frac{M_0 \Delta\sigma}{2\mu} \tag{2.9}$$

where M_0 is the seismic moment, $\Delta\sigma$ is the stress drop and μ is the shear modulus at the fault location. Knowing that the shear wave velocity writes $V_S = \sqrt{\mu/\rho}$, equation 2.9 indicates that the seismic energy is inversely proportional to V_S^2. And Figure 2.4b confirms that the PGV is negatively correlated with the velocity of S waves at the source location. The relationships between the PGV and the hypocentral distance, as well as the source S-wave velocity will be used in Chapter 3 to normalize the outputs of the surrogate model and hence, reduce the variability of ground motion amplitudes.

2.2.6 Dimensionality analyses

In supervised deep learning, it is always challenging to determine whether the size of the database (i.e. the number of samples) is sufficient to represent its variability. This questions relates to the definition of the intrinsic dimension of the dataset, which indicates the number of hidden variables that should be necessary to represent the main features of the samples. In the following, we provide insights on this question with the intrinsic dimension based on the Principal Component Analysis, the correlation dimension, the Maximum Likelihood Estimate, and the Structural Similarity Index.

For all methods, let us consider a set of N samples $\{\phi_i\}_{1\leq i\leq N} \in \mathbb{R}^D$ that can be either geological models or ground motion wavefields. We assume that there exists a continuous and sufficiently smooth mapping $g : \mathbb{R}^d \to \mathbb{R}^D$ such that each $\phi_i \in \mathbb{R}^D$ can be identified with $\widetilde{\phi}_i \in \mathbb{R}^d$ via $\phi_i = g(\widetilde{\phi}_i)$. From this formulation, d is called the intrinsic dimension.

Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) decomposes data in principal components that correspond to the directions where data vary the most. For different sizes of datasets, we compute the number of principal components required to retain 95 % of variance and define this number as the intrinsic dimension of data. The 3D geological models and the 3D ground motion wavefields are transformed into 1D vectors to perform the PCA. To reduce the memory requirements, ground motions are analyzed only on the East-West component. Geological models are represented

by 32 × 32 × 32 = 32 768 points and ground motions contain 16 × 16 × 320 = 81 920 points (16 sensors in directions x and y, and 320 time steps between 0 s and 6.4 s). To ease the computation on the large sample covariance matrix, an incremental PCA algorithm was used (Ross et al. 2008).

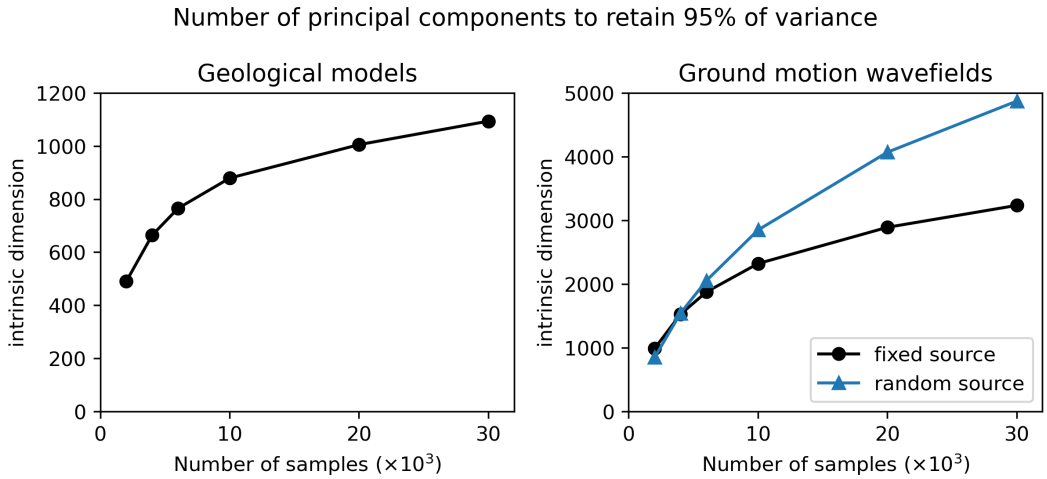


Figure 2.5: Number of principal components (y-axis) required to represent 95% of the variance in data as a function of the dataset size (x-axis) for geological models (left) and ground motion wavefields (right). For ground motion, the HEMEW-3D database is used for the fixed source (black line) and the HEMEW^S-3D database corresponds to the blue line.

Figure 2.5 shows that more than 1000 principal components are needed to reconstruct the geological models with high accuracy whereas the intrinsic dimension of ground motion wavefields is much higher. When the source is fixed (HEMEW-3D database), the intrinsic dimension is 3200 while it increases to 4900 when the source has a random location and orientation (HEMEW^S-3D database). It is reasonable that adding degrees of freedom with a random source increases the variability of data. However, due to its linearity, the PCA requires a large number of components to accurately represent complex patterns. Therefore, it may overestimate the intrinsic data dimension.

Correlation dimension

An alternative dimensionality measure was introduced by Grassberger and Procaccia 1983 as the correlation dimension, which characterizes the distance between pairs of samples. For a given *radius* r, the correlation dimension $C_N(r)$ is defined as the ratio of sample pairs $(\phi_i, \phi_j)_{i\neq j}$ being at distance less than r

$$C_N(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{1}\left(\|\phi_i - \phi_j\| \leq r\right) \tag{2.10}$$

The correlation dimension is determined as the slope of the linear part in the log-log representation of C_N (Fig. A.7). This definition is subject to some interpretation since one should determine which portion constitutes the linear part. Nevertheless, we found that small variations of the linear part limits had very little influence on the slope estimate (less than one unit).

Figure A.6 indicates a correlation dimension of 8 for the geological dataset, which is significantly lower than the PCA dimension. In fact, it is known that the correlation dimension may underestimate the intrinsic dimension, especially "when data are scattered" (Qiu et al. 2023), which is likely to be the case in high-dimensional spaces. However, the correlation dimension of the ground motion wavefields is debatable. It is evaluated around 12 when the source is fixed and it drops to 2 when the source is random (Fig. A.6). In the latter situation, Fig. A.7c shows that the log-log representation does not produce an obvious linear part.

MLE intrinsic dimension

Levina and Bickel 2004 proposed another measure of intrinsic dimension based on the Maximum Likelihood Estimator (MLE) of the distance to the closest neighbours. Similarly to the correlation dimension, this method considers

the Euclidean distance between samples $\phi_i, \phi_j \in \mathbb{R}^D$. Denoting $T_k(\phi)$ the Euclidean distance between ϕ and its k-th nearest neighbour in the dataset, the estimated intrinsic dimension at point ϕ is

$$\widehat{d}_k(\phi) = \left[\frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{T_k(\phi)}{T_j(\phi)} \right]^{-1} \tag{2.11}$$

This expression derives from the MLE estimate of dimension d when considering the observation of points in the neighbourhood of point ϕ as a Poisson process (Levina and Bickel 2004). The global intrinsic dimension $\widehat{d}$ is finally obtained by averaging over all points and a small number of neighbours $0 < k_1 < k_2$

$$\widehat{d} = \frac{1}{k_2 - k_1 + 1} \frac{1}{N} \sum_{k=k_1}^{k_2} \sum_{i=1}^{N} \widehat{d}_k(\phi_i) \tag{2.12}$$

The MLE intrinsic dimension was computed with the `scikit-dimension` package.

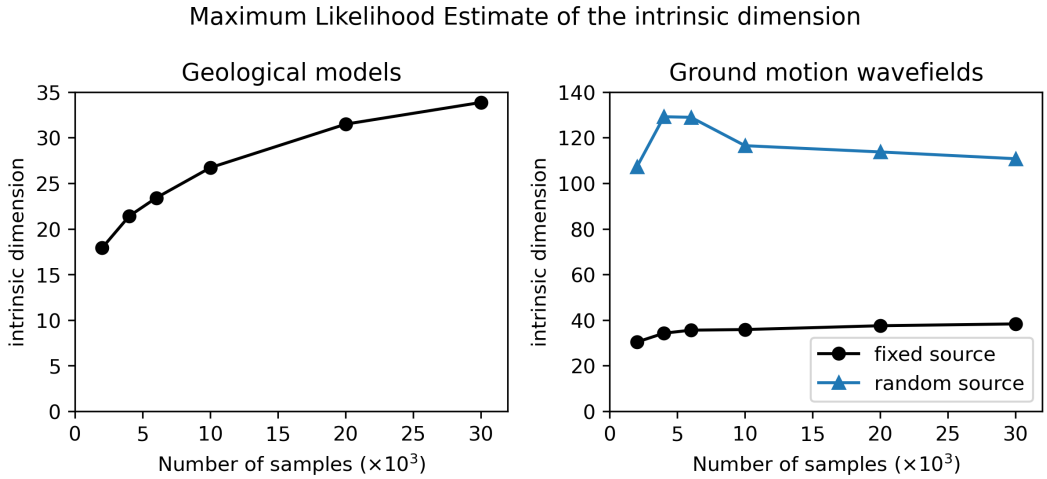


Figure 2.6: Intrinsic dimension estimated by the MLE (y-axis) as a function of the dataset size (x-axis) for geological models (left) and ground motion wavefields (right). For ground motion, the HEMEW-3D database is used for the fixed source (black line) and the HEMEW^S-3D database corresponds to the blue line.

Figure 2.6 shows the evolution of the intrinsic dimension as a function of the number of samples for geological models and ground motion wavefields. When the source is fixed, the intrinsic dimension of geological models is 34, which is similar to the one of ground motion wavefields (d=37). This is sound since variability in the wavefields is created entirely from the variability in the geological models. When the source is random, the dimension of ground motion wavefields increases significantly (d=110). Although this method may still underestimate data with high intrinsic dimensionality (Qiu et al. 2023), it provides higher estimates than the correlation dimension.

Structural similarity

The above mentioned measures of intrinsic dimension rely on the Euclidean distance which may not represent well similarities between samples. Alternative metrics such as the Structural Similarity Index Measure (SSIM) have been introduced to mimic the human perception of similarity (Z. Wang et al. 2004). The SSIM theoretically ranges from 0 to 1, with 0 indicating no similarity and 1 indicating perfectly similar samples (although values between -1 and 0 can be obtained numerically from the covariance computation). The SSIM of two samples ϕ_i, ϕ_j is defined as

$$SSIM(\phi_i, \phi_j) = \frac{(2\mu_i\mu_j + C_1)(2\sigma_{ij} + C_2)}{(\mu_i^2 + \mu_j^2 + C_1)(\sigma_i^2 + \sigma_j^2 + C_2)} \tag{2.13}$$

where μ_i and μ_j are the means of ϕ_i and ϕ_j, σ_i and σ_j are the unbiased estimators of the variance of ϕ_i and ϕ_j, σ_{ij} is the unbiased estimator of the covariance of ϕ_i and ϕ_j, C_1 and C_2 are constants determined from the range of ϕ_i and ϕ_j values.

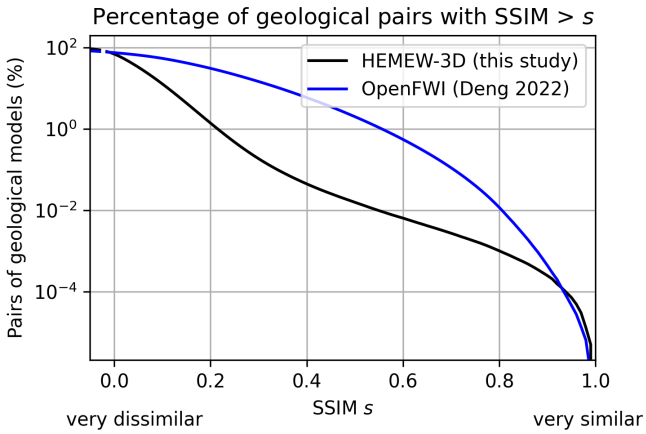


Figure 2.7: The Structural Similarity Index Measure (SSIM) quantifies the visual ressemblance between images, in a way that should mimic human perception. For each SSIM value s on the x-axis, the percentage of geological pairs being more similar than s is reported on the y-axis.

We focus on the SSIM between geological models since it indicates a notion of sparsity among the inputs of the surrogate model. Figure 2.7 shows that only 1.4 % of geological pairs have a SSIM greater than 0.2. This means that geological models are generally very distinct from each other in our database. For comparison purposes, the 2D OpenFWI dataset leads to significantly higher SSIM, with 31 % of geologies having a SSIM larger than 0.2 (3000 models were chosen from each of the 10 OpenFWI families (Deng et al. 2022)).

Partial conclusion on the intrinsic dimension

Many different methods exist to estimate the data intrinsic dimension and we exemplified the well-known fact that they can lead to very different values. Taking the MLE as a lower bound, one can argue that the intrinsic dimension of the geological database is at least 30. In addition, the low values of the SSIM indicate that geologies are sparse and quite distant from each other in the HEMEW^{S}-3D database.

Concerning the ground motion wavefields, the PCA and the MLE confirm the intuition that the intrinsic dimension is larger when the source has a random location and orientation (HEMEW^{S}-3D database) since it adds variability to the time arrival of wavefields as well as their location at the surface. In this situation, it is reasonable to consider that the intrinsic dimension of ground motion is at least on the order of 100. However, if data are decomposed with the PCA, then the number of principal components is a few thousands. The correlation dimension yields questionable estimates of the intrinsic dimension that contradict our intuition and the PCA and MLE outcomes.

It can also be noted that the intrinsic dimension increased with the number of samples, for the PCA and the MLE. This may reflect a flaw in the intrinsic dimension's definition or it may indicate that despite being already large, our database of 30 000 samples does not capture all the variability.

One should also point out the relationship between the PCA intrinsic dimension and the Kolmogorov d-width, defined as the best possible error achievable by projecting data in an original space of dimension D onto a linear subspace of dimension d (Kolmogoroff 1936). If the error is measured with the L_2 norm, then it is known that the PCA is the optimal approximation. Greif and Urban 2019 showed that the Kolmogorov d-width cannot decay faster than $\frac{1}{4}d^{-1/2}$ for an example of the second-order equation. This is in line with our observations that good approximations can only be achieved with a large number of principal components.

2.3 Reducing the dimension of geological models

Although the MLE and the correlation dimension provide small estimates of the intrinsic dimension, one cannot compute the associated reduced basis. With the PCA on the contrary, it is easy to compute the basis of principal components and hence, represent the geological models with their basis coefficients. Then, one obtains reconstructions of geological models that depend on the number of principal components retained in the basis.

It has already been mentioned that the linearity of the PCA is one factor explaining the high intrinsic dimension. Auto-encoders are among the most popular deep learning methods to reduce the dimensionality of inputs (Hinton and Salakhutdinov 2006) and they are often viewed as a nonlinear extension of the PCA. Auto-encoders have been largely used in various fields to encode complex datasets into reduced order manifolds (e.g., Kadeethum et al. 2022; Ladjal et al. 2019), but their application on 3D data remains spurious (Gangopadhyay et al. 2021; Tekawade et al. 2021; Yu et al. 2021).

In this section, we focus on the decomposition of geological models. The relationship with ground motion is only assessed afterwards from the reconstructed geology. This analysis is considered as a preliminary study to investigate the feasibility of a surrogate model built on a reduced basis.

2.3.1 3D auto-encoder with a UNet architecture

By considering geological models as 1D vectors, the PCA looses the spatial organization of 3D geologies. To obtain the same reconstruction error, one may therefore expect a greater dimensionality reduction with a 3D auto-encoder than a PCA. The auto-encoder is made of two sub-networks, namely an encoder that associates a 3D geological model with a latent representation of reduced dimension and a decoder that reconstructs a 3D geology from the latent representation while trying to minimize the error between the input and the reconstruction.

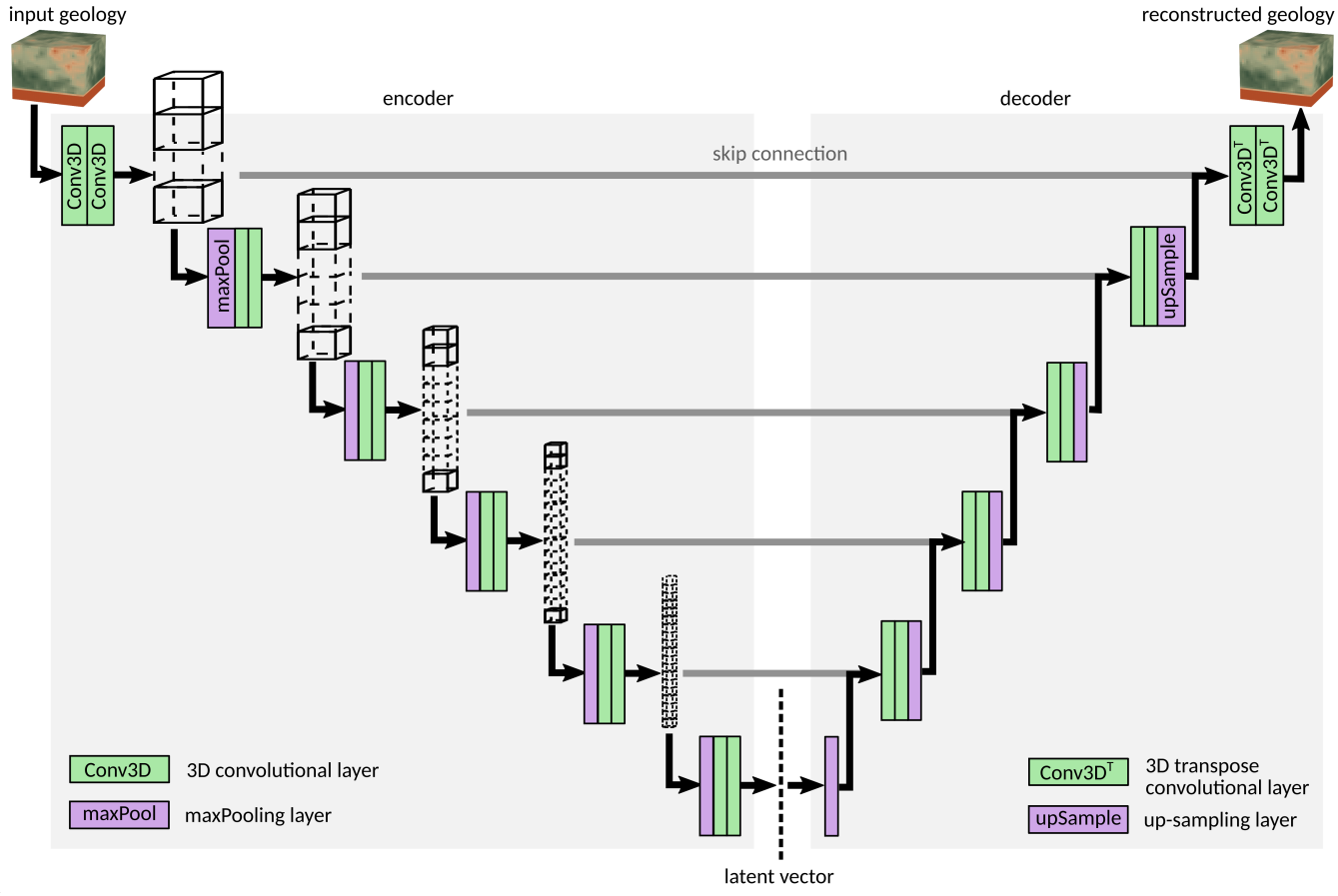


Figure 2.8: 3D UNet auto-encoder with 6 blocks of double convolutions (4 million parameters in total). Skip connections between the encoder and decoder are done via concatenation.

Our 3D auto-encoder is built from the architecture of the 3D UNet developed by Wolny et al. 2020. We conducted an exhaustive search of the main hyperparameters to adapt the proposed architecture to our objectives. Our final encoder is composed of 6 blocks that increase the number of channels of the inputs from 1 to 256. As represented in Figure 2.8, each block is composed of two convolutional layers. Blocks are separated by `max pooling` layers to reduce the dimensionality from $32 \times 32 \times 32$ with 1 channel to $1 \times 1 \times 1$ with 256 channels. A batch size of 16 was found to produce lower reconstruction errors than larger batches for both training and validation datasets. The

decoder is made of 6 blocks of 2 transposed convolutional layers separated by upsampling to increase the spatial dimensions.

The database of 40 000 geological fields was split in 90 % of training and 10 % of validation data. Inputs were centered and normalized by 4 times the standard deviation, moving values approximately into $[-0.5, 0.5]$. The training loss was simply composed of the L1 reconstruction error. The Adam optimizer was used with a learning rate of 2×10^{-4}. Training the network for 1000 epochs on 4 Nvidia Ampere A100 GPUs took 9.5 h.

2.3.2 Assessing the reconstruction accuracy

Based on the analysis of the intrinsic dimension, the PCA is applied on 40 000 geological models with 1024 components. Figure 2.9 shows that geological models reconstructed with 1024 principal components are visually close to the inputs. Indeed, the heterogeneities' size and location correspond to the inputs. However, it can be noted that geological fields reconstructed with the PCA lack some sharpness in the geological heterogeneities and can appear blurrier than the input. This is especially visible on samples with a high coefficient of variation (e.g. last row of Figure 2.9). For those samples, more than 1024 principal components are necessary to capture the small scale heterogeneities. Quantitatively, the median Root Mean Square Error (RMSE) on all samples is 90 m/s, which corresponds to 8.6 % of the minimum V_S value (Fig. 2.10).

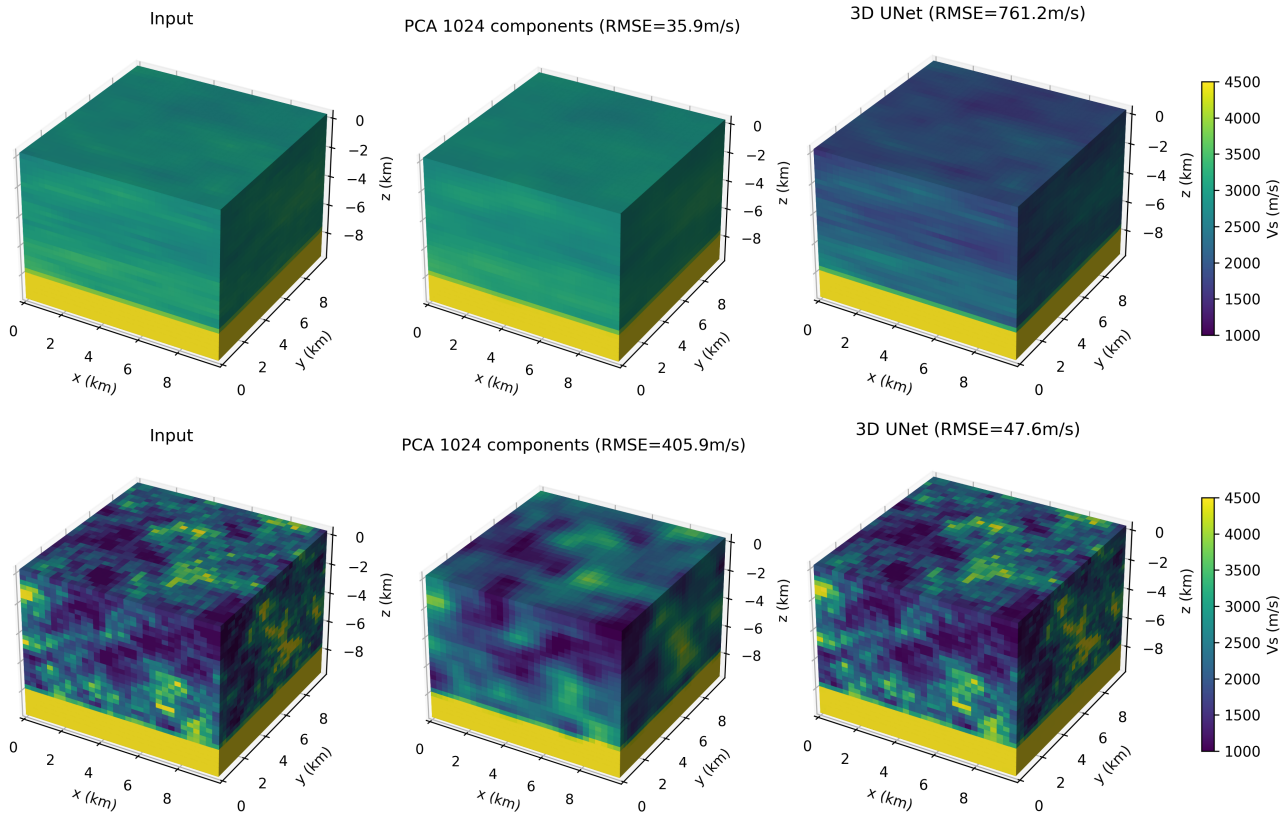


Figure 2.9: Each row represents one geological model from the test dataset of 4000 samples. (Left column) original field. (Middle column) reconstruction with a PCA with 1024 principal components. (Right column) reconstruction with a 3D UNet autoencoder. RMSE = Root Mean Square Error.

Reconstructions obtained with the 3D UNet are generally satisfying since heterogeneities are well reproduced. The median RMSE is 85 m/s, which is slightly lower than the PCA error (Fig. 2.10). Interestingly, Figure 2.10 shows that the PCA and the 3D UNet lead to very different error distributions despite having a similar median error. While the PCA reconstruction error resembles a normal distribution, the 3D UNet histogram has more values close to 0. This means that the 3D UNet is able to create more reconstructions with a very low error. As a counterpart, the 3D UNet also leads to more geological models with poorer reconstructions than the PCA. Among the geologies showing a high reconstruction error with the 3D UNet, many have a low coefficient of variation (an example is visible on the

first row of Figure 2.9). In those cases, the large error derives from a biased reconstruction by the neural network. One possible explanation is the fact that the 3D UNet was initially proposed for segmentation, which means that the architecture is well-suited to detect edges, but it may be less sensitive to the mean value when no clear edges are visible. We also found that adding a penalization term on the mean bias in the loss function did not improve the biased reconstructions.

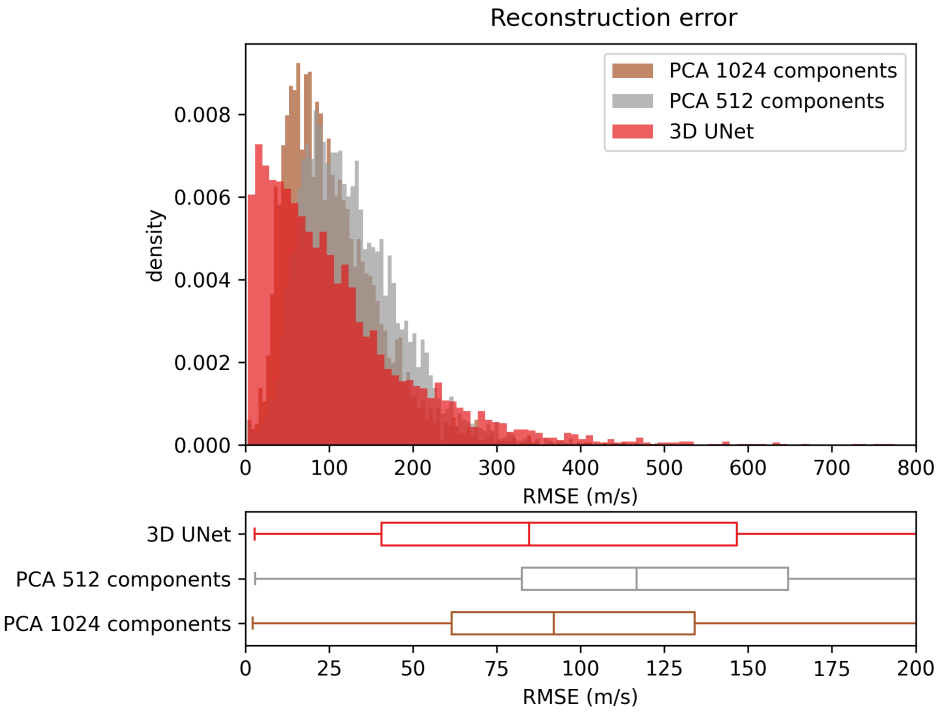


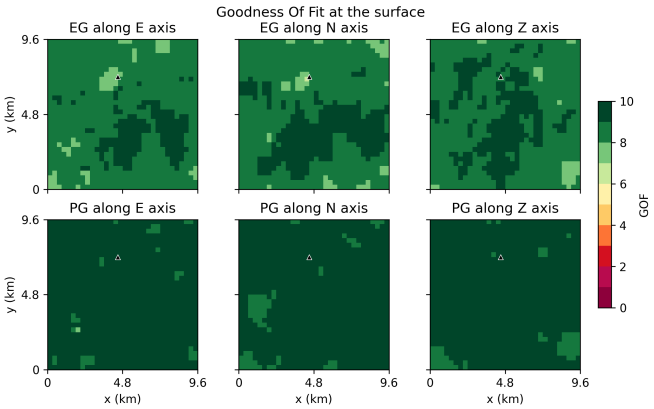
Figure 2.10: Reconstruction error on 4000 test geological fields using a PCA with 512 components (in grey) and 1024 components (in brown), and the 3D UNet (in red). Top: histogram of the error. Bottom: boxplot of the error, boxes extend from the 1st to the 3rd quartile, line shows the median.

2.3.3 Influence of the dimension on ground motion response

More importantly than the reconstruction error, we are interested in the ground motion generated from the reconstructed geological models, compared to the ones generated from the input geologies. As an illustrative example, we choose one geological model with a rather large reconstruction error, around 135 m/s for both the PCA and the 3D UNet. This example represents the general observations obtained for the ground motion of 250 reconstructed geologies. A RMSE of 135 m/s corresponds to the third quartile of the reconstruction error for both PCA and 3D UNet (Fig. 2.10). A point-wise source is located in the bottom layer and seismic waves propagate up to the surface (same methodology as the HEMEW-3D database), in the input geology as well as the geologies reconstructed by the PCA and by the 3D UNet. Figure 2.11 compares the ground motions generated through the original and reconstructed geologies by displaying GOF scores at the surface.

Figure 2.11 shows that a vast majority of sensors exhibit GOFs above 8, which is considered as an excellent agreement. Therefore, despite a large reconstruction error on the geological model, the surface ground motion is still very close to the reference one. One can also notice that the various types of errors described above for the PCA and the 3D UNet have very different consequences on the ground motion generated through the reconstructed geological models. For the PCA reconstruction, the envelope GOFs are slightly worse than the phase GOFs. This is exemplified in Figure 2.12a where the signal amplitude is higher with the reconstructed geology than the input one. This can be explained by the lack of small-scale heterogeneities that should have diffracted and reflected seismic waves. Signals propagated through the reconstructed geology are therefore less attenuated than the reference ones. However, the wave arrival times are very well reproduced since the mean velocity is correctly reconstructed by the PCA.

(a) PCA 1024 components



(b) 3D UNet

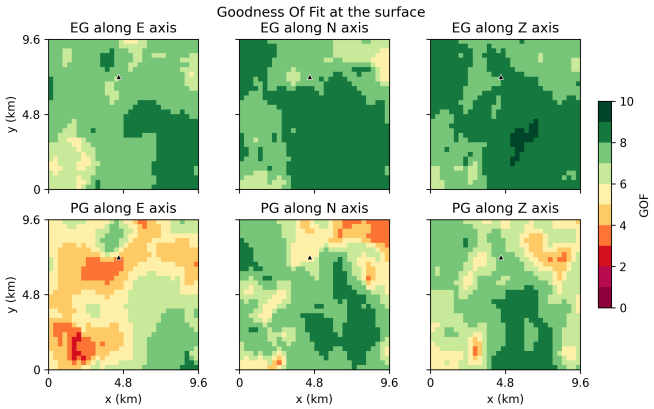


Figure 2.11: Goodness-Of-Fit (GOF) for each sensor on a 32 $\times$ 32 grid at the surface (10 means a perfect agreement). Each GOF is given for the three velocity components: East-West (E), North-South (N), Vertical (Z) axes. EG: Envelope GOF, PG: Phase GOF. The black triangles show the position of the sensor in Figure 2.12.

On the contrary, ground motion responses of the geological model reconstructed by the 3D UNet differ from the input mostly in terms of phase (Fig. 2.12b). Since the 3D UNet underestimates (for this specific sample) the mean velocity, the reconstructed signal is delayed with respect to the reference one. However, amplitudes are well reproduced thanks to the good reconstruction of small-scale heterogeneities.

2.4 Conclusion

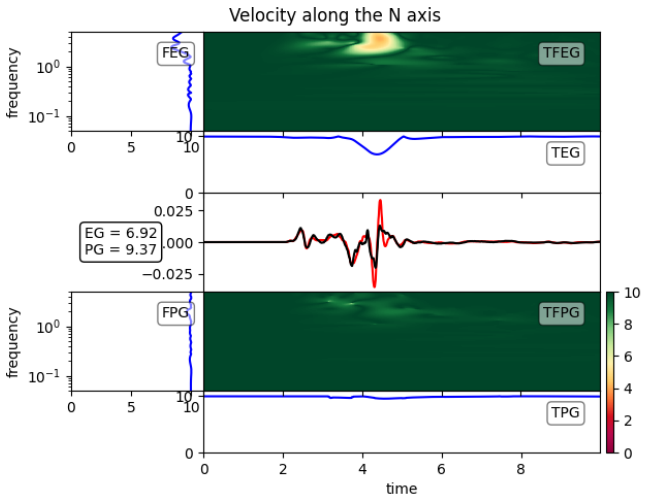
The aim of this chapter was to investigate machine learning based surrogate models that fulfill three requirements: i) a parametrization with different 3D heterogeneous geological models, ii) a dependence on the earthquake source properties, iii) the prediction of time series for each spatial point. Although several works on Gaussian processes, polynomial chaos expansion, and reduced-order models satisfy one or two of these requirements, no methodology was able to combine them all.

It then became necessary to build a database of ground motions associated with geological models that suits the framework of this work and allows us to design a new surrogate model. Our HEMEW^S-3D database is made of 30 000 3D heterogeneous geological models. It contains small-scale heterogeneities to investigate scattering effects as well as large-scale heterogeneities that represent hypothetical geological characteristics of the Earth's crust. Each geological model serves as the propagation domain for seismic waves generated from a point-wise source with a random location and orientation inside the domain.

The HEMEW^S-3D database exhibits a large variability on several aspects. The amplitude and duration of ground motion is widespread and depends, among other factors, on i) the hypocentral distance since longer propagation paths tend to reduce the PGV and increase the duration, ii) the seismic energy generated by the source, iii) the geological heterogeneities that induce dispersion and diffraction of seismic waves. This variability was quantified with several measures of intrinsic dimension in the ground motion database. Although there are several PCA methods, the analysis focused on the linear PCA. Thousands of principal components are required to capture the variability of ground motion with very good accuracy (95 % of the explained variance). Non-linear estimates of the intrinsic dimension based on the correlation dimension and the MLE indicate intrinsic dimension at least of the order of 100. While it is impossible to conclude on the exact intrinsic dimension of data, these analyses indicate that predicting the solution of the wave equation in 3D heterogeneous geologies is certainly a high-dimensional problem.

Relying on a drastic reduction of the input dimension does not seem either a very promising approach for a reduced-order surrogate model, based on our analyses of the PCA and the 3D auto-encoder. The PCA was computed with 1024 components to retain good reconstruction abilities. The latent space of the 3D UNet was designed with only 256 variables to enforce a greater reduction than the PCA. However, it should be noted that the skip connections between the encoder and the decoder transmit part of the information. In our experiments, we found that the reconstruction quality severely degrades when removing the skip connections. Without skip connections,

(a) PCA 1024 components



(b) 3D UNet

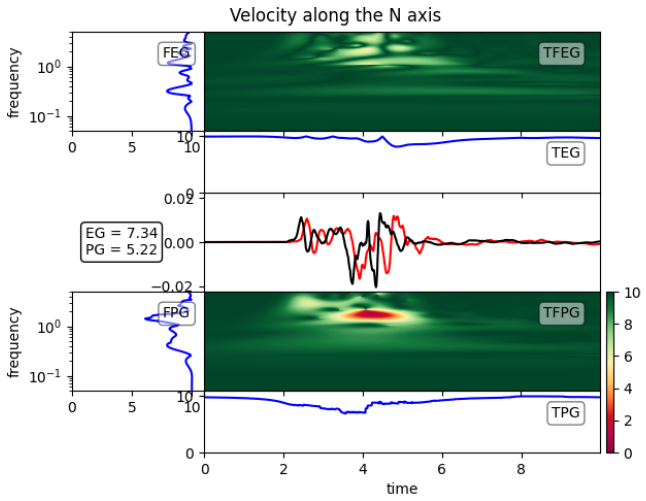


Figure 2.12: For the sensor depicted by the black triangles in Figure 2.11, velocity ground motion along the North-South component. Black: ground motion obtained with the input V_S field. Red: ground motion obtained with the V_S field reconstructed by (a) PCA and (b) 3D UNet. GOFs are represented in the frequency (FEG and FPG), time-frequency (TFEG and TFPG), and time (TEG and TPG) domain for the envelope (top) and phase (bottom).

the latent space needed to be larger than 1024 to provide correct reconstructions, which limits the benefits of the dimensionality reduction.

By propagating seismic waves through the geological models reconstructed by the PCA and the 3D UNet, we showed that the selected number of components was a minimal requirement to preserve the ground motion accuracy. It appeared that the PCA smooths small-scale heterogeneities, thereby leading to ground motions with larger amplitudes than the original ones. The only way to alleviate this issue would be to increase the number of principal components, which is not the desired direction. Different conclusions were drawn for the geological models reconstructed by the 3D UNet. The reconstruction error was somewhat similar to the one obtained with the PCA, with more geologies having an even better reconstruction. Especially, the 3D UNet was able to produce sharp outputs with small-scale heterogeneities being well preserved. However, some reconstructions were biased and yielded a time shift in the output ground motion.

In conclusion, these results show that we need another way to design a surrogate model of seismic waves propagation. Instead of trying to reduce the dimension of inputs into a 1D variable, we can take advantage of the cartesian structure of the geological models and ground motion wavefields to compute efficient Fast Fourier Transforms (FFT) of the data. Chapter 3 elaborates on the Fourier Neural Operator (FNO) that uses FFT to predict the full ground motion wavefields. Following the energy analysis of Chapter 1, one can estimate that running 30 000 SEM3D simulations to build the HEMEW^S-3D database amounts to an energy consumption of 3.3 MWh (Tab. 2.3).

Task	Number of realizations	Energy consumption (MWh)
SEM3D simulations 9.6 km $\times$ 9.6 km $\times$ 9.6 km, T_f=6.4 s	$N_{train} = 30000$	3.3

Table 2.3: Estimation of the energy consumption required to build the database. The energy consumption of 1 CPU is estimated to 4.88 Wh. One SEM3D simulation lasts in average 22.72 h CPU. Although the database contains time series of duration 8 s, we count only the 6.4 s used with the surrogate model in Chapter 3.

Chapter 3

Predicting ground motion with Fourier Neural Operators

Large datasets of internationally recorded earthquakes have enabled successful deep learning applications in seismology, in tasks as diverse as: localizing the hypocenter, characterizing the type of source (earthquake, blast, ambient noise) and the type of earthquake (subduction, crustal, etc.), picking the phase of P- and S-waves for wave arrival determination, reconstructing missing recorded data, detecting geological faults, imaging the sub-surface from ground motion recordings, etc (see Mousavi and Beroza 2023; Kubo et al. 2024 for reviews of deep learning in seismology). Focusing on the prediction of ground motion features, we should distinguish several applications. Early works start with a Multi-Layer Perceptron (MLP) that connects all artificial neurons between two consecutive layers (Fig. 3.1). Within this framework, MLPs are similar to GMPEs in the sense that the input layer gathers all parameters that are expected to influence ground motion (e.g. earthquake depth, moment magnitude, source orientation, epicentral distance, $V_{S,30}$ at the location of interest) and the output layer contains one neuron per ground motion feature to predict (e.g. PGV, PGA, PSA at various periods). Besides the numerous works based on recorded earthquakes (e.g. Alavi and Gandomi 2011; Derras et al. 2016; Kalakonas and V. Silva 2022), L. Lehmann et al. 2023 used physics-based simulations with a given 1D geological model and randomly chosen seismic sources to create a dataset of PGD, PGV, PGA from synthetic ground motions. This dataset was then used to train a MLP.

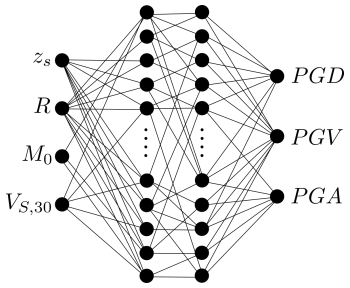


Figure 3.1: Example of a Multi Layer Perceptron (MLP) predicting ground motion features from earthquake parameters and features of the location of interest. For the sake of clarity, only a fraction of the inter-neuron connections is depicted.

Instead of considering scalar inputs that are dependent on the accuracy of the pre-processing step, ground motion features can be predicted directly from the full ground motion time series. This approach is especially useful in the context of Earthquake Early Warning (EEW) when the peak ground motion can be predicted from the first seconds of ground motion recordings. Indeed, early ground motion does not contain the peak values of the S waves, which allows to send early alerts to stations farther from the epicenter. Convolutional Neural Networks (CNN, LeCun et al. 2015) achieve this task by applying learnable kernels to the time series and extract ground motion features (Hsu and C.-W. Huang 2021; Jozinović et al. 2020), Recurrent Neural Networks (RNN) take advantage of the temporal structure of ground motion waveforms to predict peak ground motion (Mousavi and Beroza 2020; A. Wang et al. 2023), while Graph Neural Networks (GNN) benefit from the spatial arrangement of seismic sensors to learn the distance between stations (Bloemheuvel et al. 2023; Murshed et al. 2023) and Transformers learn an attention

mechanism that informs ground motion predictions with a notion of similarity between sensors (Münchmeyer et al. 2021).

However, these methods are not suitable for structural analysis since the ground motion dynamics is not learnt and only peak values are available. To predict ground motion time series, Datta et al. 2022 used RNN in the context of EEW to predict the next 10 s of ground motion from the first 15 s. Generative Adversarial Networks (GAN) have also seen several applications to create synthetic ground motion time series that resemble those from a recorded dataset (see Marano et al. 2023 for a review of GANs in seismology). In particular, the generation can be conditioned on properties such as the magnitude, epicentral distance, or $V_{S,30}$ (Esfahani et al. 2022; Florez et al. 2022). Shi et al. 2024 extended this method with a U-shaped Neural Operator (Section 3.2.2) that can handle different resolutions, although this advantage has not been investigated by the authors.

Although most of the above-mentioned works contain some information about the geological context (for instance through the $V_{S,30}$), these scalar properties give limited information on the physical propagation of seismic waves between the source and the receiver. More detailed information can be obtained when considering geological models as inputs of the neural network. Moseley et al. 2020 proposed an auto-encoder to convert a 2D geological model into a time-depth image that stacks ground motion time series recordings on a line of sensors. Geological models are randomly generated from a superposition of horizontal layers fractured by a fault and outputs are generated from the simulation of the acoustic wave equation. This example approaches our target application, even though the prediction is disconnected from the underlying Partial Differential Equation (PDE). At this point, we should also notice the numerous deep learning applications in full waveform inversion, where an image of the sub-surface properties (V_S for instance) is predicted from a time-depth image of recordings. Reviews can be found in Adler et al. 2021; Y. Lin et al. 2023 and we can mention InversionNet3D (Zeng et al. 2021) as one of the few 3D models, based on an auto-encoder made of CNN layers and invertible layers.

Contrary to data-driven methods, some neural networks are purely physics-informed and do not require any training data. Prominent representatives of these approaches are Physics-Informed Neural Networks (PINN, Raissi et al. 2019) where neurons in the input layer correspond to the spatial $\boldsymbol{x} \in \Omega$ and temporal query points $t \in [0, T_f]$ and the output layer gathers the solution $u(\boldsymbol{x}, t)$. The derivatives $\frac{\partial u}{\partial t}(\boldsymbol{x}, t), \nabla_x u(\boldsymbol{x}, t), \cdots$ can be computed efficiently with automatic differentiation. Then, the loss function constrains u to satisfy the PDE, providing a PDE solution without the need for any additional data. PINNs have seen many applications on the acoustic wave equation (Rasht-Behesht et al. 2022) and the 2D Helmholtz equation with geological models of various complexities (e.g. X. Huang and Alkhalifah 2022; Song et al. 2022; Song et al. 2023). Since PINNs are mesh-free, they can be used to solve the wave equation with non-flat surfaces and explore the influence of topography on ground motion wavefields (Song et al. 2022; Ding et al. 2023). However, training PINNs is challenging, especially for high-frequency features. These difficulties are related to the conditioning of a differential operator in the PDE one aims at solving (De Ryck et al. 2023). Several approaches have been followed to mitigate this effect, such as pre-training PINNs with lower frequency signals (X. Huang and Alkhalifah 2022) and adding a new layer of Gabor basis functions to localize the signals (Alkhalifah and X. Huang 2023). Although some advanced training procedures have been proposed to improve the flexibility of PINNs (Cheng and Alkhalifah 2024), another drawback lies in the fact that PINNs are tightened to a given geological model and training needs to restart from scratch when geology changes.

It then becomes desirable to learn the solution of a PDE for a class of PDE parameters, input and/or boundary conditions. Scientific Machine Learning (SciML) emerged in this context to design deep learning approaches that have strong relationships with the PDEs they aim at solving (N. Baker et al. 2019). In this work, we focus on the forward problem of the elastic wave equation 1.2 and rewrite it as

$$\mathcal{L}(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s, \boldsymbol{u})(\boldsymbol{x}, t) = 0, \forall \boldsymbol{x} \in \Omega, \forall t \in [0, T_f] \tag{3.1}$$

where we remind that $a : \Omega \to \mathbb{R}^+$ is the PDE parameter (S-wave velocity), $\boldsymbol{x}_s \in \mathbb{R}^3$ is the source position, $\boldsymbol{\theta}_s \in \mathbb{R}^3$ or $\boldsymbol{\theta}_s \in \mathbb{R}^6$ is the source orientation, and $\boldsymbol{u} : \Omega \times [0, T_f] \to \mathbb{R}^3$ is the solution of the elastic wave equation. The boundary condition is fixed to absorbing boundary conditions on all surfaces excepting the traction-free top surface. The solution of the elastic wave equation is then given by

$$\boldsymbol{u} = \mathcal{G}(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s) \tag{3.2}$$

so as to satisfy equation 3.1. Typically, SEM3D plays the role of the operator $\mathcal{G}$ and it is assumed that the numerical solver provides an error-free solution in the frequency range of validity. We aim at finding a surrogate model G_ϕ parametrized by ϕ that is a good approximation of $\mathcal{G}$. The surrogate model is obtained with supervised learning as the solution of a regression problem. More precisely, for a set of inputs $(a^i, \boldsymbol{x}_s^i, \boldsymbol{\theta}_s^i)_{1 \leq i \leq N}$, the corresponding

solution $(\boldsymbol{u}^i)_{1\leq i\leq N}$ is computed with the numerical solver $\mathcal{G}$. From these training samples, we aim at finding the optimal parameter ϕ^*

$$\phi^* = \arg\min_{\phi} \sum_{i=1}^{N} \|G_\phi(a^i, \boldsymbol{x}_s^i, \boldsymbol{\theta}_s^i) - \boldsymbol{u}^i\| \tag{3.3}$$

where $\|\cdot\|$ is a norm to be specified. This framework is often called a *parameter-to-solution* mapping since the surrogate seeks the solution u that corresponds to a given set of parameters $(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s)$.

In the following, we review some SciML works applicable in our context and, due to the vast and rapidly expanding literature, we give a particular emphasis to approaches that have already seen applications with wave propagation (Section 3.1). Among the wide variety of methods, the Fourier Neural Operator (FNO, Z. Li et al. 2021) fulfills our requirements and its methodology is detailed (Section 3.2). Our first study of the FNO involved the U-shaped variant of the FNO (U-NO, Rahman et al. 2023), with which we provided the first prediction of the 3D wave equation using neural operators in F. Lehmann et al. (June 2023). "Fourier Neural Operator Surrogate Model to Predict 3D Seismic Waves Propagation". In: 5th ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering. Athens, Greece. DOI: 10.7712/120223.10339.20362

Our next developments were done with the Factorized Fourier Neural Operator (F-FNO, Tran et al. 2023) that greatly reduces the number of parameters and improves the predictions accuracy. With this method, the choice of hyperparameters was thoroughly investigated and some of the analyses are reproduced (Section 3.3.3 and 3.4.1). Using the HEMEW-3D with a fixed seismic source, we showed that predictions of the 3D elastic wave equation can be efficiently obtained thanks to a depth-to-time conversion (detailed in Section 3.2). This work was published in F. Lehmann et al. (Feb. 15, 2024b). "3D Elastic Wave Propagation with a Factorized Fourier Neural Operator (F-FNO)". in: *Computer Methods in Applied Mechanics and Engineering* 420, p. 116718. ISSN: 0045-7825. DOI: 10.1016/j.cma.2023.116718

Then, we introduced our Multiple-Input Fourier Neural Operator (MIFNO) that enables the first prediction of the 3D elastic wave equation depending on both the geological model and the source characteristics. After analyzing the results accuracy (Section 3.3), we try to understand the prediction error and how it relates to input parameters (Section 3.4). We finally test the generalizability of the MIFNO in different settings to assess its robustness (Section 3.5). This work is currently under review in the *Journal of Computational Physics*: F. Lehmann et al. (2024c). *Multiple-Input Fourier Neural Operator (MIFNO) for Source-Dependent 3D Elastodynamics*. arXiv: 2404.10115. preprint.

3.1 Scientific Machine Learning for PDEs

When trying to solve PDEs with deep learning, one first approach consists in exploiting the advantages of available numerical solvers, such as their accuracy, and mitigating their weaknesses with deep learning. In particular, the computational time is often the bottleneck of numerical solvers, especially when refining the mesh, and neural networks can significantly reduce the total time (Section 3.1.1). However, it is also desirable to obtain a surrogate model that can run independently from a numerical solver, once duly trained. In this direction, a line of work relies on the assumption that PDE solutions evolve in a subspace of reduced dimension (e.g. X. Huang et al. 2022, Z. Li et al. 2023a, Bhattacharya et al. 2021, P. Chen et al. 2024, Regazzoni et al. 2023, Lee and Carlberg 2021, Venturi and Casey 2023). Building on ideas already developed in Section 2.1, these deep learning approaches rely on neural networks to exhibit a database of basis functions. Most of these approaches are auto-regressive and need to compute the solution at time $t-1$ to compute the results at time t, which can hinder the accuracy for long prediction horizons. Also relying on a form of basis functions, the Deep Operator Network (DeepONet, Lu et al. 2021) was one of the earliest methods to provide PDE solutions for a class of PDE parameters (Section 3.1.2). Neural operators are defined as an extension of neural networks when inputs and outputs are functions. Deriving from the theory of kernel integral operators and benefiting from the efficiency of the Fast Fourier Transform (FFT), the FNO has seen many developments and several applications around the wave equation (Section 3.1.3).

3.1.1 Interfacing numerical solvers with machine learning

The speed of neural networks predictions makes them ideal candidates to accelerate the most demanding operations of numerical solvers. Numerical simulations are especially costly when run for high frequencies and refined meshes (the latter being often a necessary condition for the former). In this setting, neural networks can learn the mapping from a coarse grid to a finer grid. Coarse grid solutions are typically obtained with numerical solvers at

moderate computational costs and, after training, the neural network provides the corresponding fine-grid solution. For instance, seismic inversion was tackled with 2D finite differences enhanced by a UNet auto-encoder (Gadylshin et al. 2021).

The computational costs of numerical solvers also arise from repeated iterations to reach convergence of minimization problems, such as Newton iterations. These processes become significantly faster if one is able to start from a suitable initialization value. CNN auto-encoders have been used as preconditioners for a numerical solver of the Helmholtz equation (Azulay and Treister 2023; Lerer et al. 2023). Some iterations can also be directly replaced by neural networks such as DeepONet (E. Zhang et al. 2022), whose architecture will be specified in Section 3.1.2. Substantial speed-ups can also be achieved by learning optimal parameters in standard numerical integration schemes (Mishra 2018).

3.1.2 Deep Operator Networks

The Deep Operator Network (DeepONet, Lu et al. 2021) originates from the universal approximation theorem and relies on an architecture split in two sub-networks. On the one hand, a *branch network* takes PDE parameters discretized at several points $\{a(x_1), \cdots, a(x_m)\}$ as an input and predicts a latent vector $\boldsymbol{b} = (b_1, \cdots, b_p)$. In our framework, the input would correspond to the evaluation of a geological model at some sensors but it could also be an initial or boundary condition. On the other hand, a *trunk network* predicts a latent vector $\boldsymbol{r} = (r_1, \cdots, r_p)$ from the coordinates $(\boldsymbol{x}, t)$ where one would like to obtain the solution. Then, the output is obtained as the scalar product $\langle \boldsymbol{b}, \boldsymbol{r} \rangle$ and it should approximate the solution of the PDE parametrized by a, $G(a)(\boldsymbol{x}, t)$. The choice of layers inside the branch and trunk networks is flexible but fully connected and convolutional layers are commonly found.

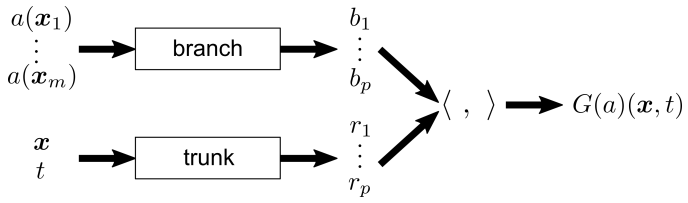


Figure 3.2: Illustration of the vanilla DeepONet (Lu et al. 2021) made of a *branch network* and a *trunk network*. For a PDE parameter a discretized at points $x_1, \cdots, x_m$ and a query point $(\boldsymbol{x}, t)$, it outputs the solution of the PDE defined by a at the query point, $G(a)(\boldsymbol{x}, t)$.

Among the many DeepONet extensions, one can notice the Physics-Informed DeepONet that constrains the solution to satisfy a given PDE with an additional term in the loss function, thereby reducing the number of training data (S. Wang et al. 2021). The MIONet augments the architecture with several branch networks to accept multiple inputs (Jin et al. 2022). A variant called GreenONet was designed to predict the Green function solution of the wave equation when the initial condition is given as an input (Aldirany et al. 2023). A 2D DeepONet was also applied for seismic inversion but it was complemented with a FNO to predict 2D maps of geological models (M. Zhu et al. 2023). However, Haghighat et al. 2024 noted that "DeepONets show weakness in learning operators in which the output solution moves spatially, corresponding to variations in the input function". Therefore, they designed an Extended-DeepONet that becomes aware of the function localization and used this model to solve the Eikonal equation. The DeepONet surrogate model was then used as a replacement of a numerical solver to find the earthquake hypocenter as the minimizer of a traveltime problem.

However, all of these works involve mapping inputs with a p-dimensional vector and to the best of our knowledge, this has not been achieved for 3D inputs. When considering 3D heterogeneous geologies, the input discretization must be fine enough to capture the small-scale geological heterogeneities, which leads to high-dimensional inputs. And the analyses conducted with the PCA and the 3D UNet have not led to any 1D geological encoding with a reasonable size. On the contrary, the FNO strongly relies on a regular discretization of the inputs.

3.1.3 Fourier Neural Operators and variants

Before explaining the FNO architecture in the next Section 3.2, we review existing works based on the FNO that tackle wave propagation problems. Similarly to the DeepONet, the FNO takes as input the initial condition, boundary condition, or parameter of a PDE and predicts its solution. For time-dependent PDEs, the FNO can either predict the

solution in an auto-regressive manner where $u(\boldsymbol{x}, t+1)$ is obtained from $u(\boldsymbol{x}, t)$, or by augmenting the problem dimension, predict the entire solution $(u(\boldsymbol{x}, t=0), \cdots, u(\boldsymbol{x}, t=T))$ in a single iteration.

The FNO and several variants were shown to be efficient surrogate models for the 2D wave equation, using the acoustic (Wei and L.-Y. Fu 2022; Y. Yang et al. 2021), elastic (O'Brien et al. 2023; Y. Yang et al. 2023; T. Zhang et al. 2023), or Helmholtz (B. Li et al. 2022) formulation (Tab. 3.1). Although the FNO was used for different objectives such as predicting high-frequency components from low-frequency features (Song and Y. Wang 2022) and predicting the following time steps based on the first iterations (Wei and L.-Y. Fu 2022; T. Zhang et al. 2023), several studies map a 2D geological model with the ground motion generated through this model. Geological models are obtained from random fields with a homogeneous background (coefficient of variation around 0.1), geometrical constructions where objects are embedded into a homogeneous background and create sharp discontinuities, or the OpenFWI dataset that contains different types of layers (Deng et al. 2022). When predicting solutions of the 2D elastic wave equation, Y. Yang et al. 2023 also consider a varying V_P/V_S ratio via random fields with large correlation lengths.

FNO applications in 3D are rare, especially for time-dependent PDEs. We were the first to show that the FNO can predict solutions of the 3D wave equation with the U-NO (F. Lehmann et al. 2023) and then, the F-FNO (F. Lehmann et al. 2024b). We proposed a depth-to-time conversion to predict the solution only at the surface of the propagation domain, where observations are usually available in real-life situations. Kong and Rodgers 2023 and C. Zou et al. 2023 instead use a 4D FNO (resp. 4D U-NO) to predict solutions of the elastic wave equation at all points in the domain but this leads to high computational and memory requirements.

In 2D and 3D, some studies consider both the source location and the geological parameters as input variables (Kong and Rodgers 2023; O'Brien et al. 2023; Y. Yang et al. 2023; C. Zou et al. 2023). Yet, these works do not take into account the different nature of the geological and source parameters: a structured 2D or 3D field for the geology and a vector of coordinates for the source. The source position is encoded as a binary field where a single pixel of value 1 describes the source location. In addition, an earthquake source cannot be described solely by the source location since its orientation also plays a crucial role, through the definition of the moment tensor. With the notable exception of O'Brien et al. 2023 that restricts their analyses to 2D and mention "significant challenges" to extend their methodology to 3D, the above-mentioned works consider an isotropic explosive source. In this situation, the moment tensor is fixed and does not need to be learnt by the surrogate model. But it prevents earthquake modelling that requires a moment tensor description of the source. We propose the MIFNO architecture to include both the source location and moment tensor by considering their vector representation.

3.2 Fourier Neural Operator for the elastic wave equation

3.2.1 Fourier Neural Operator principles

Let us consider a PDE that depends on parameters a and the source term $f(X_s) := f(\boldsymbol{x}_s, \boldsymbol{\theta}_s)$

$$\mathcal{L}_a u(\boldsymbol{x}, t) = f(X_s)(\boldsymbol{x}, t) \tag{3.4}$$

Let us introduce the variable $\xi = (\boldsymbol{x}, t)$ to ease the notation and $\Omega_\xi = \Omega \times [0, T_f]$ where $\Omega = [0; \Lambda]^3$ is the propagation domain and $T_f > 0$ is the final time. For the theoretical introduction of the surrogate model, we consider that u has values in $\mathbb{R}$ and the reasoning can be easily extended to the different components of the solution in the case of multi-dimensional outputs. If $G_a : \Omega \times \Omega \to \mathbb{R}$ is the Green function solution of

$$\mathcal{L}_a G_a(\xi, \cdot) = \delta_\xi, \tag{3.5}$$

it is well-known that the general solution u can be obtained as a convolution of the Green function and the initial condition

$$u(\xi) = \int_{\Omega_\xi} G_a(\xi, \xi') f(\xi') \mathrm{d}\xi' \tag{3.6}$$

Z. Li et al. 2020 proposed to model the Green function G_a as a neural network κ_ψ with parameters ψ

$$G_a(\xi, \xi') := \kappa_\psi \left(\xi, \xi', a(\xi), a(\xi')\right) \tag{3.7}$$

Embedding the integral formulation 3.6 with the kernel neural network, one can then introduce hidden variables $v_0, \cdots, v_\ell, \cdots, v_L$ that are obtained via

$$v_{\ell+1}(\xi) = \sigma \left(W v_\ell(\xi) + \int_{\Omega_\xi} \kappa_\psi(\xi, \xi', a(\xi), a(\xi') v_\ell(\xi')) d\xi' \right) \tag{3.8}$$

Table 3.1: Summary of studies predicting the solution of the 2D wave equation with Neural Operators. Train/test: number of training and validation data (k stands for thousand). Dimension of geological models: physical size and number of grid points in (width, depth). Dimension of seismic wavefields is implicitly the same as the geological model, with the final time T_f and the number of time steps S_t.

			Geological models			Seismic wavefields		
Dataset	train/val	Method	Domain (Resolution)	Values V_S	Construction	Time	Equation	Source
F. Chen et al. 2024	1000	NSNO	$[0;1]^2$ (256 × 256)	?	homogeneous background + geometrical objects	n-a	2D Helmholtz	random distributed sources
Garg and Chakraborty 2023	500 × 5	RP-WNO	0.69 × 0.69 km^2 (69 × 69)	?	OpenFWI (Flat-Vel-A)	n-a	2D Helmholtz	1 fixed source
B. Li et al. 2022	15k/3k	paralleled FNO (1 FNO for each of 10 frequencies)	7 × 7 km^2 (70 × 70)	1500; 4500m/s	OpenFWI	n-a	2D Helmholtz	5 surface source locations
O'Brien et al. 2023	40k	FNO	20 × 10 km^2 (200 × 100)	1000; 2333m/s	velocity gradient + random perturbations	T_f=5 s (S_t=50)	2D elastic	1 deep source with random location and random moment tensor
Ovadia et al. 2023	16k/4k	ViTO	$[0,\pi]^2$ (?)	1300; 1600 m/s	sinusoidal fluctuations	?	2D acoustic	2 sources with random location and amplitude
Wei and L.-Y. Fu 2022	800/200	recursive FNO	2.48 × 2.48 km^2 (64 × 64)	?	Velocity model is fixed to linear gradient, layered or portion of the Marmousi model	T_f=0.25 s (S_t=15)	2D acoustic	1 source with random location
Y. Yang et al. 2021	4.8k/0.2k	FNO	10.2 × 10.2 km^2 (64 × 64)	$\mathcal{N}(3000;150)$m/s	random mean V_S + von Karman random fields (ℓ_c=1.6 km, σ=0.1)	T_f=4 s	2D acoustic	1 source with random location
Y. Yang et al. 2023	18k/2k	U-NO	10.2 × 10.2 km^2 (64 × 64)	mean=3000 m/s	large-scale random fields for V_P/V_S ratio + small-scale random fields for V_S	(S_t=400)	2D elastic	1 source with random location
T. Zhang et al. 2023	90×300 / 10×300	recursive One connection FNO	10.4 × 10.4 km^2 (104 × 104)	?	100 models: homogeneous background with geometrical objects	T_f=0.6 s (S_t=300)	2D elastic	1 source with random location

Table 3.2: Continuation of Tab. 3.1 for the 3D wave equation. Dimension of geological models: physical size and number of grid points in (width, length, depth).

			Geological models			Seismic wavefields		
Dataset	train/val	Method	Domain (Resolution)	Values V_S	Construction	Time	Equation	Source
Kong and Rodgers 2023	30k	FNO	$16 \times 16 \times 16$ km^3 ($16 \times 16 \times 16$)	?	homogeneous background + von Karman random fields (σ=0.1)	(S_t=64)	3D elastic	1 explosive source with random location
C. Zou et al. 2023	27k/3k	U-NO	$5 \times 5 \times 5$ km^3 ($64 \times 64 \times 64$)	mean=3000 m/s	large-scale random fields for V_P/V_S ratio + small-scale random fields for V_S (σ=0.1)	T_f=2.5 s (S_t=50)	3D Helmholtz	1 explosive source with random location
F. Lehmann et al. 2023	27k/3k	U-NO	$9.6 \times 9.6 \times 9.6$ km^3 ($32 \times 32 \times 32$)[a]	1071; 4500m/s	random horizontal layers + von Karman random fields	T_f=6.4 s (S_t=320)	3D elastic	1 fixed source
F. Lehmann et al. 2024b	27k/3k	F-FNO	$9.6 \times 9.6 \times 9.6$ km^3 ($32 \times 32 \times 32$)[a]	1071; 4500m/s	random horizontal layers + von Karman random fields	T_f=6.4 s (S_t=320)	3D elastic	1 fixed source
F. Lehmann et al. 2024c	27k/3k	MIFNO	$9.6 \times 9.6 \times 9.6$ km^3 ($32 \times 32 \times 32$)[b]	1071; 4500m/s	random horizontal layers + von Karman random fields	T_f=6.4 s (S_t=320)	3D elastic	1 source with random location and random orientation

[a] Seismic wavefields are computed only at the surface with a resolution $16 \times 16 \times 320$
[b] Seismic wavefields are computed only at the surface with a resolution $32 \times 32 \times 320$

where W is a bias matrix and σ is the activation function (Gaussian Error Linear Unit, GELU, Hendrycks and Gimpel 2023). The integral operator appearing in Eq. 3.8 is denoted

$$K_\psi(a)v_\ell(\xi) := \int_{\Omega_\xi} \kappa_\psi\left(\xi, \xi', a(\xi), a(\xi')\right) v_\ell(\xi') d\xi' \tag{3.9}$$

K_ψ is a crucial component of the FNO and it concentrates most of the weights learnt by the network. It also requires computing an integral, which is always a numerical challenge, addressed by the Fourier formulation (detailed below).

In the parameter-to-solution framework, the surrogate model predicts the solution u from the input a. To achieve this, Z. Li et al. 2020 defines the learning $a \mapsto u$ as an iterative process

$$\begin{pmatrix} a(\xi) \\ \xi \end{pmatrix} \xrightarrow{P} v_0(\xi) \xrightarrow{F_1} v_1(\xi) \to \cdots \xrightarrow{F_\ell} v_\ell(\xi) \to \cdots \xrightarrow{F_L} v_L(\xi) \xrightarrow{Q} u(\xi) \tag{3.10}$$

Iterations can be decomposed in three main parts that are detailed hereafter: an uplift block P, a succession of Fourier layers $F_1, \cdots, F_L$, and a projection block Q (Fig. 3.3).

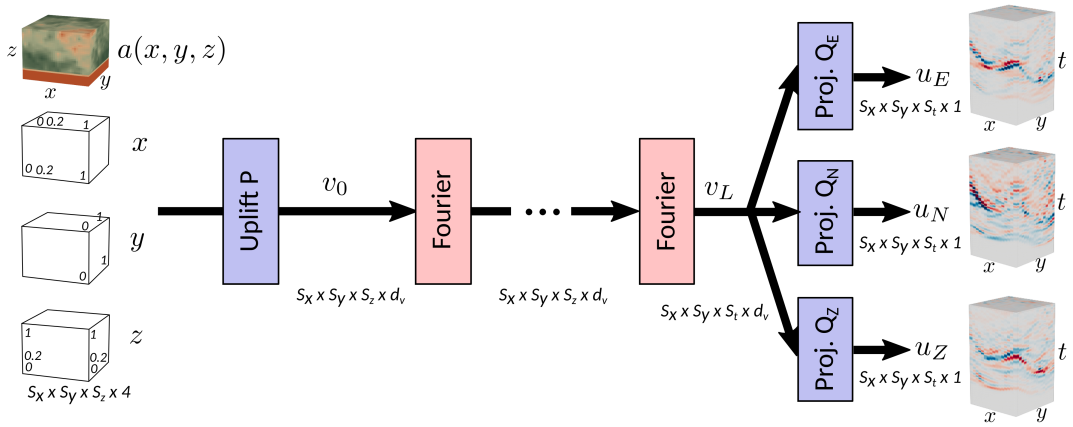


Figure 3.3: FNO architecture to predict the three components of ground motion u_E, u_N, u_Z from the geology a

It is important to note that the formulation is given for functions of ξ and this notation is preserved in the following. Indeed, neural operators predict functions, contrary to neural networks that predict scalar values. Also, we focus on the parameter-to-solution methodology since it corresponds with our objective. But the FNO can also be used to predict the solution from the initial conditions or to predict the solution at time $t+1$ from the solution at time t with the same implementation.

Uplift block P

The first step of the iterative process (Eq. 3.10) is an uplift block P that i) combines the input a with coordinates and ii) increases the data dimension by augmenting the number of channels. The input a is represented as a 3D tensor that contains the discretization of the function a on a Cartesian grid $(x_i, y_j, z_k)_{1\leq i\leq S_x, 1\leq j\leq S_y, 1\leq k\leq S_z} \subset \Omega$ with regular spacing $\Delta x, \Delta y, \Delta z$. To preserve the functional representation of the neural operator, the architecture should ensure that predictions can be queried at any point $\boldsymbol{x} \in \Omega$ at inference stage. This is achieved by combining the point's coordinates x_i, y_j, z_k with the input value at this point $a(x_i, y_j, z_k)$. In practice, the point $\boldsymbol{x} = (x_i, y_j, z_k)$ is identified with its image in the unitary cube $[0,1]^3$: $\tilde{\boldsymbol{x}} = (\Delta x \frac{i}{S_x}, \Delta y \frac{j}{S_y}, \Delta z \frac{k}{S_z})$. The four values $a(x_i, y_j, z_k), \tilde{x_i}, \tilde{y_j}, \tilde{z_k}$ are then given as inputs to the uplift block P (Fig. 3.4).

The second objective of the uplift block is to embed the inputs into a higher-dimensional space. To do so, P consists of two dense layers to produce hidden variables $v_0(x_i, y_j, z_k) \in \mathbb{R}^{d_v}$ (Fig. 3.4), where d_v is a hyperparameter to tune (the reference is chosen as d_v=16 in this work). Finally, from the concatenation of four matrices (a, x, y, z) each being of size $S_x \times S_y \times S_z$, P outputs a four-dimensional hidden variable v_0 of size $S_x \times S_y \times S_z \times d_v$. It should be emphasized that P acts point-wise and therefore, the mapping involves the same weights at each point.

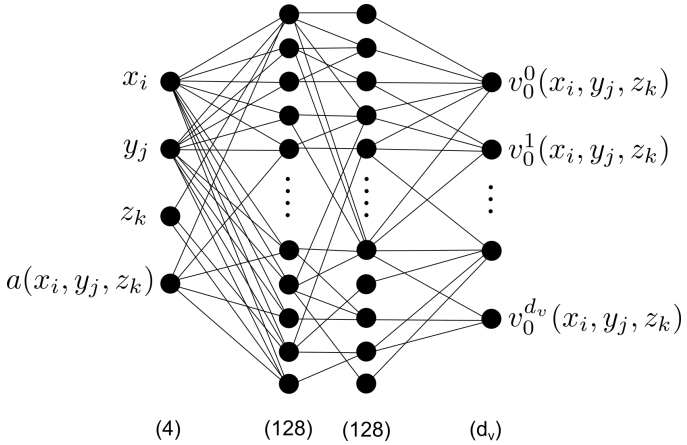


Figure 3.4: The uplift block P consists of two dense layers with 128 neurons each, followed by GELU activation. It takes as input the PDE parameter a and coordinates x, y, z. It outputs the first hidden variable v_0. Note that only a fraction of the inter-neurons connections are displayed for the sake of clarity.

Fourier layers F_ℓ

Equation 3.6 involves computing an integral on the domain Ω_ξ, which is a main computational concern. The original formulation of Z. Li et al. 2020 used a graph to compute the value at point ξ from the values of points in a neighbourhood of fixed radius. However, the computational efficiency can be significantly improved by relying on the Fast Fourier Transform (FFT). For this purpose, Z. Li et al. 2021 assumes that κ_ψ is a convolution kernel, i.e. stationary and independent from a, $\kappa_\psi(\xi, \xi', a(\xi), a(\xi')) := \kappa_\psi(\xi - \xi')$. Thanks to this assumption, the integral operator can be written as a convolution

$$K_\psi(a)v_\ell = \kappa_\psi * v_\ell \tag{3.11}$$

From the convolution theorem, the convolution can be written as a product in the Fourier space

$$K_\psi(a)v_\ell = FFT^{-1}\left(FFT(\kappa_\psi) \cdot FFT(v_\ell)\right) \tag{3.12}$$

Since κ_ψ is a neural network whose weights need to be learnt, they can be learnt directly in the Fourier space by defining $R_\psi := FFT(\kappa_\psi)$. Transforming Eq. 3.8 with this method leads to the formulation of the Fourier layer

$$v_{\ell+1} = \sigma\left(Wv_\ell + FFT^{-1}\left(R_\psi \cdot FFT(v_\ell)\right)\right) \tag{3.13}$$

To reduce the number of learnable weights, the high-order Fourier modes are discarded. Denoting M_x, M_y, M_z the number of Fourier modes preserved in each direction, R_ψ is a complex tensor of size $M_x \times M_y \times M_z \times d_v \times d_v$.

The full iterative process (Eq. 3.10) involves a succession of L Fourier layers (Tab. A.2). Although the reference FNO model has L=4 layers, this hyperparameter is optimized alongside the number of Fourier modes M_x, M_y, M_z.

Projection block Q

After the last Fourier layer F_L, the hidden variable v_L needs to be projected back to the physical space. The projection block Q is similar to the uplift P since it contains two dense layers separated by a GELU activation layer. When predicting the solution of the elastic wave equation, u has three components that correspond to the three directions of motion (East-West, North-South, vertical). They are obtained by three distinct projection blocks Q_E, Q_N, Q_Z

$$\begin{aligned} u_E &= Q_E(v_L) \\ u_N &= Q_N(v_L) \\ u_Z &= Q_Z(v_L) \end{aligned} \tag{3.14}$$

3.2.2 U-shaped Neural Operator (UNO)

The U-shaped Neural Operator (U-NO, Rahman et al. 2023) encompasses the Fourier layer into an architecture inspired from the UNet (Ronneberger et al. 2015). It comprises eight Fourier layers split into an encoder part (four

Fourier layers) with intermediate outputs of decreasing dimensions, and a decoder part (four Fourier layers) with increasing dimensions (Fig. 3.5 and Tab. A.3). Skip connections between the encoder and decoder are done via interpolation and concatenation. This architecture allows a deeper model than the FNO with a smaller number of parameters (from 58.7 million to 24.6 million) despite having twice more layers.

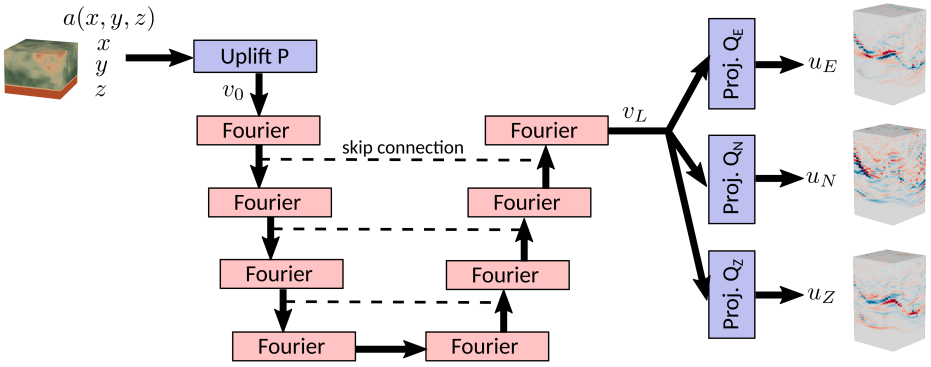


Figure 3.5: Representation of the U-NO, with four Fourier layers in the encoding part and four Fourier layers in the decoding part.

One important feature of the U-NO is its ability to modify the dimensions between v_ℓ and $v_{\ell+1}$. In the encoder part, dimensions are reduced by decreasing the number of Fourier modes in each Fourier layer. Although the original implementation suggests to increase the number of channels d_v alongside the dimension reduction (Rahman et al. 2023), this was not applied in our work. In the decoder part, dimensions are augmented by increasing the number of Fourier modes preserved inside each Fourier layer and padding with zeros when necessary. Especially, this trick allows us to obtain ground motion wavefields u_E, u_N, u_Z that have a larger dimension (S_x=32, S_y=32, S_t=320) than the geological input (S_x=32, S_y=32, S_z=32). This is applied to all neural operators used in this work.

3.2.3 Factorized Fourier Neural Operator (F-FNO)

The number of weights in the Fourier layer (Eq. 3.13) becomes very large when solving 3D PDEs. Indeed, weights R_ψ are of size $M_x \times M_y \times M_z \times d_v \times d_v$. The factorized Fourier Neural Operator (F-FNO) was introduced by Tran et al. 2023 to reduce the number of learnable parameters. The general architecture of the F-FNO is similar to the FNO and is illustrated in Fig. 3.6.

First, the structure of the factorized Fourier (F-Fourier) layer is slightly modified compared to the Fourier layer by introducing two bias matrices W_ℓ^1 and W_ℓ^2, and applying the residual connection outside the activation function

$$\begin{aligned} \boldsymbol{v}_{\ell+1} &= \boldsymbol{v}_\ell + MLP\left(\mathcal{K}_\ell(\boldsymbol{v}_\ell)\right) \\ MLP(\boldsymbol{k}) &= W_\ell^2\, \sigma\left(W_\ell^1\, \boldsymbol{k} + b_\ell^1\right) + b_\ell^2 \end{aligned} \tag{3.15}$$

where b_ℓ^1 and b_ℓ^2 are scalar biases. However, the main difference lies in the formulation of the integral operator K_ψ. Instead of computing 3D Fourier transforms as in Eq. 3.12, Fourier coefficients are computed independently along each dimension

$$K_\psi(a)v_\ell = FFT_x^{-1}\left(R_{\psi,x} \cdot FFT_x(v_\ell)\right) + FFT_y^{-1}\left(R_{\psi,y} \cdot FFT_y(v_\ell)\right) + FFT_z^{-1}\left(R_{\psi,z} \cdot FFT_z(v_\ell)\right) \tag{3.16}$$

This reduces the number of learnable weights to $(M_x + M_y + M_z) \times d_v \times d_v$. Consequently, a 4-layer F-FNO contains only 139 000 parameters, compared to 58.7 million for the FNO (Tab. A.4).

3.2.4 Multiple-Input Fourier Neural Operator (MIFNO)

The F-FNO predicts the geology-to-ground-motion mapping but does not account for the variability due to the source. Works like Kong and Rodgers 2023; C. Zou et al. 2023 add the source position as an additional input to the Neural Operator. This is done by inserting a fifth variable alongside a and the grids x, y, z. This variable is a cube full of zeros with a single pixel of 1 denoting the source position. We propose an alternative architecture that takes advantage of the vector representation of the source to include more source parameters.

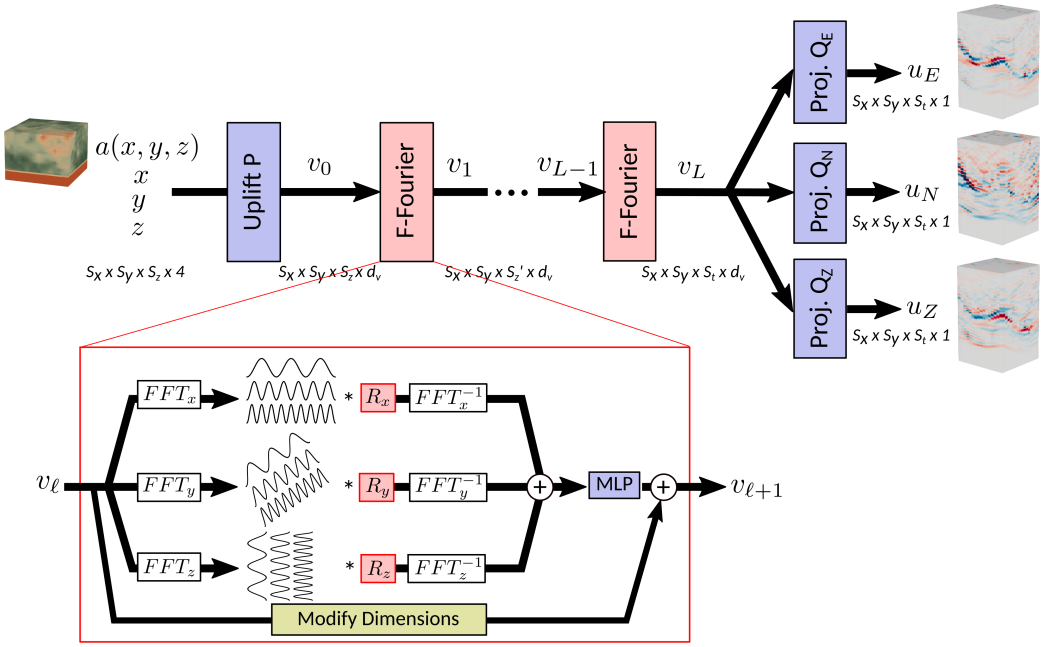


Figure 3.6: The F-FNO is made of one uplift block P, a succession of L factorized Fourier layers, and three projection sub-networks Q_E, Q_N, Q_Z. The details of a factorized Fourier layer are given to show the decomposition of the FFT along each dimension.

Inspired by the MIONet architecture that designs separate branches for each input (Jin et al. 2022), we design a Neural Operator architecture that encodes the source parameters in a specific *source branch*. Our Multiple-Input Fourier Neural Operator (MIFNO) is illustrated in Fig. 3.7.

In our MIFNO, the succession of factorized Fourier layers is split in two parts. From layers 1 to K, factorized Fourier layers act on the geology as in the original F-FNO. This forms the *geology branch*. In parallel, the *source branch* takes as input the vector of source coordinates $\boldsymbol{x_s}$ concatenated with the source orientation $\boldsymbol{\theta_s}$. Even though resolution invariance requires additional conditions (Bartolucci et al. 2023), FNOs can always be technically applied to inputs having a different resolution than the resolution used during training. This property must be preserved in the *source branch*, which is not trivial for convolutional layers whose number of weights depend on the size of the inputs. To circumvent this issue, the layers in the *source branch* are defined based on the number of modes used in the K-th factorized Fourier layer, M_x, M_y, M_z, and not the actual size of the inputs.

Several architectures were investigated to design the *source branch* (not shown) and we found that the best one was made of a perceptron and two convolutional layers separated by reshaping operations. The *source branch* can be decomposed in the following steps:

1. a 2-layer perceptron encodes the source characteristics in two dimensions. Its first layer contains 128 neurons and its second layer $4M_xM_y$ neurons. The reshaping operation leads to a 3D variable of size $2M_x \times 2M_y \times 1$.
2. a 2D CNN (with kernel size of 3) creates the third dimension. A first convolutional layer creates 8 channels and a second layer creates $2M_z$ channels. After reshaping, one gets a 4D variable of size $2M_x \times 2M_y \times 2M_z \times 1$.
3. a 3D CNN (kernel size of 3) adds the channel dimension with two convolutional layers.

At the end of the *source branch*, the `ModifyDimensions` layer transforms variables into the Fourier space where Fourier coefficients are padded with 0 if necessary, and then inverse transformed to obtain the same dimension as the output of the *geology branch* v_K, i.e. $S_x \times S_y \times S_z \times d_v$.

To combine variables from the *geology* and *source branches*, we take inspiration from Haghighat et al. 2024 who propose a 2D eXtended-DeepONet for earthquake localization. Following this idea, v_K and v_s are summed, substracted, and multiplied before being concatenated along the channel dimension. This leads to a variable of size $S_x \times S_y \times S_z \times 3d_v$ that is then sent to the remaining factorized Fourier layers, $K+1$ to L. The final projection layers Q_E, Q_N, Q_Z are identical to the F-FNO.

It can also be noted that the MIFNO is exemplified with factorized Fourier layers but the architecture can be extended to other types of layers. The MIFNO used in this work contains 16 factorized Fourier layers in total ($L = 16$), including 4 layers in the *geology branch* ($K = 4$). The number of channels is fixed to $d_v = 16$. Its total number of

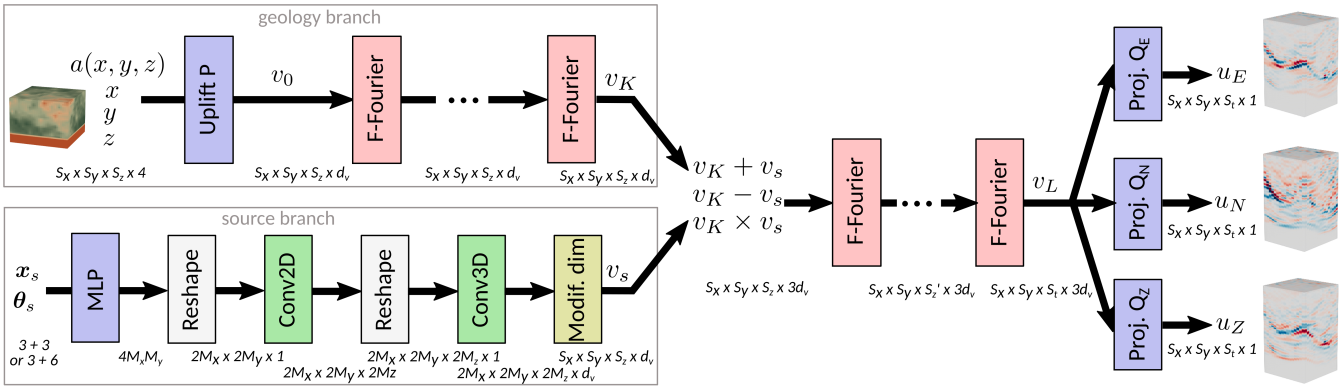


Figure 3.7: The MIFNO is made of a *geology branch* that encodes the geology with factorized Fourier (F-Fourier) layers, and a *source branch* that transforms the vector of source parameters into a 4D variable matching the dimensions of the *geology branch* output. Outputs of each branch are concatenated after elementary mathematical operations and the remaining factorized Fourier layers are applied. Uplift P and projection Q_E, Q_N, Q_Z sub-networks are the same as in the F-FNO.

parameters amounts to 3.40 million. The initial learning rate is 0.0004 and is halved on plateau, and the loss function is the relative Mean Absolute Error. The choice of hyperparameters and training strategies were investigated for the F-FNO (Section 3.3.3) and the best hyperparameters are preserved for the MIFNO. 27 000 samples were used for training and 3000 for validation. Training was performed on 4 Nvidia A100 GPUs for 200 epochs, which took 29.6 h.

Considerations related to inputs and outputs

It is important to note that the inputs and outputs of our Neural Operators do not live in the same physical space. In fact, inputs are geological models

$$a(x,y,z), \quad (x,y,z) \in \Omega.$$

while outputs are velocity wavefields defined at the surface of the propagation domain $\partial\Omega_{top}$. In Eq. 1.2, u denotes displacement at any point inside the domain while our Neural Operators predict velocity only at the surface. By abuse of notation, u denotes surface velocity $\frac{\partial u}{\partial t}|_{\partial\Omega_{top}}$ in the following and the Neural Operators' outputs are

$$\boldsymbol{u}(x,y,t) = \begin{pmatrix} u_E(x,y,t) \\ u_N(x,y,t) \\ u_Z(x,y,t) \end{pmatrix}, \quad (x,y) \in \partial\Omega_{top}, t \in [0,T_f].$$

By writing Fourier layers as functions of a hidden non-physical variable ξ, we interpret this variable ξ as transporting the depth information z contained in the third dimension of a into the temporal dimension t of $\boldsymbol{u}$. This depth-to-time conversion is a crucial feature to predict the solution of a 3D time-dependent PDE without the need for a 4D surrogate model.

Some care should be taken to normalize the inputs and outputs since their variability ranges are significantly different. Geological models are normalized to a Gaussian distribution, by centering them with respect to the mean geology and normalizing by four times the geological standard deviation. This normalization is the same as the one used to train the 3D UNet on geological reconstruction (Section 2.3.1) and yields geological values close to the interval $[-0.5; 0.5]$. The source coordinates and the source angles are mapped to the $[0,1]^3$ cube. Velocity wavefields predicted by the MIFNO are normalized with a scalar (i.e. independent of the sensor coordinates) that depends only on parameters known beforehand: the S-wave velocity at the source location $V_S(\boldsymbol{x}_s)$ and the source depth z_s. The normalization value is $c = V_s(\boldsymbol{x}_s)\sqrt{z_s^2 + (\Lambda/4)^2}$ where Λ is the length of the domain. The first term relates to the seismic energy released by an earthquake, which is inversely proportional to the shear modulus (hence, to the S-wave velocity) at the rupture location (Eq. 2.9). The second term accounts for the amplitude decrease of seismic waves generated by deep sources since they undergo more geometrical dispersion and diffraction. One should note that the normalization is applied during training to compute the loss function, but results presented hereafter have been denormalized.

3.3 Predictions accuracy

This section illustrates the MIFNO prediction ability. A representative example of MIFNO prediction is discussed in Section 3.3.1 and predictions on the test dataset are quantitatively assessed in Section 3.3.2. To justify the choice of the factorized Fourier layers and the MIFNO hyperparameters, the main results obtained for the F-FNO trained with the HEMEW-3D database are summarized in Section 3.3.3. In Section 3.3.4, the MIFNO is compared with variants of the F-FNO to ensure that the complexity added by the random source is well managed by the MIFNO architecture.

3.3.1 Prediction illustrations

Figure 3.8 illustrates the MIFNO predictions for a geology in the test dataset (Fig. 3.8a) and a source located at (3.9, 2.6, −6.2 km) with a strike of 298.7°, dip of 85.3°, and rake of 15.4°. The time series show an excellent accuracy on the three components since the wave arrival times are exactly reproduced: P-waves around 1.6 s and S-waves around 2.7 s (Fig. 3.8b). The peaks amplitude is also very similar between the prediction and the simulated ground truth. The frequency representation confirms the excellent agreement for all frequency ranges (Fig. 3.8c) and the three components.

At the depicted sensor, the envelope GOF is 8.7 and the phase GOF is 9.5. Since both scores are excellent, they reflect the visual agreement that can be observed in Fig. 3.8b. The envelope GOF is slightly lower than the phase GOF due to the presence of small-scale and small-amplitude fluctuations that are challenging to predict with the MIFNO (visible around 4 s in Fig. 3.8b).

When looking at the spatial evolution of the predictions, Fig. 3.9 shows that it is very accurate for the different time steps. The wavefront propagates from the source (epicenter denoted with the white star) with the correct speed since the arrival times are correct for all spatial locations. The source orientation is clearly visible in the first snapshot and it is correctly reproduced by the MIFNO. It is also remarkable that both P waves and S waves are well predicted, although their amplitude is very different (maximum East-West amplitude around 2 cm/s for the P waves and around 20 cm/s for the S waves). One can however notice that the predictions are smoother after the main wavefront, which reflects that small-amplitude fluctuations are under-estimated. This is visible for instance at t=2.00 s in Fig. 3.9 where the predicted wavefields show less fluctuations than the simulated ground truth.

3.3.2 Metrics analyses

To quantify the predictions accuracy more systematically, the envelope and phase GOFs were computed for samples in the training and validation datasets. Figure 3.10 shows that the GOF distributions are very similar for training and validation data. This indicates that the MIFNO is not subject to overfitting. Although the loss evolution during training shows a stable convergence with a small generalization gap (Fig. 3.11), this does not seem to impact the prediction accuracy.

The GOFs distributions confirm that the phase accuracy is better than the envelope accuracy (Fig. 3.10 and Tab. 3.3). Indeed, 80.2 % of predictions have an excellent phase GOF (phase GOF larger than 8). Envelope GOFs are good for 87.3 % of predictions (envelope GOF larger than 6), and 27.8 % are excellent (envelope GOF larger than 8). Table 3.3 indicates that the relative Root Mean Square Error (rRMSE) is 33 % on the validation dataset (first quartile). The relative RMSE is computed between predicted and simulated velocity time series at each sensor. Errors are then averaged over all sensors and samples. Relative errors on the order of 30 % are common in 3D PDE predictions, especially when complexity arises from the presence of heterogeneous parameters. The relative RMSE is complemented by the frequency biases (Section 1.3). Reminding that the frequency bias is by definition larger than −1, this explains the steep decrease of the high-frequency distribution around −1 in Fig. A.21.

Model	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$	EG	PG
MIFNO 8 layers	0.34 ; 0.59	-0.54 ; -0.09	-0.63 ; -0.15	-0.68 ; -0.18	6.32 ; 7.89	8.07 ; 9.17
MIFNO 16 layers	0.33 ; 0.59	-0.49 ; -0.05	-0.57 ; -0.09	-0.62 ; -0.11	6.57 ; 8.08	8.20 ; 9.21

Table 3.3: 1st and 3rd quartiles of the metrics computed on 1000 validation samples. rRMSE: relative RMSE (0 is best), $rFFT_{low}$: relative frequency bias 0-1Hz (0 is best), $rFFT_{mid}$: relative frequency bias 1-2Hz (0 is best), $rFFT_{high}$: relative frequency bias 2-5Hz (0 is best), EG: envelope Goodness-of-Fit (10 is best), PG: phase Goodness-of-Fit (10 is best). For frequency biases, negative values indicate underestimation.

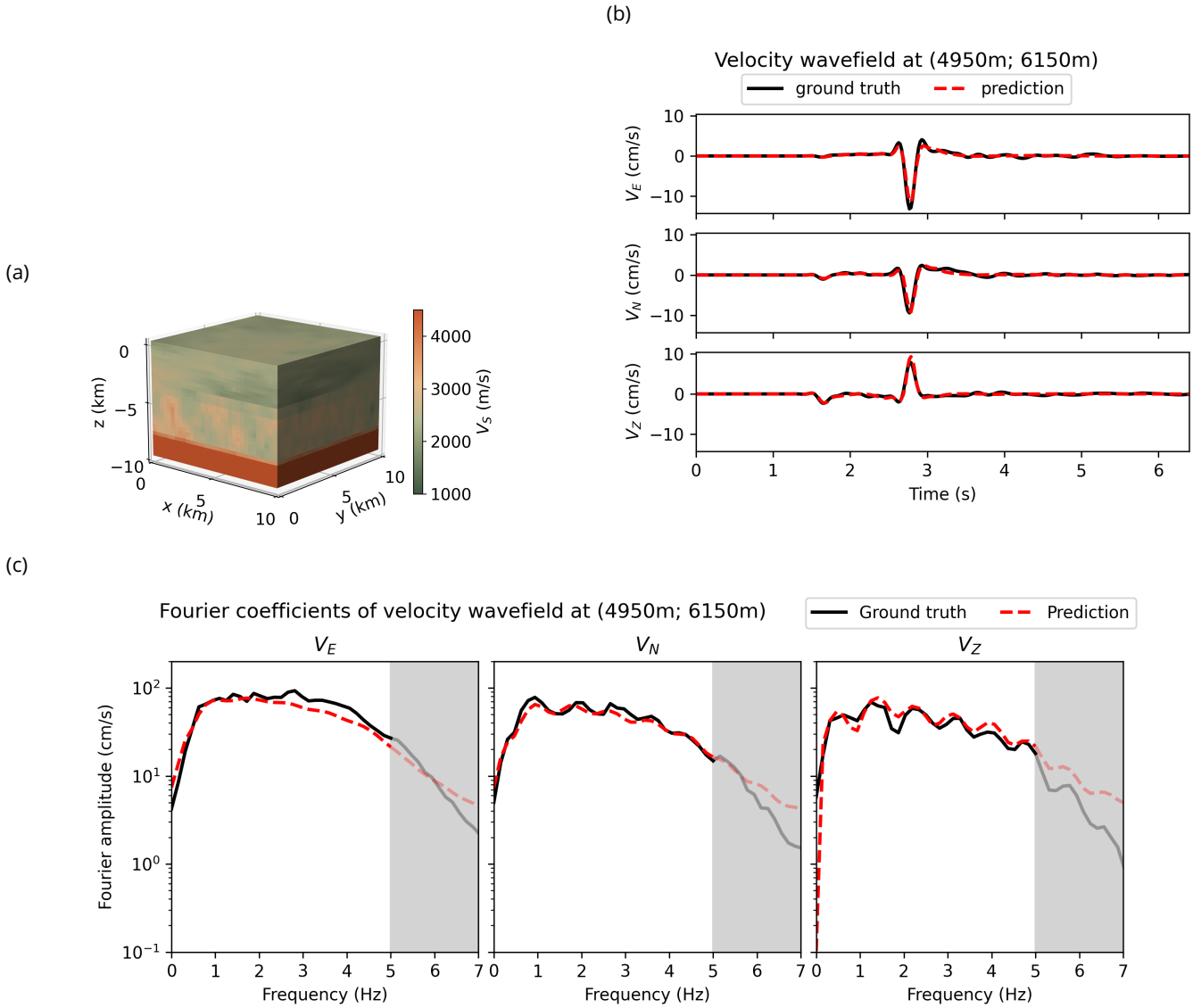


Figure 3.8: (a) Geology in the test dataset corresponding to the predictions in panels (b) and (c). The source is located at (3.9, 2.6, −6.2 km) with a strike of 298.7°, dip of 85.3°, and rake of 15.4°. (b) Velocity time series simulated (black) and predicted (dashed red line) in the three components: East-West (E), North-South (N), Vertical (Z). (c) For the same sensor as panel (b), amplitude of the Fourier coefficients of the velocity time series.

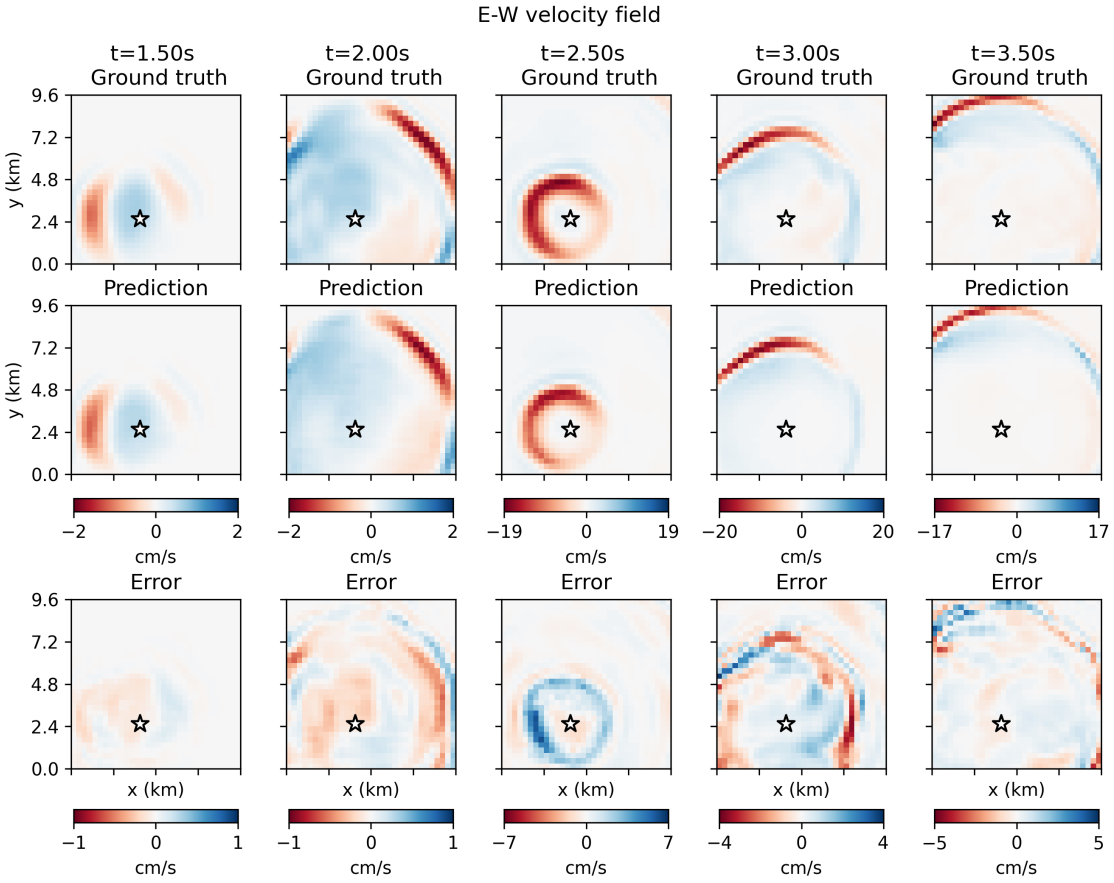


Figure 3.9: East-West component of the simulated (upper row) and predicted (center row) velocity fields for the geology illustrated in Fig. 3.8 at five time instants. The error between simulation and prediction is given in the lower row. The white star indicates the epicenter.

Table 3.3 also shows that the MIFNO mainly underestimates the frequency content of signals since frequency biases are mostly negative. This can be observed in Fig. 3.8b where the small fluctuations at the end of the signal tend to be ignored in the predictions. Findings on the F-FNO for the HEMEW-3D database with a fixed source corroborate these results (Tab. 3.4). Also, the MIFNO leads to larger underestimations for high-frequency components (the inter-quartile range is −62 % ; −11 % for high-frequency components while it reduces to −49 %; −5 % for low frequencies, Tab. 3.3). This reflects the well-known spectral bias, which states that small-scale (i.e. high-frequency) features are more difficult to predict than large-scale features (Basri et al. 2020; Rahaman et al. 2019). In addition, the late small-scale fluctuations also reflect complex physical phenomena due to the diffraction of seismic waves on scatterers. Indeed, fluctuations are mostly observed in the coda, after the S-wave arrival. Hence, they originate from the late arrivals of diffracted waves. Late velocity fields are subject to a large inter-sample variability, both between different sensors and between similar geologies with different locations of scatterers. These factors make high-frequency predictions very challenging.

Figure 3.12 gives additional insights on these results by representing the minimum and maximum GOF for each test sample. It shows that for almost all geologies, one can find a sensor that achieves an excellent GOF, and (almost) all sensors are at least fair. In addition, there is no correlation between the minimum GOF and the maximum GOF. This means that some samples have similar GOFs for all sensors while others display a large variability between sensors. The inter-sensor variability is not randomly distributed but instead shows a spatial organization that reflects underlying geological heterogeneities (see an example of the GOF maps in Fig. A.25). As will be discussed in-depth in Section 3.4.1, smaller coefficients of variation generally lead to better predictions, which is reflected by the color shades in Fig. 3.12.

One can also notice that the deeper MIFNO with 16 layers gives better predictions than the MIFNO with 8 layers (Tab. 3.3). Therefore, increasing the complexity of the model is beneficial to improve its accuracy and does not

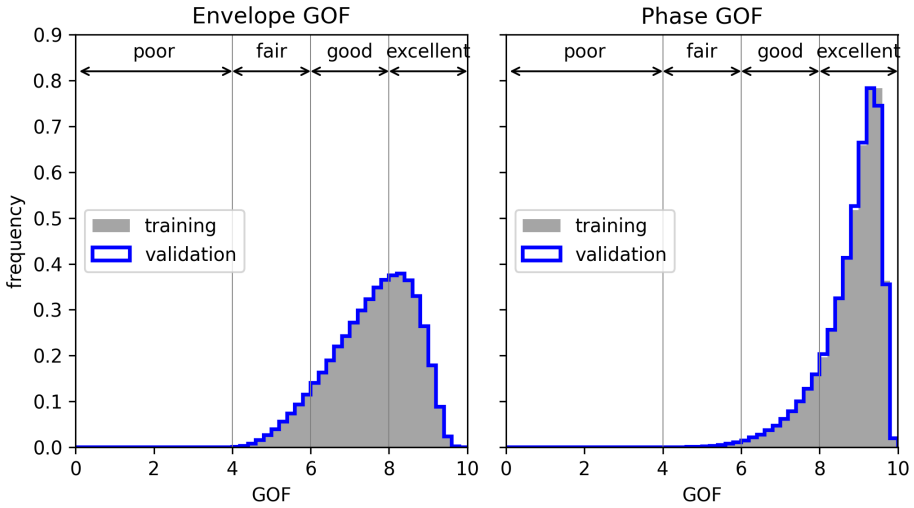


Figure 3.10: Distribution of the envelope GOF (left) and phase GOF (right) for each sensor and 1000 samples in the training (grey area) and validation (blue line) datasets.

lead to overfitting. This improvement is not reflected by the relative RMSE while it is obvious on the other metrics, emphasizing the need to consider physically-rooted metrics to evaluate the predictions accuracy. It can also be noted that there is no significant difference between the MIFNO taking as inputs the three angles describing the source orientation or the corresponding moment tensor (Tab. A.8). Therefore, the MIFNO with angles is used in the following for more direct interpretations of the results.

3.3.3 FNO, U-NO, and F-FNO comparison

The MIFNO was designed with factorized Fourier layers because our first investigations on a surrogate model predicting ground motion wavefields from the geology with a fixed source showed that the F-FNO was the most efficient and accurate architecture. This section recalls the main results obtained in this setting. All models were trained with the HEMEW-3D database, whose complexity is reduced compared to the HEMEW^S-3D database with a random source (Section 2.2.6). Therefore, errors presented in this section are smaller than the ones discussed above.

The baseline FNO model is able to predict correct surface wavefields, especially in the low-frequency range (-17% bias, Tab. 3.4). However, Fig. 3.13a shows a large generalization gap between the validation error and the training error. This may be due to the large number of parameters (58.7 million) that requires a huge training database to avoid overfitting.

HEMEW-3D database with a fixed source

Model	# parameters	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$
FNO	58.7m	0.23 $\pm$ 0.088	-0.17 $\pm$ 0.18	-0.30 $\pm$ 0.24	-0.44 $\pm$ 0.30
U-NO	24.6m	0.22 $\pm$ 0.088	-0.16 $\pm$ 0.16	-0.28 $\pm$ 0.22	-0.44 $\pm$ 0.28
F-FNO 4 layers	139k	0.23 $\pm$ 0.081	-0.29 $\pm$ 0.16	-0.42 $\pm$ 0.21	-0.55 $\pm$ 0.26
F-FNO 8 layers	246k	0.21 $\pm$ 0.083	-0.15 $\pm$ 0.14	-0.26 $\pm$ 0.19	-0.37 $\pm$ 0.26
F-FNO 28 layers	782k	**0.17** $\pm$ **0.081**	**-0.08** $\pm$ **0.10**	**-0.14** $\pm$ **0.15**	**-0.23** $\pm$ **0.22**

Table 3.4: Mean and standard deviation of relative biases for 1000 validation samples. For frequency biases $rFFT_{low}$, $rFFT_{mid}$, and $rFFT_{high}$, negative values indicate underestimation. For the number of parameters, "m" stands for million and "k" for thousand.

Despite the U-NO having been proposed to improve the performances of the FNO, our results show little differences between the two models. The medium frequency bias is −28 % for the U-NO compared to −30 % for the FNO. One can also notice on Fig. 3.13b that the generalization gap is smaller for the U-NO but still significant. The generalization gap may not be entirely related to the size of the training database. Indeed, the frequency biases are

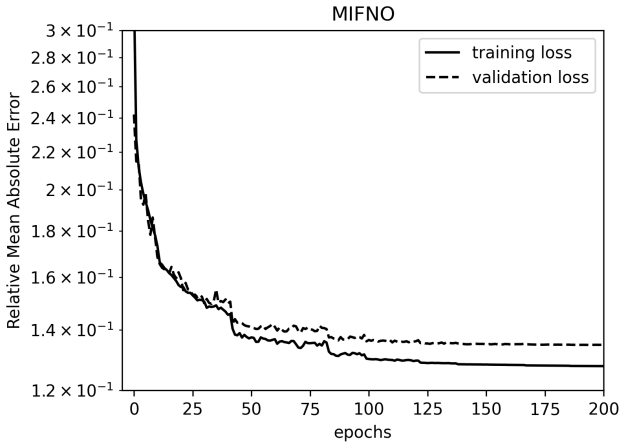


Figure 3.11: Evolution of the training loss (solid line) and validation loss (dashed line) as a function of the number of epochs.

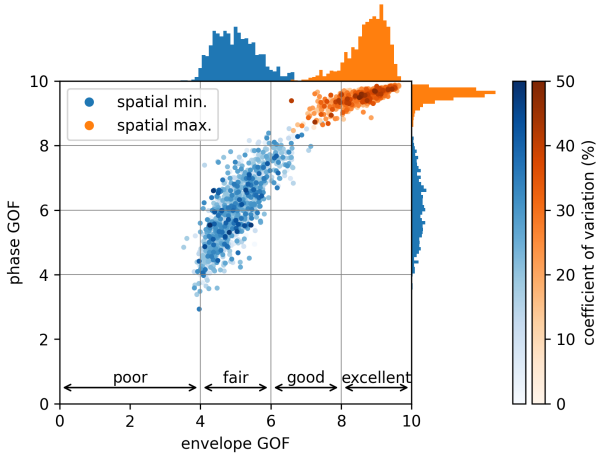


Figure 3.12: Each of the 1000 test samples is represented by the minimum (blue) and maximum (orange) GOF across all spatial points. The color shade corresponds to the coefficient of variation of the associated geology.

almost the same with 22 000 and 27 000 training samples (Fig. A.8). Though it is still possible that a database one order of magnitude larger than the one used in this study could drastically impact the results.

Additionally, we find that all skip connections are not necessary in the U-NO since the lowest errors are obtained for a model with only the deepest skip connection (Fig. A.9). The improvements obtained by removing the upper skip connections is especially visible on the high frequency biases. These findings may be explainable by the fact that inputs and outputs are not the same physical variables. Outputs of the first Fourier layer entering the upper skip connection are more similar to geological models while they get concatenated with variables being transformed into surface wavefields.

The F-FNO has the smallest number of parameters (from 139 000 with 4 layers to 782 000 with 28 layers) but its 8-layer version beats the FNO and the U-NO on all metrics reported in Tab. 3.4. Improvements are especially visible on the high-frequency bias that is reduced from −44 % to −37 %. One considerable difference between the FNO and the F-FNO also stands in the closing of the generalization gap. Independently from the number of factorized Fourier layers used in the model, the validation loss always closely follows the training loss (Fig. 3.13c).

With the F-FNO, a large number of those parameters were tested to highlight the most influential ones. All details can be found in Appendix A.3.2 and A.3.3; and the main outcomes are summarized below. The accuracy of each F-FNO variant is assessed with the frequency biases in low, medium, and high frequency ranges, on a subset of 1000 training and 1000 validation samples.

The two hyperparameters that have the most influence on the predictions' accuracy are the number of channels (d_v) and the number of layers. At least 12 channels should be considered and there is no improvement when increasing from 28 to 32 channels (Fig. 3.14a). In addition, choosing a deeper model is beneficial since the errors shrink between 4 and 16 layers and continue to reduce with more than 16 layers (Fig. 3.14b). Concerning the number of Fourier modes used in each layer, small improvements are also obtained when preserving more Fourier modes (Fig. A.10). However, effects are less pronounced for Fourier modes than for the number of layers or the number of channels. Especially, we found that increasing the number of temporal modes does not reduce the errors on the high-frequency components of the signal.

The training strategy was investigated in terms of the loss function, the size of the training database, the batch size, and the learning rate. For the loss function, the best results were obtained with the MAE rather than a combination of MAE and RMSE (Fig. A.12). Also, improvements were moderate when increasing the number of samples, which suggests that our database with 27 000 training samples is well-suited for this model (Fig. A.13). A learning rate of 6×10^{-4} was optimal to reach the lowest errors and the strategy to halve the learning rate on plateau was better than two alternative strategies (One cycle and cosine annealing with warm restarts, A.14).

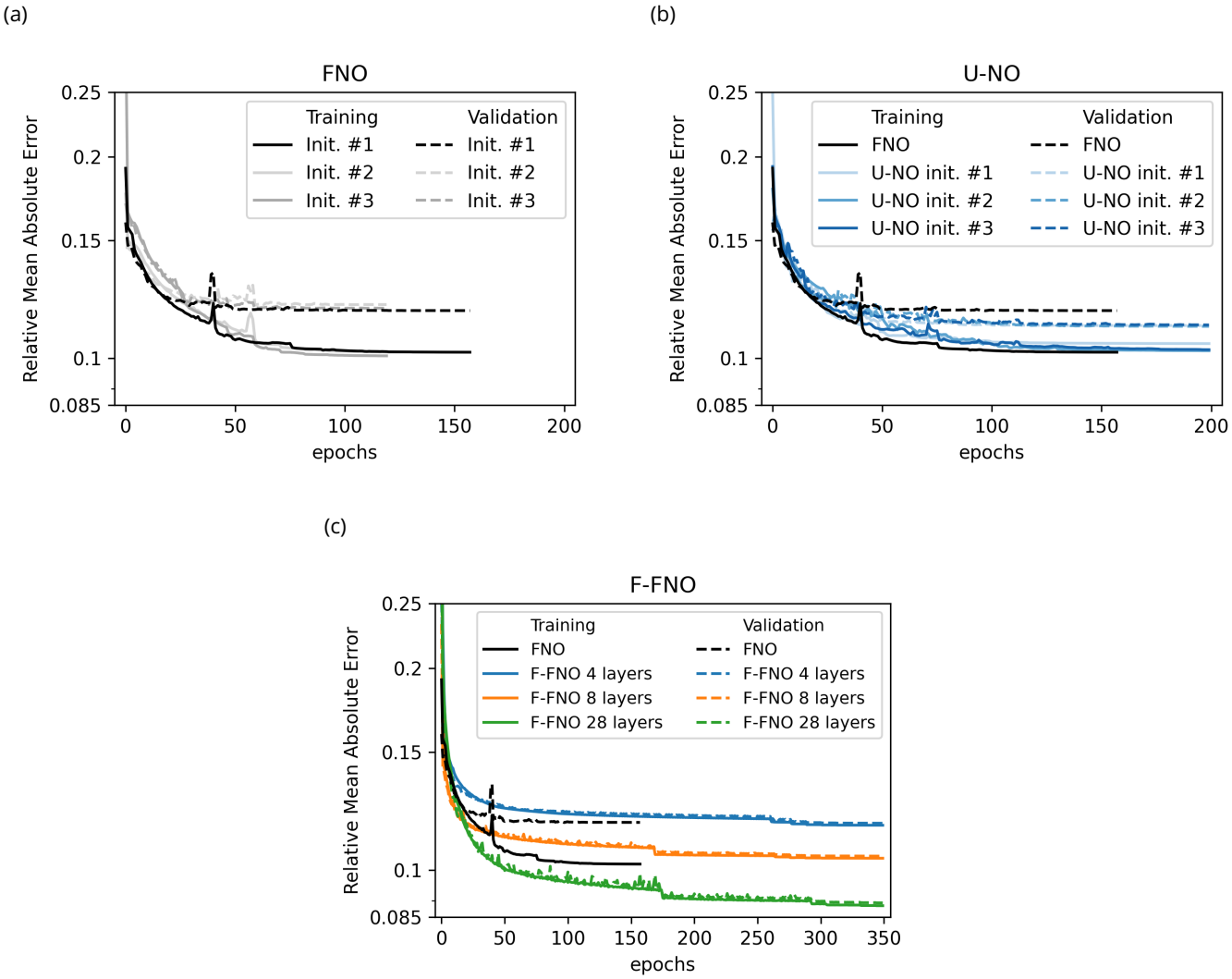


Figure 3.13: Evolution of the training (solid lines) and validation loss (dashed lines) for the three FNO variants (FNO, U-NO, F-FNO). Each panel reports the reference loss with the FNO (black). In the panels (a) and (b), the shaded lines correspond to trainings with random initializations.

3.3.4 Comparing the MIFNO with baseline models

Due to the difficulty of finding existing implementations of models for 3D and vector inputs, the MIFNO is compared to F-FNO models with less flexibility. This is meant to ensure that the additional complexity induced by the *source branch* does not deteriorate the predictions. To do so, we use three different databases that contain the same geological models but different surface velocity wavefields depending on the source parametrization:

1. in the HEMEW-3D database, both the source position and orientation are fixed. The source is located in the middle of the bottom layer: x_s=4.8 km, y_s=4.8 km, z_s=−8.4 km and its orientation is fixed to strike=48°, dip=45°, rake=88°. This database contains N_{train}=27 000 training, N_{val}=3000 validation and N_{test}=1000 test samples.
2. in an intermediate database, the source is located randomly inside the bottom layer, i.e. $x_s \sim \mathcal{U}([1200; 8100m])$, $y_s \sim \mathcal{U}([1200; 8100\ m])$, $z_s \sim \mathcal{U}([-9000; -8100\ m])$ and its orientation is fixed to the same value as the HEMEW-3D database. This database contains N_{train}=20 000 training, N_{val}=2000 validation and N_{test}=1000 test samples.
3. the HEMEW^S-3D database described in Section 2.2 contains sources located randomly anywhere in the domain and any source orientation. It contains N_{train}=27 000 training, N_{val}=3000 validation and N_{test}=1000 test samples.

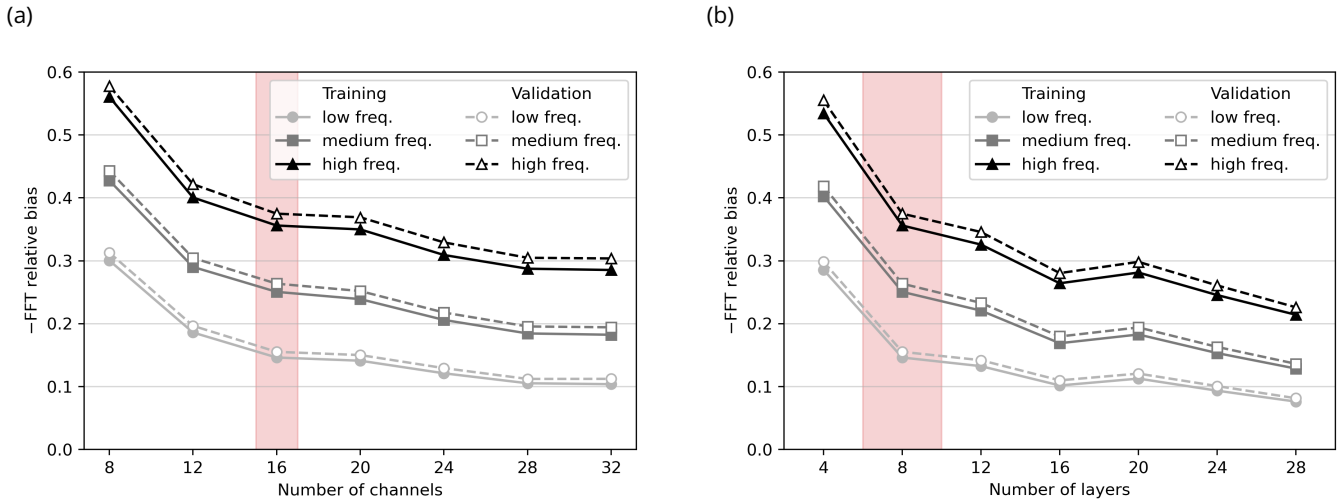


Figure 3.14: Evolution of the relative frequency bias as a function of the number of channels (d_v, panel a) and the number of layers (panel b). The mean frequency bias on training samples is shown with filled markers while unfilled markers refer to validation samples. Frequency biases are computed for low frequency (0-1Hz, light grey), medium frequency (1-2Hz, dark grey), and high frequency (2-5Hz, black). Note that the y-axis gives the opposite of the frequency biases, so that the best models have the lowest amplitude of frequency biases. The red area shows the reference model with d_v=16.

First, let us consider the prediction accuracy with respect to the geological properties. The baseline model is the 16-layer F-FNO described in Section 3.2.3. To match the number of parameters of the MIFNO, the F-FNO was designed with d_v=16 channels in layers 1 to 4, and d_v=48 channels in layers 5 to 16. The F-FNO was trained on the first database HEMEW-3D with fixed source location and orientation. The MIFNO was trained as described above on the HEMEW^S-3D database with random source location and orientation. Therefore, it should be noted that the training task of the MIFNO is much more complex than the F-FNO task.

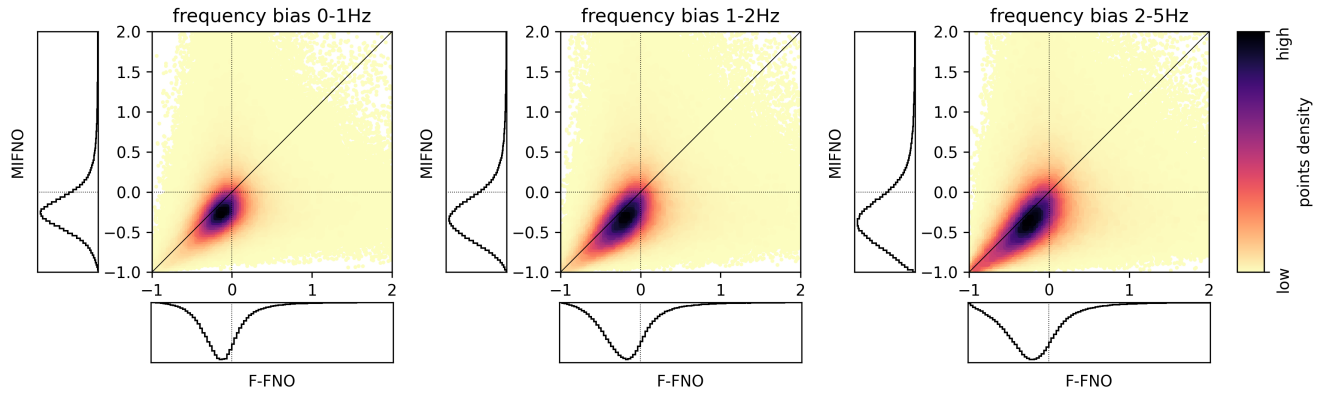


Figure 3.15: For 1000 samples in the HEMEW-3D test dataset (fixed position and fixed orientation) and each sensor, the frequency bias of the MIFNO (vertical axis) is shown against the F-FNO frequency bias (horizontal axis). Each subplot corresponds to a different frequency bias, 0-1Hz (left), 1-2Hz (middle), 2-5Hz (right). Optimal predictions correspond to a frequency bias of 0.

At inference stage, both the F-FNO and the MIFNO were used to predict surface velocity wavefields for 1000 test geologies with a fixed source corresponding to the reference source in the HEMEW-3D database. Figure 3.15 compares the F-FNO predictions with the MIFNO predictions for each sensor and each sample. One can firstly notice that points are widespread on both sides of the diagonal, meaning that the MIFNO or the F-FNO can give better predictions depending on the situation. However, points are more densely distributed in the lower right

triangle of the graph, close to the diagonal (Fig. 3.15). This indicates that the F-FNO is in average more accurate than the F-FNO when predicting only solutions with a fixed source.

However, the difference between the F-FNO and the MIFNO is around 0.8 GOF units, which remains acceptable knowing the extended complexity of the MIFNO task (Tab. A.9). Figure A.22 illustrates velocity time series predicted by the F-FNO and the MIFNO. It shows that the MIFNO predictions are already reasonable while the F-FNO locally improves the accuracy.

For the next comparison, we focus on the influence of the source location. To do so, we designed a F-FNO taking as input the geological parameters and a binary encoding of the source position, i.e. a cube full of zeros with a single 1 indicating the position of the source. The MIFNO was trained with only the source coordinates given as inputs to the source branch. In this experiment, both the F-FNO and the MIFNO were trained on the second database (random source location and fixed orientation).

Dataset with a random source position and fixed source orientation

Model	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$	EG	PG
F-FNO	0.11 ; 0.20	-0.22 ; 0.06	-0.29 ; 0.05	-0.34 ; 0.05	7.49 ; 8.63	8.74 ; 9.42
MIFNO	0.12 ; 0.21	-0.23 ; 0.05	-0.30 ; 0.04	-0.36 ; 0.04	7.36 ; 8.53	8.64 ; 9.36

Table 3.5: 1st and 3rd quartiles of the metrics computed on 1000 validation samples. rRMSE: relative RMSE (0 is best), $rFFT_{low}$: relative frequency bias 0-1Hz (0 is best), $rFFT_{mid}$: relative frequency bias 1-2Hz (0 is best), $rFFT_{high}$: relative frequency bias 2-5Hz (0 is best), EG: envelope Goodness-of-Fit (10 is best), PG: phase Goodness-of-Fit (10 is best). For frequency biases, negative values indicate underestimation. Both models contain 16 layers and were trained with 20 000 samples for 300 epochs.

When both models have 8 layers, the MIFNO predictions are slightly better than the F-FNO (Tab. A.10). With 16 layers, both models give a similar accuracy (second digit fluctuations in Tab. 3.5 are within the variability of the model due to a different random initialization). Phase GOFs are excellent for both models and the envelope GOFs are close to excellent (Tab. 3.5). Overall, these results show that the *source branch* architecture in the MIFNO preserves the expressivity of the F-FNO while providing the flexibility to add more source parameters.

3.4 Towards explainability of the prediction error

Although a surrogate model can never be expected to predict perfect results, it is desirable to understand the situations when it leads to the best accuracy, hence when one can trust its predictions the most. Section 3.4.1 investigates the influence of geological heterogeneities, in terms of the coefficients of variation and correlation length. Results in this section have been obtained with the F-FNO trained on the HEMEW-3D dataset with a fixed source. They can be extended to the MIFNO by decoupling the influence of the source, which is why we choose to present first the F-FNO results. Then, Section 3.4.2 explores the impact of the source position and source orientation on the prediction accuracy.

3.4.1 Influence of the geological heterogeneities

In this section, predictions are done with the F-FNO on the 1000 *consistent geological models* to analyse the relationship between the coefficient of variation of heterogeneities in geological models, their correlation length, and the prediction error. In the *consistent geological models*, all layers have the same coefficient of variation and the same correlation length to facilitate the interpretation. One can firstly observe in Fig. 3.16 that the prediction errors increase when coefficients of variation are larger. Additionally, the slope of the prediction error is larger when the correlation length is small. This means that there is an interaction between coefficients of variation and correlation lengths. Thus, geological models with both large coefficients of variation and small correlation lengths lead to the largest errors. Indeed, it is known that geological models with high coefficients of variation induce stronger impedance contrasts that disperse the incident wave. The subsequent multi-scattering effect is more pronounced when the coefficient of variation increases. Also, the correlation length impacts the dispersion and diffraction of incident waves at different wavelengths.

These findings reflect previous remarks on the spectral bias inherent to neural networks, which implies that high-frequency features are less well predicted than lower frequencies. It can be emphasized that these prediction

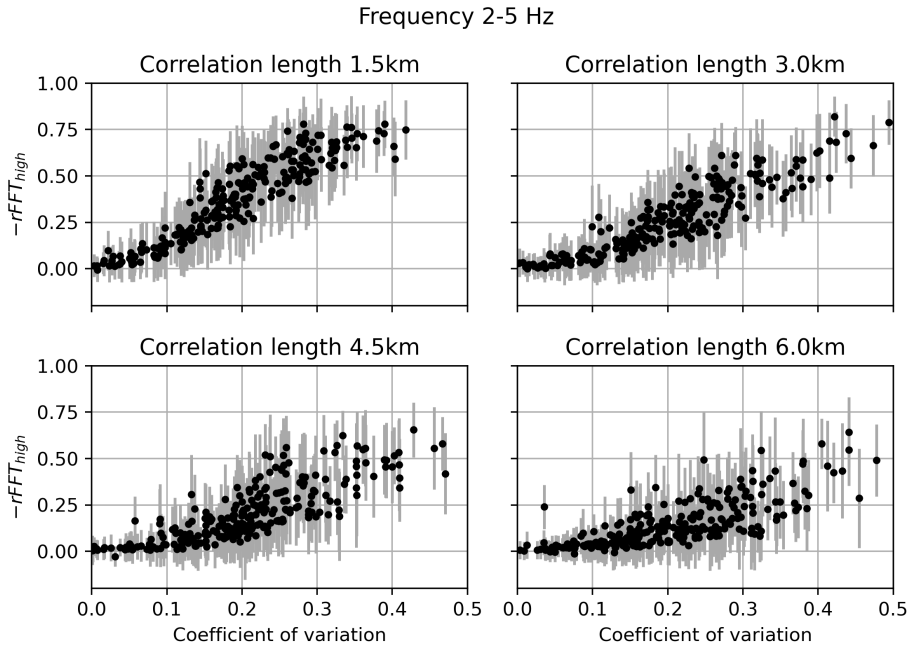


Figure 3.16: For 1000 *consistent geological models* in the test database, the high-frequency relative frequency bias is computed for each spatial point where mean (dots) and standard deviation (error bars) are extracted. Each geological model is characterized by its coefficient of variation (x-axis). Velocity models are stratified by their correlation length.

errors are not due to some imbalance in the training dataset. Indeed, coefficients of variation of geological models follow a normal distribution of mean 0.2. This implies that there are statistically as many geological models with coefficients of variation around 0.1 than there are around 0.3. Although it is very clear from Fig. 3.16 that frequency biases are significantly larger for coefficients of variation around 0.3.

For the MIFNO, the interpretation is challenging since more parameters influence the predictions accuracy but Fig. 3.12 suggests that geologies with the largest coefficients of variation lead to lower GOFs.

3.4.2 Influence of the source parameters

In this section, we highlight the influence of the source parameters on the MIFNO predictions and on the error. For a given geology, Figure 3.17a compares the predictions and the ground truth for three different positions of the source (with the same orientation). One can firstly observe that the predicted wavefields are very similar to the ground truth for all source positions. Especially, their spatial location closely follows the source position. This confirms that the source position encoding in the *source branch* is accurate. Figure 3.17a also illustrates the variability of surface wavefields that can be obtained by moving the source in a heterogeneous geology. Indeed, the same geology was used for all three snapshots and the source orientation is fixed but the surface wavefields are distorted differently due to the heterogeneities encountered between the source and the surface.

Next, we focus on the source orientation while fixing the position of the source. A geology with low heterogeneities is chosen to emphasize the source orientation. Three sets of (strike, dip, rake) angles are examined in Fig. 3.17b. The selected time is just after the P-wave arrival time. For all orientations, the signs of the predicted ground motion closely match the ground truth, meaning that the MIFNO can predict the influence of the source orientation. As already mentioned, a slight amplitude underestimation is visible in some regions but this is not due to the source orientation.

For more systematic analyses, the 1-2 Hz frequency bias is evaluated on 1000 samples with various geologies, source locations, and source orientations. Figure 3.18 shows the frequency bias against the source coordinates. Similar errors are obtained for all source longitudes and latitudes, meaning that the prediction accuracy is independent from the epicenter location. There is a slightly positive correlation between the source depth and the frequency bias but the inter-sample variability is of the same order. Figure 3.18 suggests that the underestimation of frequency

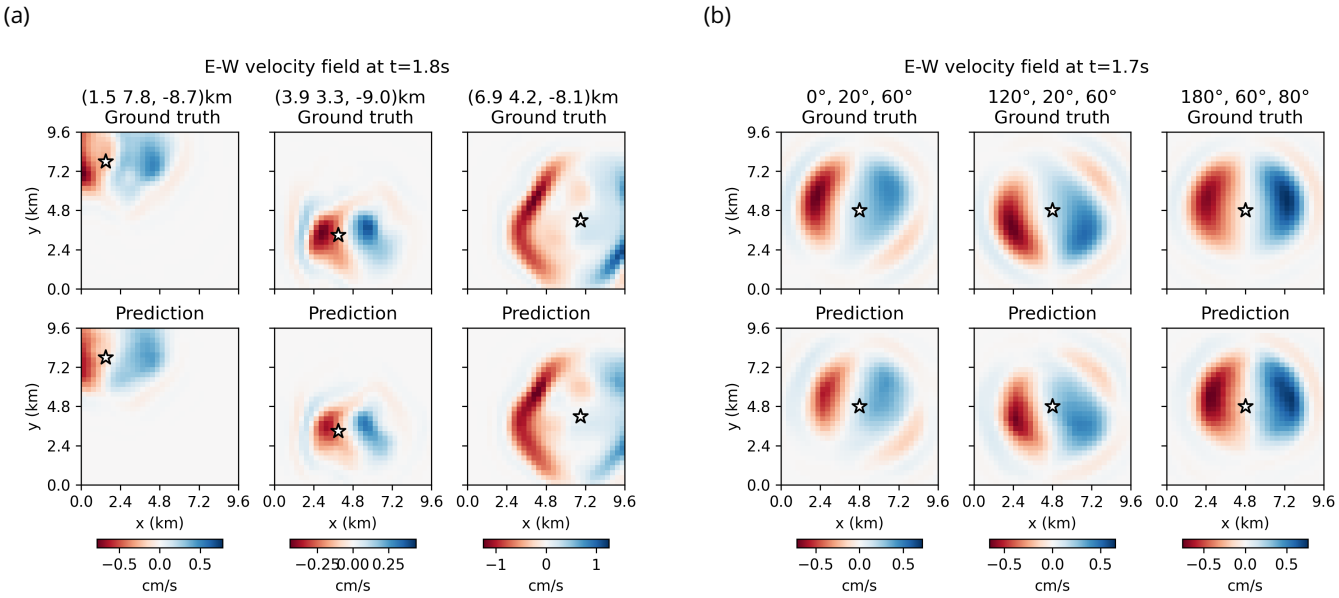


Figure 3.17: Comparison of the MIFNO predictions (lower rows) with the simulated ground truth (upper rows) for a given geology and time. (a) the source position changes (indicated by the white star) but the orientation remains constant. (b) the source position is fixed but the orientation changes and is indicated with the (strike, dip, rake) angles.

content is greater when the source is deeper while errors are smaller and overestimation is more frequent for shallow sources. Since both the phase and envelope GOFs indicate better results for shallow sources (Fig. A.24), this effect is likely due to the increased complexity of the wavefields generated by a long propagation of waves. Indeed, seismic waves originating from deep sources statistically encounter more heterogeneities on their path, which leads to diffraction and dispersion and hence, perturbs the surface wavefields. Therefore, it is generally more complex to predict accurate wavefields generated by deep sources.

Figure 3.19 corroborates this interpretation since, for most predictions, the GOF is inversely correlated with the epicentral distance, defined as the horizontal distance between the source and the sensor. This effect is especially pronounced for the envelope GOF since we have already commented that longer propagation paths lead to more complex ground motion time series, with small fluctuations. In conclusion, both the source depth and the epicentral distance are pertinent explanatory variables to understand the prediction accuracy. On the contrary, it can be observed that the source orientation parameters (strike, dip, and rake angles) have no effect on the prediction accuracy (Fig. A.23). This confirms that the MIFNO produces equally accurate predictions for all types of sources.

3.4.3 Relationship with the energy integral

We have already mentioned that wavefields with small-scale fluctuations are predicted less accurately. To quantify this phenomenon, let us introduce the energy integral IE at sensor $\boldsymbol{x}$ and the normalized energy integral $\widetilde{IE}$ with respect to all surface sensors of a given realization

$$IE(\boldsymbol{x}) = \int_0^T \frac{u_E(\boldsymbol{x},t)^2 + u_N(\boldsymbol{x},t)^2 + u_Z(\boldsymbol{x},t)^2}{3} dt \tag{3.17}$$

$$\widetilde{IE}(\boldsymbol{x}) = \frac{IE(\boldsymbol{x})}{\max_{\boldsymbol{x}' \in \partial\Omega_{top}} IE(\boldsymbol{x}')} \tag{3.18}$$

Figure 3.20 displays the envelope and phase GOF as a function of the energy integral. There is a significant positive correlation between the GOF and the energy integral, meaning that the most energetic signals lead to the best predictions. The correlation is stronger for the envelope GOF than the phase GOF. This reflects previous observations that the envelope GOF is hindered by errors on small-amplitude fluctuations. From an engineering

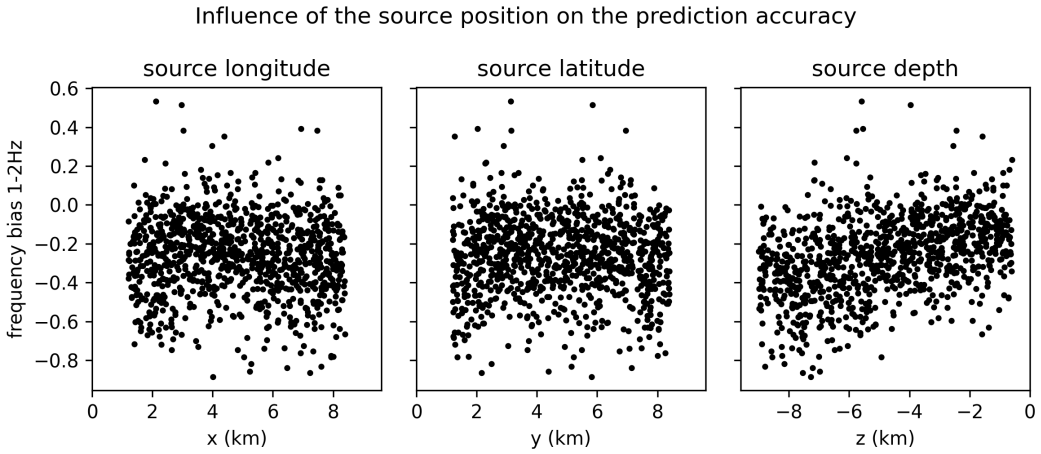


Figure 3.18: For 1000 samples, the 1-2 Hz frequency bias is shown against the source position (x, y, z). Negative (resp. positive) values indicate underestimation (resp. overestimation) of the frequency content.

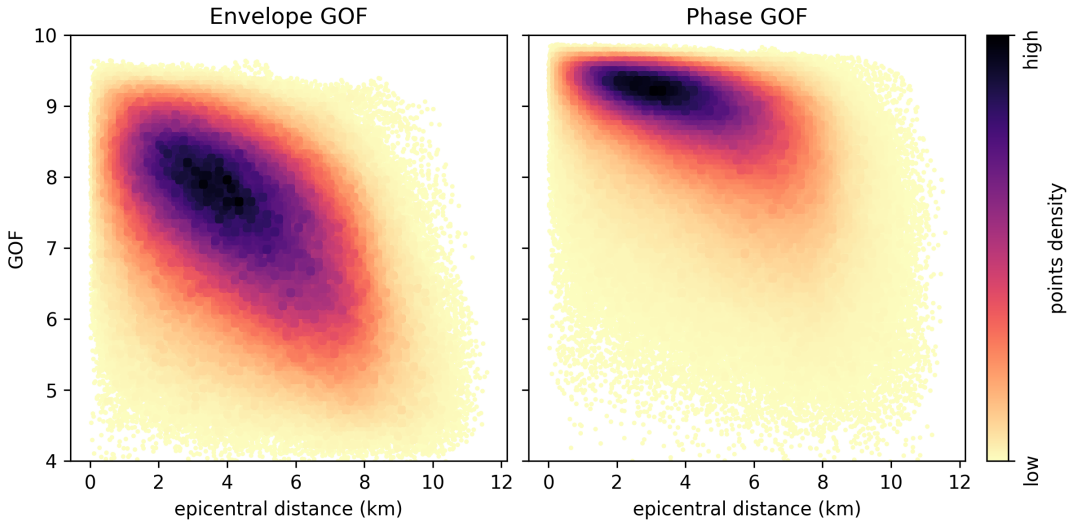


Figure 3.19: Envelope GOF and phase GOF represented for each sensor and 1000 test samples as a function of the epicentral distance. Colors indicate the density of points.

perspective, this association is important as it means that the most damaging ground motions (i.e. with the highest energy) are generally well predicted.

3.5 Generalizability

Neural networks and neural operators are not meant to extrapolate data beyond the training distribution. However, when evaluating predictions on real data, one cannot guarantee that test data are perfectly in-distribution. Although predictions are not expected to be excellent in this situation, it is crucial that they remain reasonable. For the elastic wave equation, this especially means that predicted amplitudes should be of the right order of magnitude and wave arrival times should be close to the reference.

Generalization was firstly assessed on the F-FNO by adding white noise to the geological models (Appendix A.3.4). This shows that predictions are stable and the F-FNO acts as a low-pass filter to remove the influence of non-coherent geological fluctuations. Then, the MIFNO generalization ability is assessed on out-of-distribution sources (Section 3.5.1) and geologies with a real overthrust geological model (Section 3.5.2). We also test the MIFNO ability to perform zero-shot super resolution by increasing the spatial resolution of the prediction (Section 3.5.3).

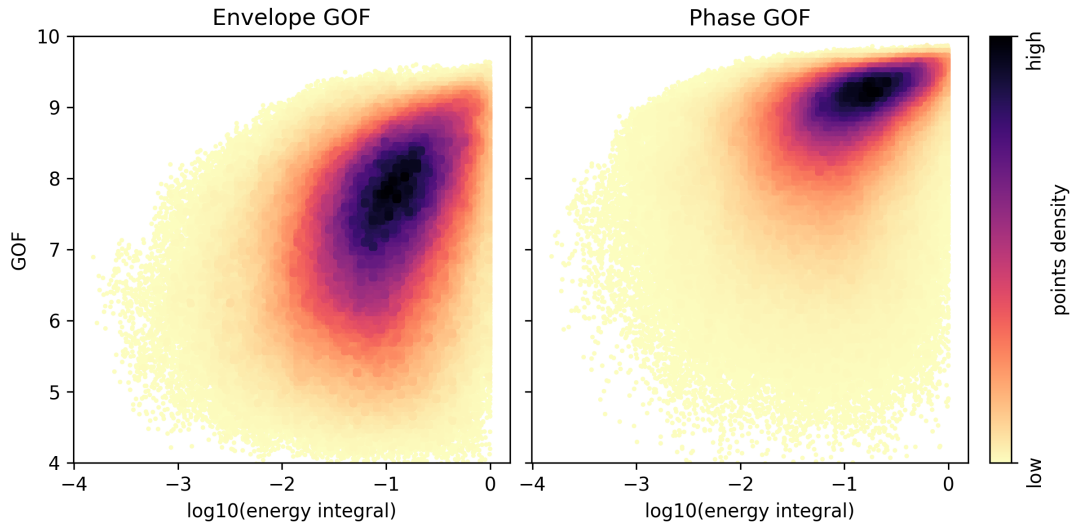


Figure 3.20: Envelope GOF and phase GOF represented for each sensor and 1000 test samples as a function of the normalized energy integral $\widetilde{IE}$ at each sensor. Colors indicate the density of points.

3.5.1 Generalization to out-of-distribution sources

In the HEMEW^S-3D database, the propagation domain extends from 0 km to 9.6 km and the sources are located between 1.2 km and 8.4 km. However, since the source location is provided as a vector of coordinates, the MIFNO is not constrained to sources located inside the domain. Thus, in this section, we explore its ability to predict surface ground motion originating from a source outside the training domain.

To simulate the reference ground motion, we create 1000 new geologies that are twice larger than the original propagation domains and have the same depth. They are all heterogeneous geologies with parameters following the same distributions as the HEMEW^S-3D database (Tab. 2.2). In these large geologies, sources are located randomly within $[-3.6\text{km}; 13.2\text{km}]$ and their orientation is also randomly chosen. The reference ground motion is acquired only at the surface of the original propagation domains, i.e. $[0\text{km}; 9.6\text{km}]$.

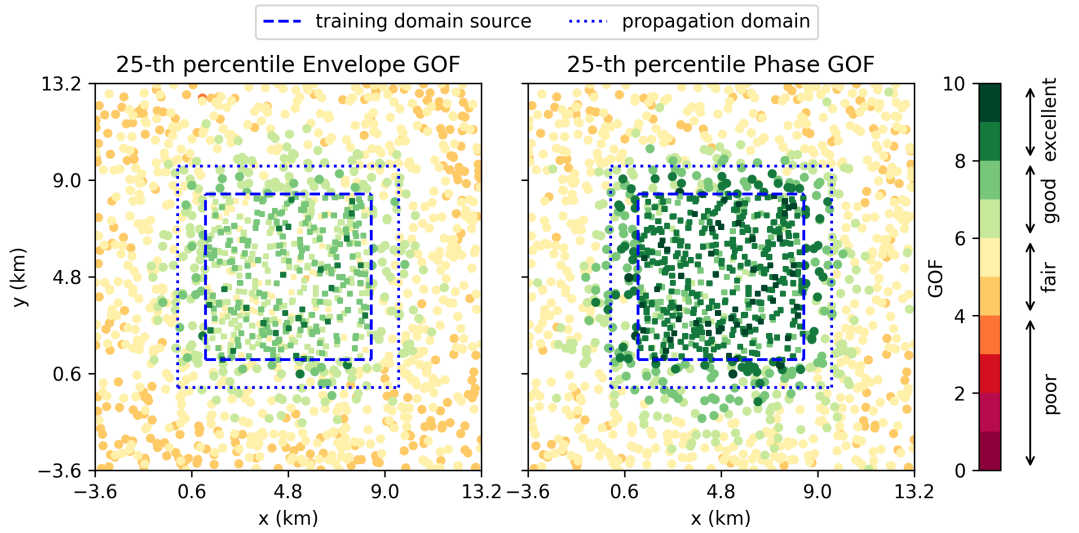


Figure 3.21: 25-th percentile of envelope GOF (left) and phase GOF (right) for 1000 samples where the source is located outside the training domain (dashed square) or even outside the physical domain (dotted square). For reference, 1000 samples with a source inside the training domain are shown.

Figure 3.21 shows that the MIFNO maintains very good predictions when the source is located outside the train-

ing domain, both for the envelope GOF and the phase GOF. When the source remains inside the propagation domain (dotted square in Fig. 3.21), the inter-quartile range is 6.24-7.70 for the envelope GOF and 7.41-8.77 for the phase GOF. These good to excellent results are satisfying for out-of-distribution predictions since they are not too far from the reference results on in-distribution data (Tab. 3.3). When the source locations are farther from the training domain, predictions worsen. This can be seen in Fig. 3.21 for sources out of the training domain and it is perfectly expected since neural operators are not designed for severe extrapolation tasks.

3.5.2 Generalization to out-of-distribution geologies

To test the generalization ability of the MIFNO to out-of-distribution geologies, a 3D overthrust model was adapted to the size of our geological domain (Aminzadeh et al. 1997). This model is a 3D extension of the well-used Marmousi model. It shows a succession of thin layers with contrasting velocities and several folds. It is downscaled to a spatial resolution of 300 m to match the resolution of the HEMEW^S-3D database (Fig. 3.22a). The source is located at (1.2 km, 4.8 km, −9.0 km) with strike = 48°, dip = 45°, rake = 88°.

Figure 3.22 shows that the predicted wavefronts are close to the simulations, in all parts of the domain. In particular, the wave arrival times are accurately predicted and the influence of the source orientation is well preserved since the phases correspond to the reference. The amplitude of the main fluctuations is also accurate. The P wave peaks tend to be overestimated (see t=2.50 s in Fig. 3.22c) but they have a small amplitude on the East-West component. The focus is instead on the S waves, for which the amplitudes agree well with the reference simulation (Fig. A.26a).

The envelope GOFs are good for a large majority of sensors and phase GOFs are very good to excellent (Fig. 3.22b). The GOF distributions are concentrated around the mean of the GOFs obtained on the HEMEW^S-3D database, thereby showing the generalization ability of the MIFNO to real complex geologies. Figure 3.22b also illustrates the large GOF variability between sensors for the same realization (around 2 GOF units). This variability is similar to the HEMEW^S-3D database as shown by the difference between the worst and best sensor in Fig. 3.12. Differences are not randomly distributed but instead follow a coherent spatial organization that is related to the geological heterogeneities. Figure A.25 indeed shows that areas with a GOF lower than 6 are rare and most of the sensors achieved at least a good GOF.

One can however notice the absence of late fluctuations in the predictions (mainly visible for low amplitudes around t=3 s). This was expected from our previous analyses since these ground motion fluctuations are created by the multiple wave reflections and refractions on the thin geological layers and they are difficult to predict. These differences are exacerbated when the simulations are run with a high resolution geological model (50 m) instead of using the same resolution as the prediction (300 m). Figure A.26 indeed shows that ground motions obtained through a high-resolution geology contain more complex small-scale patterns.

3.5.3 Generalization to higher resolution

Although zero-shot super resolution can always be technically performed with FNOs, FNOs are not invariant with respect to the resolution in the general framework (Bartolucci et al. 2023). This means that high-resolution predictions may be less accurate than predictions at the training resolution. Since there is no theoretical result for FNOs including source and PDE parameter, we investigate zero-shot super resolution experimentally with the MIFNO. While the MIFNO was trained with inputs of spatial resolution of $32 \times 32 \times 32$, geologies are interpolated to $64 \times 64 \times 32$ to obtain high-resolution predictions. The vertical dimension is preserved to maintain the same depth-to-time conversion. Since the source is given with absolute coordinates, there is no need to modify the *source branch* when modifying the resolution.

Figure 3.23 shows that it is possible to obtain accurate velocity wavefields with a resolution of 64 that has not been seen during training. The wavefronts are better defined with the increased resolution, which allows to capture peaks more accurately in this example (see snapshots at t=1.40 s in Fig. 3.23 for instance). However, some discrepancies are visible at higher resolution, especially on the boundaries. These artefacts can be explained by the fact that high-resolution predictions are in advance at the edge of the domain, which induces a phase difference between the prediction and the reference (Fig. A.27).

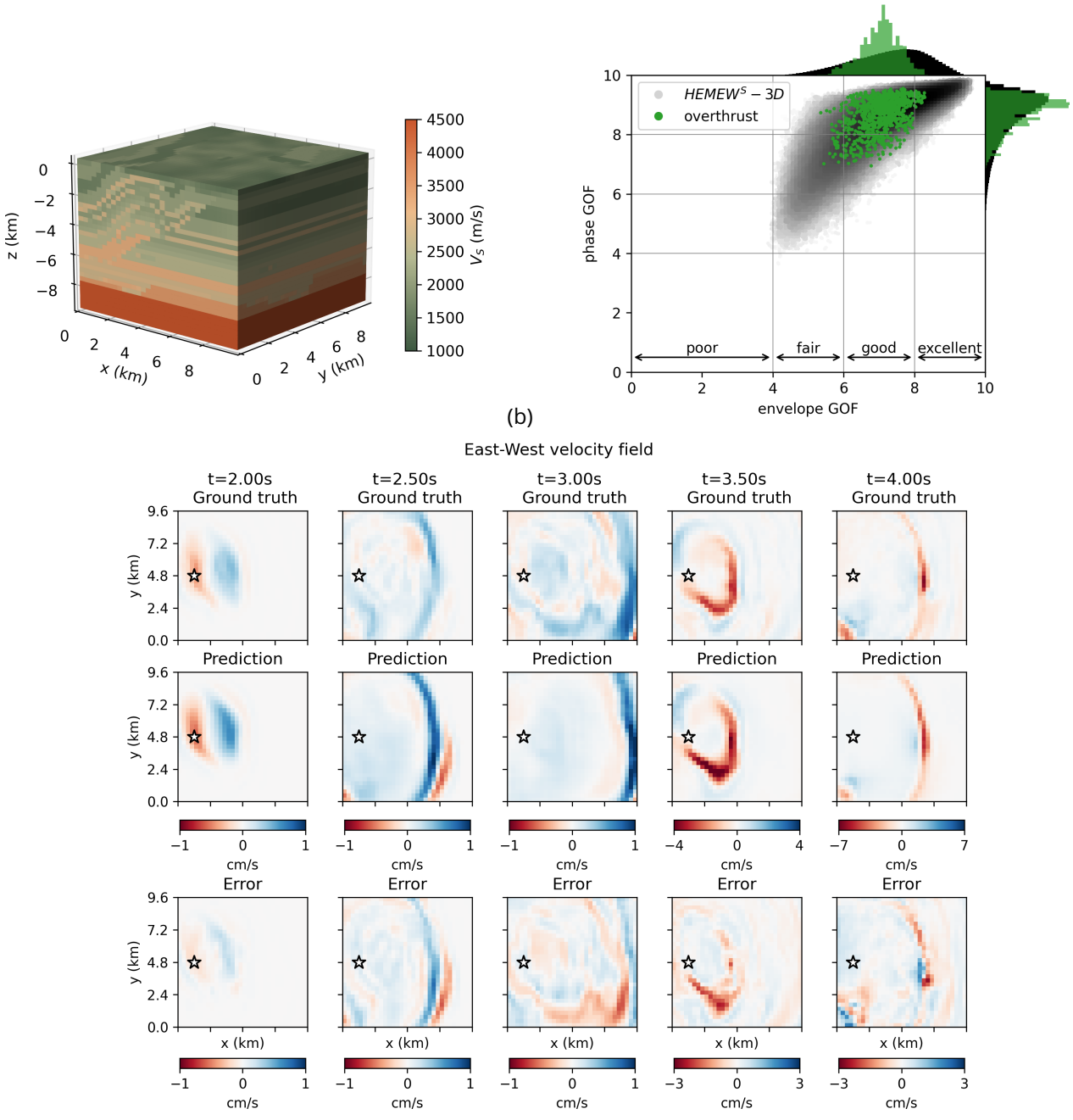


Figure 3.22: (a) Overthrust geology downscaled to 300 m. (b) Distribution of the envelope and phase GOF for the HEMEW^S-3D database (grey, darker colors indicate higher points density) and for each surface sensor predicted with the overthrust geology (green). (c) East-West component of the simulated (upper row) and predicted (middle row) velocity fields at five time instants. The white star denotes the position of the source (depth=−9 km). The error between simulation and prediction is given in the lower row.

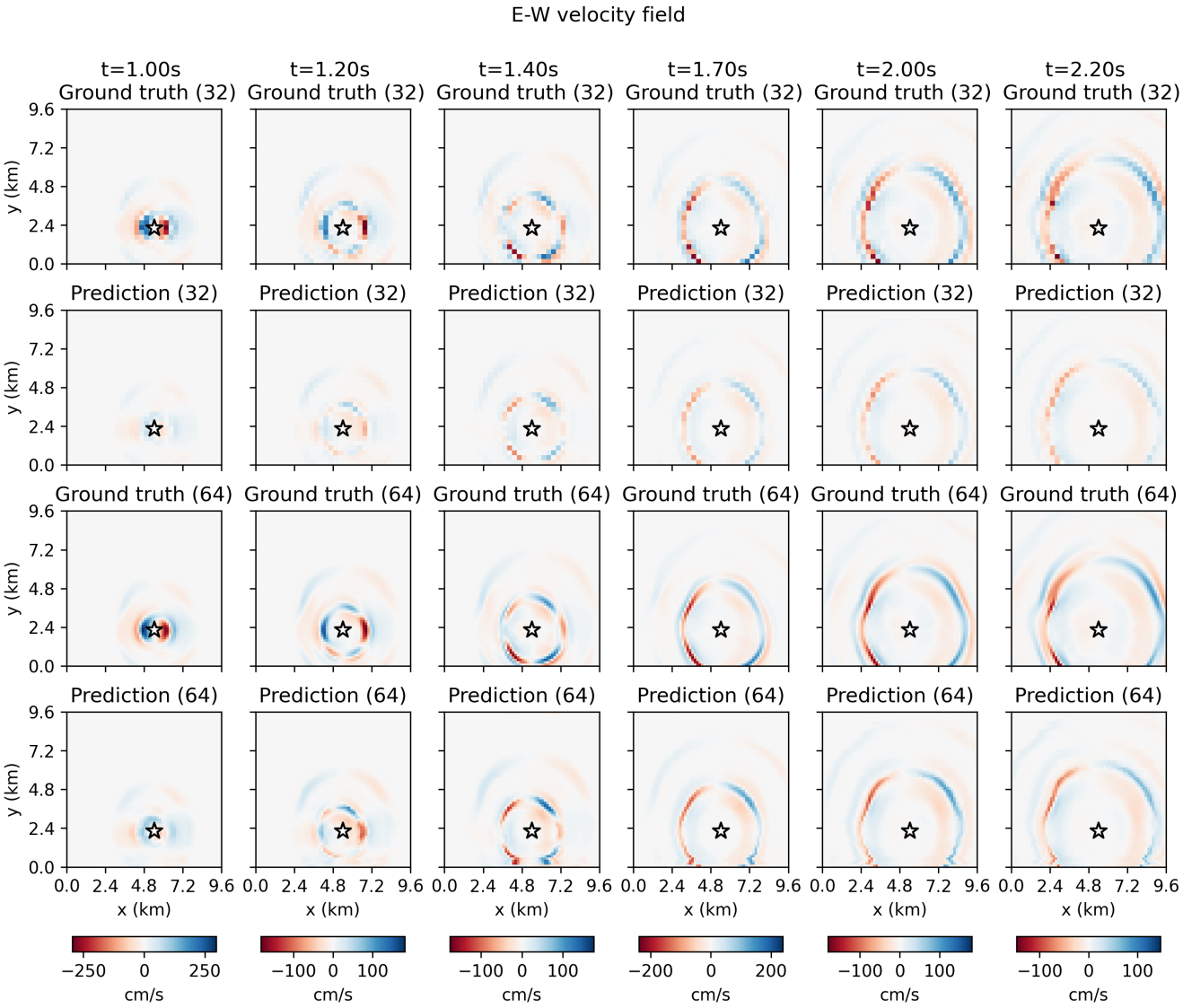


Figure 3.23: East-West component of the velocity wavefield at five time instants. first row: simulation with 32×32 sensors. second row: prediction with 32×32 sensors. third row: simulation with 64×64 sensors. fourth row: prediction with 64×64 sensors. The white star denotes the position of the source (depth=−1.3 km).

3.6 Conclusion

We proposed an implementation of the factorized Fourier Neural Operator (F-FNO) that predicts surface ground motion generated in 3D geologies thanks to a depth-to-time conversion. By focusing on the surface of the domain, i.e. the only location where observations are commonly available, we can provide the solution of a 3D time-dependent PDE without a 4D surrogate model. The factorized Fourier layer is efficient for 3D PDEs as it significantly reduces the number of parameters compared to the baseline FNO. In particular, this allows to obtain a deep and wide F-FNO that does not lead to overfitting, contrary to the FNO that suffers from a large number of parameters.

Building on these findings, we introduced a Multiple-Input Fourier Neural Operator (MIFNO) that takes as input a 3D geological domain and a vector describing the source position and orientation. The structured geology is processed with factorized Fourier layers, and the source parameters are transformed via convolutional layers while maintaining resolution invariance properties. The MIFNO was trained on 30 000 earthquake simulations from the HEMEW^S-3D database covering a large variety of heterogeneous geological models, source positions, and source orientations. It predicts 3-component surface ground motion between 0 s and 6.4 s for frequencies up to 5 Hz.

The MIFNO performs on par with the reference F-FNO model, as was verified by restraining the MIFNO training task to a dataset with i) a fixed source and ii) a randomly located source with a fixed orientation. This shows that the flexibility offered by our proposed architecture preserves the accuracy of the F-FNO.

MIFNO predictions are considered good to excellent for most sensors, following the common understanding of Goodness-Of-Fit (GOF) criteria. GOFs are better for the phase than for the envelope, which indicates that the MIFNO is well suited to predict wave arrival times and the spatial propagation of wavefronts. With 87 % of predictions having a good envelope GOF, the main peaks are correctly predicted, which is crucial for engineering applications. Complex wave propagation phenomena due to heterogeneities yield small-scale fluctuations with small amplitude that are difficult to predict accurately and negatively impact the envelope GOF. Indeed, large coefficients of variation and small correlation lengths lead to larger prediction errors.

These difficulties are exacerbated by the design choices of the HEMEW^S-3D database that purposely contains geologies with strong heterogeneities (mean coefficient of variation of 0.2). However, such large coefficients of variations are rarely found in reality for large correlation lengths. They are included in the HEMEW^S-3D database to enhance its variability in a Scientific Machine Learning perspective. Therefore, predictions on common 1D geological models augmented by moderate heterogeneities will fall in the most accurate range of the MIFNO predictions.

We found that the MIFNO tends to underestimate the frequency content of the ground motion fluctuations. Geological heterogeneities lead to the dispersion of wave energy over time. Since the low-frequency components are affected only by the equivalent mean geological field, they are not influenced by small heterogeneities. Therefore, the influence of dispersion is visible only on the high-frequency components. Since dispersed ground motion wavefields are more challenging to predict, this explains that the prediction errors are larger for high frequencies. However, GOF scores are positively associated with the signal energy, meaning that the most energetic ground motions are generally well predicted, which is crucial for engineering applications.

Using a more complex MIFNO (i.e. with more layers and more channels) is a privileged approach to increase the accuracy, as illustrated by the difference between the 8-layer and 16-layer MIFNO. Investigations on the F-FNO also showed that this architecture scales well when the number of parameters increases and we hypothesize that similar results hold true for the MIFNO. As a consequence, provided that one can obtain enough training data and bear the cost of training deeper models, the MIFNO would become more accurate.

The MIFNO is also able to generalize well to different out-of-distribution data. The source was located out of the training domain with only a moderate degradation of the accuracy. Although it could also be placed outside the propagation domain, this large distribution shift worsened the predictions. In addition, the MIFNO can be applied to geologies that have a different resolution than the training resolution. Although there is no guarantee that the accuracy will systematically be maintained at higher resolutions, the illustrated example showed that finer predictions improve the predictions of peaks amplitudes. These results show the robustness of the MIFNO when applied on test data, which is crucial to move towards real-case applications.

Generalization to geologies that are far from the training database is challenging but the MIFNO predicts accurate ground motion for an overthrust geology with thin folded layers. The GOF distributions with the overthrust geology follow the distributions on the HEMEW^S-3D database. These good to excellent GOFs illustrate the good generalization ability of the MIFNO for real complex geologies. However, the thin layers in the overthrust geology create dispersed wavefields, which are challenging to predict with a generic surrogate model.

When aiming for more accurate predictions in complex settings, transfer learning should be used to specialize the MIFNO on a smaller database tailored to the region under study. Once duly trained, the MIFNO produces quasi real-time evaluations of surface ground motion, thereby paving the way for 3D uncertainty quantification analysis,

design optimization, and inverse problems that are beyond reach with traditional methods.

Indeed, the MIFNO provides 6.4 s of ground motion in around 10 ms (with one GPU) while numerical simulations take around 1 day (1 CPU equivalent). The energy consumption of the MIFNO should be distinguished between training and inference. The training phase amounts to 22 kWh and thanks to the speed of the MIFNO, the inference phase represents only 52 Wh to obtain $N_{MC} = 10^5$ predictions (Tab. 3.6).

Task	Number of realizations	Energy consumption (MWh)
MIFNO training	$N_{train} = 30000$	0.022
MIFNO prediction	$N_{MC} = 10^5$	5.2×10^{-5}

Table 3.6: Energy consumption of the MIFNO training and prediction. The energy consumption of 1 GPU is estimated to 187 Wh. Training the MIFNO on 4 A100 GPUs lasts for 29.6 h. One MIFNO prediction takes around 10 ms on 1 A100 GPU.

Chapter 4

Application to the Le Teil earthquake: transfer learning and uncertainty quantification

Seismic hazard analyses are subject to many sources of uncertainty, ranging from fault rupture to site effects and geological properties, not withstanding uncertainties in the models providing ground motions. To provide reliable estimates of seismic hazard, it is crucial to study how these uncertainties propagate to ground motion intensity. Seismic hazard analysis can be classified in two broad categories, namely deterministic (or scenario-based) seismic hazard analysis and Probabilistic Seismic Hazard Analysis (PSHA). Although formalizing a full PSHA framework is out of the scope of this work, we explore how our surrogate model can be used in some of the PSHA components.

Seismic hazard aims at determining the rate of exceedance of an Intensity Measure (IM), i.e. $\lambda(IM > im)$ where im is a user-defined threshold. IMs can be seen as quantities of interest computed from ground motion $\boldsymbol{u}$ through $IM = h(\boldsymbol{u})$, where h is a deterministic function. Common IMs are the Peak Ground Velocity (PGV) and Pseudo-Spectral Acceleration (PSA) at specific periods. PSHA computes the rate of exceedance as

$$\lambda(IM > im) = \sum_{rup \in \mathcal{R}} \mathbb{P}(IM > im | rup, m) \lambda(rup) \tag{4.1}$$

where $\mathcal{R}$ is the set of possible fault ruptures generating earthquakes in the region under study, $\mathbb{P}(IM > im|rup, m)$ is the probability that the IM exceeds the threshold im when the earthquake is generated by rupture rup and propagates with model parameters m (m includes the geological properties, the ground motion model parameters, and the site condition), and $\lambda(rup)$ is the rate of occurrence of rupture rup within a specific time window.

We choose as a study region the surroundings of the Le Teil earthquake. Seismological and geological investigations provide reference geological models for this region and we adapt the 1D model of Causse et al. 2021 to the design of the HEMEW^{S}-3D database. This model is complemented by 3D geological heterogeneities to provide a *prior distribution* of geological properties $p_a(a)$. The fault ruptures are described by the possible source locations $\boldsymbol{x}_s$ on the La Rouvière fault (Vallage et al. 2021) and the source orientations $\boldsymbol{\theta}_s$ compatible with the seismological context (Delouis et al. 2021). Seismotectonics knowledge therefore yields a prior distribution of the source properties $p_s(\boldsymbol{x}_s, \boldsymbol{\theta}_s)$ (Section 4.1).

For any geological model $a \sim p_a$ and any source $(\boldsymbol{x}_s, \boldsymbol{\theta}_s) \sim p_s$, our surrogate model G_ϕ trained with data $\mathcal{D}$ gives an estimation of ground motion at location $\boldsymbol{x}$

$$\hat{\boldsymbol{u}}(\boldsymbol{x}) = G_\phi(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s | \mathcal{D})(\boldsymbol{x}) \tag{4.2}$$

One non-negligible source of uncertainty is the prediction error of the surrogate model, namely $e(\boldsymbol{x}) := \hat{\boldsymbol{u}}(\boldsymbol{x}) - u(\boldsymbol{x})$ where $\boldsymbol{u}$ is assumed to be the error-free ground motion time series $\boldsymbol{u} = \mathcal{G}(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s)$. In reality, $\boldsymbol{u}$ suffers from measurements error if acquired from recordings, or from numerical errors if obtained with physics-based simulations. As a first approximation, these uncertainties are neglected in this work.

Chapter 3 showed that the MIFNO is an accurate surrogate model of ground motion propagation but prediction errors can be significant in some situations. One efficient method to reduce the prediction error of deep learning-based surrogate models is to train them on data that are specific to the target task. In our framework, this pertains

to geological models and source properties specific to the Le Teil earthquake. Transfer learning is widely used in the machine learning community to transfer knowledge i) from a specific task to a different task, or ii) from a generic task to a specific task (Quiñonero-Candela 2009; Pan and Q. Yang 2010; Weiss et al. 2016). This work takes advantage of the latter framework. For a regression problem like the one we are concerned about, a task should be understood as the prediction of an output $\boldsymbol{u}$ given inputs $(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s)$ with training data $\mathcal{D} = \{a^i, \boldsymbol{x}_s^i, \boldsymbol{\theta}_s^i, \boldsymbol{u}^i\}_{1 \leq i \leq N}$. Then, transfer learning means that a surrogate model $G_\phi(\cdot|\mathcal{D}_1)$ trained on the generic dataset $\mathcal{D}_1$ will be further trained (or *fine-tuned*) on the specific dataset $\mathcal{D}_2$. Depending on the distributional shift between $\mathcal{D}_1$ and $\mathcal{D}_2$, and the transfer learning strategies, the fine-tuned surrogate model should be (significantly) more accurate than the generic one.

In the following, Section 4.1 details the ground motion databases designed for the Le Teil earthquake and our transfer learning methodology for the MIFNO. Section 4.2 then presents the ground motion predictions and illustrates the benefits of transfer learning to improve the MIFNO accuracy. Although the MIFNO was trained only with point-wise sources, we show that it predicts the ground motion generated by a kinematic fault rupture with excellent accuracy. With this improved surrogate model, we conduct uncertainty analyses on several intensity measures (Section 4.3). The influence of geological heterogeneities on PGV is firstly assessed (Sections 4.3.1). This work was conducted with the F-FNO fine-tuned with a fixed source and was published as a conference paper for the 2023 TINCE conference (Technological Innovations in Nuclear Civil Engineering) with the associated special issue F. Lehmann et al. (n.d.). "Quantifying Uncertainties in Seismic Hazard Analysis with a Fourier Neural Operator". In: *European Journal of Environmental Civil Engineering* (under review) ()

Excepting this Section 4.3.1, all the other results were obtained with the MIFNO. We focus on the PSA and perform sensitivity analyses (Section 4.3.2) to highlight the influence of several parameters. We show that our surrogate model allows to constrain the source properties in the presence of uncertainties, based on PSA *observations* (Section 4.3.3). We finally take advantage of the speed of the MIFNO predictions to estimate confidence intervals of the high PSA percentiles (Section 4.3.4).

4.1 Transfer learning: data and methods

4.1.1 Le Teil geological database

Seismological studies based on inversion methods provide a reference 1D model of V_P and V_S in the vicinity of the Le Teil earthquake (Causse et al. 2021). After some slight adaptations to the design of the HEMEW^S-3D database, the 1D model contains six horizontal layers and the S-wave velocities described in Table 4.1. The ratio of P- and S-waves velocities (V_P and V_S) is assumed constant and equal to 1.7 Density and attenuation factors are computed from V_P and V_S (equation 2.7). This model serves as the *reference geology* $\overline{a}$ in our Le Teil database.

A set of 4000 heterogeneous geologies was then derived by superposing log-normal random fields δ to the *reference geology*

$$a(\boldsymbol{x}) = \overline{a}(\boldsymbol{x}) \cdot \delta(\boldsymbol{x}), \quad \text{with } \delta(\boldsymbol{x}) \sim \mathcal{LN}(0, k_{\sigma,\ell_x,\ell_y,\ell_z}(\boldsymbol{x}, \boldsymbol{x'})) \tag{4.3}$$

where k is the von Karman correlation kernel depending on the coefficient of variation σ and correlation lengths ℓ_x, ℓ_y, ℓ_z (Section 2.2.2). It should be noted that random fields are sampled independently inside each of the five upper layers. To enable an easier interpretation of the results, parameters $\sigma, \ell_x, \ell_y, \ell_z$ are the same in all layers and they follow the same distribution as the HEMEW^S-3D database

$$\begin{aligned} p_\sigma &= |\mathcal{N}(0.2, 0.1)| \\ p_{\ell_x} &= \mathcal{U}(\{1.5, 3, 4.5, 6 \text{ km}\}) \\ p_{\ell_y} &= \mathcal{U}(\{1.5, 3, 4.5, 6 \text{ km}\}) \\ p_{\ell_z} &= \mathcal{U}(\{1.5, 3, 4.5, 6 \text{ km}\}) \end{aligned} \tag{4.4}$$

The prior distribution on geologies can then be expressed as

$$p_a = p_{\delta,\sigma,\ell_x,\ell_y,\ell_z} = p_{\delta|\sigma,\ell_x,\ell_y,\ell_z} \cdot p_\sigma \cdot p_{\ell_x} \cdot p_{\ell_y} \cdot p_{\ell_z} \tag{4.5}$$

The Le Teil database finally contains 4000 heterogeneous geologies that are all plausible variations of the *reference geology*. Figure 4.1 illustrates the vertical distribution of V_S .

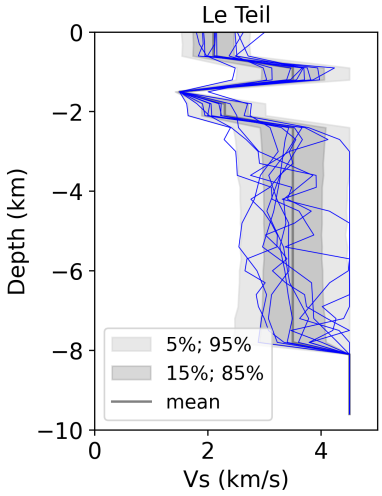


Figure 4.1: V_S distribution in the Le Teil database. For each of the 32 vertical levels, the mean and percentiles are computed over all horizontal points and samples (shaded areas). Blue lines show some individual vertical profiles.

Thickness (m)	V_S (m/s)	V_P (m/s)	ρ (kg/m^3)
600	2100	3570	2329
600	3500	5950	2706
300	1200	2040	1923
600	2300	3910	2380
5700	3500	5950	2706
1800	4500	7650	3170

Table 4.1: Reference 1D geological model for the LeTeil region. Each layer from top to bottom is described by its thickness, S-wave velocity (V_S), P-wave velocity (V_P), and density ρ (adapted from (Smerzini et al. 2023)).

4.1.2 Le Teil source database

Several studies estimated the source parameters of the Le Teil earthquake, both in the form of a fault plane and an equivalent point-source. In our Le Teil database, we choose the source locations close to the fault plane obtained by Vallage et al. 2021. The fault is 7 km-wide, 4 km-high and extends between −780 m and −4210 m. Since the source must be located close to the fault plane, we firstly define the three normal vectors that characterize the fault plane $\overrightarrow{\alpha}$ along strike, $\overrightarrow{\beta}$ along dip and $\overrightarrow{n}$ normal to the fault plane (Fig. 4.2). The source location is then sampled with a Latin Hypercube Sampling (LHS) in the $(\overrightarrow{\alpha}, \overrightarrow{\beta}, \overrightarrow{n})$ frame as three independent variables

$$\begin{aligned} s_\alpha &\sim \mathcal{U}([0, 7\text{ km}]) \\ s_\beta &\sim \mathcal{U}([0, 4\text{ km}]) \\ s_n &\sim \mathcal{U}([-0.1, 0.1\text{ km}]) \end{aligned} \tag{4.6}$$

The absolute source position is then obtained as

$$\begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} = s_\alpha \overrightarrow{\alpha} + s_\beta \overrightarrow{\beta} + s_n \overrightarrow{n} \tag{4.7}$$

Each point source is assigned a random moment tensor, whose angles (ϕ, δ, λ) are chosen from the moment tensor inversion conducted by Delouis et al. 2021 and sampled with a LHS

$$\begin{aligned} \phi &\sim \mathcal{U}([30°, 70°]) \\ \delta &\sim \mathcal{U}([20°, 70°]) \\ \lambda &\sim \mathcal{U}([70°, 120°]) \end{aligned} \tag{4.8}$$

Since all source parameters are chosen independently, the source probability distribution can be written

$$p_s = p_{s_\alpha} \cdot p_{s_\beta} \cdot p_{s_\gamma} \cdot p_\phi \cdot p_\delta \cdot p_\lambda \tag{4.9}$$

Finally, we follow the same methodology as the HEMEW^S-3D database to generate 6.4 s of ground motion inside each geological domain. Ground motion is synthesized at the surface by 32 × 32 sensors. Within this framework, we obtain 4000 samples $(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s, \boldsymbol{u})$ tailored to the Le Teil earthquake.

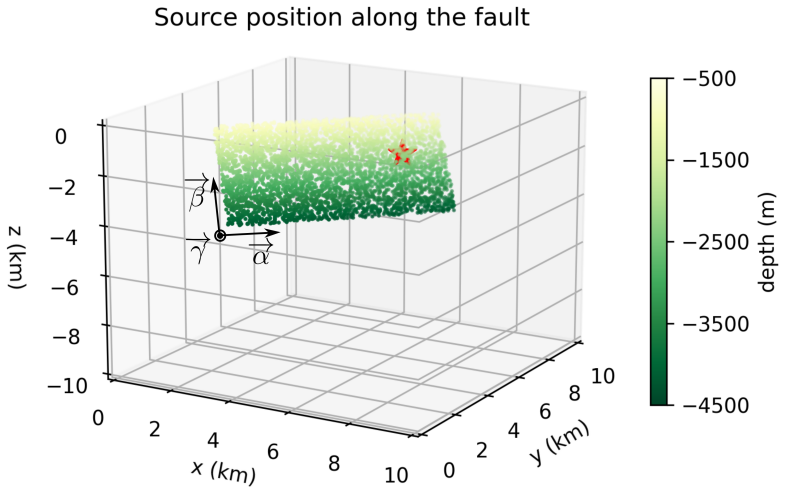


Figure 4.2: Position of the 4000 sources in the Le Teil^S database, colored by their depth. The red star denotes the source that generated the ground motion in Fig. 4.3

4.1.3 Transfer learning

Among the 4000 pairs of geologies and velocity wavefields in the Le Teil database, up to 3000 were used for training, 300 for validation, and 700 for testing. With transfer learning, the neural operator's weights are initialized with the weights of the MIFNO trained on the HEMEW^S-3D database instead of starting from a random initialization. This way, the model starts from a solution where it is already able to predict the propagation of seismic waves in generic geologies.

Then, the model is trained solely on samples from the Le Teil database, forcing it to specialize to the Le Teil context. The number of samples used for the second training is called the number of *transfer learning samples* and is denoted N_{TL}. We used between N_{TL}=100 and N_{TL}=3000 transfer learning samples.

Most of the following results were obtained with the 16-layer MIFNO described in Section 3.2.4. The MIFNO was fine-tuned with one A100 GPU, a learning rate of 1×10^{-4}, a batch size of 16 and an early stopping criterion when the validation loss does not improve. Transfer learning with 100 samples took 0.5 h for 150 epochs (i.e. 11 s per epoch) and 11 h with 3000 samples for 200 epochs (i.e. 198 s per epoch). We also tried to freeze the upper layers of the MIFNO but it degraded the MIFNO accuracy on training and test data. Therefore, all weights are updated during transfer learning.

Only when focusing on the influence of geological heterogeneities on ground motion with a fixed source (Section 4.3.1), a 20-layer F-FNO with 16 channels was trained on the HEMEW-3D database and then fine-tuned with a database of Le Teil geologies and ground motion generated from the fixed source of the HEMEW-3D database.

4.2 Prediction results

4.2.1 Ground motion prediction illustrations

Figure 4.3 illustrates the velocity time series obtained for a test sample in the Le Teil database. Predictions are very close to the simulated ground truth for all wave arrivals and amplitudes. One can especially notice the numerous high fluctuations that are very well captured at all sensors. The MIFNO is also able to reproduce the late small-scale fluctuations (between 3 s and 5 s, Fig. 4.3) thanks to transfer learning. Phase GOFs are excellent or close to excellent for all sensors, while envelope GOFs are excellent for the sensors with the largest fluctuations (two bottom rows in Fig. 4.3) and good for the other sensors. These metrics confirm the excellent visual agreement.

The temporal evolution of the wavefield can also be observed on the snapshots (Fig. 4.4). The MIFNO predictions are very close to the simulations and the spatial propagation of the wavefronts is very well reproduced. As already mentioned for the generic MIFNO, the late fluctuations are less localized than the reference and appear blurry (see

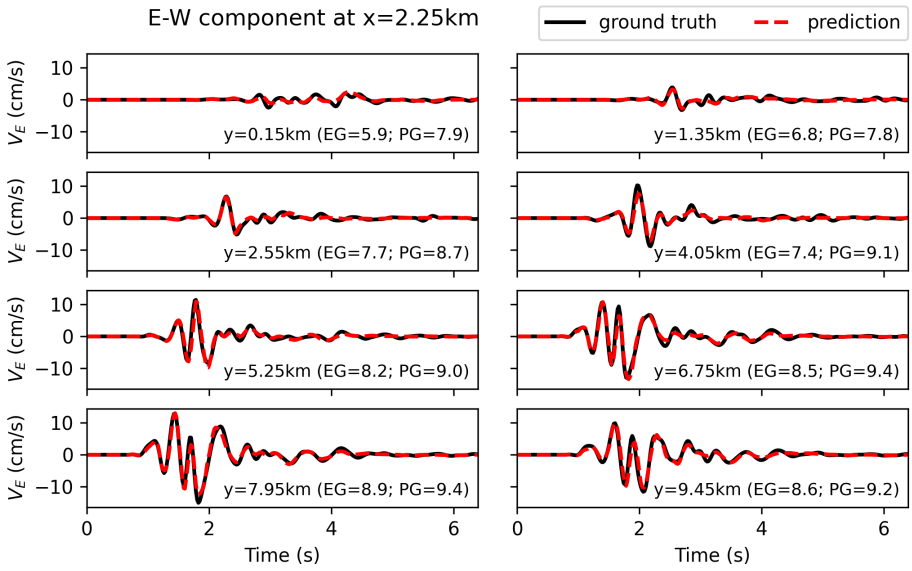


Figure 4.3: East-West velocity time series at 8 sensors aligned at x=2.25 km for a test sample. Predictions (red dashed lines) were obtained with the MIFNO fine-tuned on 2000 samples. The reference ground motion is given by the numerical simulation (black lines). For each sensor, its position along the y axis is indicated, and the envelope GOF (EG) and phase GOF (PG) are also reported. The source is located in the upper part of the fault plane (5.71 km, 6.94 km, −2.15 km) and its orientation is defined as strike=68.6°, dip=43.9°, rake=102.7°. The corresponding geology has a 5.2 % coefficient of variation with correlation lengths of 1.5 km, 6.0 km, and 3.0 km in directions x, y, and z.

the bottom left corner at t=3.0 s). However, one should also notice the amplitude scale which is almost ten times smaller than the wavefront at t=1.0 s. Therefore, the late errors are much less significant for seismic hazard analyses.

4.2.2 Improved accuracy with transfer learning

Figure A.28 compares the evolution of the loss function for different numbers of transfer learning samples. It shows that the validation loss always converges and errors decrease when using more data. For comparison purposes, the MIFNO trained only with 3000 samples from the Le Teil database (i.e. without pretraining) is also given (black dashed line in Fig. A.28). One can observe that its loss evolution is more chaotic and reaches a plateau higher than the loss with transfer learning (except for N_{TL}=100). This exemplifies the benefits of transfer learning to improve the accuracy of the MIFNO.

Quantitatively, 700 predictions in the test dataset were assessed with the relative RMSE, the frequency biases and the GOFs. When predicting ground motion directly with the generic MIFNO (i.e. without transfer learning, line N_{TL}=0 in Tab. 4.2), metrics are in line with the general results on the HEMEW^S-3D database (Tab. 3.3), although slightly better due to the absence of deep sources in the Le Teil database. With only 100 transfer learning samples, the GOFs already improve by 1 to 2 units, illustrating the considerable benefits of transfer learning, even with a limited number of data. After even a light transfer learning (N_{TL}=100), 68 % predictions reach a good envelope GOF and 61 % an excellent phase GOF. When training the MIFNO solely on 3000 samples from the Le Teil database (line N_{TL}=3000 in Tab. 4.2), the predictions accuracy was worse than the predictions obtained with 250 transfer learning samples. This shows the major benefits of initializing the weights with the pre-trained MIFNO.

Table 4.2 also shows that the metrics continuously improve when the number of transfer learning samples increases. The first quartile of frequency biases improves by 23 % between N_{TL}=100 and N_{TL}=3000. The GOFs also benefit from larger transfer learning databases and the improvement is mainly visible on the first quartile, meaning that good predictions with N_{TL}=100 tend towards excellent with N_{TL}=3000 This is illustrated by comparing the MIFNO predictions obtained with 2000 transfer learning samples (Fig. 4.3) and 500 samples (Fig. A.29). While the main fluctuations are already accurate with 500 transfer learning samples, increasing the number of samples allows to better capture the late fluctuations with a small amplitude.

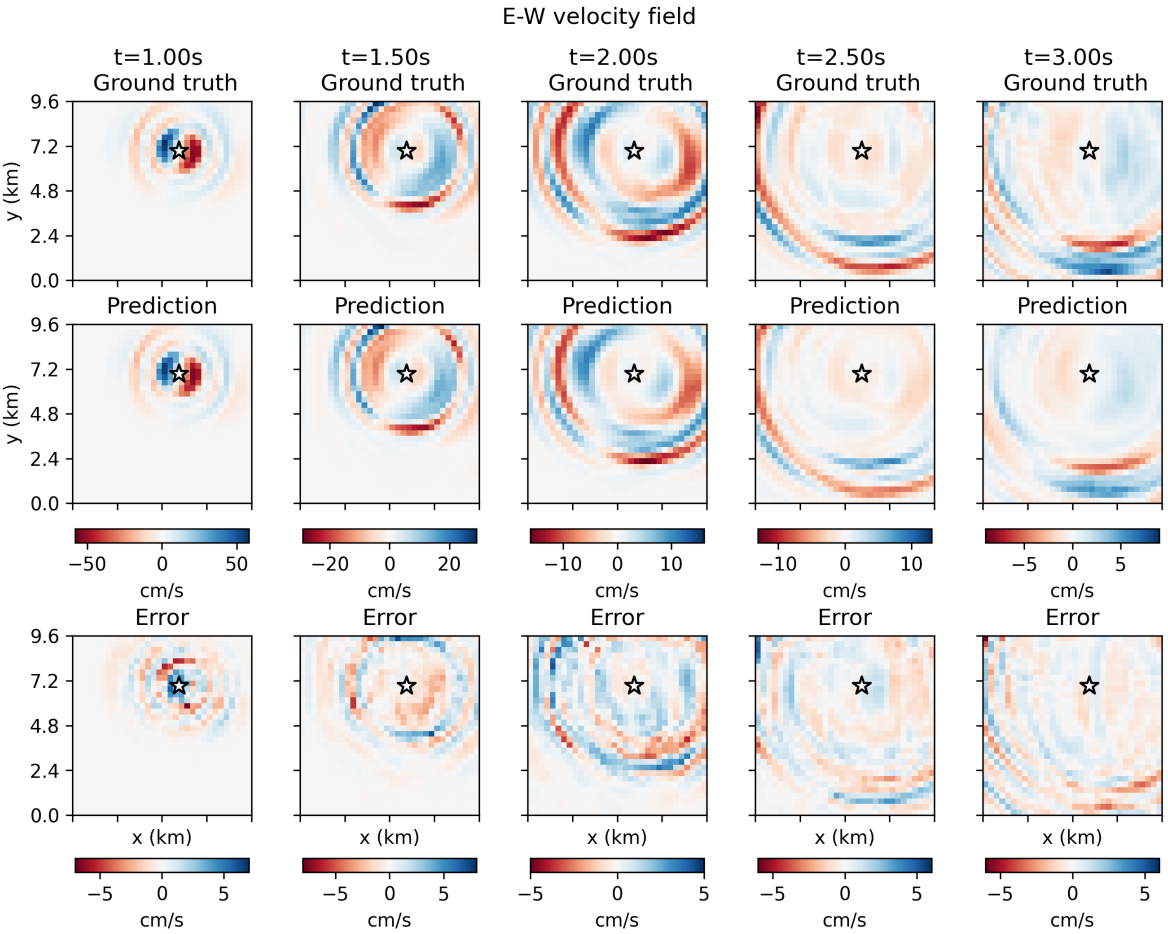


Figure 4.4: East-West component of the simulated (upper row) and predicted (center row) velocity fields for a geology in the Le Teil database. The error between the prediction and the simulation is shown in the lower row. The white star indicates the epicenter.

4.2.3 Variability with the choice of the transfer learning samples

Since the full MIFNO is fine-tuned with only a limited number of transfer learning samples, it is important to ensure that the results do not depend on the sample choice. For this purpose, the MIFNO was trained with six non-overlapping sets of 100 to 500 samples and predictions were assessed on a fixed set of 700 test samples. Figure A.30 indicates a moderate inter-set variability, slightly higher for the envelope GOF than the phase GOF. Both the mean and extreme GOFs improve when the number of transfer learning samples increases, showing that the better accuracy observed for larger transfer learning datasets in Tab. 4.2 is not due to some favorable sample choice.

Figure 4.5 additionally shows that increasing the number of transfer learning samples does not compensate a limited pre-training. Indeed, reducing the number of pre-training samples from 27 000 to 18 000 systematically worsened the GOFs (0.6 units reduction for the envelope and 0.4 units for the phase GOF). This reduction was much less pronounced on the whole HEMEW^S-3D database (0.1 GOF units worsening when training the MIFNO with 18 000 samples instead of 27 000). This advocates for the need of a large pre-training database to obtain good generic predictions and the benefits of fine-tuning on a small database to reach excellent accuracy.

4.2.4 Fault modelling

Since the wave equation 1.2 is linear with respect to the source, it is well-known that the ground motion generated by an extended fault can be computed as the sum of the ground motion generated individually by each point source on the fault. Hence, the dynamic rupture of the fault determined by Vallage et al. 2021 was represented by 16 660 point sources with different source time functions describing the slip of each fault patch (Figure 4.6).

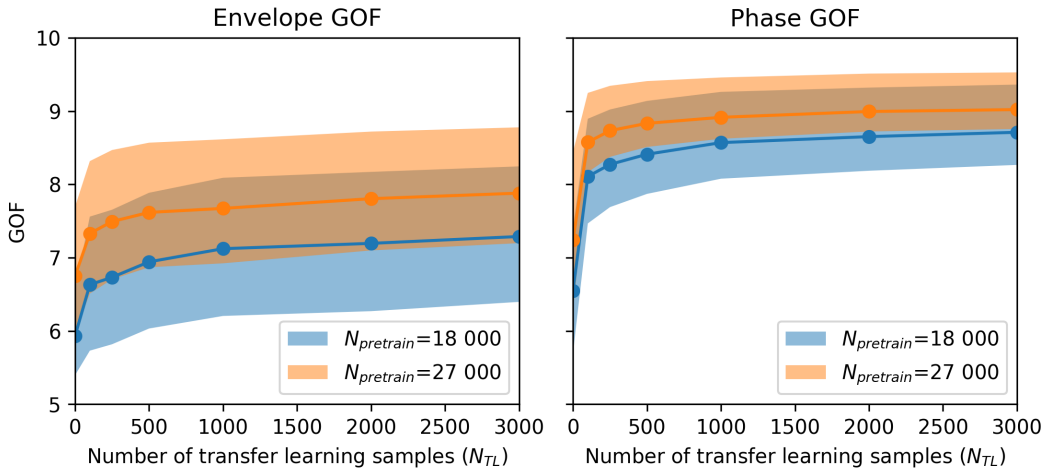


Figure 4.5: Envelope and phase GOF computed for different number of transfer learning samples when the MIFNO was pre-trained with 18 000 (blue) and 27 000 (orange) HEMEW^S-3D samples. The solid line indicates the mean GOF and the shaded area extends from the first to the third quartile.

In the HEMEW^S-3D database, all point sources have the same source time function (a 10 Hz spice bench, Fig. 2.2). To correctly model the fault rupture, one needs to predict the ground motion generated from the source time functions at different points (Fig. 4.6b). Following the methodology of Section 1.1.3, this can be done easily by i) deconvolving the ground motion response by the spice bench source time function (this gives the approximated response to a Dirac pulse) and ii) convolving the impulse response with the target source time function. In summary, the ground motion generated by the fault was predicted following the steps below:

1. select one geology in the Le Teil database
2. for each point on the fault plane, predict the surface wavefields with the fine-tuned MIFNO
3. deconvolve all surface wavefields by the spice bench source time function
4. convolve each surface wavefield by the source time function corresponding to its associated point source
5. sum all surface wavefields

With this setting, the geology is chosen with a 0.11 coefficient of variation, correlation length of 4.5 km along x, 3 km along y, 1.5 km along z. The fault is characterized by a 45° strike, 108° rake, and 60° dip (Fig. 4.6a).

Figure 4.7 illustrates the spatial propagation of waves originating from the extended fault. One can first observe that predictions agree very well with the reference simulation since the wave fronts are accurately reproduced, both for their arrival times and their amplitude. Despite the first wave arrivals being more complex than the pulse-like ground motion generated by individual point sources, Fig. 4.8 shows that their temporal evolution is very accurate. Thanks to the contribution of the different point sources, detailed ground motion features are also correctly captured by the predictions. Time series in Fig. 4.8 and snapshots around 1.8 s in Fig. 4.7 indeed show that small-amplitude fluctuations are well reproduced by the MIFNO. This visual agreement is confirmed by the GOF scores since phase GOFs are excellent at all spatial locations and envelope GOFs are excellent for 85.6 % of points.

4.3 Uncertainty propagation

Section 4.3.1 focuses on the influence of geological heterogeneities on PGV and PSA. These results were obtained with the F-FNO fine-tuned on the Le Teil database (i.e. with a fixed source). In the following sections, all results are provided by the MIFNO.

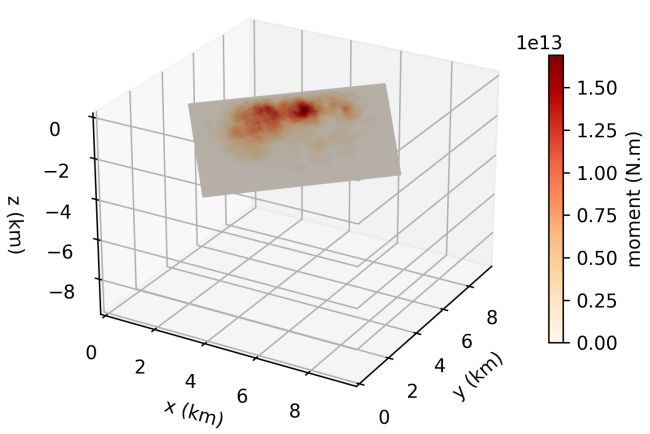
4.3.1 Influence of geological heterogeneities on Peak Ground Velocity (PGV)

Several intensity measures are of interest for engineering applications. Therefore, one should ensure that the F-FNO can provide accurate estimates of the different intensity measures. In this section, we focus on the Peak Ground Velocity (PGV), computed as the maximum value over time of the geometric mean of the two horizontal components.

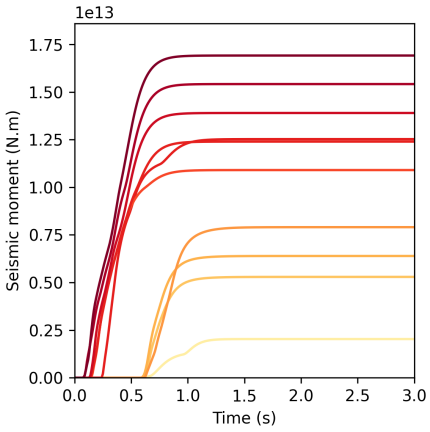
Le Teil database with a random source along the fault plane

# samples	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$	EG	PG
N_{train}=3000	0.40 ; 0.73	-0.49 ; -0.02	-0.60 ; -0.06	-0.67 ; -0.09	6.60 ; 8.43	8.25 ; 9.31
N_{TL}=0	0.64 ; 0.98	-0.44 ; 0.33	-0.64 ; 0.06	-0.68 ; -0.05	5.98 ; 7.69	6.30 ; 8.44
N_{TL}=100	0.41 ; 0.78	-0.44 ; 0.08	-0.56 ; 0.00	-0.66 ; -0.05	6.51 ; 8.32	8.17 ; 9.25
N_{TL}=250	0.38 ; 0.75	-0.41 ; 0.09	-0.52 ; 0.03	-0.61 ; -0.01	6.70 ; 8.47	8.38 ; 9.35
N_{TL}=500	0.37 ; 0.74	-0.38 ; 0.11	-0.49 ; 0.06	-0.58 ; 0.02	6.87 ; 8.57	8.51 ; 9.41
N_{TL}=1000	0.35 ; 0.69	-0.42 ; 0.03	-0.52 ; -0.02	-0.59 ; -0.04	6.92 ; 8.62	8.62 ; 9.46
N_{TL}=2000	0.33 ; 0.68	-0.38 ; 0.05	-0.47 ; 0.01	-0.54 ; -0.01	7.10 ; 8.72	8.72 ; 9.51
N_{TL}=3000	0.33 ; 0.68	-0.34 ; 0.08	-0.43 ; 0.04	-0.51 ; 0.02	7.20 ; 8.78	8.76 ; 9.53

Table 4.2: 1st and 3rd quartiles of the metrics computed on 700 test samples specific to the Le Teil region. (upper row): training with only 3000 specific data. In other experiments, transfer learning was used with 100 to 3000 samples (N_{TL}= number of transfer learning samples). rRMSE: relative RMSE (0 is best), $rFFT_{low}$: relative frequency bias 0-1Hz (0 is best), $rFFT_{mid}$: relative frequency bias 1-2Hz (0 is best), $rFFT_{high}$: relative frequency bias 2-5Hz (0 is best), EG: envelope Goodness-of-Fit (10 is best), PG: phase Goodness-of-Fit (10 is best). For frequency biases, negative values indicate underestimation.



(a) 3D representation of the fault plane



(b) Seismic moment function for some points on the fault plane

Figure 4.6: Kinematic model of the fault that generated the Le Teil earthquake computed by Vallage et al. 2021. Colors indicate the final seismic moment of each point.

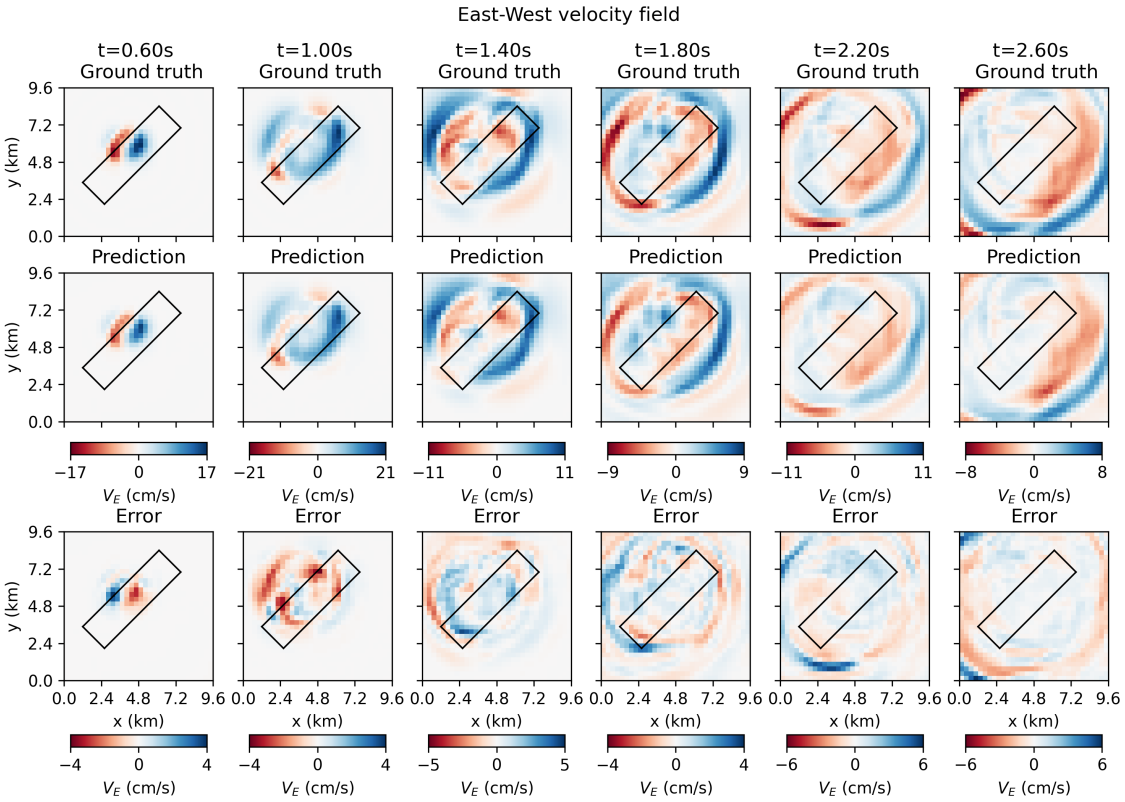


Figure 4.7: East-West component of the simulated (upper row) and predicted (center row) velocity fields for a heterogeneous geology in the Le Teil database. The error between simulation and prediction is given in the lower row. The black rectangle indicates the fault extend.

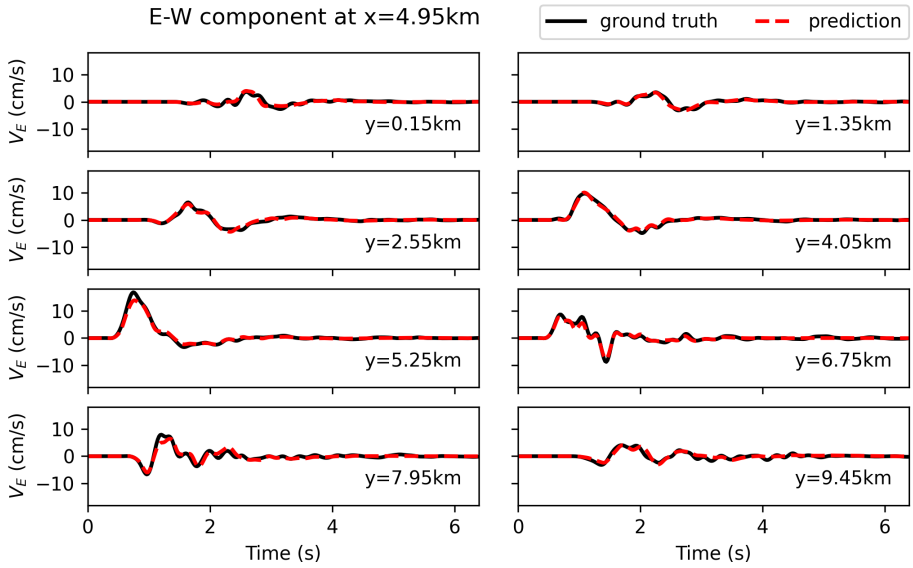
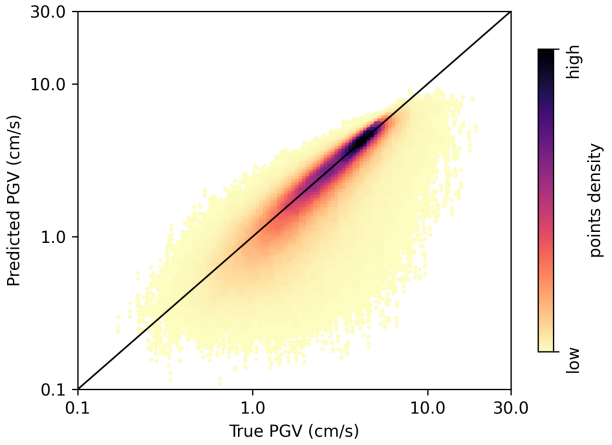
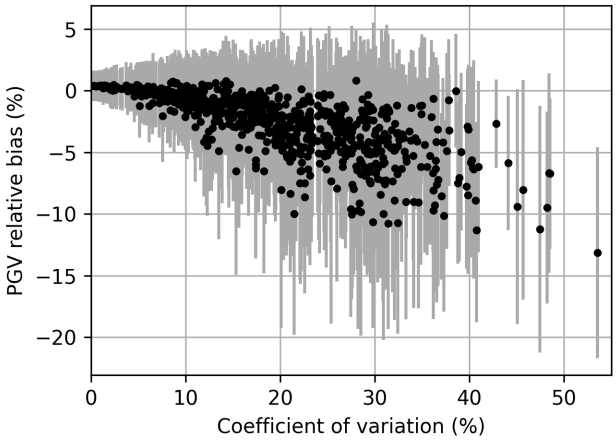


Figure 4.8: East-West velocity time series at 8 sensors aligned at x=2.25 km generated from the extended fault. Predictions (red dashed lines) were obtained with the MIFNO fine-tuned on 2000 samples. The reference ground motion is given by the numerical simulation (black lines). For each sensor, its position along the y axis is indicated.



(a) For each test sample and each sensor point, the predicted PGV is shown against the true PGV. Colors denote the density of points.



(b) Each test sample is represented by the geological coefficient of variation (x axis). For each sample, the PGV bias is given as the mean of all sensors (black dot) and the standard deviation across samples (grey error bar).

Figure 4.9: PGV prediction with the fine-tuned F-FNO

Figure 4.9a first compares the predicted PGV with the simulated PGV for 700 test samples and all spatial points. One can observe that the predictions align well with the ground truth for a large majority of points (dense dark area). This indicates that the F-FNO accurately predicts the peak values, despite not being targeted for this specific quantity of interest. When predictions are inaccurate, they are more often underestimated than overestimated. This is especially true for large PGVs (larger than 10 cm/s) that represent rare events where predictions are very challenging.

Quantitatively, the PGV accuracy can be defined as the relative error between the prediction and the simulation. Figure 4.9b then shows the PGV relative bias as a function of the coefficient of variation for the associated geology. One can notice that predictions are excellent when heterogeneities have a low amplitude. Indeed, the PGV relative bias is lower than 2.5 % for coefficients of variation smaller than 5 %. Then, in average, the F-FNO tends to slightly underestimate the PGV. Indeed, the PGV relative bias sensors is between −7.4 % and 3.3 % for coefficients of variation around 10 % (one standard deviation around the mean, Fig. 4.9b). Across all samples, the PGV relative bias is comprised between −17 % and 14 % for coefficients of variation lower than 10 %. When heterogeneities have large amplitudes (coefficients of variation larger than 20 %), the underestimation generally worsens. Thus, the F-FNO provides accurate PGV estimates for low to moderate coefficients of variation that are commonly encountered in practice at our scale of interest (few kilometers). Increasing the number of transfer learning samples improves the PGV predictions.

4.3.2 Sensitivity analysis

In the following, we analyze the relationships between the geological/source parameters and the PSA. To ease the interpretation, the analysis is restricted to six receivers R_0, $R_1, \cdots, R_5$ with various azimuths from the fault (Fig. 4.10).

Global sensitivity analysis amounts to quantifying the influence of individual parameters $X_1, \cdots, X_q$ on a model outcome $Y \in \mathbb{R} = \mathcal{M}(X_1, \cdots, X_q)$. In our framework, parameters $X_1, \cdots, X_q$ are the coefficients of variation σ_k and the correlation lengths in the three directions $\ell_{x,k}$, $\ell_{y,k}$, $\ell_{z,k}$ for each of the 5 layers $k \in \{1, \cdots, 5\}$ in the geological model. Parameters also contain the source position relative to the fault s_α, s_β, s_n and the source orientation ϕ, δ, λ. The outcome Y is the PSA at a given receiver $\boldsymbol{x}_R$ and a given period T. Denoting h the deterministic function that computes the PSA from the velocity time series, the model is $\mathcal{M} := h \circ G_\phi$.

One common method in global sensitivity analysis is the Sobol analysis (Sobol 2001). It provides first-order Sobol' indices S_i that quantify the influence of parameter X_i on the outcome Y via

$$S_i = \frac{Var(\mathbb{E}[Y|X_i])}{Var(Y)} \tag{4.10}$$

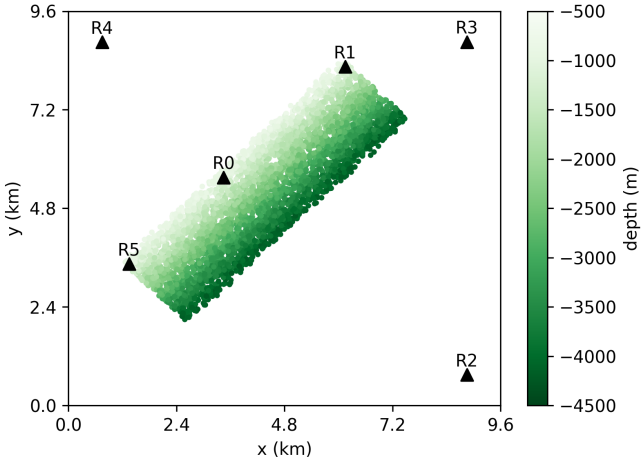


Figure 4.10: Six receivers R_0, R_1, $\cdots$, R_5 and the vertical projection of the fault plane colored by depth.

To account for interactions between parameters, let us introduce

$$\begin{aligned} D_i(Y) &:= Var(\mathbb{E}[Y|X_i]) \\ D_{ij}(Y) &:= Var(\mathbb{E}[Y|X_i, X_j]) - D_i(Y) - D_j(Y), \end{aligned} \tag{4.11}$$

with natural extensions to higher-order interactions. The total indices S_{T_i} can then be computed on all sets of parameters I including i

$$S_{T_i} = \sum_{I \subset \{1,\cdots,q\} | i \in I} S_I \tag{4.12}$$

To compute the expectation in the Sobol' indices, we rely on a Saltelli sampler with 2^{13} Monte Carlo samples (Saltelli 2002). Since our model depends on $q = 26$ parameters, the Sobol analysis requires $2^{13}(2q+2) = 442\,368$ samples. All geological and source parameters are independent and they follow the same distribution as the Le Teil database (equations 4.4, 4.6, 4.8), with the exception of the coefficient of variation that was limited to smaller values, common in the literature (Section 1.2) to avoid large deviations from the *reference geology*. Coefficients of variation σ_k follow a truncated normal distribution between 0 and 0.15 with mean 0.1 and standard deviation 0.04

$$\sigma_k \sim p_{\sigma_k} = \mathcal{N}_{[0,0.15]}(0.1, 0.04) \tag{4.13}$$

From the 442 368 parameter values determined by the Saltelli sampler, we designed the corresponding 3D heterogeneous geological models and predicted the velocity wavefields with the MIFNO. The PSA is finally computed at the six receivers of interest (Fig. 4.10), leading to 442 368 predictions $\times$ 29 PSA periods $\times$ 6 receivers = 76 972 032 values to analyze.

Figure 4.11 shows the first-order and total Sobol indices for the PSA computed at receiver R_3 at all periods between 0.2 s and 3 s. The total indices are the largest for the two parameters s_α and s_β corresponding to the source position on the fault plane. They are also the largest first-order indices and this observation holds for all receivers (Tab A.12). Note that s_β accounts for the source depth and (s_α, s_β) is correlated with the source-to-receiver distance. Therefore, it is perfectly expected that they strongly influence the PSA.

The source orientation (ϕ, δ, λ) has moderate total indices (smaller than 0.1) and the relative influence of each angle varies between the receivers (Tab A.13). The first-order indices of the geological parameters $\sigma_k, \ell_{x,k}, \ell_{y,k}, \ell_{z,k}$ are non-significant but their total indices are of the same order as the total indices of the source orientation. Therefore, it means that heterogeneities influence the PSA with high-order interactions between several parameters (second-order interactions account only for a small portion of the total indices).

It is interesting to notice that indices are generally similar for all periods, excepting at receivers R_2 and R_4 where total indices of geological parameters are significantly larger at small periods (Fig. A.31). Both receivers are located along the fault normal, which may explain that they exhibit a slightly different behaviour than the receivers parallel to the fault plane.

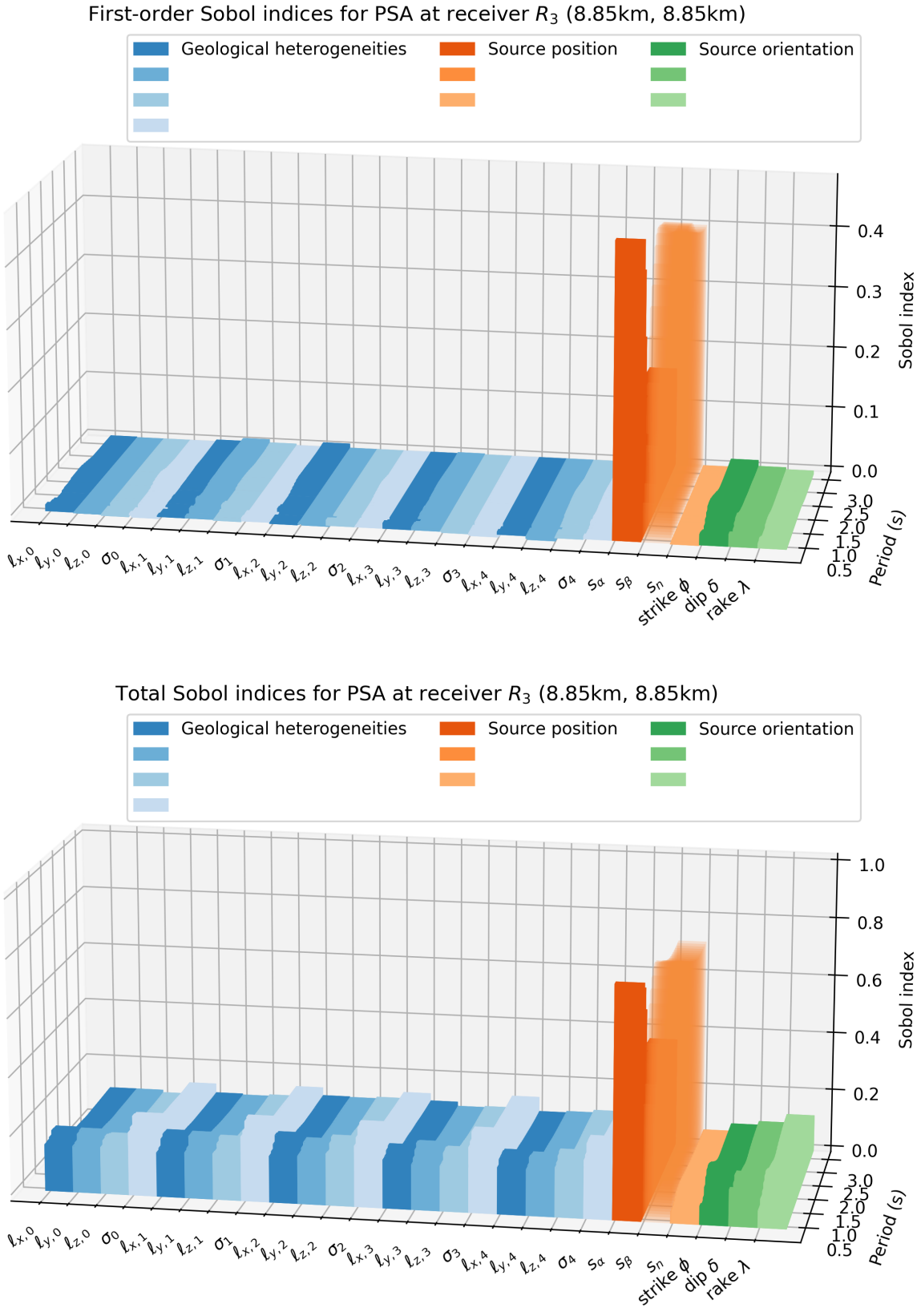


Figure 4.11: First-order (top) and total (bottom) Sobol indices for the PSA on the East-West component computed at receiver R_3. Each of the 26 parameters is reported on the x axis and the y axis denotes the period.

4.3.3 Parameters conditioning from fictive observations

Although all geological models designed in the previous section are plausible realizations in the Le Teil region, they may contradict the real geology, for instance if high-velocity heterogeneities have been located in an area with low-velocity fluctuations. To better inform the PSA predictions, we want to focus on *admissible* geologies that approach the real geology. One possibility to discriminate these *admissible* geologies is to create or record a small-magnitude event with a controlled source (e.g. microtremor) and record the PSA at a reference receiver. In the absence of such recordings, we use a fictive *observation* obtained from numerical simulations.

Let us denote $\mathcal{S} = \{a_i, \boldsymbol{x}_{s,i}, \boldsymbol{\theta}_{s,i}\}_{1 \leq i \leq 442368}$ the set of geologies, source positions, and source orientations sampled with the Saltelli sampler. The *observed* ground motion $\boldsymbol{u}^*$ is simulated with a geology a^* and source parameters $(\boldsymbol{x}_s^*, \boldsymbol{\theta}_s^*)$ that do not belong to $\mathcal{S}$. The *observed* PSA is computed at receiver R_3. This quantity of interest is denoted $y^* := h(\boldsymbol{u}^*(\boldsymbol{x}_{R_3}))$. Figure 4.12a shows the PSA distribution on the set $\mathcal{S}$ (by abuse of notation, this implies the set of PSA $\hat{y} = h(\hat{\boldsymbol{u}}) = h(G_\phi(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s))$ computed from parameters $(a, \boldsymbol{x}_s, \boldsymbol{\theta}_s)$ in $\mathcal{S}$). The *observed* PSA y^* is close to the median on the East-West and vertical components, and higher than the median on the North-South component (Fig. 4.12a).

(a) Distribution on $\mathcal{S}$

(b) Distribution on $\mathcal{S}_{|y^*}$

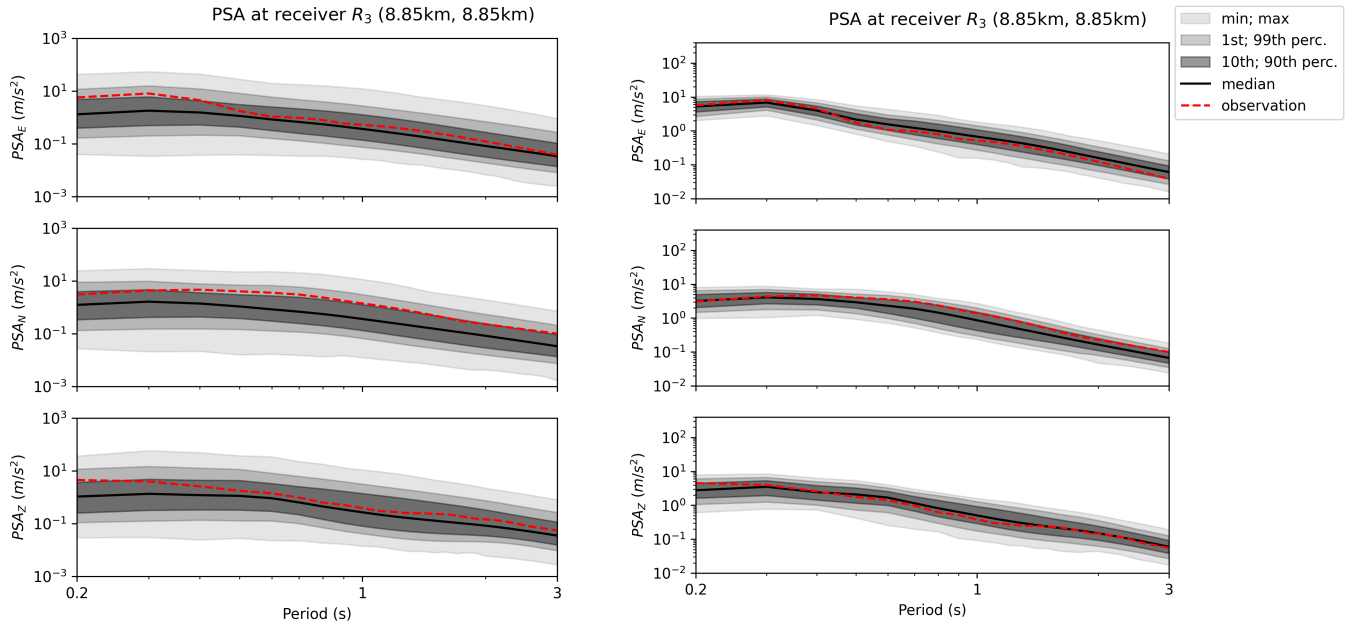
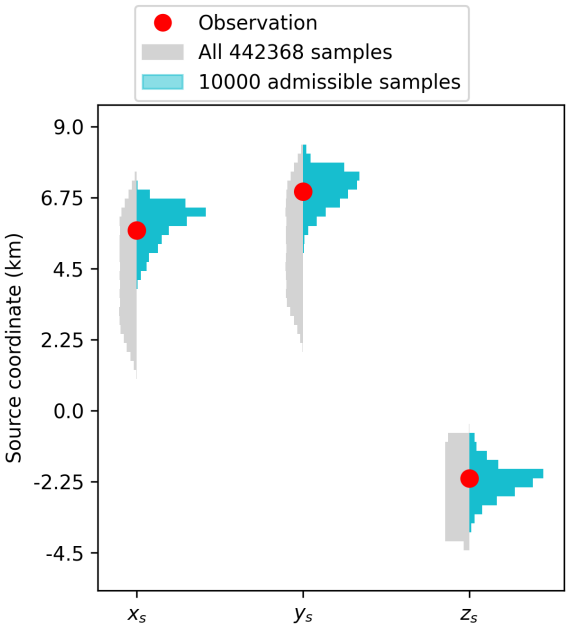


Figure 4.12: Distribution of predicted PSA $\hat{y}$ on the three components at receiver R_3 (grey shaded areas) and *observation* y^* (red dashed line)

The difference between one PSA $\hat{y}$ obtained from the set $\mathcal{S}$ and the *observed* PSA y^* is computed as the Euclidean distance $\|\hat{y} - y^*\|$ (more specific metrics could be chosen as well but we aim at demonstrating the applicability of the method). Then, we consider that the 10 000 realizations with the smallest difference are *admissible* and this set is denoted $\mathcal{S}_{|y^*}$ (the choice of 10 000 *admissible* realizations corresponds to a rupture point in the cumulative distribution of $\|\hat{y} - y^*\|$). Figure 4.12b illustrates that the PSA variance reduces significantly when considering the set of *admissible* geologies.

One can then consider the set of source positions $\boldsymbol{x}_s \in \mathcal{S}_{|y^*}$. This yields posterior distributions of the source parameters constrained by the observation y^*. Figure 4.13a shows that the source position is well constrained by the observation since the variance significantly reduces between $p(x_s)$ and $p(x_s|y^*)$ (similarly for $p(y_s|y^*)$ and $p(z_s|y^*)$). Additionally, the distribution concentrates around the source position that generated the observation. Although the posterior distributions of the strike and dip angles are different from the prior distributions, they are less concentrated around the true values (Fig. 4.13b). And there is no significant difference between the prior and posterior distributions of the rake angle. This indicates that it is more difficult to constrain the source orientation than the source position. Moreover, the observation does not allow to constrain the geological parameters, as

(a) Distribution of source positions



(b) Distribution of source orientations

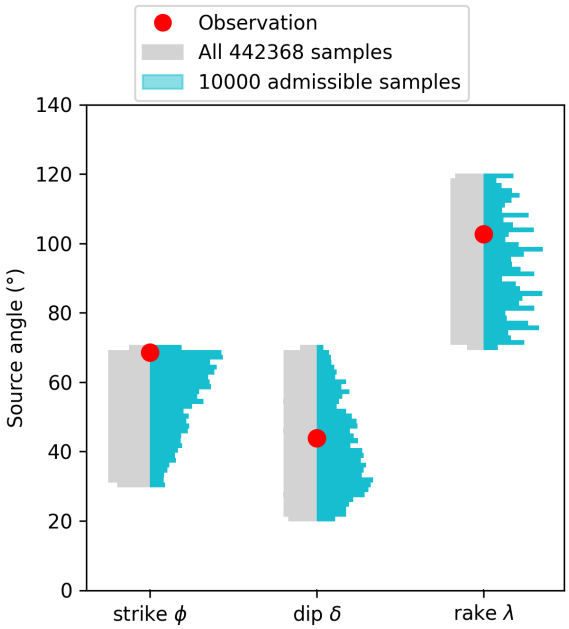


Figure 4.13: Distribution of source positions (a) and source orientations (b) for samples in $\mathcal{S}$ (grey) and *admissible* samples in $\mathcal{S}_{|y^*}$ (blue). The parameters corresponding to the *observation* $\boldsymbol{x}_s^*$, $\boldsymbol{\theta}_s^*$ are shown with the red dots.

illustrated with the posterior distributions of the coefficients of variation that are similar to the prior distributions (Fig. A.32).

Finally, it is interesting to note that the parameters that are better constrained by the observation correspond to the ones with the largest first-order Sobol indices. Indeed, large Sobol indices mean that a small parameter variation induces a large outcome change. Therefore, one is able to discriminate parameters that do not lead to a PSA close to the observation.

4.3.4 Uncertainty quantification on the PSA

The set of 10 000 *admissible* geologies provides distributions of predicted PSA at each receiver $p(\hat{y}(\boldsymbol{x}_R)|y^*)$. Civil engineering focuses on the highest percentiles of this distribution since structures are most threatened by high ground motion intensity. In this percentile estimation, one should also i) account for the MIFNO prediction error, ii) compare the predicted PSA percentile with simulations, and iii) estimate the number of samples necessary to reach convergence.

The 4000 training samples provide a reference to calibrate the MIFNO prediction error on the PSA. This error $e_{\boldsymbol{x}_R,T}$ is estimated at each receiver $\boldsymbol{x}_R$ and each period T

$$\log_{10}(y(\boldsymbol{x}_R,T)) = e_{\boldsymbol{x}_R,T}(\log_{10}(\hat{y}(\boldsymbol{x}_R,T))) + \epsilon \tag{4.14}$$

$e_{\boldsymbol{x}_R,T}$ is a skew normal distribution from which we find the parameters that minimize the residual ϵ (Fig. A.33).

Then, we choose randomly a set of n samples in $\mathcal{S}_{|y^*}$ and compute the 90th percentile of the predicted PSA in this set. For n varying from 10 to 10 000 this procedure is repeated 100 times to quantify the variability (orange boxplots in Fig. 4.14). For each PSA prediction, the correction $e_{\boldsymbol{x}_R,T}$ is also applied before computing the 90th percentile. Figure 4.14 shows that the variability decreases when the number of samples n increases (red boxplots). Indeed, the percentile estimation becomes more robust when the set is large. With correction, the predicted PSA converges towards 2.16 m/s^2. The confidence interval estimated with 100 repetitions of 1000 samples is [2.01 m/s^2; 2.34 m/s^2].

Without the surrogate model, the 90th percentile of PSA could also be estimated from simulations, in which case the geological models are not constrained. In the Le Teil database of 4000 samples, 944 geologies satisfy the conditions of the sensitivity analysis (coefficient of variation smaller than 0.15). Therefore, the 90th percentile

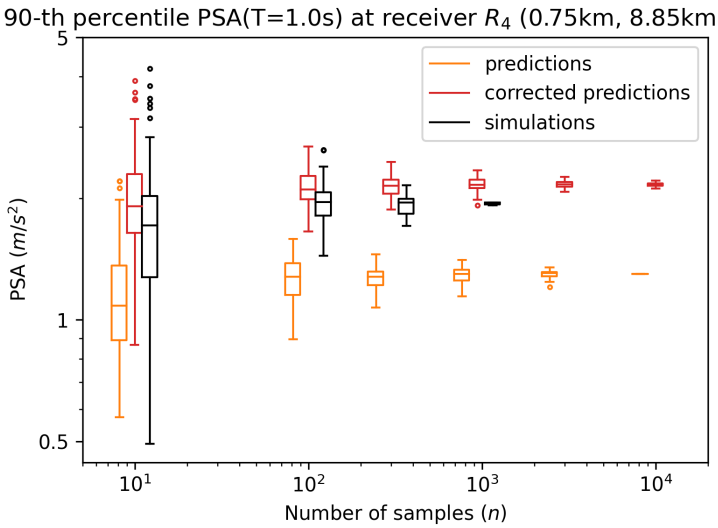


Figure 4.14: Variability of the 90th percentile computed on n samples, repeated 100 times. Prediction $\hat{y}$ (orange), corrected predictions $10^{e_{\boldsymbol{x}_R,T}(\log_{10}(\hat{y}))}$ (red), and simulation y (black)

is evaluated on n samples among the 944. Except for very small sets (n=10), the PSA variability is comparable between simulations and predictions (Fig. 4.14). However, the simulated PSA converges towards 1.94 m/s^2, which is lower than the lower bound of the predicted confidence interval.

Finally, one should note that these results have been obtained at receiver R_4, which is not the receiver used to constrain the PSA with the *observation*. Despite this, the predicted PSA is coherent with the simulation. Similar conclusions can be drawn at other receivers (Fig. A.34) and other periods (Fig. A.35).

4.4 Conclusion

Using a database of geological models and source properties tailored to the Le Teil earthquake, the MIFNO was fine-tuned to predict ground motions specific to this region. We showed that transfer learning is an efficient approach to improve the MIFNO accuracy when focusing on a target area. As few as 250 transfer learning samples were sufficient to obtain good envelope GOFs and excellent phase GOFs. This training took only 1 h with one A100 GPU, making it affordable for other site-specific applications.

We proved that the MIFNO has an excellent phase accuracy for 69% of predictions (GOF larger than 8) and good amplitude accuracy for more than 73% of predictions (GOF larger than 6). In particular, the wave arrival times are well reproduced, the spatial propagation of the waves is correctly captured, and the amplitude of the main fluctuations is accurate. Thanks to this good accuracy, we applied the superposition principle of the elastic wave equation to predict ground motion generated by an extended fault source. Results were excellent on the vast majority of sensors, despite the temporal complexity of the ground motion time series.

The influence of geological uncertainties on ground motion intensity was further assessed with the Peak Ground Velocity (PGV). Results indicate that the F-FNO predictions are excellent when the geologies have moderate amplitude heterogeneities ; the mean PGV relative bias was between −3 % and 7 % when geologies have a coefficient of variation smaller than 0.1. When geologies have stronger heterogeneities, the F-FNO tends to slightly underestimate the PGV when comparing predictions and simulated ground truths for the same geology.

Thanks to the MIFNO speed, a sensitivity analysis was performed on more than 440 000 ground motion predictions. First-order Sobol indices showed that the PSA is largely influenced by the source position. Geological parameters have non-negligible total indices but they reflect high-order interactions. Using a fictive *observation* at a reference receiver, we were able to constrain the source position and obtain meaningful posterior distributions. However, constraining the source orientation was less obvious and a single observation was not enough to infer the geological parameters that generated the *observation*. These results show that the MIFNO is an appropriate surrogate model to address inverse problems and we have only provided a first insight into the possible applications.

More advanced approaches could be explored to study the optimal position of reference receivers, choose a more specific definition of the *admissible* geologies, or formalize a rigorous inverse problem.

Our method already provides confidence intervals of the 90th PSA percentile. By selecting only *admissible* geologies, the MIFNO predictions rely on realistic geological and source parameters. If one were to compute confidence intervals only with numerical simulations, it would not be possible to restrain the variability of configurations and extreme unrealistic PSA values could be found. For real-life applications where security is a major concern, it would be ideal to run a small number of simulations wisely sampled among the set of *admissible* geologies to ensure that the confidence intervals match the MIFNO estimates.

The influence of the MIFNO error could also be addressed in more details. Especially, there is a wide literature on neural networks that include an intrinsic notion of uncertainty. Extending those methods to neural operators is not an easy task due to high computational requirements but several studies have already shown promising results (e.g. Agata et al. 2024; Cao et al. 2024; Guilhoto and Perdikaris 2024; Ma et al. 2024; Z. Zou et al. 2022). They provide a complimentary view to our approach where uncertainty evaluation does not rely on repeated calls to the surrogate model.

Conclusion and perspectives

Physics-based simulations

Physics-based simulations are convenient methods to evaluate ground motion intensity. Thanks to advanced numerical schemes, they solve the elastic wave equation with high fidelity. However, their outcomes strongly depend on the simulation parameters such as the geological models and source properties. When simulating the ground motion produced by the Le Teil earthquake, we showed in Chapter 1 that adding random heterogeneities to the reference geological model was an adequate way to compensate for the general lack of geological knowledge. Especially, heterogeneities attenuated the ground motion peaks coming from the absence of scatterers in the smooth reference model. Yet different heterogeneities samplings led to significant differences in ground motion intensity and the computational costs of large-scale 3D physics-based simulations prevent uncertainty quantification analyses.

Simulations of the Le Teil earthquake indicated the key elements to include in our database of ground motions. The HEMEW^S-3D database contains 30 000 geological models of size 9.6 km $\times$ 9.6 km $\times$ 9.6 km accompanied by the surface ground motion generated by random point-wise sources. The geological models are based on a random arrangement of horizontal layers onto which random heterogeneities are added. The choice of horizontal layers is lightly constrained to offer flexibility to the machine learning model. Random heterogeneities are stronger than the typical values observed at our scale of interest to create geological peculiarities that cannot be represented by horizontal layers. In Chapter 2, we showed that this design leads to a large intrinsic dimension of the geological database, at least of the order of 30.

Inside each geological domain, seismic waves were propagated from a point-wise source having a random location and random orientation. The propagation was computed with the SEM3D earthquake simulator based on the Spectral Element Method. The 3D simulations are valid up to a frequency of 5 Hz and are conducted for 6.4 s. The 30 000 simulations amount to a simulation time of 6.8×10^5 h equivalent CPU. To reduce the memory requirements, seismic waves are recorded only at the surface of the propagation domain on a grid of 32×32 virtual sensors. Then, each ground motion sample in the HEMEW^S-3D database is a set of three 3D variables (one for each component of ground motion). The intrinsic dimension of the ground motion database is at least on the order of 100, suggesting the high complexity of the ground motion prediction task by a surrogate model.

The Multiple Input Fourier Neural Operator (MIFNO)

We proposed an extension of the Factorized Fourier Neural Operator (F-FNO) called Multiple Input Fourier Neural Operator (MIFNO) that predicts surface ground motion depending on a 3D geology and a vector of source properties. This surrogate model derives from successive extensions of the Fourier Neural Operator (FNO) and achieves the following objectives:

1. the geological model is given as a 3D variable, which allows any parametrization of geological interfaces and small-scale heterogeneities
2. ground motion is a time series predicted for any point at the surface
3. although ground motion is a time-dependent 3D variable, we avoid the need for a huge 4D surrogate model with a depth-to-time conversion between inputs and outputs
4. the earthquake source can be placed at any location inside the domain and parametrized by any moment tensor orientation

The MIFNO predicts accurate ground motion for a vast majority of samples. It is especially accurate to estimate the wave arrival times and the phase of ground motion signals, as illustrated by the excellent phase Goodness-Of-

Fit (GOF). Envelope GOFs quantify the errors on ground motion amplitudes and are considered good to excellent. Overall, these results demonstrate that the MIFNO is an accurate surrogate model for ground motion prediction.

Predictions are also robust to out-of-distribution data as quantified by the good predictions on realistic geologies such as the Marmousi model and the good predictions when the source is located outside the training domain. The MIFNO can also be readily applied to inputs of higher resolution than the training resolution. Although it does not systematically improve the results, predictions are stable for higher resolutions.

Prediction errors can be interpreted from several physical variables, which helps to understand the MIFNO limitations. We showed in Chapter 3 that errors increase for high-frequency components, reflecting the well-known spectral bias inherent to neural networks. Larger errors are also obtained for geological models displaying small correlation lengths and large coefficients of variation. Indeed, geological heterogeneities induce dispersion that create ground motion fluctuations. Ground motion then becomes intrinsically harder to predict. Similarly, deeper sources and larger epicentral distances generally lead to more complex ground motion. However, the most energetic, and hence damaging, ground motions tend to be well predicted, which is of the utmost importance for engineering applications.

Training the MIFNO on the HEMEW^S-3D database provides a generic surrogate model that gives good predictions for a wide variety of geological models and source parameters. However, training on a large database is computationally demanding and it becomes difficult to significantly improve the accuracy without massive investments in data collection and multi-GPU training. In this context, transfer learning is an efficient approach to specialize the MIFNO to a target task with light computational requirements. In Chapter 4, we designed a transfer learning database tailored to the Le Teil earthquake. The variability of geological models is reduced since they all originate from a reference 1D model and sources correspond to the La Rouvière fault that created the Le Teil earthquake. We showed that as few as 250 samples from this tailored database were sufficient to improve the MIFNO accuracy by 1 to 2 GOF units. One could even achieve almost excellent predictions with a few thousands transfer learning samples.

We illustrated the benefits of the fine-tuned MIFNO on several applications that are relevant for seismic hazard analyses. When geological models present moderate heterogeneities (i.e. in line with observations at our scale of interest), the F-FNO predicts accurate Peak Ground Velocity (PGV). It can also predict Pseudo Spectral Acceleration (PSA) distributions that match well with the simulations. Thanks to the speed of the MIFNO predictions, confidence intervals can be obtained by predicting ground motion for hundreds of thousands of geological and source configurations. They give higher estimates than the simulations, thereby providing safety margins.

Societal implications

Uncertainty quantification is unreachable with high-fidelity 3D simulations due to the computational time. However, deep learning surrogate models require large training databases that need to be acquired from numerical simulations (at least in the supervised learning framework adopted in this work). Throughout this thesis, the energy consumption of the different tasks was assessed to evaluate whether the MIFNO is beneficial compared to simulations in this respect. We summarize the results from Tab. 1.3, Tab. 2.3, and Tab. 3.6 in the following Tab. 5.1 and Fig. 5.1.

Conducting the uncertainty analysis of Chapter 4 with $N_{MC}=10^5$ realizations would represent 11 MWh with CPU simulations and preliminary benchmarks on SEM3D give an equivalent projection of 1.9 MWh with GPU simulations. Assuming that the generic MIFNO would be applied at 100 different locations where transfer learning would be used to fine-tune the surrogate model, the same uncertainty analysis would amount to 0.14 MWh. Therefore, the MIFNO reduces the total energy consumption by two orders of magnitude compared to the current framework.

We have mentioned in Chapter 4 that the number of initial training samples was crucial to allow accurate predictions after transfer learning. By factorizing the costs of the database simulation across several studies, one can make the corresponding energy consumption reasonable. Results in Tab. 5.1 assume that a large number of transfer learning samples would be necessary to reach excellent accuracy on the target task and we would argue that they give a conservative evaluation of the MIFNO energy consumption. Most importantly, the MIFNO prediction time are so negligible that the MIFNO energy consumption does not depend on the number of predictions. Though the costs scale linearly in the case of numerical simulations. Therefore, demanding Monte Carlo frameworks with millions of iterations, or more, will benefit even more from the MIFNO (Fig. 5.1).

Finally, transferring knowledge between a generic surrogate model and target tasks require that the database, the neural operator architecture, and the weights of the pre-trained model are publicly available and usable. The

Task	Number of realizations	Energy consumption (MWh)
SEM3D simulation (CPU)	$N_{MC} = 10^5$	**11**
SEM3D simulation (projection GPU)	$N_{MC} = 10^5$	1.9
Database simulation (CPU)	$N_{train} = 30\,000$ shared in 100	0.033
Transfer learning database (CPU)	$N_{TL} = 1000$	0.11
MIFNO training (GPU)	$N_{train} = 30\,000$ shared in 100	0.000 22
MIFNO fine-tuning (GPU)	$N_{TL} = 1000$	0.000 74
MIFNO prediction (GPU)	$N_{MC} = 10^5$	0.000 052
Total MIFNO (CPU + GPU)	$N_{MC} = 10^5$	**0.14**

Table 5.1: Estimation of the energy consumption for all the tasks presented in this thesis. The energy consumption of 1 CPU is estimated to 4.88 Wh. The energy consumption of 1 GPU is estimated to 187 Wh. In all situations, the domain has a size of 9.6 km $\times$ 9.6 km $\times$ 9.6 km and time series have a duration T_f=6.4 s. From preliminary benchmarks, SEM3D simulation time is estimated to 0.10 h one 1 GPU. The costs of building the HEMEW^S-3D database and training the MIFNO are shared between 100 studies at different sites of interest.

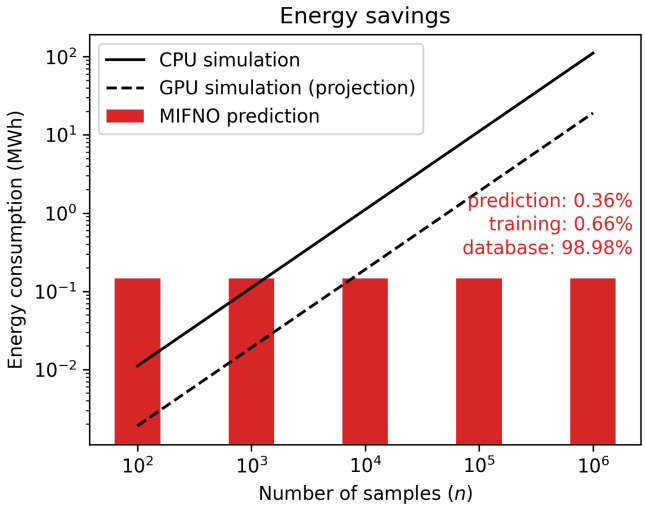


Figure 5.1: Energy consumption of n ground motion evaluations using numerical simulations on CPUs (black line), GPUs (black dashed line), and the MIFNO (red bards).

HEMEW-3D and HEMEW^S-3D databases can be downloaded on Recherche Data Gouv[1] and the code to train, re-use, and fine-tune the neural operators is available on Github[2].

Perspectives

The field of Scientific Machine Learning (SciML) evolves rapidly and new neural operators regularly improve state-of-the-art benchmarks. Several techniques could improve the MIFNO accuracy, such as adding constraints on the derivative in the loss function (Cao et al. 2024; Cho et al. 2024) or focusing on extreme values that are the most threatening for infrastructures (Lagerquist and Ebert-Uphoff 2022; Xu et al. 2024). Building on alternative architectures that alleviate the requirements of a Cartesian grid (H. Lin et al. 2022; Z. Li et al. 2023b), one could also envision a surrogate model that considers a non-flat topography. This would allow statistical quantification of site effects due to 3D topography.

The MIFNO can also be extended to more flexible geological parameters. Especially, one could add the P-wave velocity and density as additional inputs. We do not expect a massive increase in computational requirements since the architecture would require only small changes in the uplift network. However, it requires designing a new

[1] https://doi.org/10.57745/LAI6YU
[2] https://github.com/lehmannfa/HEMEW3D

training database. In this framework, it would be interesting to add more physical constraints to the geological models. In particular, one may want to reach lower surface velocities and avoid strong velocity inversion between layers. It it also possible to include anisotropy in the random fields (Ta et al. 2010).

Considering the benefits of transfer learning, it is desirable to apply this method with other geological models. For instance, it would be interesting to investigate the site effects created by sedimentary basins with this approach. Thanks to the 3D representation of geological models, the MIFNO allows a rich parametrization of the basin shapes and it could complement existing works based on numerical simulations (e.g. Soto Moncada 2023).

The MIFNO could also be coupled with generative methods that increase the frequency content of physics-based simulations to reach frequencies of interest in engineering applications (e.g. Gatti and Clouteau 2020). Then, it could be embedded in a domain reduction method to analyze soil-structure interactions in deeper extents than currently achieved with numerical simulations (Korres et al. 2022).

Finally, many physical problems rely on parametrized PDEs with various inputs and wave propagation appears in several contexts others than seismology. Although the MIFNO was inspired by seismic wave propagation challenges, its architecture is more general and can certainly benefit other scientific communities.

Bibliography

Adler, A., M. Araya-Polo, and T. Poggio (2021). "Deep Learning for Seismic Inverse Problems: Toward the Acceleration of Geophysical Analysis Workflows". In: *IEEE Signal Processing Magazine* 38.2, pp. 89–119. ISSN: 1053-5888, 1558-0792. DOI: 10.1109/MSP.2020.3037429.

Agata, R., K. Shiraishi, and G. Fujie (2024). *Physics-Informed Learning Fully Quantifies Uncertainty in Seismic Structure and Source Estimate*. arXiv: 2402.08854. URL: http://arxiv.org/abs/2402.08854. preprint.

Aki, K. and P. G. Richards (1980). *Quantitative Seismology*. San Francisco: Freeman.

— (2002). *Quantitative Seismology*.

Alavi, A. H. and A. H. Gandomi (2011). "Prediction of Principal Ground-Motion Parameters Using a Hybrid Method Coupling Artificial Neural Networks and Simulated Annealing". In: *Computers & Structures* 89.23-24, pp. 2176–2194. ISSN: 00457949. DOI: 10.1016/j.compstruc.2011.08.019.

Aldirany, Z., R. Cottereau, M. Laforest, and S. Prudhomme (2023). "Operator Approximation of the Wave Equation Based on Deep Learning of Green's Function". In: DOI: 10.48550/ARXIV.2307.13902.

Alimoradi, A. (2011). *Earthquake Ground Motion Simulation Using Novel Machine Learning Tools*. EERL 2011-01. Pasadena, California: California institute of technology.

Alimoradi, A. and J. L. Beck (2015). "Machine-Learning Methods for Earthquake Ground Motion Analysis and Simulation". In: *Journal of Engineering Mechanics* 141.4, p. 04014147. ISSN: 0733-9399, 1943-7889. DOI: 10.1061/(ASCE)EM.1943-7889.0000869.

Alkhalifah, T. and X. Huang (2023). *Physics-Informed Neural Wavefields with Gabor Basis Functions*. arXiv: 2310.10602. URL: http://arxiv.org/abs/2310.10602. preprint.

Aminzadeh, F., J. Brac, and T. Kunz (1997). *3-D Salt and Overthrust Models*. SEG/EAGE 3-D Modeling Series No.1.

Ancheta, T., Y. Bozorgnia, R. Darragh, W. Silva, B. Chiou, J. Stewart, et al. (2012). "PEER NGA-West2 Database: A Database of Ground Motions Recorded in Shallow Crustal Earthquakes in Active Tectonic Regions". In: *Proceedings, 15th World Conference on Earthquake Engineering*, pp. 24–28.

Anderson, J. G. (2004). "Quantitative Measure Of The Goodness-Of-Fit of Synthetic Seismograms". In: Proceedings of the 13th World Conference on Earthquake Engineering. Vol. 243. Vancouver, Canada, p. 14.

Annon (2022). *Simulate CO2 Flow with Open Porous Media*. URL: https://github.com/microsoft/AzureClusterlessHPC.jl/tree/main/examples/opm.

Arias, A. (1970). "A Measure of Earthquake Intensity". In: *Seismic design for nuclear plants*, pp. 438–483.

Arroucau, P. (2020). *A Preliminary Three-Dimensional Seismological Model of the Crust and Uppermost Mantle for Metropolitan Franc*. SIGMA2-2018-D2014. URL: https://www.sigma-2.net/medias/files/sigma2-2018-d2-014-3d-velocity-model-france-approved-public-.pdf.

Azulay, Y. and E. Treister (2023). "Multigrid-Augmented Deep Learning Preconditioners for the Helmholtz Equation". In: *SIAM Journal on Scientific Computing* 45.3, S127–S151. DOI: 10.1137/21M1433514.

Bahrampouri, M., A. Rodriguez-Marek, S. Shahi, and H. Dawood (2021). "An Updated Database for Ground Motion Parameters for KiK-net Records". In: *Earthquake Spectra* 37.1, pp. 505–522. ISSN: 8755-2930, 1944-8201. DOI: 10.1177/8755293020952447.

Baker, J. W. and C. A. Cornell (2006). "Which Spectral Acceleration Are You Using?" In: *Earthquake Spectra* 22.2, pp. 293–312.

Baker, N., F. Alexander, T. Bremer, A. Hagberg, Y. Kevrekidis, H. Najm, et al. (2019). "Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence". In: DOI: 10.2172/1478744.

Bartolucci, F., E. de Bézenac, B. Raonić, R. Molinaro, S. Mishra, and R. Alaifari (2023). *Are Neural Operators Really Neural Operators? Frame Theory Meets Operator Learning*. arXiv: 2305.19913. URL: http://arxiv.org/abs/2305.19913. preprint.

Basri, R., M. Galun, A. Geifman, D. Jacobs, Y. Kasten, and S. Kritchman (2020). "Frequency Bias in Neural Networks for Input of Non-Uniform Density". In: *Proceedings of the 37th International Conference on Machine Learning*. ICML'20. JMLR.org.

Berenger, J.-P. (1994). "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves". In: *Journal of Computational Physics* 114.2, pp. 185–200. ISSN: 00219991. DOI: 10.1006/jcph.1994.1159.

Berge-Thierry, C., F. Voldoire, F. Ragueneau, F. Lopez-Caballero, and A. Le Maoult (2020). "Main Achievements of the Multidisciplinary SINAPS@ Research Project: Towards an Integrated Approach to Perform Seismic Safety Analysis of Nuclear Facilities". In: *Pure and Applied Geophysics* 177.5, pp. 2299–2351. ISSN: 0033-4553, 1420-9136. DOI: 10.1007/s00024-019-02194-4.

Berge-Thierry, C., F. Cotton†, O. Scotti, D.-A. Griot-Pommera, and Y. Fukushima (2003). "New Empirical Response Spectral Attenuation Laws for Moderate European Earthquakes". In: *Journal of Earthquake Engineering* 7.2, pp. 193–222. ISSN: 1363-2469, 1559-808X. DOI: 10.1080/13632460309350446.

Bhattacharya, K., B. Hosseini, N. B. Kovachki, and A. M. Stuart (2021). "Model Reduction And Neural Networks For Parametric PDEs". In: *The SMAI journal of computational mathematics* 7, pp. 121–157. ISSN: 2426-8399. DOI: 10.5802/smai-jcm.74.

Bloemheuvel, S., J. Van Den Hoogen, D. Jozinović, A. Michelini, and M. Atzmueller (2023). "Graph Neural Networks for Multivariate Time Series Regression with Application to Seismic Data". In: *International Journal of Data Science and Analytics* 16.3, pp. 317–332. ISSN: 2364-415X, 2364-4168. DOI: 10.1007/s41060-022-00349-6.

Brandstetter, J., D. Worrall, and M. Welling (2022). "Message Passing Neural PDE Solvers". In: International Conference on Learning Representations. URL: https://openreview.net/forum?id=vSix3HPYKSU.

Bravard, J.-P. and P. Gaydou (2015). "Historical Development and Integrated Management of the Rhône River Floodplain, from the Alps to the Camargue Delta, France". In: *Geomorphic Approaches to Integrated Floodplain Management of Lowland Fluvial Systems in North America and Europe*. Ed. by P. F. Hudson and H. Middelkoop. New York, NY: Springer New York, pp. 289–320. ISBN: 978-1-4939-2379-3 978-1-4939-2380-9. URL: http://link.springer.com/10.1007/978-1-4939-2380-9_12.

Cao, L., T. O'Leary-Roseberry, and O. Ghattas (2024). *Efficient Geometric Markov Chain Monte Carlo for Nonlinear Bayesian Inversion Enabled by Derivative-Informed Neural Operators*. arXiv: 2403.08220. URL: http://arxiv.org/abs/2403.08220. preprint.

Capdeville, Y. and J.-J. Marigo (2007). "Second Order Homogenization of the Elastic Wave Equation for Non-Periodic Layered Media". In: *Geophysical Journal International* 170.2, pp. 823–838. ISSN: 0956540X, 1365246X. DOI: 10.1111/j.1365-246X.2007.03462.x.

Causse, M., C. Cornou, E. Maufroy, J.-R. Grasso, L. Baillet, and E. El Haber (2021). "Exceptional Ground Motion during the Shallow Mw 4.9 2019 Le Teil Earthquake, France". In: *Communications Earth & Environment* 2.1, p. 14. ISSN: 2662-4435. DOI: 10.1038/s43247-020-00089-0.

Chaljub, E. (2006). "Spectral Element Modeling of 3d Wave Propagation in the Alpine Valley of Grenoble, France". In: *Third International Symposium on the Effects of Surface Geology on Seismic Motion*. Third International Symposium on the Effects of Surface Geology on Seismic Motion. Vol. S04. Grenoble, France, p. 9.

Chaljub, E., M. Celorio, C. Cornou, F. D. Martin, E. E. Haber, L. Margerin, et al. (2021). "Numerical Simulation of Wave Propagation in Heterogeneous and Random Media for Site Effects Assessment in the Grenoble Valley". In: He 6th IASPEI/IAEE International Symposium: Effects of Surface Geology on Seismic Motion.

Chen, F., Z. Liu, G. Lin, J. Chen, and Z. Shi (2024). *NSNO: Neumann Series Neural Operator for Solving Helmholtz Equations in Inhomogeneous Medium*. arXiv: 2401.13494. URL: http://arxiv.org/abs/2401.13494. preprint.

Chen, P., T. Hu, and J. Guilleminot (2024). "A Nonlinear-Manifold Reduced-Order Model and Operator Learning for Partial Differential Equations with Sharp Solution Gradients". In: *Computer Methods in Applied Mechanics and Engineering* 419, p. 116684. ISSN: 00457825. DOI: 10.1016/j.cma.2023.116684.

Cheng, S. and T. Alkhalifah (2024). *Meta-PINN: Meta Learning for Improved Neural Network Wavefield Solutions*. arXiv: 2401.11502. URL: http://arxiv.org/abs/2401.11502. preprint.

Cho, N., J. Ryu, and H. J. Hwang (2024). *Sobolev Training for Operator Learning*. arXiv: 2402.09084. URL: http://arxiv.org/abs/2402.09084. preprint.

Cole, H. M., W. L. Yeck, and H. M. Benz (2023). "MLAAPDE: A Machine Learning Dataset for Determining Global Earthquake Source Parameters". In: *Seismological Research Letters*. ISSN: 0895-0695. DOI: 10.1785/0220230021.

Cottereau, R. (2016). "Scale Coupling and Upscaling Techniques in Elastostatics and Elastodynamics in Random Media". Université Pierre et Marie Curie, Paris VI. URL: https://theses.hal.science/tel-01374413/document.

Cruz-Jiménez, H., G. Li, P. M. Mai, I. Hoteit, and O. M. Knio (2018). "Bayesian Inference of Earthquake Rupture Models Using Polynomial Chaos Expansion". In: *Geoscientific Model Development* 11.7, pp. 3071–3088. ISSN: 1991-9603. DOI: 10.5194/gmd-11-3071-2018.

Cupillard, P. and Y. Capdeville (2018). “Non-Periodic Homogenization of 3-D Elastic Media for the Seismic Wave Equation”. In: *Geophysical Journal International* 213.2, pp. 983–1001. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggy032.

Datta, A., D. J. Wu, W. Zhu, M. Cai, and W. L. Ellsworth (2022). “DeepShake: Shaking Intensity Prediction Using Deep Spatiotemporal RNNs for Earthquake Early Warning”. In: *Seismological Research Letters*. ISSN: 0895-0695, 1938-2057. DOI: 10.1785/0220210141.

De Martin, F. (2011). “Verification of a Spectral-Element Method Code for the Southern California Earthquake Center LOH.3 Viscoelastic Case”. In: *Bulletin of the Seismological Society of America* 101.6, pp. 2855–2865. ISSN: 0037-1106. DOI: 10.1785/0120100305.

De Novellis, V., V. Convertito, S. Valkaniotis, F. Casu, R. Lanari, M. F. Monterroso Tobar, et al. (2020). “Coincident Locations of Rupture Nucleation during the 2019 Le Teil Earthquake, France and Maximum Stress Change from Local Cement Quarrying”. In: *Communications Earth & Environment* 1.1, p. 20. ISSN: 2662-4435. DOI: 10.1038/s43247-020-00021-6.

De Ryck, T., F. Bonnet, S. Mishra, and E. de Bézenac (2023). *An Operator Preconditioning Perspective on Training in Physics-Informed Machine Learning*. arXiv: 2310.05801. URL: http://arxiv.org/abs/2310.05801. preprint.

De Carvalho Paludo, L., V. Bouvier, and R. Cottereau (2019). “Scalable Parallel Scheme for Sampling of Gaussian Random Fields over Very Large Domains: Parallel Scheme for Sampling of Random Fields over Very Large Domains”. In: *International Journal for Numerical Methods in Engineering* 117.8, pp. 845–859. ISSN: 00295981. DOI: 10.1002/nme.5981.

Delouis, B., E. Oral, M. Menager, J.-P. Ampuero, A. Guilhem Trilla, M. Régnier, et al. (2021). “Constraining the Point Source Parameters of the 11 November 2019 Mw 4.9 Le Teil Earthquake Using Multiple Relocation Approaches, First Motion and Full Waveform Inversions”. In: *Comptes Rendus. Géoscience* 353.S1, pp. 1–24. ISSN: 1778-7025. DOI: 10.5802/crgeos.78.

Deng, C., S. Feng, H. Wang, X. Zhang, P. Jin, Y. Feng, et al. (2022). “OpenFWI: Large-scale Multi-Structural Benchmark Datasets for Full Waveform Inversion”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. Vol. 35. Curran Associates, Inc., pp. 6007–6020. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/27d3ef263c7cb8d542c4f9815a49b69b-Paper-Datasets_and_Benchmarks.pdf.

Derras, B., P.-Y. Bard, and F. Cotton (2016). “Site-Condition Proxies, Ground Motion Variability, and Data-Driven GMPEs: Insights from the NGA-West2 and RESORCE Data Sets”. In: *Earthquake Spectra* 32.4, pp. 2027–2056. ISSN: 8755-2930, 1944-8201. DOI: 10.1193/060215EQS082M.

Ding, Y., S. Chen, X. Li, S. Wang, S. Luan, and H. Sun (2023). “Self-Adaptive Physics-Driven Deep Learning for Seismic Wave Modeling in Complex Topography”. In: *Engineering Applications of Artificial Intelligence* 123, p. 106425. ISSN: 09521976. DOI: 10.1016/j.engappai.2023.106425.

Dreger, D. S., G. Hurtado, A. K. Chopra, and S. Larsen (2007). *Near-Fault Seismic Ground Motions*. EERC 2007-03. University of California, Berkeley. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.434.578&rep=rep1&type=pdf.

Duverger, C., G. Mazet-Roux, L. Bollinger, A. G. Trilla, A. Vallage, B. Hernandez, et al. (2021). “A Decade of Seismicity in Metropolitan France (2010–2019): The CEA/LDG Methodologies and Observations”. In: p. 25.

El Haber, E., C. Cornou, D. Jongmans, D. Youssef Abdelmassih, F. Lopez-Caballero, and T. AL-Bittar (2019). “Influence of 2D Heterogeneous Elastic Soil Properties on Surface Ground Motion Spatial Variability”. In: *Soil Dynamics and Earthquake Engineering* 123, pp. 75–90. ISSN: 02677261. DOI: 10.1016/j.soildyn.2019.04.014.

El Haber, E., C. Smerzini, M. Fasan, L. Saint Mard, M. Vanini, P. Traversa, et al. (2021). *Simulation Techniques Benchmark, the Test Case of the November 11, 2019 Mw4.9 Le Teil Earthquake*. SIGMA2-2021-D3-082. SIGMA2.

Equinor (2020). *Sleipner 2019 Benchmark Model*. DOI: 10.11582/2020.00004.

Esfahani, R. D. D., F. Cotton, M. Ohrnberger, and F. Scherbaum (2022). “TFCGAN: Nonstationary Ground-Motion Simulation in the Time–Frequency Domain Using Conditional Generative Adversarial Network (CGAN) and Phase Retrieval Methods”. In: *Bulletin of the Seismological Society of America*. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120220068.

Faccioli, E., F. Maggio, R. Paolucci, and A. Quarteroni (1997). “2d and 3D Elastic Wave Propagation by a Pseudo-Spectral Domain Decomposition Method”. In: *Journal of Seismology* 1.3, pp. 237–251. ISSN: 1573-157X. DOI: 10.1023/A:1009758820546.

Feng, S., H. Wang, C. Deng, Y. Feng, Y. Liu, M. Zhu, et al. (2023). $\mathbb{E}^{FWI}$ *: Multi-parameter Benchmark Datasets for Elastic Full Waveform Inversion of Geophysical Properties*. arXiv: 2306.12386. URL: http://arxiv.org/abs/2306.12386. preprint.

Festa, G. and J.-P. Vilotte (2005). "The Newmark Scheme as Velocity-Stress Time-Staggering: An Efficient PML Implementation for Spectral Element Simulations of Elastodynamics". In: *Geophysical Journal International* 161.3, pp. 789–812. ISSN: 0956540X, 1365246X. DOI: 10.1111/j.1365-246X.2005.02601.x.

Florez, M. A., M. Caporale, P. Buabthong, Z. E. Ross, D. Asimaki, and M.-A. Meier (2022). "Data-Driven Synthesis of Broadband Earthquake Ground Motions Using Artificial Intelligence". In: *Bulletin of the Seismological Society of America* 112.4, pp. 1979–1996. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120210264.

Fu, H., C. He, B. Chen, Z. Yin, Z. Zhang, W. Zhang, et al. (2017). "18.9-Pflops Nonlinear Earthquake Simulation on Sunway TaihuLight: Enabling Depiction of 18-Hz and 8-Meter Scenarios". In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. SC '17. New York, NY, USA: Association for Computing Machinery. ISBN: 978-1-4503-5114-0. DOI: 10.1145/3126908.3126910.

Gadylshin, K., V. Lisitsa, K. Gadylshina, D. Vishnevsky, and M. Novikov (2021). "Machine Learning-Based Numerical Dispersion Mitigation in Seismic Modelling". In: *Computational Science and Its Applications – ICCSA 2021*. Ed. by O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, et al. Cham: Springer International Publishing, pp. 34–47. ISBN: 978-3-030-86653-2.

Gallovič, F. (2016). "Modeling Velocity Recordings of the M_w 6.0 South Napa, California, Earthquake: Unilateral Event with Weak High-Frequency Directivity". In: *Seismological Research Letters* 87.1, pp. 2–14. ISSN: 0895-0695, 1938-2057. DOI: 10.1785/0220150042.

Gangopadhyay, T., V. Ramanan, A. Akintayo, P. K Boor, S. Sarkar, S. R. Chakravarthy, et al. (2021). "3D Convolutional Selective Autoencoder for Instability Detection in Combustion Systems". In: *Energy and AI* 4, p. 100067. ISSN: 26665468. DOI: 10.1016/j.egyai.2021.100067.

Garg, S. and S. Chakraborty (2023). *Randomized Prior Wavelet Neural Operator for Uncertainty Quantification*. arXiv: 2302.01051. URL: http://arxiv.org/abs/2302.01051. preprint.

Gatti, F., L. D. Carvalho Paludo, A. Svay, F. Lopez-Caballero, R. Cottereau, and D. Clouteau (2017). "Investigation of the Earthquake Ground Motion Coherence in Heterogeneous Non-Linear Soil Deposits". In: *Procedia Engineering* 199, pp. 2354–2359. ISSN: 18777058. DOI: 10.1016/j.proeng.2017.09.232.

Gatti, F. (2017). "Analyse Physics-Based de Scénarios Sismiques "de La Faille Au Site" : Prédiction de Mouvement Sismique Fort Pour l'étude de Vulnérabilité Sismique de Structures Critiques". Paris: Université Paris Saclay. URL: http://www.theses.fr/2017SACLC051.

Gatti, F. and D. Clouteau (2020). "Towards Blending Physics-Based Numerical Simulations and Seismic Databases Using Generative Adversarial Network". In: *Computer Methods in Applied Mechanics and Engineering* 372, p. 113421. ISSN: 00457825. DOI: 10.1016/j.cma.2020.113421.

Gatti, F., S. Touhami, F. Lope-Caballero, and D. Pitilakis (2018). "3-D Source-to-Site Numerical Investigation on the Earthquake Ground Motion Coherency in Heterogeneous Soil Deposits". In: *Numerical Methods in Geotechnical Engineering IX*. 9th European Conference on Numerical Methods in Geotechnical Engineering. Porto, Portugal: CRC Press, pp. 829–835. ISBN: 978-1-351-00362-9. DOI: 10.1201/9781351003629-104.

Ghanem, R. G. and P. D. Spanos (1991). *Stochastic Finite Elements: A Spectral Approach*. New York, NY: Springer New York. ISBN: 978-1-4612-7795-8 978-1-4612-3094-6. URL: http://link.springer.com/10.1007/978-1-4612-3094-6.

Ghattas, O. and K. Willcox (2021). "Learning Physics-Based Models from Data: Perspectives from Inverse Problems and Model Reduction". In: *Acta Numerica* 30, pp. 445–554. ISSN: 0962-4929, 1474-0508. DOI: 10.1017/S0962492921000064.

Grady, T. J., R. Khan, M. Louboutin, Z. Yin, P. A. Witte, R. Chandra, et al. (2023). "Model-Parallel Fourier Neural Operators as Learned Surrogates for Large-Scale Parametric PDEs". In: *Computers & Geosciences* 178, p. 105402. ISSN: 00983004. DOI: 10.1016/j.cageo.2023.105402.

Grassberger, P. and I. Procaccia (1983). "Measuring the Strangeness of Strange Attractors". In: *Physica D: nonlinear phenomena* 9.1-2, pp. 189–208.

Greif, C. and K. Urban (2019). "Decay of the Kolmogorov N -Width for Wave Problems". In: *Applied Mathematics Letters* 96, pp. 216–222. ISSN: 08939659. DOI: 10.1016/j.aml.2019.05.013.

Guilhoto, L. F. and P. Perdikaris (2024). *Composite Bayesian Optimization In Function Spaces Using NEON – Neural Epistemic Operator Networks*. arXiv: 2404.03099. URL: http://arxiv.org/abs/2404.03099. preprint.

Haghighat, E., U. B. Waheed, and G. Karniadakis (2024). "En-DeepONet: An Enrichment Approach for Enhancing the Expressivity of Neural Operators with Applications to Seismology". In: *Computer Methods in Applied Mechanics and Engineering* 420, p. 116681. ISSN: 00457825. DOI: 10.1016/j.cma.2023.116681.

Hartzell, S., S. Harmsen, and A. Frankel (2010). "Effects of 3D Random Correlated Velocity Perturbations on Predicted Ground Motions". In: *Bulletin of the Seismological Society of America* 100.4, pp. 1415–1426. ISSN: 0037-1106. DOI: 10.1785/0120090060.

Hawkins, R., M. H. Khalid, K. Smetana, and J. Trampert (2023). "Model Order Reduction for Seismic Waveform Modelling: Inspiration from Normal Modes". In: *Geophysical Journal International* 234.3, pp. 2255–2283. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggad195.

Heinecke, A., A. Breuer, S. Rettenberger, M. Bader, A. .-.-A. Gabriel, C. Pelties, et al. (2014). "Petascale High Order Dynamic Rupture Earthquake Simulations on Heterogeneous Supercomputers". In: SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. New Orleans, LA, USA, pp. 3–14. ISBN: 2167-4337. DOI: 10.1109/SC.2014.6.

Heller, G. J. (2021). "Vers Une Meilleure Estimation de La Magnitude à Partir de La Coda Sismique". Toulouse: Université Paul Sabatier.

Henderson, I., P. Noble, and O. Roustant (2023). *Wave Equation-Tailored Gaussian Process Regression with Applications to Related Inverse Problems*. URL: https://hal.science/hal-03941939. preprint.

Hendrycks, D. and K. Gimpel (2023). *Gaussian Error Linear Units (GELUs)*. arXiv: 1606.08415. URL: http://arxiv.org/abs/1606.08415. preprint.

Hinton, G. E. and R. R. Salakhutdinov (2006). "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786, pp. 504–507. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1127647.

Hsu, T.-Y. and C.-W. Huang (2021). "Onsite Early Prediction of PGA Using CNN With Multi-Scale and Multi-Domain P-Waves as Input". In: *Frontiers in Earth Science* 9, p. 626908. ISSN: 2296-6463. DOI: 10.3389/feart.2021.626908.

Hu, J., Y. Ding, S. Lin, H. Zhang, and C. Jin (2023). "A Machine-Learning-Based Software for the Simulation of Regional Characteristic Ground Motion". In: *Applied Sciences* 13.14, p. 8232. ISSN: 2076-3417. DOI: 10.3390/app13148232.

Huang, X., Z. Ye, H. Liu, B. Shi, Z. Wang, K. Yang, et al. (2022). *Meta-Auto-Decoder for Solving Parametric Partial Differential Equations*. arXiv: 2111.08823. URL: http://arxiv.org/abs/2111.08823. preprint.

Huang, X. and T. Alkhalifah (2022). "PINNup: Robust Neural Network Wavefield Solutions Using Frequency Upscaling and Neuron Splitting". In: *Journal of Geophysical Research: Solid Earth* 127.6, e2021JB023703. ISSN: 2169-9313. DOI: 10.1029/2021JB023703.

Imai, R., N. Kasui, A. Iwaki, and H. Fujiwara (2021). "A Sample Generation of Scenario Earthquake Shaking Maps via a Combination of Modal Decomposition and Empirical Copula toward Seismic Hazard Assessment". In: *Bulletin of the Seismological Society of America* 111.6, pp. 3341–3355. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120210086.

Imperatori, W. and P. M. Mai (2013). "Broad-Band near-Field Ground Motion Simulations in 3-Dimensional Scattering Media". In: *Geophysical Journal International* 192.2, pp. 725–744. ISSN: 1365-246X, 0956-540X. DOI: 10.1093/gji/ggs041.

Jessell, M., J. Guo, Y. Li, M. Lindsay, R. Scalzo, J. Giraud, et al. (2022). "Into the Noddyverse: A Massive Data Store of 3D Geological Models for Machine Learning and Inversion Applications". In: *Earth System Science Data* 14.1, pp. 381–392. ISSN: 1866-3516. DOI: 10.5194/essd-14-381-2022.

Jin, P., S. Meng, and L. Lu (2022). "MIONet: Learning Multiple-Input Operators via Tensor Product". In: *SIAM Journal on Scientific Computing* 44.6, A3490–A3514. DOI: 10.1137/22M1477751.

Jozinović, D., A. Lomax, I. Štajduhar, and A. Michelini (2020). "Rapid Prediction of Earthquake Ground Shaking Intensity Using Raw Waveform Data and a Convolutional Neural Network". In: *Geophysical Journal International* 222.2, pp. 1379–1389. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggaa233.

Kadeethum, T., F. Ballarin, Y. Choi, D. O'Malley, H. Yoon, and N. Bouklas (2022). "Non-Intrusive Reduced Order Modeling of Natural Convection in Porous Media Using Convolutional Autoencoders: Comparison with Linear Subspace Techniques". In: *Advances in Water Resources* 160, p. 104098. ISSN: 03091708. DOI: 10.1016/j.advwatres.2021.104098.

Kalakonas, P. and V. Silva (2022). "Earthquake Scenarios for Building Portfolios Using Artificial Neural Networks: Part I—Ground Motion Modelling". In: *Bulletin of Earthquake Engineering*. ISSN: 1570-761X, 1573-1456. DOI: 10.1007/s10518-022-01598-3.

Kawase, H. (2003). "61 Site Effects on Strong Ground Motions". In: *International Geophysics*. Vol. 81. Elsevier, pp. 1013–1030. ISBN: 978-0-12-440658-2. URL: https://linkinghub.elsevier.com/retrieve/pii/S0074614203801754.

Khan, S., M. Van Der Meijde, H. Van Der Werff, and M. Shafique (2020). "The Impact of Topography on Seismic Amplification during the 2005 Kashmir Earthquake". In: *Natural Hazards and Earth System Sciences* 20.2, pp. 399–411. ISSN: 1684-9981. DOI: 10.5194/nhess-20-399-2020.

Kolmogoroff, A. (1936). "Uber Die Beste Annaherung von Funktionen Einer Gegebenen Funktionenklasse". In: *Annals of Mathematics*, pp. 107–110.

Komatitsch, D. and J. Tromp (1999). "Introduction to the Spectral Element Method for Three-Dimensional Seismic Wave Propagation". In: *Geophysical Journal International* 139.3, pp. 806–822. ISSN: 0956540X, 1365246X. DOI: 10.1046/j.1365-246x.1999.00967.x.

Kong, Q. and A. Rodgers (2023). *Feasibility of Using Fourier Neural Operators for 3D Elastic Seismic Simulations*. LLNL-TR-854521. Lawrence Livermore National Laboratory. URL: https://www.osti.gov/servlets/purl/2001189.

Korres, M., F. Lopez-Caballero, V. Alves Fernandes, F. Gatti, I. Zentner, F. Voldoire, et al. (2022). "Enhanced Seismic Response Prediction of Critical Structures via 3D Regional Scale Physics-Based Earthquake Simulation". In: *Journal of Earthquake Engineering*, pp. 1–29. ISSN: 1363-2469, 1559-808X. DOI: 10.1080/13632469.2021.2009061.

Kristeková, M., J. Kristek, and P. Moczo (2009). "Time-Frequency Misfit and Goodness-of-Fit Criteria for Quantitative Comparison of Time Signals". In: *Geophysical Journal International* 178.2, pp. 813–825. ISSN: 0956540X, 1365246X. DOI: 10.1111/j.1365-246X.2009.04177.x.

Kubo, H., M. Naoi, and M. Kano (2024). "Recent Advances in Earthquake Seismology Using Machine Learning". In: *Earth, Planets and Space* 76.1, p. 36. ISSN: 1880-5981. DOI: 10.1186/s40623-024-01982-0.

Lacour, M. (2023). "Efficient Non-Ergodic Ground-Motion Prediction for Large Datasets". In: *Bulletin of Earthquake Engineering* 21.11, pp. 5209–5232. ISSN: 1570-761X, 1573-1456. DOI: 10.1007/s10518-022-01402-2.

Ladjal, S., A. Newson, and C.-H. Pham (2019). "A PCA-like Autoencoder". In: *arXiv*. DOI: 1904.01277.

Lagerquist, R. and I. Ebert-Uphoff (2022). "Can We Integrate Spatial Verification Methods into Neural Network Loss Functions for Atmospheric Science?" In: *Artificial Intelligence for the Earth Systems* 1.4, e220021. DOI: 10.1175/AIES-D-22-0021.1.

Laporte, M., J. Letort, M. Bertin, and L. Bollinger (2024). "Understanding Earthquake Location Uncertainties Using Global Sensitivity Analysis Framework". In: *Geophysical Journal International*, ggae093. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggae093.

Larroque, C., S. Baize, J. Albaric, H. Jomard, J. Trévisan, M. Godano, et al. (2021). "Seismotectonics of Southeast France: From the Jura Mountains to Corsica". In: *Comptes Rendus. Géoscience* 353.S1, pp. 105–151. ISSN: 1778-7025. DOI: 10.5802/crgeos.69.

LeCun, Y., Y. Bengio, and G. Hinton (2015). "Deep Learning". In: *Nature* 521.7553, pp. 436–444. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14539.

Lee, K. and K. T. Carlberg (2021). "Deep Conservation: A Latent-Dynamics Model for Exact Satisfaction of Physical Conservation Laws". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.1, pp. 277–285. ISSN: 2374-3468, 2159-5399. DOI: 10.1609/aaai.v35i1.16102.

Lehmann, F., F. Gatti, M. Bertin, and D. Clouteau (2024a). "Synthetic Ground Motions in Heterogeneous Geologies: The HEMEW-3D Dataset for Scientific Machine Learning". In: *Earth System Science Data Discussions* 2024, pp. 1–26. DOI: 10.5194/essd-2023-470.

Lehmann, F., F. Gatti, M. Bertin, and D. Clouteau (2022). "Machine Learning Opportunities to Conduct High-Fidelity Earthquake Simulations in Multi-Scale Heterogeneous Geology". In: *Frontiers in Earth Science* 10. ISSN: 2296-6463. DOI: 10.3389/feart.2022.1029160.

— (2023). "Fourier Neural Operator Surrogate Model to Predict 3D Seismic Waves Propagation". In: 5th ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering. Athens, Greece. DOI: 10.7712/120223.10339.20362.

— (2024b). "3D Elastic Wave Propagation with a Factorized Fourier Neural Operator (F-FNO)". In: *Computer Methods in Applied Mechanics and Engineering* 420, p. 116718. ISSN: 0045-7825. DOI: 10.1016/j.cma.2023.116718.

Lehmann, F., F. Gatti, M. Bertin, D. Grenié, and D. Clouteau (n.d.). "Quantifying Uncertainties in Seismic Hazard Analysis with a Fourier Neural Operator". In: *European Journal of Environmental Civil Engineering* (under review) ().

Lehmann, F., F. Gatti, and D. Clouteau (2024c). *Multiple-Input Fourier Neural Operator (MIFNO) for Source-Dependent 3D Elastodynamics*. arXiv: 2404.10115. preprint.

Lehmann, L., M. Ohrnberger, M. Metz, and S. Heimann (2023). "Accelerating Low-Frequency Ground Motion Simulation for Finite Fault Sources Using Neural Networks". In: *Geophysical Journal International*, ggad239. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggad239.

Lerer, B., I. Ben-Yair, and E. Treister (2023). *Multigrid-Augmented Deep Learning for the Helmholtz Equation: Better Scalability with Compact Implicit Layers*. arXiv: 2306.17486. URL: http://arxiv.org/abs/2306.17486. preprint.

Levina, E. and P. Bickel (2004). "Maximum Likelihood Estimation of Intrinsic Dimension". In: *Advances in Neural Information Processing Systems*. Ed. by L. Saul, Y. Weiss, and L. Bottou. Vol. 17. MIT Press. URL: https://proceedings.neurips.cc/paper_files/paper/2004/file/74934548253bcab8490ebd74afed7031-Paper.pdf.

Li, B., H. Wang, X. Yang, and Y. Lin (2022). *Solving Seismic Wave Equations on Variable Velocity Models with Fourier Neural Operator*. arXiv: 2209.12340. preprint.

Li, Z., S. Patil, D. Shu, and A. B. Farimani (2023a). "Latent Neural PDE Solver for Time-Dependent Systems". In: *NeurIPS 2023 AI for Science Workshop*. URL: https://openreview.net/forum?id=iJfPFUvFfy.

Li, Z., N. B. Kovachki, K. Azizzadenesheli, B. liu, K. Bhattacharya, A. Stuart, et al. (2021). "Fourier Neural Operator for Parametric Partial Differential Equations". In: *International Conference on Learning Representations*. DOI: 10.48550/arXiv.2010.08895.

Li, Z., N. B. Kovachki, C. Choy, B. Li, J. Kossaifi, S. P. Otta, et al. (2023b). *Geometry-Informed Neural Operator for Large-Scale 3D PDEs*. arXiv: 2309.00583. URL: http://arxiv.org/abs/2309.00583. preprint.

Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, et al. (2020). *Neural Operator: Graph Kernel Network for Partial Differential Equations*. arXiv: 2003.03485. URL: http://arxiv.org/abs/2003.03485. preprint.

Lin, H., L. Wu, Y. Xu, Y. Huang, S. Li, G. Zhao, et al. (2022). *Non-Equispaced Fourier Neural Solvers for PDEs*. URL: https://arxiv.org/abs/2212.04689. preprint.

Lin, Y., J. Theiler, and B. Wohlberg (2023). "Physics-Guided Data-Driven Seismic Inversion: Recent Progress and Future Opportunities in Full-Waveform Inversion". In: *IEEE Signal Processing Magazine* 40.1, pp. 115–133. ISSN: 1053-5888, 1558-0792. DOI: 10.1109/MSP.2022.3217658.

Liu, B., S. Yang, Y. Ren, X. Xu, P. Jiang, and Y. Chen (2021). "Deep-Learning Seismic Full-Waveform Inversion for Realistic Structural Models". In: *GEOPHYSICS* 86.1, R31–R44. ISSN: 0016-8033, 1942-2156. DOI: 10.1190/geo2019-0435.1.

Loshchilov, I. and F. Hutter (2017). "SGDR: Stochastic Gradient Descent with Warm Restarts". In: ICLR. DOI: arXiv:1608.03983v5.

Lu, L., P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis (2021). "Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators". In: *Nature Machine Intelligence* 3.3, pp. 218–229. ISSN: 2522-5839. DOI: 10.1038/s42256-021-00302-5.

Luzi, L., R. Puglia, E. Russo, M. D'Amico, C. Felicetta, F. Pacor, et al. (2016). "The Engineering Strong-Motion Database: A Platform to Access Pan-European Accelerometric Data". In: *Seismological Research Letters* 87.4, pp. 987–997. ISSN: 0895-0695, 1938-2057. DOI: 10.1785/0220150278.

Ma, Z., K. Azizzadenesheli, and A. Anandkumar (2024). *Calibrated Uncertainty Quantification for Operator Learning via Conformal Prediction*. arXiv: 2402.01960. URL: http://arxiv.org/abs/2402.01960. preprint.

Madariaga, R. (2015). "Seismic Source Theory". In: *Treatise on Geophysics*. Elsevier, pp. 51–71. ISBN: 978-0-444-53803-1. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444538024000701.

Makra, K. and F. J. Chávez-García (2016). "Site Effects in 3D Basins Using 1D and 2D Models: An Evaluation of the Differences Based on Simulations of the Seismic Response of Euroseistest". In: *Bulletin of Earthquake Engineering* 14.4, pp. 1177–1194. ISSN: 1570-761X, 1573-1456. DOI: 10.1007/s10518-015-9862-7.

Mansoor, K., T. Buscheck, X. Yang, S. Carroll, and X. Chen (2020). *LLNL Kimberlina 1.2 NUFT Simulations June 2018 (V2)*. DOI: 10.18141/1603336.

Marano, G. C., M. M. Rosso, A. Aloisio, and G. Cirrincione (2023). "Generative Adversarial Networks Review in Earthquake-Related Engineering Fields". In: *Bulletin of Earthquake Engineering*. ISSN: 1570-761X, 1573-1456. DOI: 10.1007/s10518-023-01645-7.

Marconato, L., P. H. Leloup, C. Lasserre, R. Jolivet, S. Caritg, R. Grandin, et al. (2022). "Insights on Fault Reactivation during the 2019 November 11, Mw 4.9 Le Teil Earthquake in Southeastern France, from a Joint 3-D Geological Model and InSAR Time-Series Analysis". In: *Geophysical Journal International* 229.2, pp. 758–775. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggab498.

Maufroy, E., E. Chaljub, F. Hollender, P.-Y. Bard, J. Kristek, P. Moczo, et al. (2016). "3D Numerical Simulation and Ground Motion Prediction: Verification, Validation and beyond – Lessons from the E2VP Project". In: *Soil Dynamics and Earthquake Engineering* 91, pp. 53–71. ISSN: 02677261. DOI: 10.1016/j.soildyn.2016.09.047.

Michelini, A., S. Cianetti, S. Gaviano, C. Giunchi, D. Jozinović, and V. Lauciani (2021). "INSTANCE – the Italian Seismic Dataset for Machine Learning". In: *Earth System Science Data* 13.12, pp. 5509–5544. ISSN: 1866-3516. DOI: 10.5194/essd-13-5509-2021.

Mishra, S. (2018). "A Machine Learning Framework for Data Driven Acceleration of Computations of Differential Equations". In: *Mathematics in Engineering* 1.1, pp. 118–146. ISSN: 2640-3501. DOI: 10.3934/Mine.2018.1.118.

Molinari, I. and A. Morelli (2011). "EPcrust: A Reference Crustal Model for the European Plate: EPcrust". In: *Geophysical Journal International* 185.1, pp. 352–364. ISSN: 0956540X. DOI: 10.1111/j.1365-246X.2011.04940.x.

Moseley, B., T. Nissen-Meyer, and A. Markham (2020). "Deep Learning for Fast Simulation of Seismic Waves in Complex Media". In: *Solid Earth* 11.4, pp. 1527–1549. ISSN: 1869-9529. DOI: 10.5194/se-11-1527-2020.

Mousavi, S. M. and G. C. Beroza (2020). "A Machine-Learning Approach for Earthquake Magnitude Estimation". In: *Geophysical Research Letters* 47.1. ISSN: 0094-8276, 1944-8007. DOI: 10.1029/2019GL085976.

— (2023). "Machine Learning in Earthquake Seismology". In: *Annual Review of Earth and Planetary Sciences* 51.1, pp. 105–129. ISSN: 0084-6597, 1545-4495. DOI: 10.1146/annurev-earth-071822-100323.

Mousavi, S. M., Y. Sheng, W. Zhu, and G. C. Beroza (2019). "STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI". In: *IEEE Access* 7. DOI: 10.1109/ACCESS.2019.2947848.

Münchmeyer, J., D. Bindi, U. Leser, and F. Tilmann (2021). "The Transformer Earthquake Alerting Model: A New Versatile Approach to Earthquake Early Warning". In: *Geophysical Journal International* 225.1, pp. 646–656. ISSN: 0956-540X. DOI: 10.1093/gji/ggaa609.

Murshed, R. U., K. Noshin, M. A. Zakaria, M. F. Uddin, A. F. M. S. Amin, and M. E. Ali (2023). "Real-Time Seismic Intensity Prediction Using Self-supervised Contrastive GNN for Earthquake Early Warning". In: DOI: 10.48550/ARXIV.2306.14336.

O'Brien, G. S., C. J. Bean, H. Meiland, and P. Witte (2023). "Imaging and Seismic Modelling inside Volcanoes Using Machine Learning". In: *Scientific Reports* 13.1, p. 630. ISSN: 2045-2322. DOI: 10.1038/s41598-023-27738-6.

Olsen, K. B. and J. E. Mayhew (2010). "Goodness-of-Fit Criteria for Broadband Synthetic Seismograms, with Application to the 2008 Mw 5.4 Chino Hills, California, Earthquake". In: *Seismological Research Letters* 81.5, pp. 715–723. ISSN: 0895-0695. DOI: 10.1785/gssrl.81.5.715.

Ovadia, O., A. Kahana, P. Stinis, E. Turkel, and G. E. Karniadakis (2023). *ViTO: Vision Transformer-Operator*. arXiv: 2303.08891. URL: http://arxiv.org/abs/2303.08891. preprint.

Pan, S. J. and Q. Yang (2010). "A Survey on Transfer Learning". In: *IEEE Transactions on Knowledge and Data Engineering* 22.10, pp. 1345–1359. ISSN: 1041-4347. DOI: 10.1109/TKDE.2009.191.

Paolucci, R., C. Smerzini, and M. Vanini (2021). "BB-SPEEDset: A Validated Dataset of Broadband Near-Source Earthquake Ground Motions from 3D Physics-Based Numerical Simulations". In: *Bulletin of the Seismological Society of America* 111.5, pp. 2527–2545. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120210089.

Pecker, A., E. Faccioli, A. Gurpinar, C. Martin, and P. Renault (2017). *An Overview of the SIGMA Research Project*. Springer Cham. Geotechnical, Geological, and Earthquake Engineering. ISBN: 978-3-319-58153-8. URL: https://doi.org/10.1007/978-3-319-58154-5.

Poursartip, B., A. Fathi, and J. L. Tassoulas (2020). "Large-Scale Simulation of Seismic Wave Motion: A Review". In: *Soil Dynamics and Earthquake Engineering* 129, p. 105909. ISSN: 02677261. DOI: 10.1016/j.soildyn.2019.105909.

Qiu, H., Y. Yang, and H. Pan (2023). "Underestimation Modification for Intrinsic Dimension Estimation". In: *Pattern Recognition* 140, p. 109580. ISSN: 00313203. DOI: 10.1016/j.patcog.2023.109580.

Quiñonero-Candela, J., ed. (2009). *Dataset Shift in Machine Learning*. Neural Information Processing Series. Cambridge, Mass: MIT Press. 229 pp. ISBN: 978-0-262-17005-5 978-0-262-54587-7.

Rahaman, N., A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, et al. (2019). "On the Spectral Bias of Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, pp. 5301–5310. URL: https://proceedings.mlr.press/v97/rahaman19a.html.

Rahman, M. A., Z. E. Ross, and K. Azizzadenesheli (2023). "U-NO: U-shaped Neural Operators". In: *Transactions on Machine Learning Research*. ISSN: 2835-8856. URL: https://openreview.net/forum?id=j3oQF9coJd.

Raissi, M., P. Perdikaris, and G. Karniadakis (2019). "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations". In: *Journal of Computational Physics* 378, pp. 686–707. ISSN: 00219991. DOI: 10.1016/j.jcp.2018.10.045.

Rasht-Behesht, M., C. Huber, K. Shukla, and G. E. Karniadakis (2022). "Physics-Informed Neural Networks (PINNs) for Wave Propagation and Full Waveform Inversions". In: *Journal of Geophysical Research: Solid Earth* 127.5. ISSN: 2169-9313, 2169-9356. DOI: 10.1029/2021JB023120.

Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, Mass: MIT Press. 248 pp. ISBN: 978-0-262-18253-9.

Regazzoni, F., S. Pagani, M. Salvador, L. Dede', and A. Quarteroni (2023). *Latent Dynamics Networks (LDNets): Learning the Intrinsic Dynamics of Spatio-Temporal Processes*. arXiv: 2305.00094. URL: http://arxiv.org/abs/2305.00094. preprint.

Rekoske, J. M., A. Gabriel, and D. A. May (2023). "Instantaneous Physics-Based Ground Motion Maps Using Reduced-Order Modeling". In: *Journal of Geophysical Research: Solid Earth* 128.8, e2023JB026975. ISSN: 2169-9313, 2169-9356. DOI: 10.1029/2023JB026975.

Ritz, J.-F., S. Baize, M. Ferry, C. Larroque, L. Audin, B. Delouis, et al. (2020). "Surface Rupture and Shallow Fault Reactivation during the 2019 Mw 4.9 Le Teil Earthquake, France". In: *Communications Earth & Environment* 1.1, p. 10. ISSN: 2662-4435. DOI: 10.1038/s43247-020-0012-z.

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi. Cham: Springer International Publishing, pp. 234–241. ISBN: 978-3-319-24574-4.

Ross, D. A., J. Lim, R.-S. Lin, and M.-H. Yang (2008). "Incremental Learning for Robust Visual Tracking". In: *International Journal of Computer Vision* 77.1, pp. 125–141. ISSN: 1573-1405. DOI: 10.1007/s11263-007-0075-7.

Rosti, A., C. Smerzini, R. Paolucci, A. Penna, and M. Rota (2023). "Validation of Physics-Based Ground Shaking Scenarios for Empirical Fragility Studies: The Case of the 2009 L'Aquila Earthquake". In: *Bulletin of Earthquake Engineering* 21.1, pp. 95–123. ISSN: 1573-1456. DOI: 10.1007/s10518-022-01554-1.

Ruiz, J. A., D. Baumont, P. Bernard, and C. Berge-Thierry (2011). "Modelling Directivity of Strong Ground Motion with a Fractal, K-2, Kinematic Source Model: Modelling Directivity of Strong Ground Motion". In: *Geophysical Journal International* 186.1, pp. 226–244. ISSN: 0956540X. DOI: 10.1111/j.1365-246X.2011.05000.x.

Saltelli, A. (2002). "Making Best Use of Model Evaluations to Compute Sensitivity Indices". In: *Computer Physics Communications* 145.2, pp. 280–297. ISSN: 0010-4655. DOI: 10.1016/S0010-4655(02)00280-1.

Scalise, M., A. Pitarka, J. N. Louie, and K. D. Smith (2021). "Effect of Random 3D Correlated Velocity Perturbations on Numerical Modeling of Ground Motion from the Source Physics Experiment". In: *Bulletin of the Seismological Society of America* 111.1, pp. 139–156. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120200160.

Seriani, G. and S. P. Oliveira (2020). "Numerical Modeling of Mechanical Wave Propagation". In: *La Rivista del Nuovo Cimento* 43.9, pp. 459–514. ISSN: 0393-697X, 1826-9850. DOI: 10.1007/s40766-020-00009-0.

Shearer, P. M. (2019). *Introduction to Seismology*. 3rd ed. Cambridge: Cambridge University Press. ISBN: 978-1-107-18447-3. URL: https://www.cambridge.org/core/books/introduction-to-seismology/C1471C1B553C05997E2BC7EB26D

Shi, Y., G. Lavrentiadis, D. Asimaki, Z. E. Ross, and K. Azizzadenesheli (2024). "Broadband Ground-Motion Synthesis via Generative Adversarial Neural Operators: Development and Validation". In: *Bulletin of the Seismological Society of America*. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120230207.

Shinozuka, M. and G. Deodatis (1991). "Simulation of Stochastic Processes by Spectral Representation". In: *Applied Mechanics Reviews* 44.4, pp. 191–204. ISSN: 0003-6900. DOI: 10.1115/1.3119501.

Smerzini, C., M. Vanini, R. Paolucci, P. Renault, and P. Traversa (2023). "Regional Physics-Based Simulation of Ground Motion within the Rhône Valley, France, during the Mw 4.9 2019 Le Teil Earthquake". In: *Bulletin of Earthquake Engineering* 21.4, pp. 1747–1774. ISSN: 1570-761X, 1573-1456. DOI: 10.1007/s10518-022-01591-w.

Smerzini, C., R. Paolucci, and M. Stupazzini (2011). "Comparison of 3D, 2D and 1D Numerical Approaches to Predict Long Period Earthquake Ground Motion in the Gubbio Plain, Central Italy". In: *Bulletin of Earthquake Engineering* 9.6, pp. 2007–2029. ISSN: 1573-1456. DOI: 10.1007/s10518-011-9289-8.

Smith, L. N. and N. Topin (2018). *Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates*. preprint.

Sobol, I. (2001). "Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates". In: *Mathematics and Computers in Simulation* 55.1, pp. 271–280. ISSN: 0378-4754. DOI: 10.1016/S0378-4754(00)00270-6.

Sochala, P., F. De Martin, and O. Le Maître (2020). "Model Reduction for Large-Scale Earthquake Simulation in an Uncertain 3D Medium". In: *International Journal for Uncertainty Quantification* 10.2, pp. 101–127. ISSN: 2152-5080. DOI: 10.1615/Int.J.UncertaintyQuantification.2020031165.

Soize, C. (2008). "Tensor-Valued Random Fields for Meso-Scale Stochastic Model of Anisotropic Elastic Microstructure and Probabilistic Analysis of Representative Volume Element Size". In: *Probabilistic Engineering Mechanics* 23.2-3, pp. 307–323. ISSN: 02668920. DOI: 10.1016/j.probengmech.2007.12.019.

Song, C., T. Alkhalifah, and U. B. Waheed (2022). "A Versatile Framework to Solve the Helmholtz Equation Using Physics-Informed Neural Networks". In: *Geophysical Journal International* 228.3, pp. 1750–1762. ISSN: 0956-540X. DOI: 10.1093/gji/ggab434.

Song, C., Y. Liu, P. Zhao, T. Zhao, J. Zou, and C. Liu (2023). "Simulating Multi-Component Elastic Seismic Wavefield Using Deep Learning". In: *IEEE Geoscience and Remote Sensing Letters*, pp. 1–1. ISSN: 1545-598X, 1558-0571. DOI: 10.1109/LGRS.2023.3250522.

Song, C. and Y. Wang (2022). "High-Frequency Wavefield Extrapolation Using the Fourier Neural Operator". In: *Journal of Geophysics and Engineering* 19.2, pp. 269–282. ISSN: 1742-2132. DOI: 10.1093/jge/gxac016.

Soto Moncada, V. P. (2023). "Numerical Simulation of Wave Propagation in 3D Basin Models and Their Effects on Nonlinear Structures Response". Paris-Saclay: Université Paris Saclay.

Stupazzini, M., R. Paolucci, and H. Igel (2009). "Near-Fault Earthquake Ground-Motion Simulation in the Grenoble Valley by a High-Performance Spectral Element Code". In: *Bulletin of the Seismological Society of America* 99.1, pp. 286–301. ISSN: 0037-1106. DOI: 10.1785/0120080274.

Sudret, B. (2008). "Global Sensitivity Analysis Using Polynomial Chaos Expansions". In: *Reliability Engineering & System Safety* 93.7, pp. 964–979. ISSN: 09518320. DOI: 10.1016/j.ress.2007.04.002.

Svay, A., V. Perron, A. Imtiaz, I. Zentner, R. Cottereau, D. Clouteau1, et al. (2017). "Spatial Coherency Analysis of Seismic Ground Motions from a Rock Site Dense Array Implemented during the Kefalonia 2014 Aftershock Sequence". In: *Earthquake Engineering & Structural Dynamics*. Publisher: John Wiley & Sons, Ltd 46.12, pp. 1895–1917. ISSN: 0098-8847. DOI: 10.1002/eqe.2881.

Ta, Q.-A., D. Clouteau, and R. Cottereau (2010). "Modeling of Random Anisotropic Elastic Media and Impact on Wave Propagation". In: *European Journal of Computational Mechanics* 19.1-3, pp. 241–253. ISSN: 1779-7179, 1958-5829. DOI: 10.3166/ejcm.19.241-253.

Tamhidi, A., N. Kuehn, S. F. Ghahari, A. J. Rodgers, M. D. Kohler, E. Taciroglu, et al. (2022). "Conditioned Simulation of Ground-Motion Time Series at Uninstrumented Sites Using Gaussian Process Regression". In: *Bulletin of the Seismological Society of America* 112.1, pp. 331–347. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120210054.

Tekawade, A., Z. Liu, P. Kenesei, T. Bicer, F. D. Carlo, R. Kettimuthu, et al. (2021). "3d Autoencoders For Feature Extraction In X-Ray Tomography". In: *2021 IEEE International Conference on Image Processing (ICIP)*. 2021 IEEE International Conference on Image Processing (ICIP), pp. 3477–3481. ISBN: 2381-8549. DOI: 10.1109/ICIP42928.2021.9506494.

Thomas, J. W. (1995). "Hyperbolic Equations". In: *Numerical Partial Differential Equations: Finite Difference Methods*. Ed. by J. W. Thomas. New York, NY: Springer New York, pp. 205–259. ISBN: 978-1-4899-7278-1. URL: https://doi.org/10.1007/978-1-4899-7278-1_6.

Thompson, E. M., L. G. Baise, and R. E. Kayen (2007). "Spatial Correlation of Shear-Wave Velocity in the San Francisco Bay Area Sediments". In: *Soil Dynamics and Earthquake Engineering* 27.2, pp. 144–152. ISSN: 02677261. DOI: 10.1016/j.soildyn.2006.05.004.

Touhami, S. (2020). "Numerical Modeling of Seismic Field and Soil Interaction : Application to the Sedimentary Basin of Argostoli (Greece)". Paris: Université Paris Saclay. URL: http://www.theses.fr/2020UPASC007.

Touhami, S., F. Gatti, F. Lopez-Caballero, R. Cottereau, L. de Abreu Corrêa, L. Aubry, et al. (2022). "SEM3D: A 3D High-Fidelity Numerical Earthquake Simulator for Broadband (0–10 Hz) Seismic Response Prediction at a Regional Scale". In: *Geosciences* 12.3, p. 112. ISSN: 2076-3263. DOI: 10.3390/geosciences12030112.

Touhami, S., F. Lopez-Caballero, and D. Clouteau (2021). "A Holistic Approach of Numerical Analysis of the Geology Effects on Ground Motion Prediction: Argostoli Site Test". In: *Journal of Seismology* 25.1, pp. 115–140. ISSN: 1383-4649, 1573-157X. DOI: 10.1007/s10950-020-09961-0.

Tran, A., A. Mathews, L. Xie, and C. S. Ong (2023). "Factorized Fourier Neural Operators". In: *The Eleventh International Conference on Learning Representations*. URL: https://openreview.net/forum?id=tmIiMPl4IPa.

Vallage, A., L. Bollinger, J. Champenois, C. Duverger, A. G. Trilla, B. Hernandez, et al. (2021). "Multitechnology Characterization of an Unusual Surface Rupturing Intraplate Earthquake: The *M* L 5.4 2019 Le Teil Event in France". In: *Geophysical Journal International* 226.2, pp. 803–813. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggab136.

Veinante-Delhaye, A. and J. P. Santoire (1980). "Sismicite Recente de l'Arc Sud-Armoricain et Du Nord-Ouest Du Massif Central; Mecanismes Au Foyer et Tectonique". In: *Bulletin de la Société Géologique de France* S7-XXII.1, pp. 93–102. ISSN: 0037-9409. DOI: 10.2113/gssgfbull.S7-XXII.1.93.

Venturi, S. and T. Casey (2023). "SVD Perspectives for Augmenting DeepONet Flexibility and Interpretability". In: *Computer Methods in Applied Mechanics and Engineering* 403, p. 115718. ISSN: 00457825. DOI: 10.1016/j.cma.2022.115718.

Vergniault, C. and J.-L. Mari (2020). "1 Shear Velocity Measurement in Boreholes". In: Mari, J.-L. and C. Vergniault. *Well Seismic Surveying and Acoustic Logging*. EDP Sciences, pp. 15–48. ISBN: 978-2-7598-2263-8. URL: https://www.degruyter.com/document/doi/10.1051/978-2-7598-2263-8.c003/html.

Vyas, J. C., P. M. Mai, M. Galis, E. M. Dunham, and W. Imperatori (2018). "Mach Wave Properties in the Presence of Source and Medium Heterogeneity". In: *Geophysical Journal International* 214.3, pp. 2035–2052. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggy219.

Wang, A., S. Li, J. Lu, H. Zhang, B. Wang, and Z. Xie (2023). "Prediction of PGA in Earthquake Early Warning Using a Long Short-Term Memory Neural Network". In: *Geophysical Journal International*, ggad067. ISSN: 0956-540X, 1365-246X. DOI: 10.1093/gji/ggad067.

Wang, F. and K. Sett (2016). "Time-Domain Stochastic Finite Element Simulation of Uncertain Seismic Wave Propagation through Uncertain Heterogeneous Solids". In: *Soil Dynamics and Earthquake Engineering* 88, pp. 369–385. ISSN: 02677261. DOI: 10.1016/j.soildyn.2016.07.011.

— (2019). "Time Domain Stochastic Finite Element Simulation towards Probabilistic Seismic Soil-Structure Interaction Analysis". In: *Soil Dynamics and Earthquake Engineering* 116, pp. 460–475. ISSN: 02677261. DOI: 10.1016/j.soildyn.2018.10.021.

Wang, S., H. Wang, and P. Perdikaris (2021). "Learning the Solution Operator of Parametric Partial Differential Equations with Physics-Informed DeepONets". In: *Science Advances* 7.40, p. 9. DOI: 10.1126/sciadv.abi8605.

Wang, Z., A. Bovik, H. Sheikh, and E. Simoncelli (2004). "Image Quality Assessment: From Error Visibility to Structural Similarity". In: *IEEE Transactions on Image Processing* 13.4, pp. 600–612. ISSN: 1057-7149. DOI: 10.1109/TIP.2003.819861.

Wei, W. and L.-Y. Fu (2022). “Small-Data-Driven Fast Seismic Simulations for Complex Media Using Physics-Informed Fourier Neural Operators”. In: *Geophysics* 87.6, T435–T446. ISSN: 0016-8033, 1942-2156. DOI: 10.1190/geo2021-0573.1.

Weiss, K., T. M. Khoshgoftaar, and D. Wang (2016). “A Survey of Transfer Learning”. In: *Journal of Big Data* 3.1, p. 9. ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6.

Wellmann, F. and G. Caumon (2018). “3-D Structural Geological Models: Concepts, Methods, and Uncertainties”. In: *Advances in Geophysics*. Vol. 59. Elsevier, pp. 1–121. ISBN: 978-0-12-815208-9. URL: https://doi.org/10.1016/bs.agph.2018.09.001.

Wen, G., Z. Li, Q. Long, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson (2023). “Real-Time High-Resolution CO2 Geological Storage Prediction Using Nested Fourier Neural Operators”. In: *Energy & Environmental Science* 16.4, pp. 1732–1741. ISSN: 1754-5692. DOI: 10.1039/d2ee04204e.

Wiener, N. (1938). “The Homogeneous Chaos”. In: *American Journal of Mathematics* 60.4, pp. 897–936.

Witte, P. A., T. Konuk, E. Skjetne, and R. Chandra (2023). “Fast CO2 Saturation Simulations on Large-Scale Geomodels with Artificial Intelligence-Based Wavelet Neural Operators”. In: *International Journal of Greenhouse Gas Control* 126, p. 103880. ISSN: 1750-5836. DOI: 10.1016/j.ijggc.2023.103880.

Wolny, A., L. Cerrone, A. Vijayan, R. Tofanelli, A. V. Barro, M. Louveaux, et al. (2020). “Accurate and Versatile 3D Segmentation of Plant Tissues at Cellular Resolution”. In: *eLife* 9. Ed. by C. S. Hardtke, D. C. Bergmann, D. C. Bergmann, and M. Graeff, e57613. ISSN: 2050-084X. DOI: 10.7554/eLife.57613.

Wu, M.-H., J. Wang, and P.-E. Chiang (2022). “Cumulative Absolute Velocity (CAV) Seismic Hazard Assessment for Taiwan”. In: *Journal of Earthquake Engineering* 26.7, pp. 3440–3460. ISSN: 1363-2469, 1559-808X. DOI: 10.1080/13632469.2020.1803161.

Wu, R.-S. and K. Aki (1988). “Introduction: Seismic Wave Scattering in Three-Dimensionally Heterogeneous Earth”. In: *pure and applied geophysics* 128.1, pp. 1–6. ISSN: 1420-9136. DOI: 10.1007/BF01772587.

Xu, W., K. Chen, T. Han, H. Chen, W. Ouyang, and L. Bai (2024). “ExtremeCast: Boosting Extreme Value Prediction for Global Weather Forecast”. In: DOI: 10.48550/ARXIV.2402.01295.

Yang, Y., A. F. Gao, K. Azizzadenesheli, R. W. Clayton, and Z. E. Ross (2023). “Rapid Seismic Waveform Modeling and Inversion with Neural Operators”. In: *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1. ISSN: 0196-2892, 1558-0644. DOI: 10.1109/TGRS.2023.3264210.

Yang, Y., A. F. Gao, J. C. Castellanos, Z. E. Ross, K. Azizzadenesheli, and R. W. Clayton (2021). “Seismic Wave Propagation and Inversion with Neural Operators”. In: *The Seismic Record* 1.3, pp. 126–134. ISSN: 2694-4006. DOI: 10.1785/0320210026.

Yu, J., L. Zhang, Q. Li, Y. Li, W. Huang, Z. Sun, et al. (2021). “3D Autoencoder Algorithm for Lithological Mapping Using ZY-1 02D Hyperspectral Imagery: A Case Study of Liuyuan Region”. In: *Journal of Applied Remote Sensing* 15.04. ISSN: 1931-3195. DOI: 10.1117/1.JRS.15.042610.

Zeng, Q., S. Feng, B. Wohlberg, and Y. Lin (2021). “InversionNet3D: Efficient and Scalable Learning for 3D Full Waveform Inversion”. In: *CoRR*. URL: https://arxiv.org/abs/2103.14158.

Zhang, E., A. Kahana, E. Turkel, R. Ranade, J. Pathak, and G. E. Karniadakis (2022). *A Hybrid Iterative Numerical Transferable Solver (HINTS) for PDEs Based on Deep Operator Network and Relaxation Methods*. preprint.

Zhang, T., D. Trad, and K. Innanen (2023). “Learning to Solve the Elastic Wave Equation with Fourier Neural Operators”. In: *Geophysics*, pp. 1–63. ISSN: 0016-8033, 1942-2156. DOI: 10.1190/geo2022-0268.1.

Zhang, W., Z. Zhang, H. Fu, Z. Li, and X. Chen (2019). “Importance of Spatial Resolution in Ground Motion Simulations With 3-D Basins: An Example Using the Tangshan Earthquake”. In: *Geophysical Research Letters* 46.21, pp. 11915–11924. ISSN: 0094-8276, 1944-8007. DOI: 10.1029/2019GL084815.

Zhu, C., G. Weatherill, F. Cotton, M. Pilz, D. Y. Kwak, and H. Kawase (2021). “An Open-Source Site Database of Strong-Motion Stations in Japan: K-NET and KiK-net (v1.0.0)”. In: *Earthquake Spectra* 37.3, pp. 2126–2149. ISSN: 8755-2930, 1944-8201. DOI: 10.1177/8755293020988028.

Zhu, M., S. Feng, Y. Lin, and L. Lu (2023). “Fourier-DeepONet: Fourier-enhanced Deep Operator Networks for Full Waveform Inversion with Improved Accuracy, Generalizability, and Robustness”. In: *Computer Methods in Applied Mechanics and Engineering* 416. ISSN: 0045-7825. DOI: 10.1016/j.cma.2023.116300.

Zou, C., K. Azizzadenesheli, Z. E. Ross, and R. W. Clayton (2023). *Deep Neural Helmholtz Operators for 3D Elastic Wave Propagation and Inversion*. arXiv: 2311.09608. URL: http://arxiv.org/abs/2311.09608. preprint.

Zou, Z., X. Meng, A. F. Psaros, and G. E. Karniadakis (2022). *NeuralUQ: A Comprehensive Library for Uncertainty Quantification in Neural Differential Equations and Operators*. arXiv: 2208.11866. URL: http://arxiv.org/abs/2208.11866. preprint.

Appendix A

Appendices

A.1 Le Teil simulations

	network	latitude	longitude	elevation	type	epicentral distance (km)	PGA (cm/s^2)	PGV (cm/s)
VIVF	LDG	44.86° N	4.67° E	618 m	velocimeter	37 km	-	-
OGCB	FR	44.85° N	5.1° E	773 m	velocimeter	50 km	1.80	0.11
P53	OHP	44.74° N	4.97° E	157 m	velocimeter	34 km	7.53	0.75
CAI535	SNCF	44.66° N	4.92° E	212 m	velocimeter	25 km	16.32	0.64
CRU1	3C	44.64° N	4.76° E	77 m	accelerometer	14 km	44.91	1.21
P54	SNCF	44.61° N	4.88° E	148 m	velocimeter	19 km	19.29	0.80
OGDF	FR	44.56° N	5.07° E	709 m	velocimeter	32 km	4.12	0.35
CAI550	SNCF	44.53° N	4.82° E	129 m	velocimeter	11 km	96.00	2.76
P55	SNCF	44.48° N	4.77° E	133 m	velocimeter	8 km	69.65	3.18
OGCC	FR	44.43° N	5.55° E	1250 m	velocimeter	70 km	0.77	0.21
CAI567	SNCF	44.4° N	4.74° E	60 m	velocimeter	14 km	15.14	1.76
ADHE	3C	44.37° N	4.77° E	90 m	velocimeter	17 km	4.17	0.31
BANN	FR	44.37° N	4.16° E	343 m	accelerometer	44 km	4.96	0.42
P56	SNCF	44.32° N	4.7° E	48 m	velocimeter	22 km	8.24	1.15
OGLP	RA	44.31° N	4.69° E	46 m	accelerometer	23 km	6.10	1.15
P60	SNCF	44.16° N	4.74° E	42 m	velocimeter	39 km	6.78	0.82
CAI604	SNCF	44.09° N	4.77° E	25 m	velocimeter	48 km	2.92	0.15
P61	SNCF	43.99° N	4.74° E	44 m	velocimeter	59 km	7.38	0.77
SAUF	FR	43.94° N	5.11° E	230 m	accelerometer	73 km	1.33	0.14
P71	SNCF	43.94° N	4.72° E	109 m	velocimeter	64 km	0.69	0.09
P62	SNCF	43.93° N	4.78° E	17 m	velocimeter	66 km	3.72	0.28
CAI633	SNCF	43.9° N	4.89° E	38 m	velocimeter	70 km	1.58	0.15

Table A.1: Stations with available records ordered by decreasing latitude. PGA (resp. PGV): maximum of the East-West and North-South components for the Peak Ground Acceleration (resp. Velocity). VIVF is a vertical sensor only. LDG = Geophysical and Detection Laboratory of the French Alternative Energies and Atomic Energy Commission (CEA). FR, OHP, 3C, RA are networks of RAP-Résif (French Permanent Accelerometric Network). SNCF = French National Railway Company

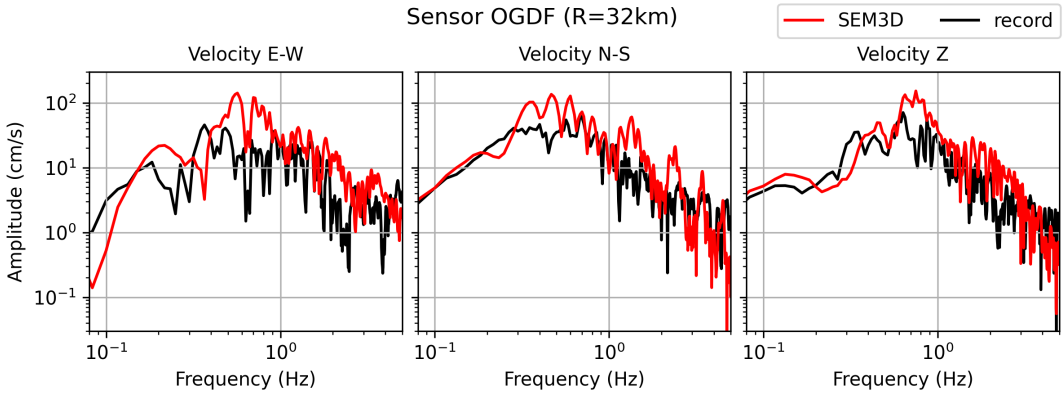


Figure A.1: Frequency response spectra of the numerical simulation (in red) obtained with the 1D geological model and a point source. Comparison with seismograms records (in black). Velocities are given in the East-West (E-W), North-South (N-S), and vertical (Z) directions for the station OGDF. R denotes the epicentral distance.

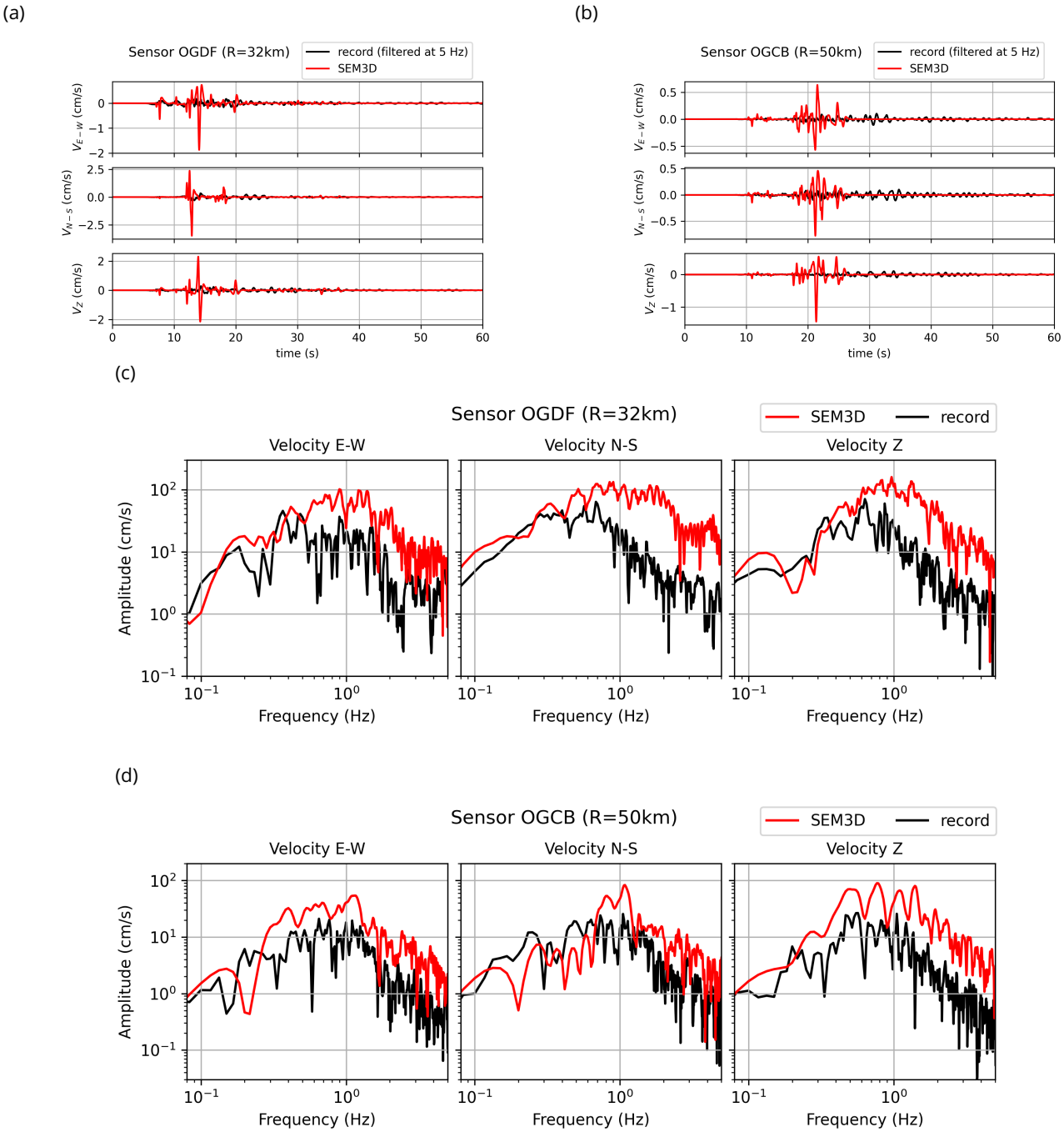


Figure A.2: Results of the numerical simulation (in red) obtained with the 3D geological model and a point source. Comparison with seismograms records (in black) filtered at 5 Hz. Velocities are given in the East-West (E-W), North-South (N-S), and vertical (Z) directions.

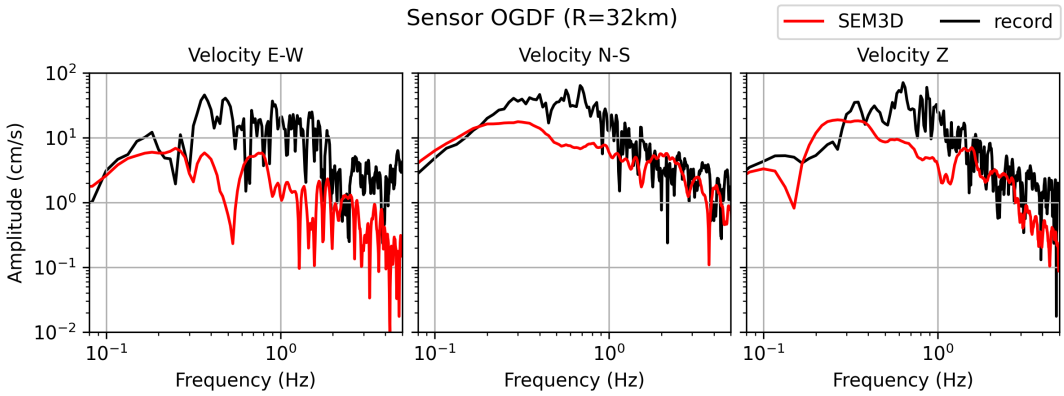


Figure A.3: Frequency response spectra of the numerical simulation (in red) obtained with the 3D geological model and a kinematic fault model. Comparison with seismograms records (in black). Velocities are given in the East-West (E-W), North-South (N-S), and vertical (Z) directions for the station OGDF.

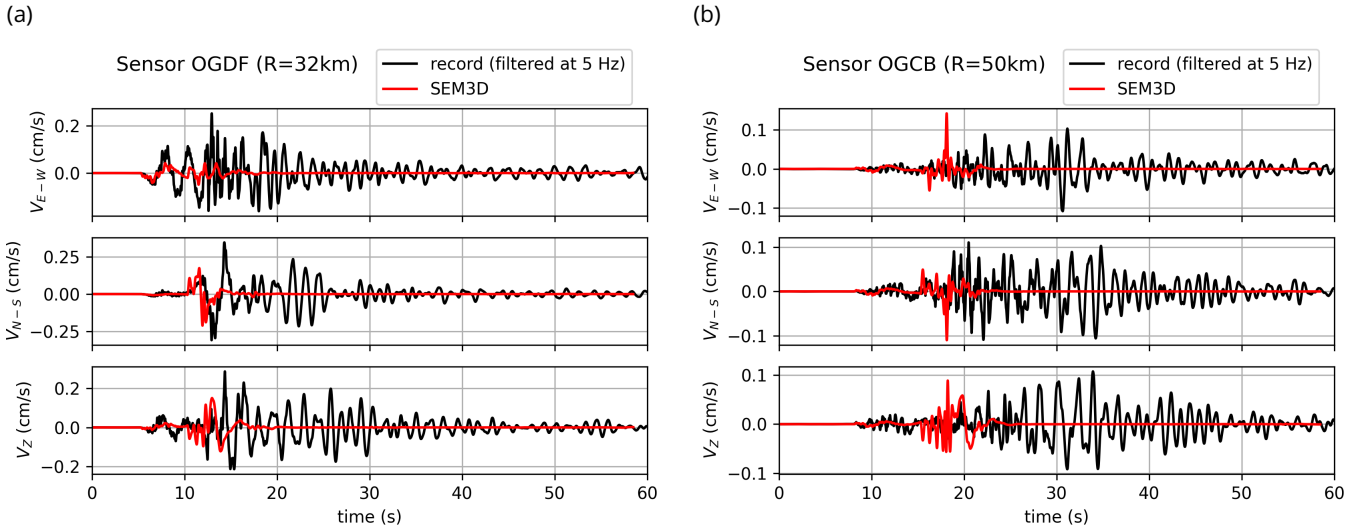


Figure A.4: Results of the numerical simulation (in red) obtained with the 3D geological model and a kinematic fault model. Comparison with seismograms records (in black) filtered at 5 Hz. Velocities are given in the East-West (E-W), North-South (N-S), and vertical (Z) directions in stations OGDF A.4a and OGCB A.4b

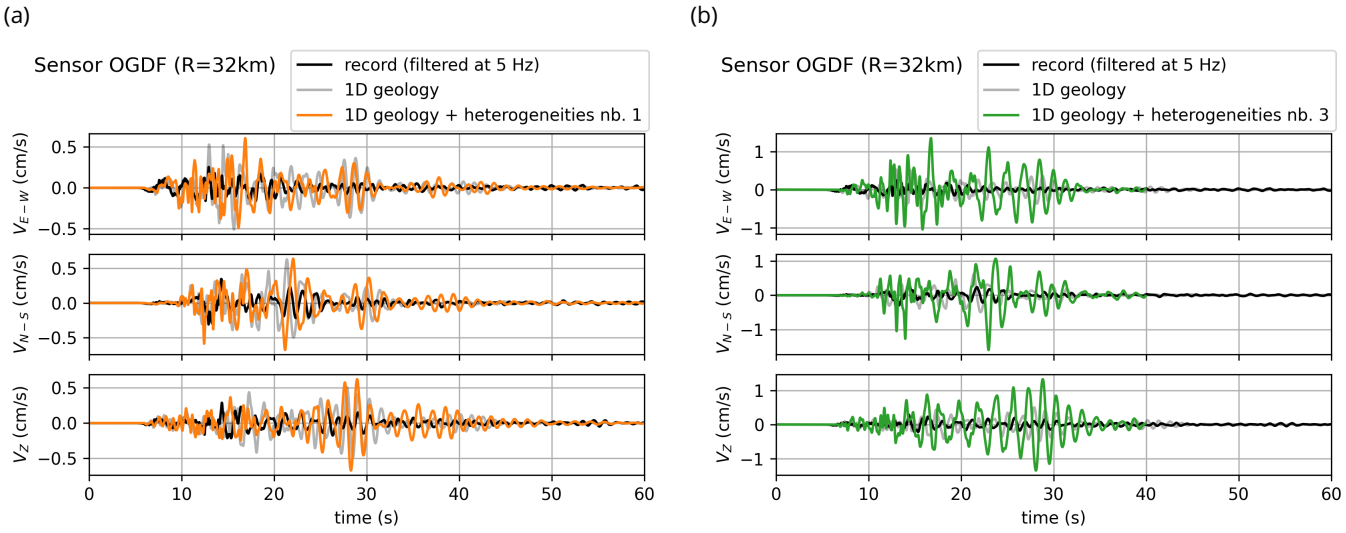


Figure A.5: Velocities obtained with two random fields added to the 1D geological model in OGDF station. The third random field is shown in Figure 1.16.

A.2 Dimensionality reduction

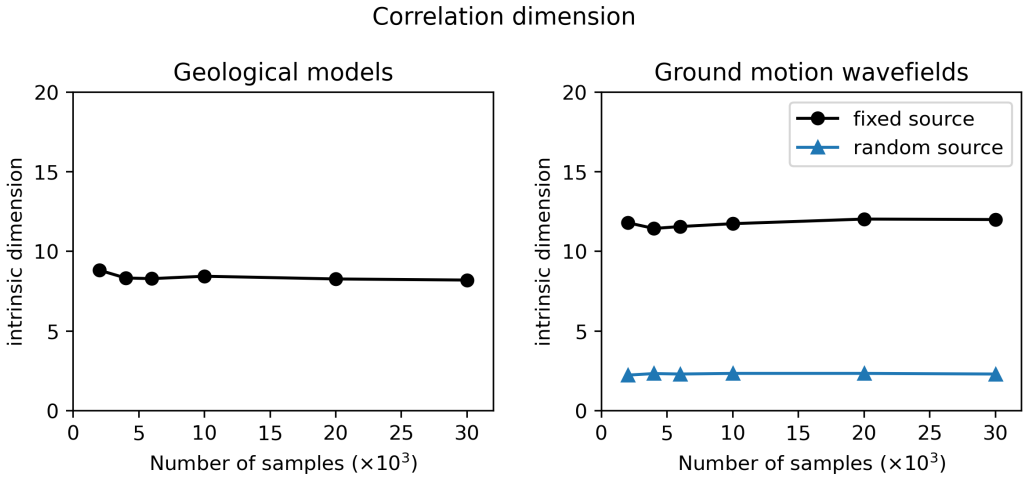
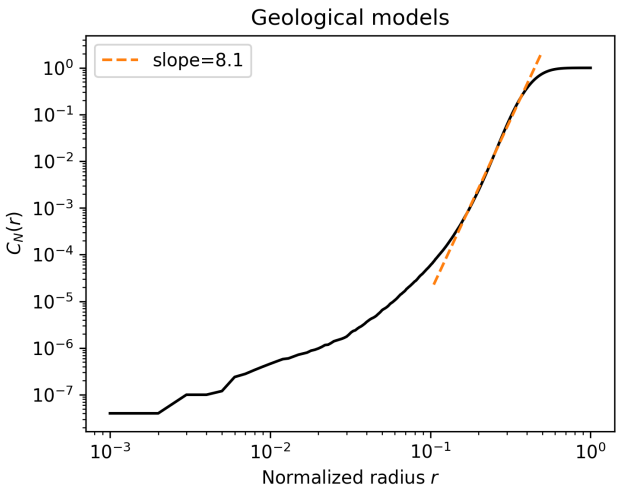
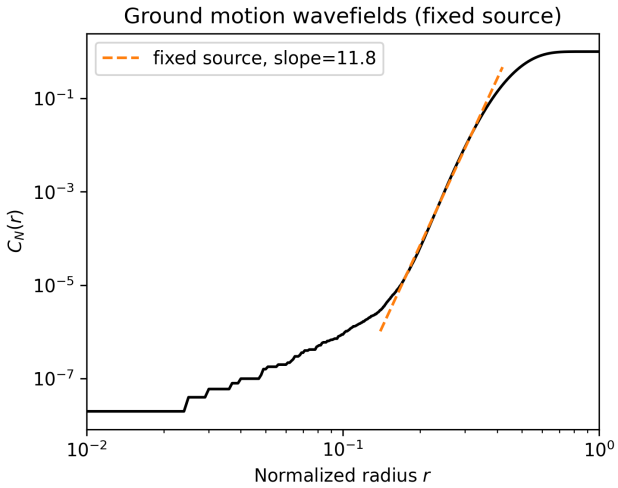


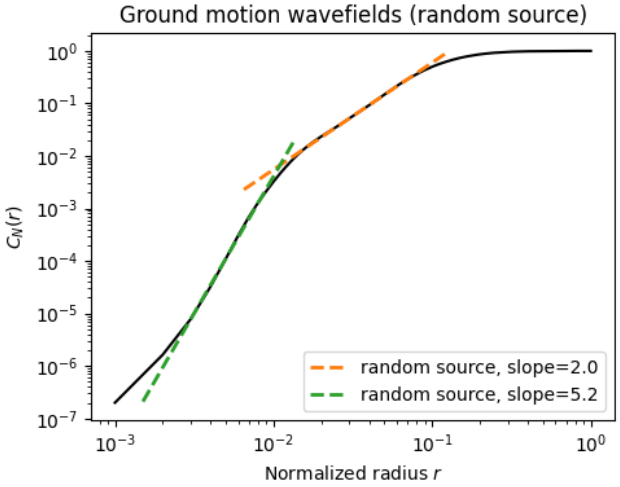
Figure A.6: Correlation dimension (y-axis) as a function of the dataset size (x-axis) for geological models (left) and ground motion wavefields (right). For ground motion, the HEMEW-3D database is used for the fixed source (black line) and the HEMEW^S-3D database corresponds to the blue line.



(a) Correlation dimension for 30 000 geological models



(b) Correlation dimension for 30 000 ground motion wavefields with a fixed source (HEMEW-3D database)



(c) Correlation dimension for 30 000 ground motion wavefields with a random source (HEMEW^S-3D database)

Figure A.7: The correlation dimension $C_N(r)$ is computed from the number of samples being at (Euclidean) distance smaller than r for different values of r (Equation 2.10). Then, the correlation dimension is obtained as the slope of the linear part in the log-log representation.

A.3 Fourier Neural Operators architectures

This section provides details on the layers in each FNO variant: FNO (Tab. A.2), U-NO (Tab. A.3), and F-FNO (Tab. A.4). The number of parameters is indicated for each layers ("k" denotes thousands and "m" denote millions).

	output shape	# parameters
	b, 32, 32, 32, 1	
concatenate grid	b, 32, 32, 32, 6	
uplift	b, 32, 32, 32, 16	200
permute and optional padding	b, 16, 32, 32, 32	
Fourier layer 1 (modes=16, 16, 16)	b, 16, 32, 32, 64	8.39m
Fourier layer 2 (modes=16, 16, 32)	b, 16, 32, 32, 128	16.8m
Fourier layer 3 (modes=16, 16, 32)	b, 16, 32, 32, 256	16.8m
Fourier layer 4 (modes=16, 16, 32)	b, 16, 32, 32, 320	16.8m
unpadding and permute	b, 32, 32, 320, 16	
projection Q_E	b, 32, 32, 320, 1	1.1k
total		58.7m

Table A.2: For each group of operations in the FNO, output shape and number of parameters in this group. In the output shape column, dimensions refer to (batch size, x, y, z or t, channels). The number of channels is 16 in all layers. The number of modes in (x, y, z or t) is indicated for each Fourier layer.

	output shape	# parameters
	b, 32, 32, 32, 1	
concatenate grid	b, 32, 32, 32, 6	
uplift	b, 32, 32, 32, 16	200
permute and optional padding	b, 16, 32, 32, 32	
Fourier layer 1 (modes=12, 12, 12)	b, 16, 24, 24, 24	3.54m
Fourier layer 2 (modes=9, 9, 9)	b, 16, 18, 18, 18	1.49m
Fourier layer 3 (modes=6, 6, 7)	b, 16, 13, 13, 13	516k
Fourier layer 4 (modes=4, 4, 5)	b, 16, 8, 8, 8	164k
Fourier layer 5 (modes=4, 4, 5)	b, 16, 13, 13, 17	164k
Fourier layer 6 (modes=6, 6, 9)	b, 16, 18, 18, 34	1.33m
Fourier layer 7 (modes=9, 9, 17)	b, 16, 24, 24, 64	5.64m
Fourier layer 8 (modes=12, 12, 20)	b, 16, 32, 32, 320	11.8m
unpadding and permute	b, 32, 32, 320, 16	
projection Q_E	b, 32, 32, 320, 1	2.17k
total		24.6m

Table A.3: For each group of operations in the U-NO, output shape and number of parameters in this group. In the output shape column, dimensions refer to (batch size, x, y, z or t, channels). The number of channels is 16 in all layers. The number of modes in (x, y, z or t) is indicated for each Fourier layer.

	output shape	# parameters
	b, 32, 32, 32, 1	
concatenate grid	b, 32, 32, 32, 4	
uplift	b, 32, 32, 32, 16	96
permute and optional padding	b, 16, 32, 32, 32	
F-Fourier layer 1 (modes=16, 16, 16)	b, 16, 32, 32, 32	26.8k
F-Fourier layer 2 (modes=16, 16, 16)	b, 16, 32, 32, 32	26.8k
F-Fourier layer 3 (modes=16, 16, 16)	b, 16, 32, 32, 32	26.8k
F-Fourier layer 4 (modes=16, 16, 16)	b, 16, 32, 32, 32	26.8k
F-Fourier layer 5 (modes=16, 16, 16)	b, 16, 32, 32, 64	26.8k
F-Fourier layer 6 (modes=16, 16, 32)	b, 16, 32, 32, 128	35.0k
F-Fourier layer 7 (modes=16, 16, 32)	b, 16, 32, 32, 256	35.0k
F-Fourier layer 8 (modes=16, 16, 32)	b, 16, 32, 32, 320	35.0k
unpadding and permute	b, 32, 32, 320, 16	
projection Q_E	b, 32, 32, 320, 1	2.4k
total		246k

Table A.4: F-FNO with 8 layers. For each group of operations in the F-FNO, output shape and number of parameters in this group. In the output shape column, dimensions refer to (batch size, x, y, z or t, channels). The number of channels is 16 in all layers. The number of modes in (x, y, z or t) is indicated for each factorized Fourier layer (F-Fourier layer).

A.3.1 Main results on the Factorized Fourier Neural Operator (F-FNO)

The influence of the training database size was investigated by reducing the database of 27 000 training samples to 12 000 samples. Figure A.8 shows that a small database (N_{train}=12 000) is detrimental but increasing from 22 000 to 27 000 training samples has no visible effect on the prediction error.

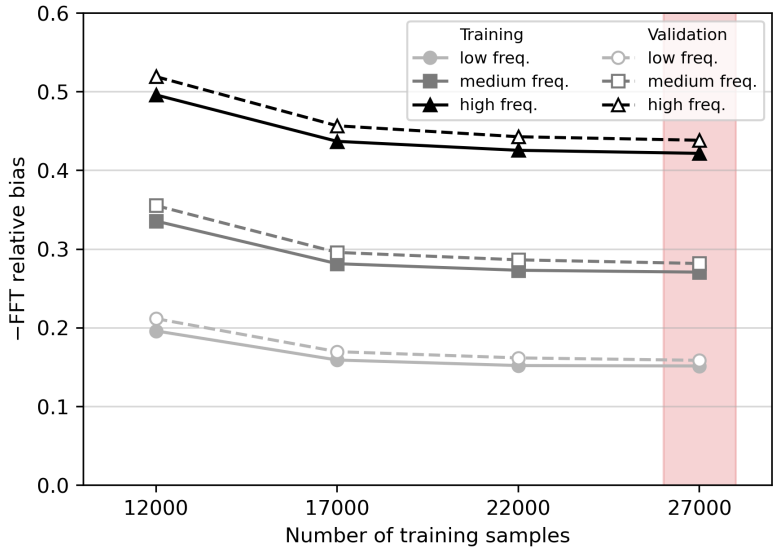


Figure A.8: frequency bias as a function of the number of training samples for the U-NO. frequency bias is given for low frequency (0-1Hz, light grey), medium frequency (1-2 Hz, drak grey), and high frequency (2-5Hz, high frequency). Average results are reported for 1000 training samples (filled markers and solid lines) and 1000 validation samples (unfilled markers and dashed lines). Note that the y-axis gives the opposite of the frequency bias.

The original U-NO model has three skip connections (Fig. 3.5). The importance of the skip connections was quantified by removing them progressively from top to bottom. Results for two skip connections in Fig. A.9 correspond to a model without the upper skip connection (between the Fourier layers F_1 and F_8).

Evaluating the accuracy of predictions with frequency biases suggests that the two upper connections are not useful. The lowest errors are even found for the U-NO with a single skip connection (between the Fourier layers F_3 and F_6).

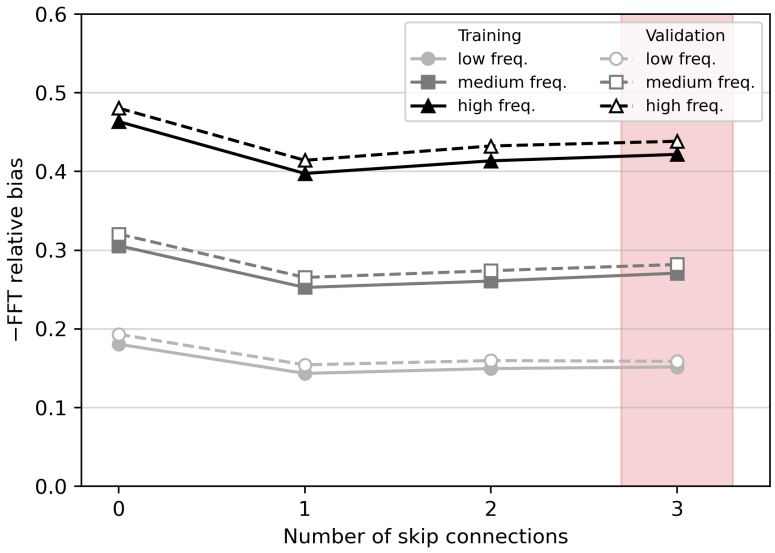


Figure A.9: frequency bias as a function of the number of skip connections in the U-NO. frequency bias is given for low frequency (0-1Hz, light grey), medium frequency (1-2 Hz, drak grey), and high frequency (2-5Hz, high frequency). Average results are reported for 1000 training samples (filled markers and solid lines) and 1000 validation samples (unfilled markers and dashed lines). Note that the y-axis gives the opposite of the frequency bias.

A.3.2 F-FNO hyperparameters

This section presents the influence of hyperparameters on the prediction accuracy and complements the main text in Section 3.3.3. The default F-FNO has 8 layers, 16 channels ($d_v = 16$), 16 Fourier modes in the first and second dimension ($M_{\ell,1} = M_{\ell,2} = 16, \forall \ell$) and 32 Fourier modes in the third dimension except for the first layer ($M_{1,3} = 16$, $M_{\ell,3} = 32, \forall \ell \geq 2$). With nominal settings, training was performed on 4 Nvidia A100 GPUs for 350 epochs with a learning rate of 0.0006 halved on plateau, and the relative Mean Absolute Error as loss function. Activation functions are Gaussian Error Linear Units (GeLU, Hendrycks and Gimpel 2023). Early stopping was used when the validation loss did not improve. The default configuration is identified with the red area in the following figures.

The first hyperparameter tested is the number of channels (d_v). It has a significant effect on the predictions accuracy since Fig. 3.14a shows a steep reduction of the frequency biases when the number of channels increases. Improvements are similar for all frequency ranges. For high frequencies, the relative frequency bias shrinks from −58 % to −30 % when increasing the number of channels from 8 to 28. However, it does not seem beneficial to go beyond 28 channels. Note that since all models underestimate the frequency content, y-axes display the oppositive of the frequency biases. Also, the best models have the lowest amplitude of frequency biases.

As already seen in Tab. 3.4, the number of layers considerably influences the F-FNO accuracy. Figure 3.14a details the relative frequency biases for models having between 4 and 28 layers. The most striking improvements occur between 4 and 16 layers but the biases keep reducing when the number of layers is increased beyond 16. The linear decrease between 20 and 28 layers even suggests that deeper models could be beneficial.

The number of Fourier modes selected in each factorized Fourier layer is a key hyperparameter but it seems to have much less influence than the number of layers or the number of channels. The maximum number of Fourier modes is determined by the variables size. Since inputs and outputs have a spatial size of 32, at most 16 spatial Fourier modes can be computed. Models with 8 (Fig. A.10 top left), 12 (Fig. A.10 top right), and 16 (Fig. A.10 bottom) spatial Fourier modes were tested. Along the third dimension, inputs are of size 32 (depth of the geological model) and outputs are of size 320 (number of time steps of the surface wavefields). Table A.5 indicates the number of points and modes in the third dimension for the reference model. When testing a model with 20 Fourier modes in the third dimension for instance, the first five layers were preserved with 16 modes and the three last layers were reduced to 20 modes.

	layer 1	layer 2	layer 3	layer 4	layer 5	layer 6	layer 7	layer 8
Third dimension	32	32	32	32	64	128	256	320
Number of modes $M_{\ell,3}$	16	16	16	16	16	32	32	32

Table A.5: Number of points in the third dimension and number of Fourier modes in the reference F-FNO model.

Figure A.10 shows that 16 spatial Fourier modes are slightly preferable than 12 or 8. Indeed, with 8 modes in the third dimension, the high-frequency frequency bias reduces from −43 % at 8 spatial modes to −39 % at 16 spatial modes. When using 8 or 12 spatial modes, the frequency bias slightly decreases when increasing the number of modes in the third dimension. However, with 16 spatial modes, increasing the number of modes in the third dimension does not improve the results (bottom panel in Fig. A.10). In particular, this means that the larger errors on the high-frequency components of the time series are not due to a limited representation of temporal patterns.

Some other hyperparameters had almost no impact on the predictions' accuracy; their detailed results are given in the appendix. Firstly, padding is often added on the sides of the domain to account for the lack of periodicity in the inputs. While the reference model has no padding, we tested adding 2 to 6 zero pixels to the x and y sides of the inputs. Figure A.11 shows that the frequency biases are the same for all padding values. Therefore, padding is not necessary in our model.

Some versions of the FNO also proposed modifying the grid of coordinates that are concatenated with the inputs. In the original implementation, each coordinate is represented by a linear grid between 0 and 1. This configuration was used in the reference model. An alternative method relies on a sinusoidal decomposition of the coordinates where, for instance, the x coordinate will be described by two grids: $(\cos x_i, \sin x_i)$, $0 \leq x_i \leq 2\pi$. Our experiments show that the sinusoidal decomposition slightly altered the predictions accuracy (Tab. A.6).

Finally, we mention that our model requires increasing the third dimension from a size of 32 (corresponding to the depth of the geological models) to a size of 320 (corresponding to the number of time steps in the surface wavefields). Since this is unusual practice, we tested three different strategies of increasing the dimension to ensure that this did not impact the results (Tab. A.7).

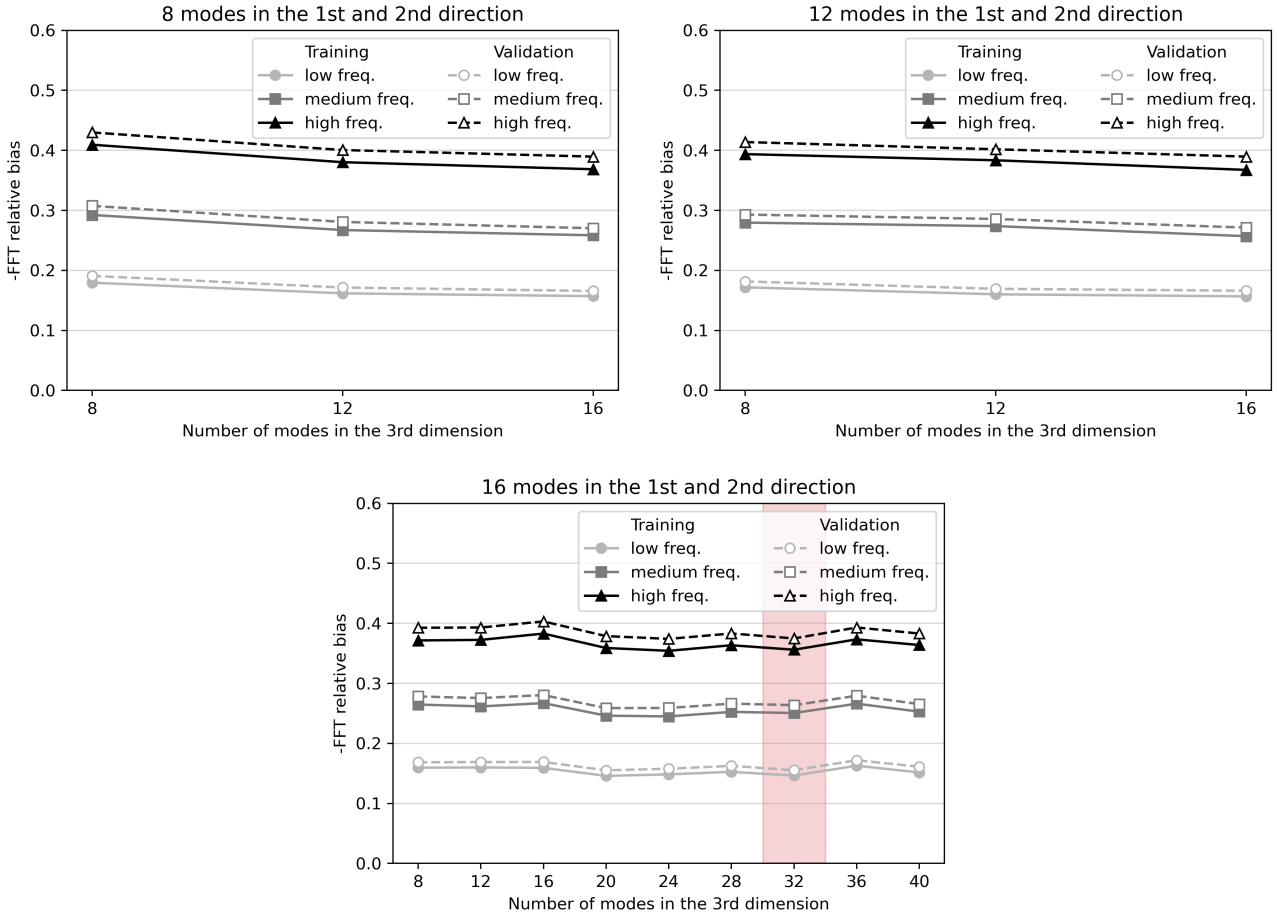


Figure A.10: Relative frequency bias depending on the number of Fourier modes. In each panel, the number of spatial modes (first and second dimensions) is fixed. The reference model has 16 spatial modes and 32 modes in the third dimension (red area).

A.3.3 F-FNO training strategy

Besides considerations on the model hyperparameters, training choices such as the loss function, the number of training samples, the batch size, and the learning rate are crucial to obtain a model with a satisfying accuracy.

Common loss functions to train neural operators involve pixel-wise metrics such as the (relative) MAE and (relative) MSE. We investigated linear combinations of these two loss metrics by constructing the loss function as a linear combination of relative MAE and relative MSE:

$$\mathcal{L}(\cdot) = \omega\, rMAE(\cdot) + (1-\omega)\, rMSE(\cdot), \quad 0 \leq \omega \leq 1$$

It is important to note that predictions are assessed with frequency biases, which are independent from the loss function. Figure A.12 shows a clear improvement for all frequency ranges when using only the relative MAE.

When training deep neural networks, larger datasets (almost) always improve the predictions' accuracy. However, in the context of supervised learning when data are expensive to acquire, one tries to limit the size of datasets. Our dataset contains 27 000 training data and Fig. A.13 shows similar results for models trained with 22 000 and 27 000 samples. This suggests that our database is well-suited for training the 3D F-FNO.

The influence of the learning rate was firstly evaluated by varying the initial learning rate between 1×10^{-4} and 1×10^{-3}. For small learning rates (1×10^{-4} and 2×10^{-4}), the convergence rate is too low and the loss function reaches a sub-optimal plateau (Fig. A.17). For larger learning rates, it is noticeable that the training remains very stable for learning rates up to 1×10^{-3} (we observed divergence issues above this value, not shown). Analysing the Fourier relative biases in Fig. A.14 (left) suggests that a learning rate of 6×10^{-4} may be preferred.

In the default training configuration, the learning rate was halved when the validation loss did not improve for 10

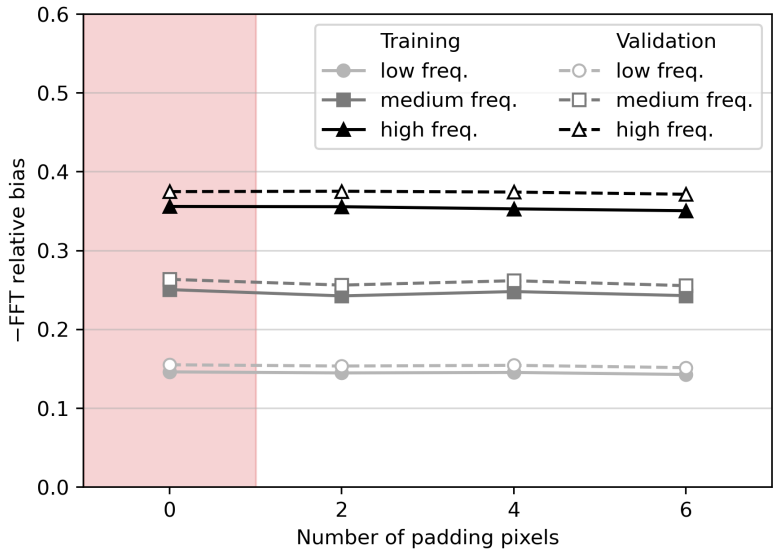


Figure A.11: frequency bias as a function of the number of padding pixels. frequency bias is given for low frequency (0-1Hz, light grey), medium frequency (1-2 Hz, drak grey), and high frequency (2-5Hz, high frequency). Average results are reported for 1000 training samples (filled markers and solid lines) and 1000 validation samples (unfilled markers and dashed lines). Note that the y-axis gives the opposite of the frequency bias.

Model	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$
Cartesian grid (reference)	-0.15 $\pm$ 0.14	-0.26 $\pm$ 0.19	-0.37 $\pm$ 0.26
Sinus grid	-0.18 $\pm$ 0.15	-0.29 $\pm$ 0.20	-0.40 $\pm$ 0.26

Table A.6: Mean and standard deviation of relative frequency biases for 1000 validation samples.

epochs. This well-known strategy aims at exploring potentially close local minima by reducing the gradient descent step once the algorithm starts converging. Among the diversity of alternative learning rate schedulers, we tested the cosine annealing with warm restarts (Loshchilov and Hutter 2017) and the one cycle scheduler (Smith and Topin 2018). The cosine annealing with warm restarts begins with a large learning rate (6×10^{-4}) that is sinusoidally reduced for 70 epochs where it is brutally increased to its initial value (see Fig. A.18). The periodic increases aim at helping the gradient descent algorithm escape local minima where it could get trapped during training. Our experiments show that the loss function decreased more slowly than the reference loss during the first period due to the diminishing learning rate (Fig. A.14, right). Moreover, the successive restarts did not allow the loss to reduce significantly, thereby reaching a plateau higher than the reference.

Contrary to the cosine annealing scheduler, the one cycle scheduler starts with a very low learning rate that should allow the gradient descent to reach some initial stability. Then, the learning rate increases up to 2.5×10^{-4} to explore vast portions of the loss landscape. Figure A.14 (right) shows that the loss function degrades and becomes noisy at this point. Then, the learning rate is slowly reduced to let the gradient descent converge to a local minima (see Fig. A.18). However, the convergence value was significantly higher than the value obtained with the reference scheduler. Therefore, with the parameters used in these experiments, halving the learning rate on plateaus remains the best strategy.

Lastly, we explored the variability of results due to the weights' initialization. With the default configuration, models having 4, 8, 12, and 16 layers were trained with five different seeds for the random distribution of initial weights. The largest variability in relative frequency biases was 0.07, which remains smaller than the differences between two models with a distinct number of layers (Fig. A.15).

Model (3rd dimension)	rFFT_{low}	rFFT_{mid}	$\text{rFFT}_{\text{high}}$
32-32-32-32-64-128-256-320 (reference)	-0.15 $\pm$ 0.14	-0.26 $\pm$ 0.19	-0.37 $\pm$ 0.26
32-48-64-96-128-192-256-320	-0.16 $\pm$ 0.14	-0.27 $\pm$ 0.19	-0.38 $\pm$ 0.26
32-320-320-320-320-320-320-320	-0.16 $\pm$ 0.14	-0.27 $\pm$ 0.20	-0.38 $\pm$ 0.26

Table A.7: Mean and standard deviation of relative frequency biases for 1000 validation samples.

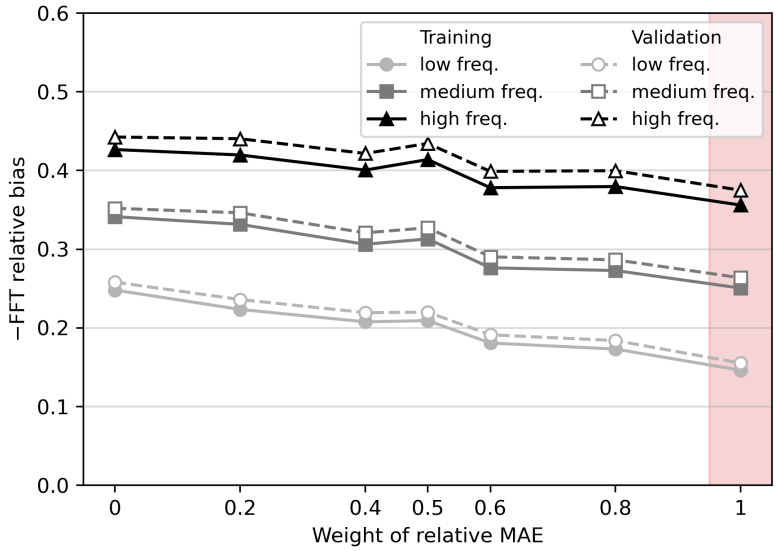


Figure A.12: Relative frequency bias as a function of the weight ω of the relative MAE in the loss function. The reference model is evaluated with only the rMAE (red area).

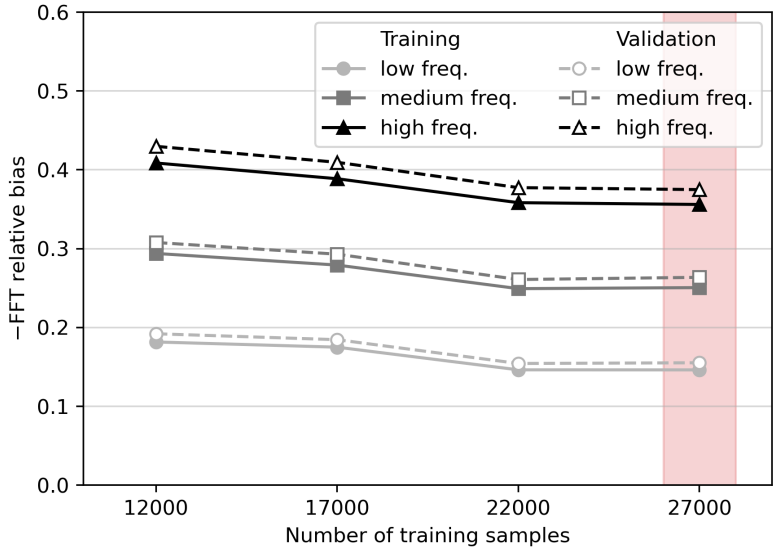


Figure A.13: Relative frequency bias as a function of the number of training samples N_{train}. The reference model is trained with 27 000 samples (red area).

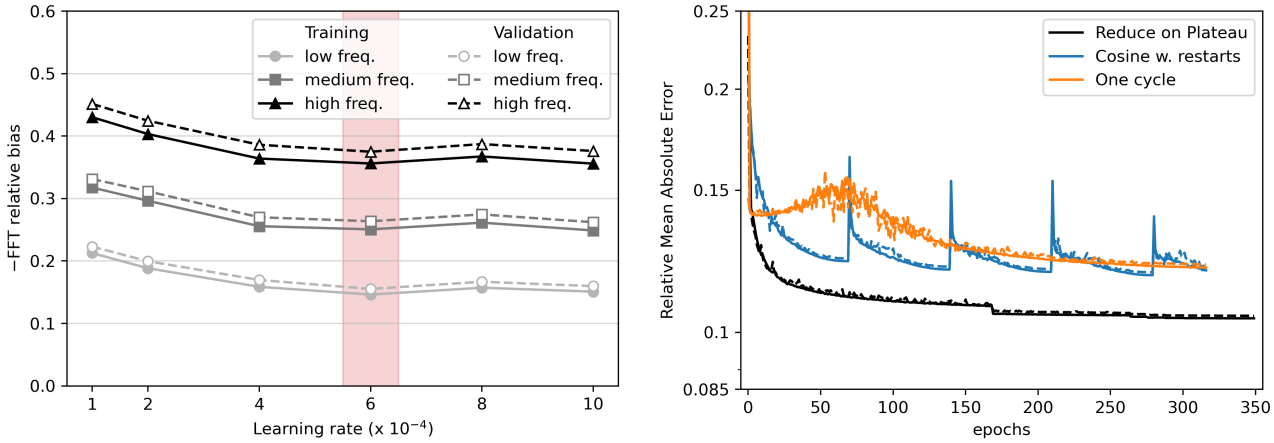


Figure A.14: Influence of the learning rate on the predictions. (left): relative frequency biases for initial learning rates between 1×10^{-4} and 1×10^{-3} halved on plateau. (right): Evolution of the loss function during training for three learning rate schedulers: reduce on plateau (black, reference model), cosine annealing with warm restarts (blue), one cycle (orange). Solid lines show training loss functions and dashed lines show validation loss.

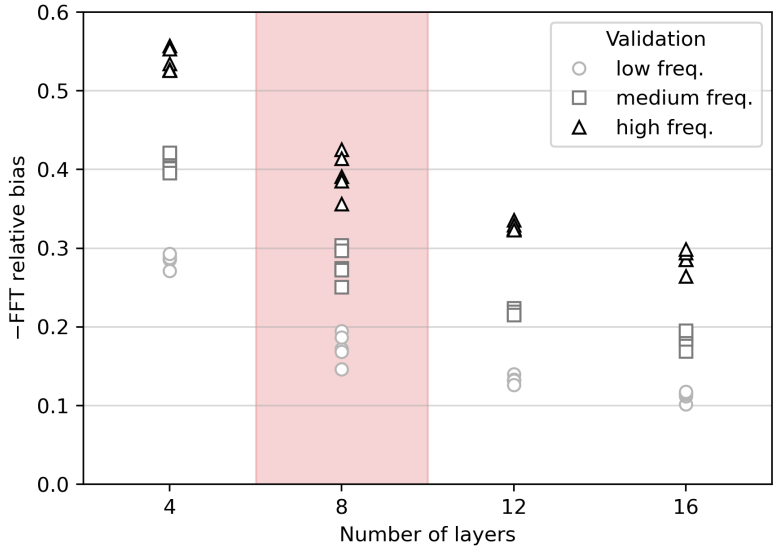


Figure A.15: Relative frequency bias for models with 4, 8, 12, and 16 layers when using five different initialization values for each model.

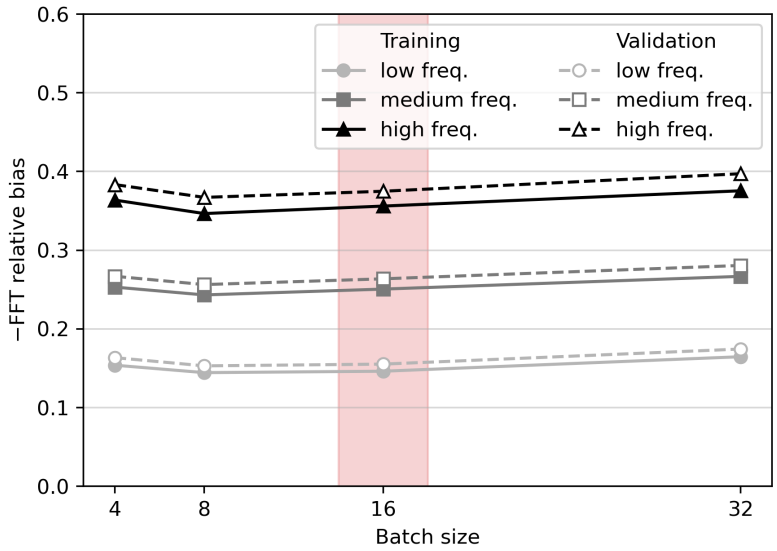


Figure A.16: Relative frequency bias as a function of the number of elements per mini-batch. The reference model is trained with mini-batches of 16 (red area).

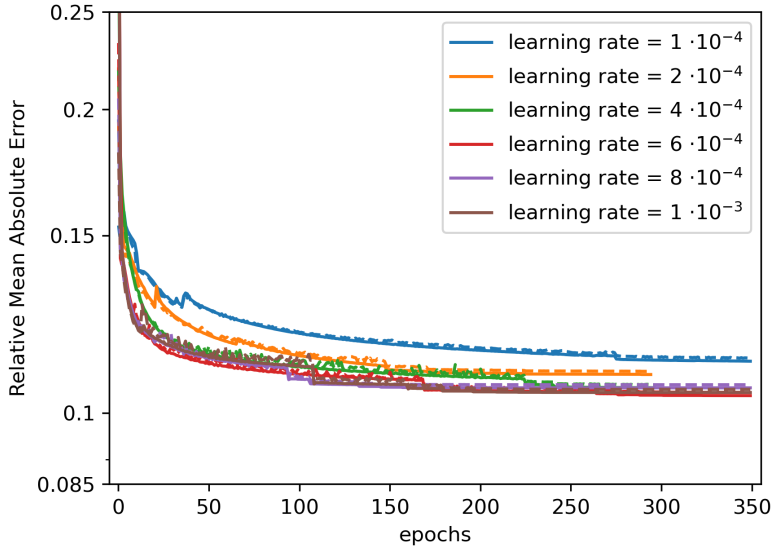


Figure A.17: Evolution of the training loss for several initial rates between 1×10^{-4} and 1×10^{-3}

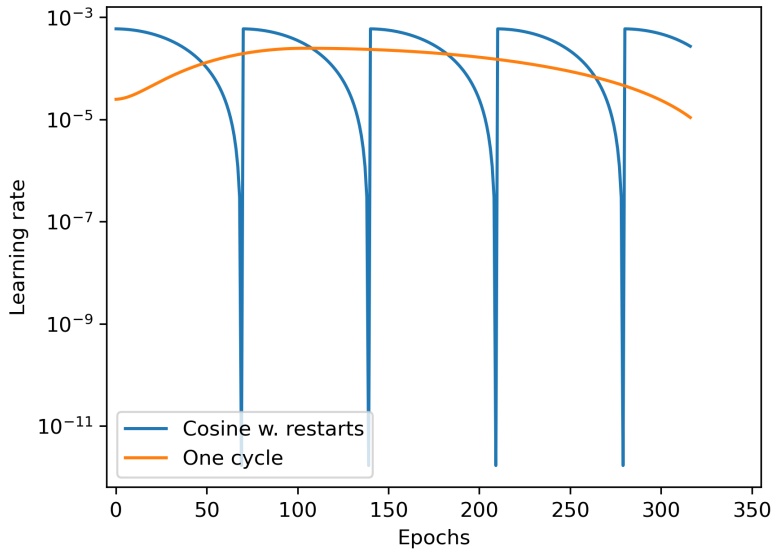


Figure A.18: Evolution of the learning rate using the cosine annealing with warm restarts scheduler (blue) and the one cycle scheduler (orange).

A.3.4 F-FNO robustness to noise

Velocity models in the HEMEW-3D database were created from random fields with spatially coherent features. In reality, they could be corrupted by some white noise originating, for instance, from measurement errors. In this case, it is desirable that predictions are not too sensitive to noise in the velocity models. To test this, we added white noise as a centered Gaussian distribution $\mathcal{N}(0, \sigma\overline{\sigma_{V_S}})$ with a noise level σ varying from 0.01 to 0.25 and $\overline{\sigma_{V_S}}$ being the mean standard deviation of velocity models.

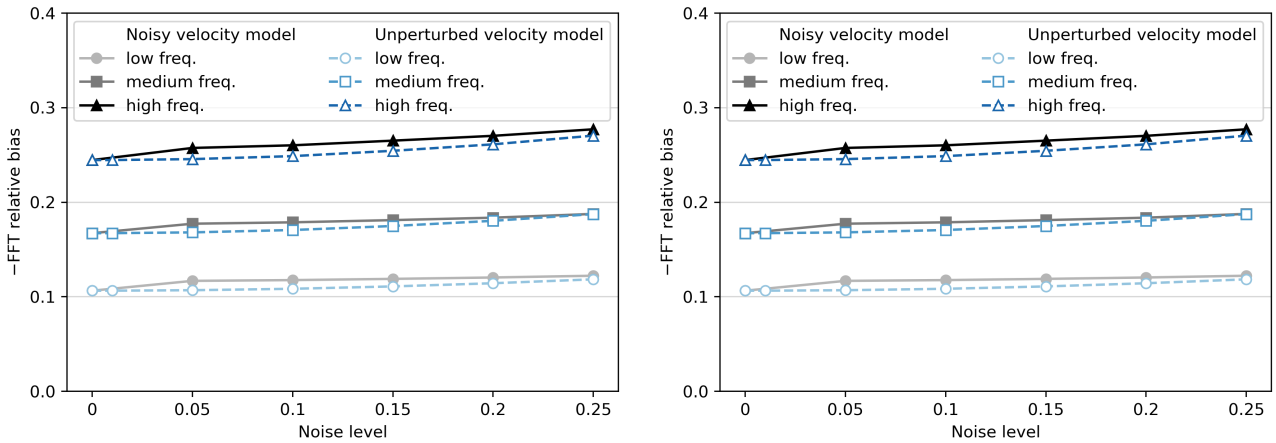


Figure A.19: Relative frequency bias (left) and relative RMSE (right) for increasing noise level in the velocity models. Predictions are compared with surface wavefields generated in unperturbed velocity models (blue lines) and noisy velocity models (black lines).

Figure A.19 firstly compares the F-FNO predictions on noisy velocity models with the ground truth simulations in the unperturbed velocity models (blue dashed curves). The relative frequency bias degrades slightly but the degradation remains very limited, from −24.4 % without noise to −27.0 % with a 0.25 noise level (for the high-frequency bias). For the relative RMSE, this represents an increase from 16.6 % to 17.1 %. These results show that the F-FNO is very robust to noise added in the inputs.

Secondly, the predictions on noisy velocity models are compared with the true surface wavefields generated in the same noisy velocity models (black curves in Fig. A.20). Then, the prediction error increases slightly more, especially in the high-frequency range (relative high-frequency frequency bias degrades from −24.4 % without noise to −27.7 % with a 0.25 noise level). This can be better understood by examining the time series in Fig. A.20. When simulating the true wavefield in the noisy velocity model, the small-scale heterogeneities added by the noise disperse and diffract seismic waves. This is reflected by the presence of small-scale fluctuations after 3.5 s in Fig. A.20 (black curves, right panels). It has already been mentionned that the F-FNO predictions have limited high-frequency accuracy, thereby explaining that prediction errors are higher when comparing with simulations in noisy velocity models.

However, when studying noise robustness, the main interest remains in having predictions on noisy inputs close to predictions on unperturbed inputs. This is achieved by the F-FNO to a great extent, as shown by our first experiment.

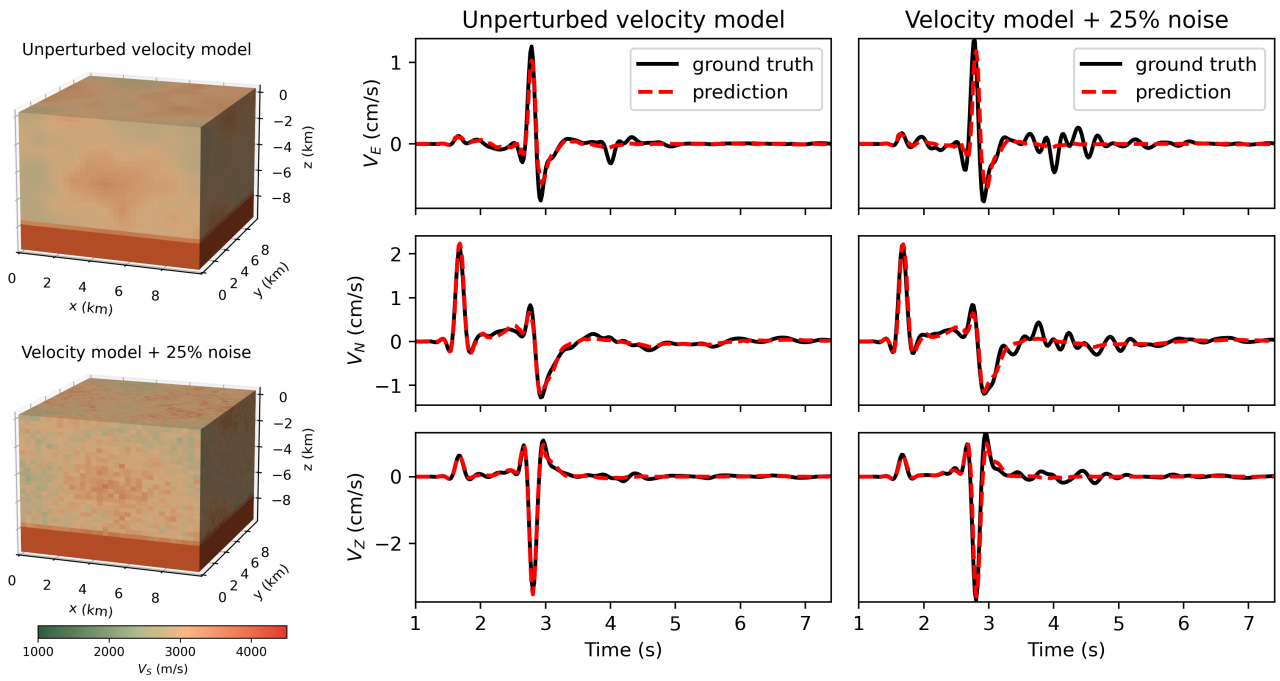


Figure A.20: Surface wavefield at one sensor synthetized with the unperturbed velocity model (black curves, left panel) and the noisy velocity model (black curves, right panel). F-FNO predictions are shown for the unperturbed velocity model (red dashed curves, left panel) and the noisy velocity model (red dashed curves, right panel). The corresponding velocity models are illustrated on the left. Time series of velocity wavefield are given in the three components (East-West, North-South, vertical).

A.4 Multiple Input Fourier Neural Operator (MIFNO)

A.4.1 Metrics analyses

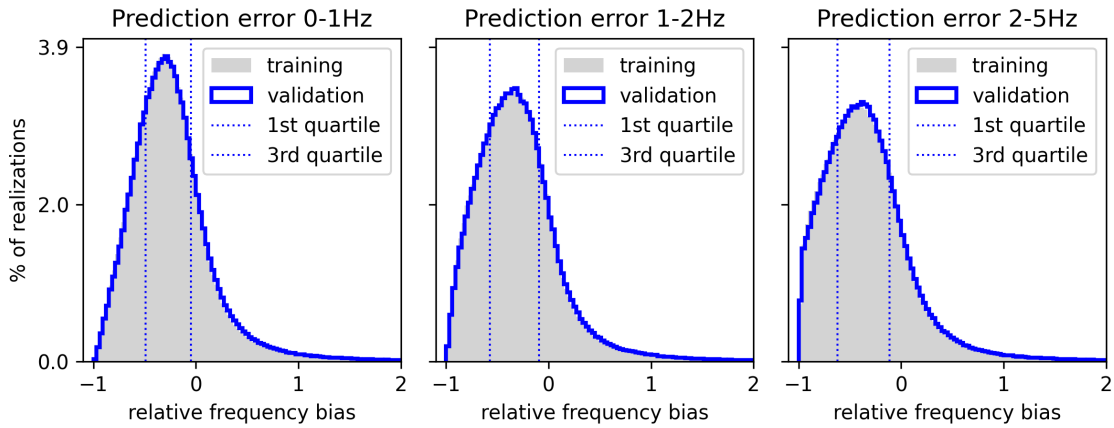


Figure A.21: Distribution of the frequency bias for each sensor and each sample in the training (grey area) and validation (blue line) dataset.

Table A.8 compares the 8-layer MIFNO taking as input the 6 components of the moment tensor (M_{xx}, M_{yy}, M_{zz}, M_{xy}, M_{xz}, M_{yz}) or the equivalent representation with 3 angles (strike, dip, rake).

Model	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$	EG	PG
angle	0.341 ; 0.589	-0.545 ; -0.095	-0.63 ; -0.15	-0.68 ; -0.18	6.32 ; 7.89	8.07 ; 9.17
moment	0.338 ; 0.589	-0.536 ; -0.091	-0.64 ; -0.16	-0.70 ; -0.20	6.29 ; 7.86	8.06 ; 9.16

Table A.8: 1st and 3rd quartiles of the metrics computed on 1000 validation samples. rRMSE: relative RMSE (0 is best), $rFFT_{low}$: relative frequency bias 0-1Hz (0 is best), $rFFT_{mid}$: relative frequency bias 1-2Hz (0 is best), $rFFT_{high}$: relative frequency bias 2-5Hz (0 is best), EG: envelope Goodness-of-Fit (10 is best), PG: phase Goodness-of-Fit (10 is best). For frequency biases, negative values indicate underestimation.

A.4.2 Comparison with baseline models

Table A.9 compares a 16-layer F-FNO with a 16-layer MIFNO predicting surface velocity wavefields when the source has a fixed position and orientation. It should be noted that the MIFNO was not specifically trained on this database.

Dataset with a fixed source position and fixed source orientation

Model	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$	EG	PG
F-FNO	0.14 ; 0.25	-0.30 ; 0.03	-0.39 ; 0.01	-0.44 ; 0.01	7.09 ; 8.37	8.49 ; 9.32
MIFNO	0.20 ; 0.31	-0.40 ; -0.03	-0.51 ; -0.06	-0.56 ; -0.08	6.36 ; 7.65	7.71 ; 8.83

Table A.9: 1st and 3rd quartiles of the metrics computed on 1000 validation samples. rRMSE: relative RMSE (0 is best), $rFFT_{low}$: relative frequency bias 0-1Hz (0 is best), $rFFT_{mid}$: relative frequency bias 1-2Hz (0 is best), $rFFT_{high}$: relative frequency bias 2-5Hz (0 is best), EG: envelope Goodness-of-Fit (10 is best), PG: phase Goodness-of-Fit (10 is best). For frequency biases, negative values indicate underestimation.

Figure A.22 illustrates the F-FNO and MIFNO predictions when the source has a fixed position and orientation. Envelope GOFs of the F-FNO are between 8.3 and 8.7, MIFNO ones are between 7.5 and 8.7. Phase GOFs of the F-FNO are between 9.2 and 9.6, MIFNO ones between 8.2 and 9.4.

Dataset with a random source position and fixed source orientation

Model	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$	EG	PG
F-FNO 8 layers	0.14 ; 0.24	-0.31 ; 0.01	-0.41 ; -0.03	-0.48 ; -0.04	6.71 ; 8.05	8.26 ; 9.19
MIFNO 8 layers	0.13 ; 0.23	-0.27 ; 0.05	-0.37 ; -0.00	-0.43 ; -0.00	6.93 ; 8.24	8.40 ; 9.26

Table A.10: 1st and 3rd quartiles of the metrics computed on 1000 validation samples. rRMSE: relative RMSE (0 is best), $rFFT_{low}$: relative frequency bias 0-1Hz (0 is best), $rFFT_{mid}$: relative frequency bias 1-2Hz (0 is best), $rFFT_{high}$: relative frequency bias 2-5Hz (0 is best), EG: envelope Goodness-of-Fit (10 is best), PG: phase Goodness-of-Fit (10 is best). For frequency biases, negative values indicate underestimation. Both models were trained with 20 000 samples for 300 epochs.

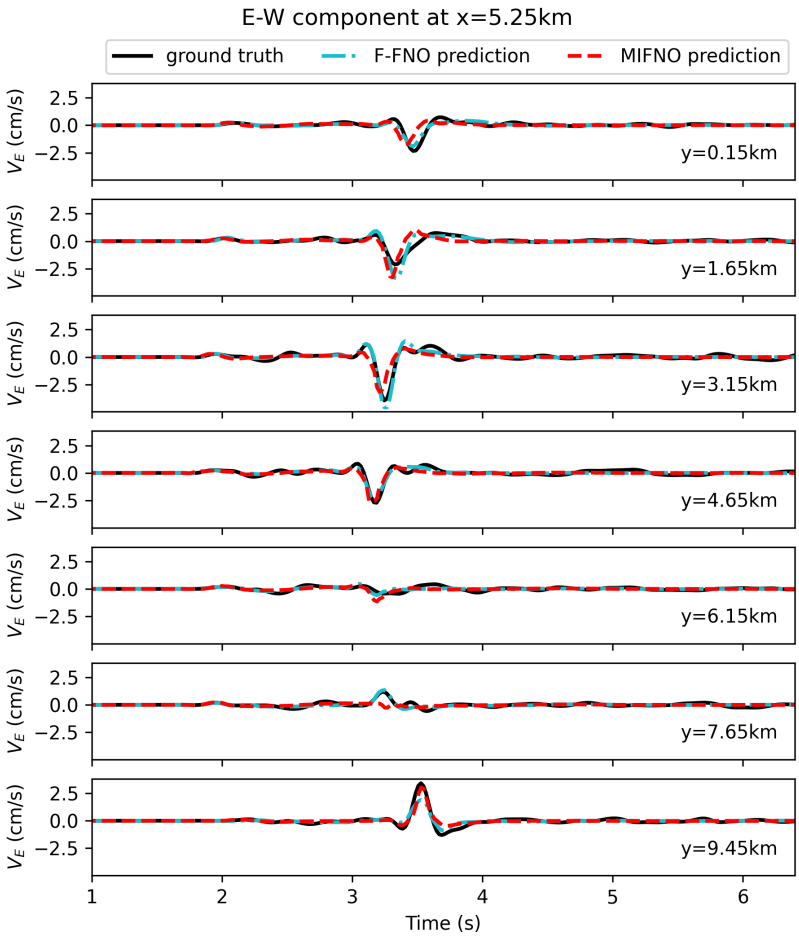


Figure A.22: East-West component of ground motion from simulations (black solid line), F-FNO predictions (blue dashed line), and MIFNO predictions (red dashed line)

A.4.3 Influence of the source parameters

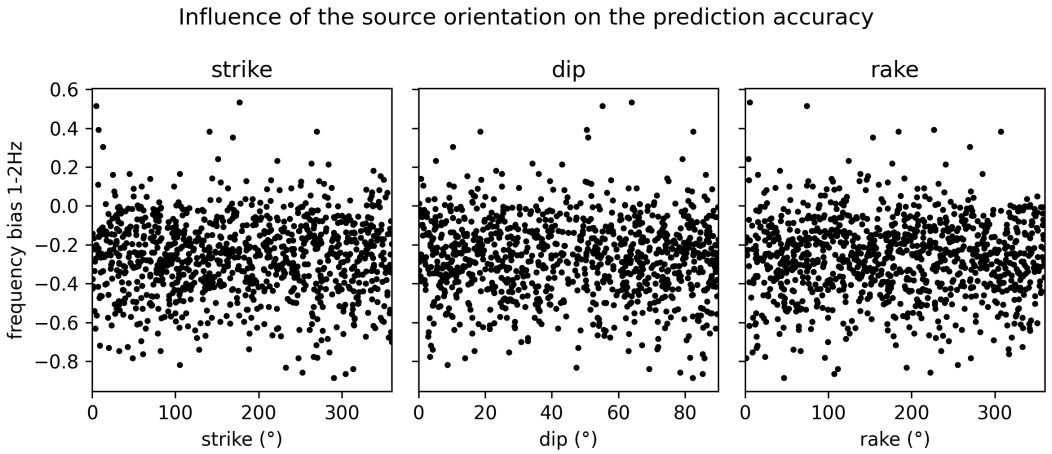


Figure A.23: For 1000 samples, the 1-2 Hz frequency bias is shown against the source orientation (strike, dip, rake).

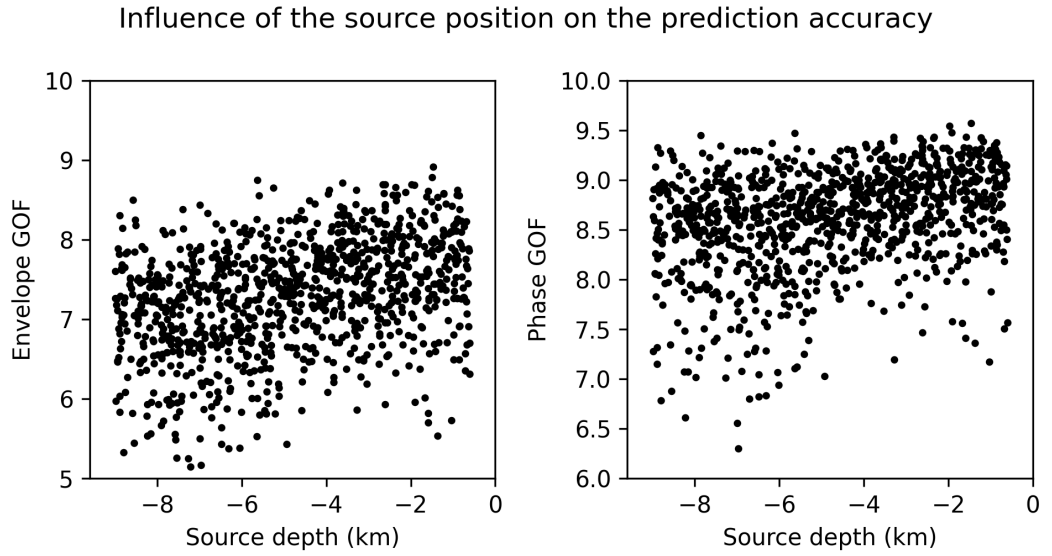


Figure A.24: For 1000 samples, the envelope and phase GOF of MIFNO predictions is shown against the source depth.

A.4.4 Out-of-distribution data

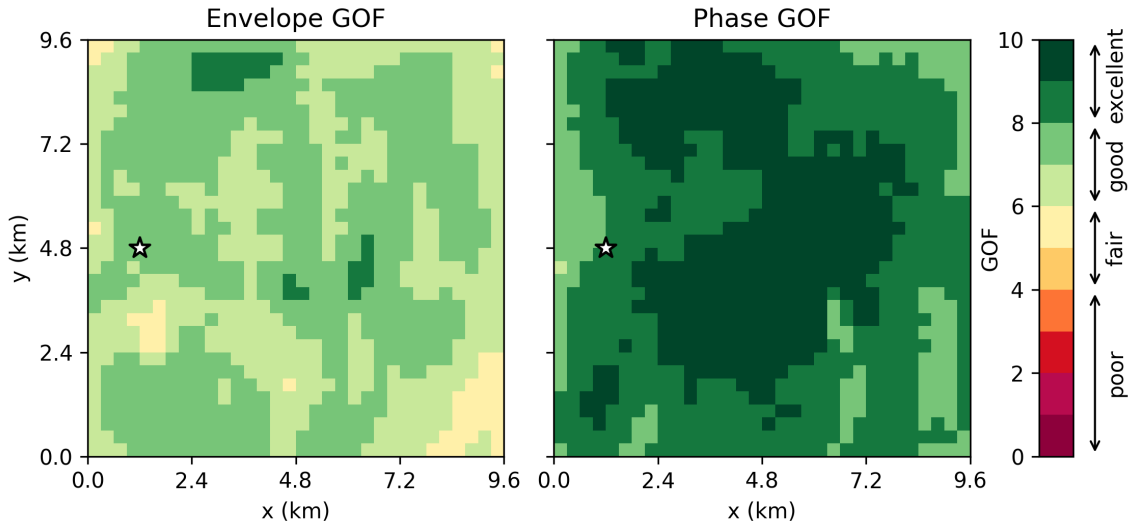
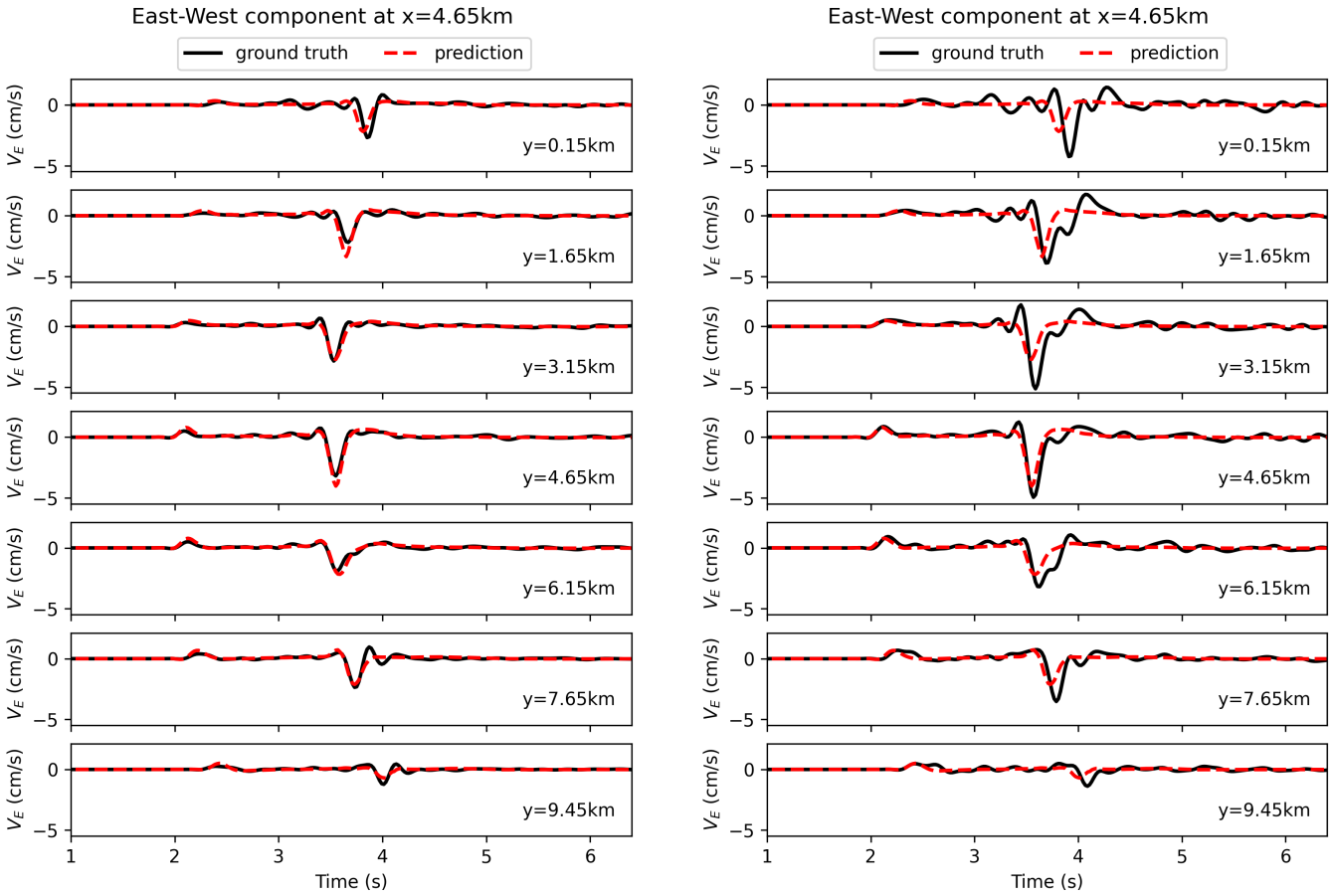


Figure A.25: Envelope and phase GOF for the geology depicted in Fig. 3.22. The white star denotes the epicenter.



(a) Simulations with a low-resolution geology (300 m)

(b) Simulations with a high-resolution geology (50 m)

Figure A.26: East-West component of ground motion from simulations (black solid line) and MIFNO predictions (red dashed line) for the overthrust geology depicted in Fig. 3.22

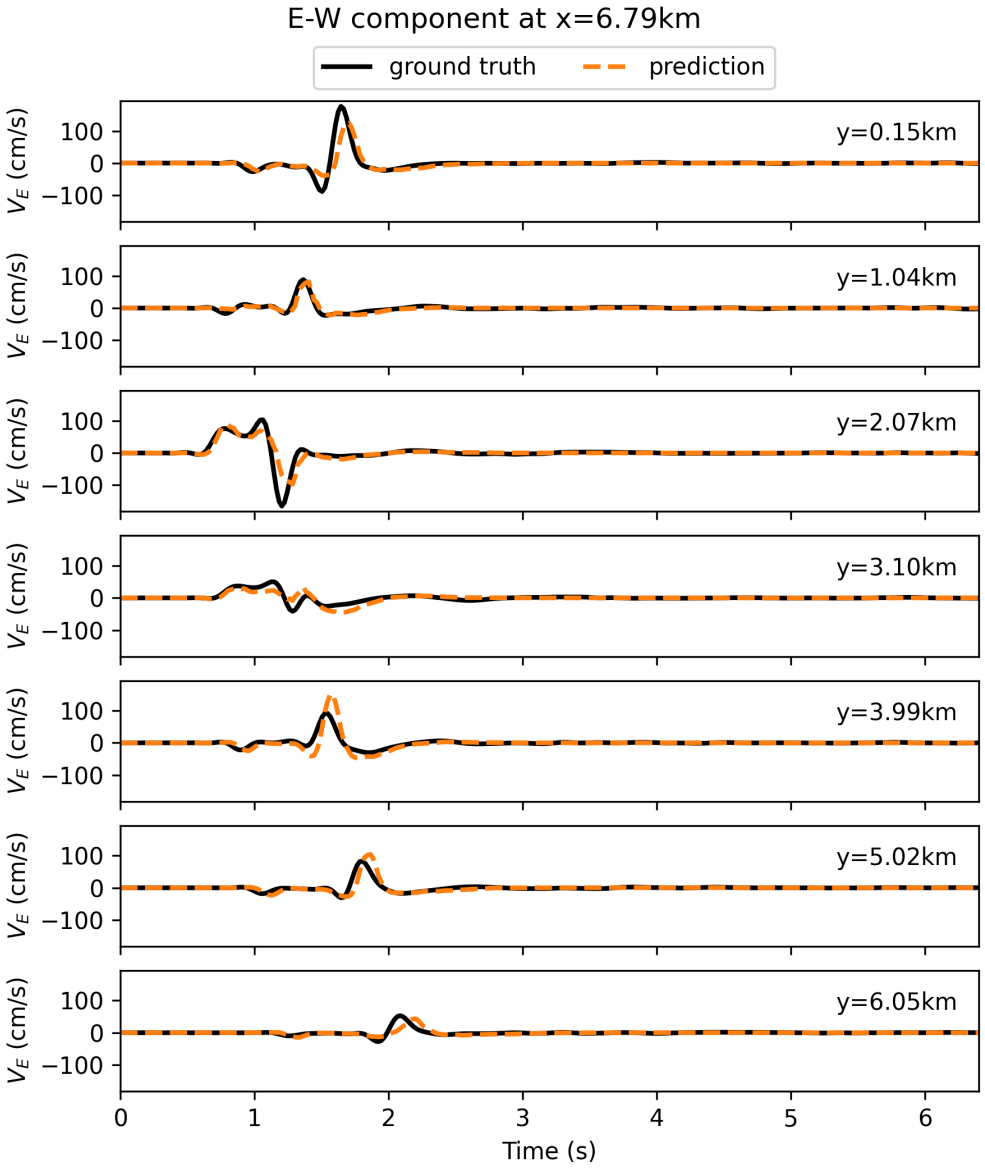


Figure A.27: East-West component of ground motion from simulations (black solid line) and MIFNO predictions with resolution 64 (orange dashed line)

A.5 Applications to the Le Teil earthquake

A.5.1 Transfer learning

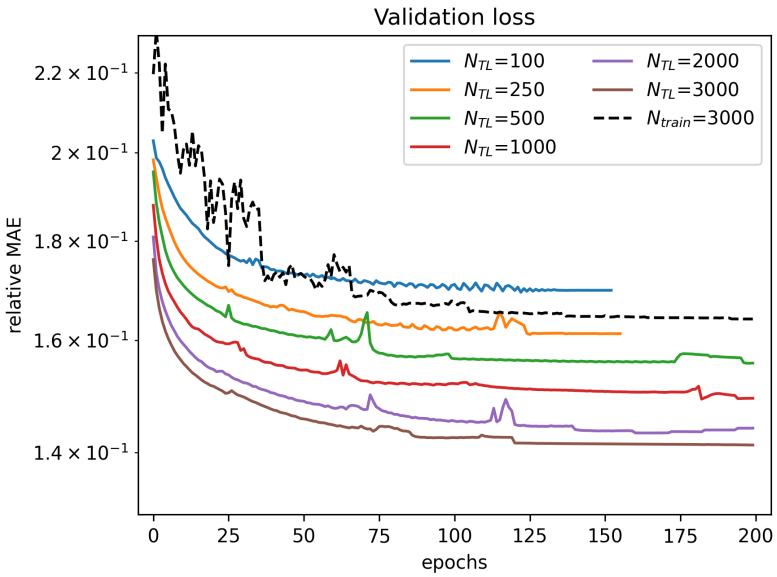


Figure A.28: Evolution of the validation loss computed as the relative Mean Absolute Error when the MIFNO is trained on the Le Teil database. The black line corresponds to a random initialization. Colored lines illustrate trainings with different number of training samples.

Le Teil database with a fixed source

Model	rRMSE	$rFFT_{low}$	$rFFT_{mid}$	$rFFT_{high}$	EG	PG
N_{train}=3000	0.17 ; 0.28	-0.45 ; -0.04	-0.64 ; -0.14	-0.65 ; -0.12	5.98 ; 7.76	7.49 ; 9.00
N_{TL}=0	0.31 ; 0.45	-0.28 ; 0.49	-0.59 ; 0.39	-0.70 ; -0.22	5.47 ; 6.32	5.43 ; 6.33
N_{TL}=10	0.17 ; 0.29	-0.29 ; 0.14	-0.47 ; 0.02	-0.48 ; 0.05	6.35 ; 7.84	7.50 ; 8.95
N_{TL}=50	0.14 ; 0.26	-0.29 ; 0.08	-0.43 ; 0.03	-0.45 ; 0.03	6.61 ; 8.20	7.95 ; 9.20
N_{TL}=100	0.14 ; 0.25	-0.28 ; 0.07	-0.43 ; 0.02	-0.44 ; 0.02	6.68 ; 8.25	8.04 ; 9.23
N_{TL}=250	0.12 ; 0.23	-0.24 ; 0.08	-0.38 ; 0.04	-0.39 ; 0.04	7.02 ; 8.54	8.32 ; 9.37
N_{TL}=500	0.11 ; 0.22	-0.24 ; 0.06	-0.36 ; 0.05	-0.38 ; 0.03	7.14 ; 8.66	8.42 ; 9.43
N_{TL}=1000	0.10 ; 0.22	-0.22 ; 0.07	-0.35 ; 0.06	-0.37 ; 0.04	7.25 ; 8.75	8.50 ; 9.47
N_{TL}=2000	0.10 ; 0.21	-0.21 ; 0.07	-0.34 ; 0.06	-0.36 ; 0.05	7.35 ; 8.85	8.57 ; 9.51
N_{TL}=3000	0.10 ; 0.21	-0.22 ; 0.06	-0.34 ; 0.05	-0.37 ; 0.03	7.36 ; 8.86	8.60 ; 9.53

Table A.11: 1st and 3rd quartiles of the F-FNO metrics computed on 700 test samples. rRMSE: relative RMSE (0 is best), $rFFT_{low}$: relative frequency bias 0-1Hz (0 is best), $rFFT_{mid}$: relative frequency bias 1-2Hz (0 is best), $rFFT_{high}$: relative frequency bias 2-5Hz (0 is best), EG: enveloppe Goodness-of-Fit (10 is best), PG: phase Goodness-of-Fit (10 is best). For frequency biases, negative values indicate underestimation.

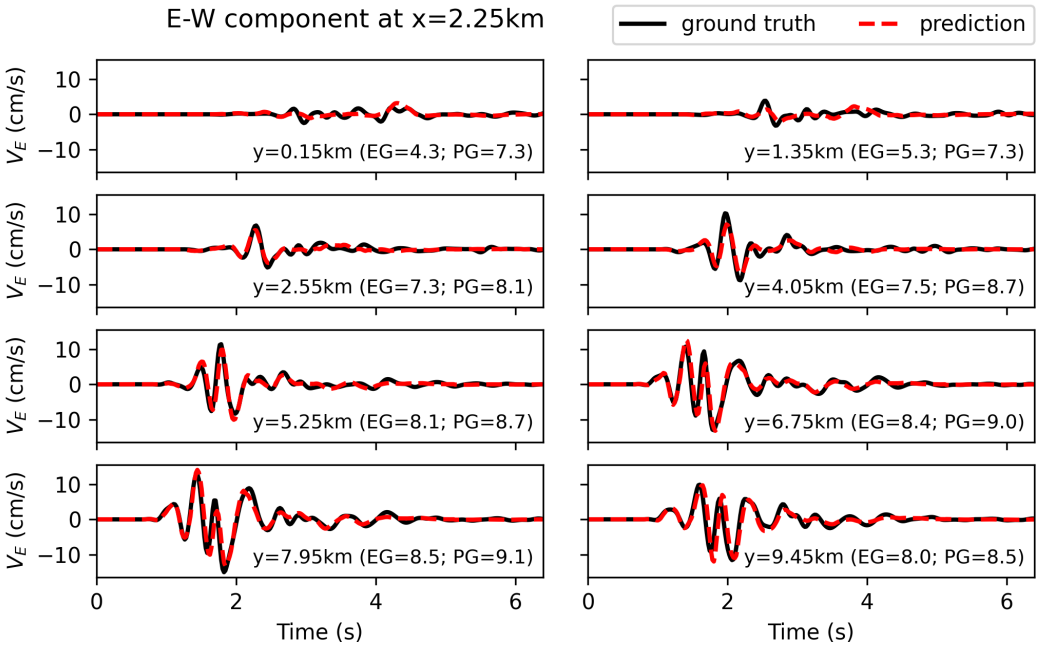


Figure A.29: Same as Figure 4.3 but only 500 transfer learning samples

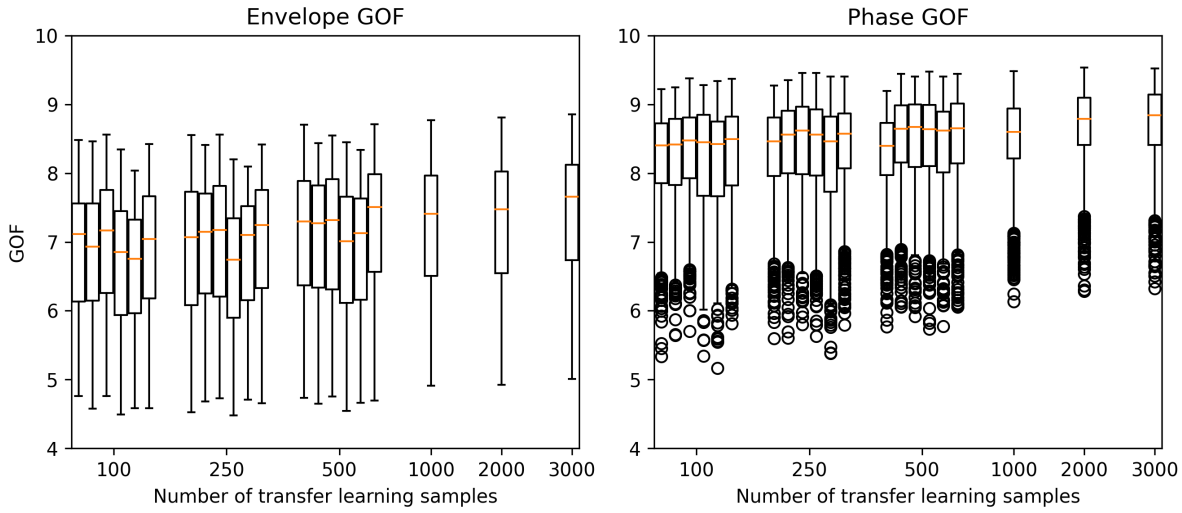


Figure A.30: Envelope and phase GOF computed on 700 test samples when the MIFNO is trained on non-overlapping sets of transfer learning samples. Each boxplot corresponds to the training with one set of transfer learning samples.

A.5.2 Le Teil earthquake: uncertainty quantification

Parameter	R_0	R_1	R_2	R_3	R_4	R_5
$\ell_{x,0}$	-0.010 $\pm$ 0.010	0.006 $\pm$ 0.007	0.000 $\pm$ 0.017	0.011 $\pm$ 0.013	0.014 $\pm$ 0.016	-0.000 $\pm$ 0.!005
$\ell_{y,0}$	0.003 $\pm$ 0.007	-0.002 $\pm$ 0.006	-0.007 $\pm$ 0.020	-0.001 $\pm$ 0.012	0.011 $\pm$ 0.017	0.004 $\pm$ 0.006
$\ell_{z,0}$	-0.003 $\pm$ 0.008	-0.002 $\pm$ 0.008	-0.005 $\pm$ 0.019	-0.003 $\pm$ 0.012	-0.004 $\pm$ 0.014	-0.000 $\pm$ 0.006
σ_0	-0.006 $\pm$ 0.013	0.009 $\pm$ 0.010	0.010 $\pm$ 0.025	0.003 $\pm$ 0.014	0.014 $\pm$ 0.017	0.004 $\pm$ 0.007
$\ell_{x,1}$	0.005 $\pm$ 0.009	0.001 $\pm$ 0.007	-0.008 $\pm$ 0.017	0.002 $\pm$ 0.011	0.017 $\pm$ 0.015	-0.002 $\pm$ 0.008
$\ell_{y,1}$	-0.003 $\pm$ 0.010	0.010 $\pm$ 0.011	-0.024 $\pm$ 0.021	0.004 $\pm$ 0.011	0.005 $\pm$ 0.014	-0.001 $\pm$ 0.007
$\ell_{z,1}$	0.000 $\pm$ 0.008	-0.000 $\pm$ 0.008	-0.014 $\pm$ 0.020	-0.001 $\pm$ 0.012	0.014 $\pm$ 0.016	0.004 $\pm$ 0.008
σ_1	-0.004 $\pm$ 0.011	0.003 $\pm$ 0.008	-0.006 $\pm$ 0.020	0.006 $\pm$ 0.017	0.008 $\pm$ 0.018	0.002 $\pm$ 0.008
$\ell_{x,2}$	-0.001 $\pm$ 0.008	0.001 $\pm$ 0.007	-0.003 $\pm$ 0.018	-0.002 $\pm$ 0.012	0.014 $\pm$ 0.015	0.002 $\pm$ 0.007
$\ell_{y,2}$	-0.003 $\pm$ 0.010	0.001 $\pm$ 0.011	-0.011 $\pm$ 0.021	0.001 $\pm$ 0.013	0.012 $\pm$ 0.016	-0.002 $\pm$ 0.007
$\ell_{z,2}$	0.002 $\pm$ 0.008	0.002 $\pm$ 0.007	-0.001 $\pm$ 0.020	0.009 $\pm$ 0.015	0.012 $\pm$ 0.016	0.001 $\pm$ 0.006
σ_2	0.000 $\pm$ 0.008	-0.001 $\pm$ 0.010	-0.018 $\pm$ 0.023	0.001 $\pm$ 0.014	0.005 $\pm$ 0.020	-0.002 $\pm$ 0.006
$\ell_{x,3}$	-0.000 $\pm$ 0.011	-0.001 $\pm$ 0.008	0.010 $\pm$ 0.017	0.009 $\pm$ 0.011	0.011 $\pm$ 0.016	-0.004 $\pm$ 0.006
$\ell_{y,3}$	-0.001 $\pm$ 0.010	0.005 $\pm$ 0.010	-0.010 $\pm$ 0.018	0.005 $\pm$ 0.012	0.019 $\pm$ 0.016	0.004 $\pm$ 0.007
$\ell_{z,3}$	-0.002 $\pm$ 0.009	-0.005 $\pm$ 0.006	-0.008 $\pm$ 0.018	-0.000 $\pm$ 0.010	0.004 $\pm$ 0.016	-0.001 $\pm$ 0.005
σ_3	0.007 $\pm$ 0.012	-0.007 $\pm$ 0.008	-0.001 $\pm$ 0.022	-0.003 $\pm$ 0.014	0.003 $\pm$ 0.016	0.001 $\pm$ 0.010
$\ell_{x,4}$	0.000 $\pm$ 0.012	-0.002 $\pm$ 0.007	-0.012 $\pm$ 0.016	0.005 $\pm$ 0.011	0.017 $\pm$ 0.019	0.004 $\pm$ 0.007
$\ell_{y,4}$	0.002 $\pm$ 0.012	-0.000 $\pm$ 0.007	-0.018 $\pm$ 0.020	-0.004 $\pm$ 0.013	0.011 $\pm$ 0.014	0.001 $\pm$ 0.005
$\ell_{z,4}$	0.001 $\pm$ 0.009	0.005 $\pm$ 0.011	-0.024 $\pm$ 0.017	-0.001 $\pm$ 0.013	0.008 $\pm$ 0.015	-0.000 $\pm$ 0.007
σ_4	0.006 $\pm$ 0.010	0.002 $\pm$ 0.011	-0.010 $\pm$ 0.024	0.005 $\pm$ 0.013	0.021 $\pm$ 0.016	0.006 $\pm$ 0.007
s_α	0.209 $\pm$ 0.033	0.349 $\pm$ 0.049	0.073 $\pm$ 0.018	0.475 $\pm$ 0.036	0.080 $\pm$ 0.017	0.349 $\pm$ 0.035
s_β	0.231 $\pm$ 0.041	0.112 $\pm$ 0.032	0.116 $\pm$ 0.024	0.059 $\pm$ 0.019	0.359 $\pm$ 0.041	0.099 $\pm$ 0.026
s_n	-0.002 $\pm$ 0.006	-0.000 $\pm$ 0.004	-0.002 $\pm$ 0.006	-0.001 $\pm$ 0.005	-0.006 $\pm$ 0.007	0.004 $\pm$ 0.005
strike ϕ	0.011 $\pm$ 0.006	0.001 $\pm$ 0.005	0.020 $\pm$ 0.010	0.017 $\pm$ 0.007	0.005 $\pm$ 0.005	0.009 $\pm$ 0.006
dip δ	0.030 $\pm$ 0.018	0.021 $\pm$ 0.014	0.038 $\pm$ 0.014	0.020 $\pm$ 0.013	0.010 $\pm$ 0.011	0.014 $\pm$ 0.010
rake λ	0.001 $\pm$ 0.010	-0.006 $\pm$ 0.006	0.024 $\pm$ 0.009	0.008 $\pm$ 0.009	0.001 $\pm$ 0.005	0.007 $\pm$ 0.006

Table A.12: First-order Sobol indices of the PSA at T=0.2 s for the six receivers

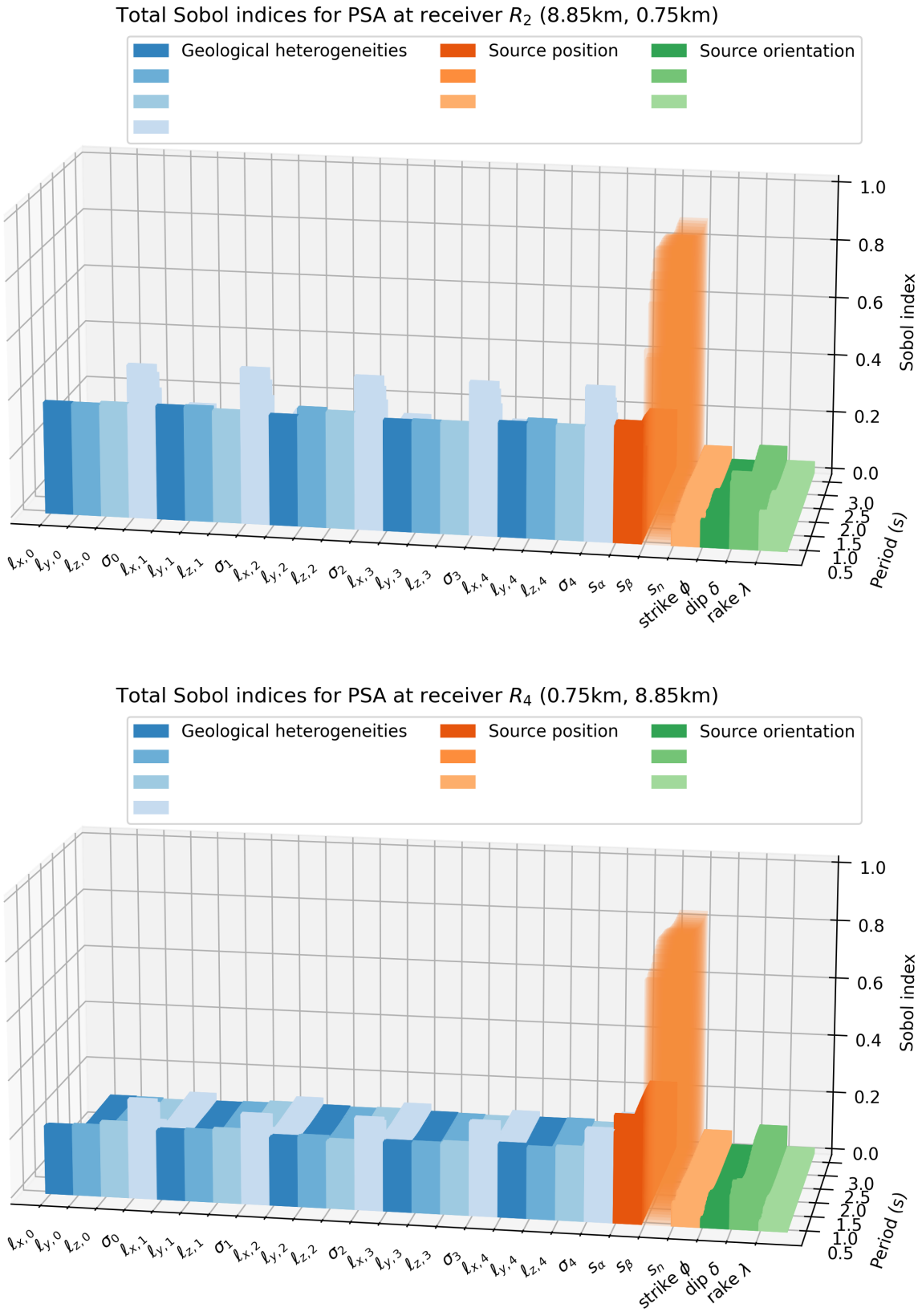


Figure A.31: Total Sobol indices for the PSA on the East-West component computed at receiver R_2 (top) and R_4 (bottom). Each of the 26 parameters is reported on the x axis and the y axis denotes the period.

Parameter	R_0	R_1	R_2	R_3	R_4	R_5
$\ell_{x,0}$	0.07 ± 0.01	0.08 ± 0.01	0.38 ± 0.02	0.16 ± 0.02	0.23 ± 0.02	0.07 ± 0.01
$\ell_{y,0}$	0.07 ± 0.01	0.07 ± 0.01	0.38 ± 0.02	0.16 ± 0.02	0.24 ± 0.02	0.06 ± 0.01
$\ell_{z,0}$	0.07 ± 0.01	0.07 ± 0.01	0.39 ± 0.02	0.16 ± 0.02	0.25 ± 0.02	0.06 ± 0.01
σ_0	0.10 ± 0.01	0.10 ± 0.02	0.52 ± 0.02	0.21 ± 0.02	0.34 ± 0.02	0.08 ± 0.01
$\ell_{x,1}$	0.07 ± 0.01	0.07 ± 0.01	0.39 ± 0.02	0.15 ± 0.02	0.24 ± 0.02	0.07 ± 0.01
$\ell_{y,1}$	0.06 ± 0.01	0.07 ± 0.02	0.39 ± 0.02	0.16 ± 0.02	0.24 ± 0.02	0.07 ± 0.01
$\ell_{z,1}$	0.07 ± 0.01	0.07 ± 0.01	0.38 ± 0.02	0.16 ± 0.02	0.25 ± 0.03	0.06 ± 0.01
σ_1	0.10 ± 0.01	0.09 ± 0.01	0.53 ± 0.03	0.22 ± 0.02	0.31 ± 0.02	0.08 ± 0.01
$\ell_{x,2}$	0.06 ± 0.01	0.07 ± 0.01	0.37 ± 0.02	0.16 ± 0.02	0.23 ± 0.02	0.06 ± 0.01
$\ell_{y,2}$	0.07 ± 0.01	0.08 ± 0.02	0.40 ± 0.02	0.16 ± 0.02	0.24 ± 0.02	0.06 ± 0.01
$\ell_{z,2}$	0.07 ± 0.01	0.07 ± 0.01	0.39 ± 0.02	0.18 ± 0.04	0.23 ± 0.02	0.06 ± 0.01
σ_2	0.09 ± 0.01	0.09 ± 0.01	0.52 ± 0.02	0.21 ± 0.02	0.31 ± 0.02	0.08 ± 0.01
$\ell_{x,3}$	0.08 ± 0.01	0.08 ± 0.02	0.37 ± 0.02	0.17 ± 0.02	0.24 ± 0.02	0.07 ± 0.01
$\ell_{y,3}$	0.07 ± 0.01	0.07 ± 0.01	0.38 ± 0.02	0.15 ± 0.02	0.24 ± 0.02	0.06 ± 0.01
$\ell_{z,3}$	0.07 ± 0.01	0.06 ± 0.01	0.38 ± 0.02	0.15 ± 0.02	0.24 ± 0.02	0.06 ± 0.01
σ_3	0.09 ± 0.01	0.10 ± 0.02	0.52 ± 0.02	0.22 ± 0.02	0.32 ± 0.02	0.08 ± 0.01
$\ell_{x,4}$	0.07 ± 0.01	0.06 ± 0.01	0.39 ± 0.02	0.16 ± 0.02	0.25 ± 0.02	0.06 ± 0.01
$\ell_{y,4}$	0.07 ± 0.01	0.06 ± 0.01	0.40 ± 0.02	0.16 ± 0.02	0.24 ± 0.02	0.06 ± 0.01
$\ell_{z,4}$	0.07 ± 0.01	0.07 ± 0.01	0.39 ± 0.02	0.17 ± 0.03	0.25 ± 0.02	0.06 ± 0.01
σ_4	0.10 ± 0.01	0.10 ± 0.02	0.52 ± 0.02	0.20 ± 0.02	0.31 ± 0.02	0.08 ± 0.01
s_α	0.62 ± 0.05	0.78 ± 0.08	0.39 ± 0.02	0.79 ± 0.05	0.36 ± 0.02	0.78 ± 0.07
s_β	0.69 ± 0.07	0.55 ± 0.06	0.65 ± 0.03	0.39 ± 0.04	0.82 ± 0.05	0.51 ± 0.05
s_n	0.06 ± 0.01	0.05 ± 0.01	0.07 ± 0.01	0.04 ± 0.01	0.08 ± 0.01	0.05 ± 0.01
strike ϕ	0.04 ± 0.00	0.04 ± 0.01	0.09 ± 0.01	0.06 ± 0.01	0.03 ± 0.00	0.03 ± 0.01
dip δ	0.26 ± 0.03	0.29 ± 0.04	0.22 ± 0.01	0.12 ± 0.01	0.11 ± 0.01	0.18 ± 0.02
rake λ	0.07 ± 0.01	0.08 ± 0.02	0.13 ± 0.01	0.06 ± 0.01	0.03 ± 0.00	0.05 ± 0.01

Table A.13: Total indices of the PSA at T=0.2 s for the six receivers

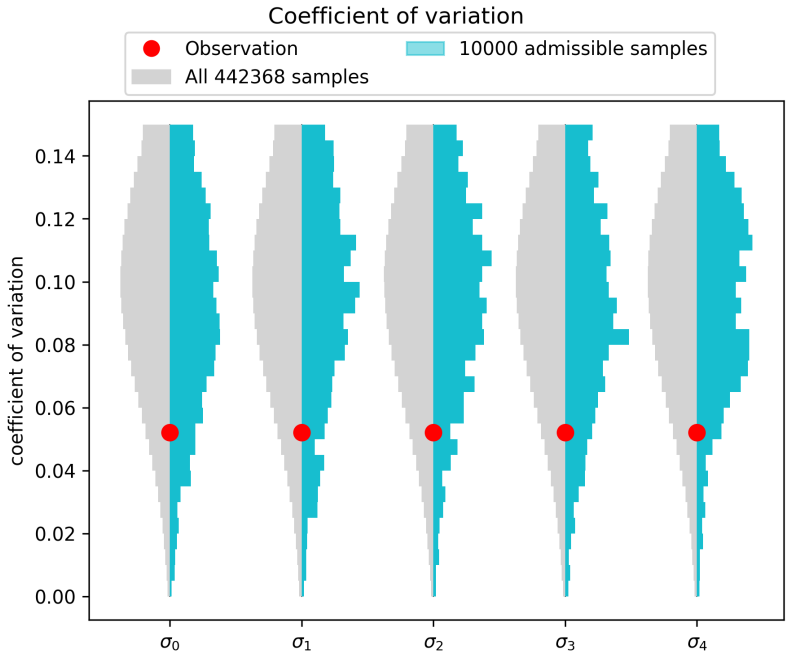


Figure A.32: Distribution of coefficients of variation inside each layer for samples in $\mathcal{S}$ (grey) and *admissible* samples in $\mathcal{S}_{|y^*}$ (blue). The coefficients of variation corresponding to the *observation* $\boldsymbol{x}_s^*$, $\boldsymbol{\theta}_s^*$ are shown with the red dots.

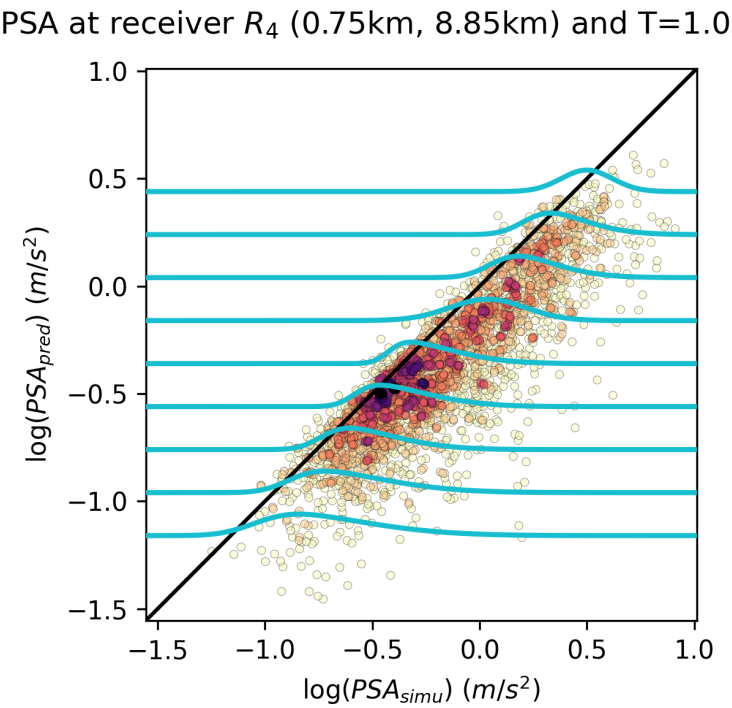


Figure A.33: Predicted PSA at receiver R_4 shown against the simulated PSA (colored dots, where darker colors correspond to higher points density). Blue lines show the skewed normal distribution that approximates the prediction error.

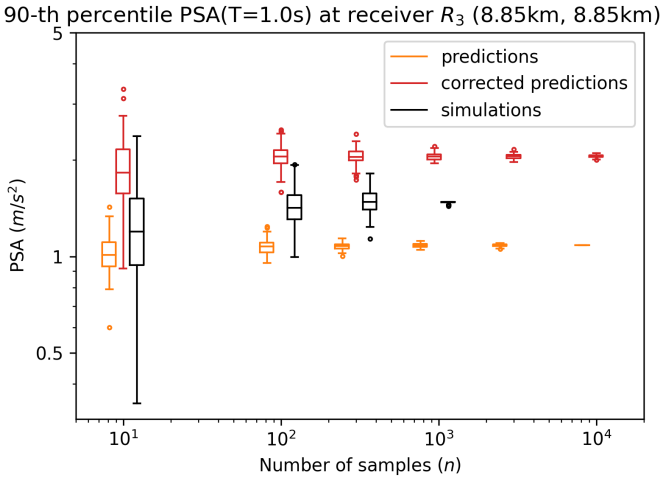


Figure A.34: Same as Fig. 4.14 for receiver R_3

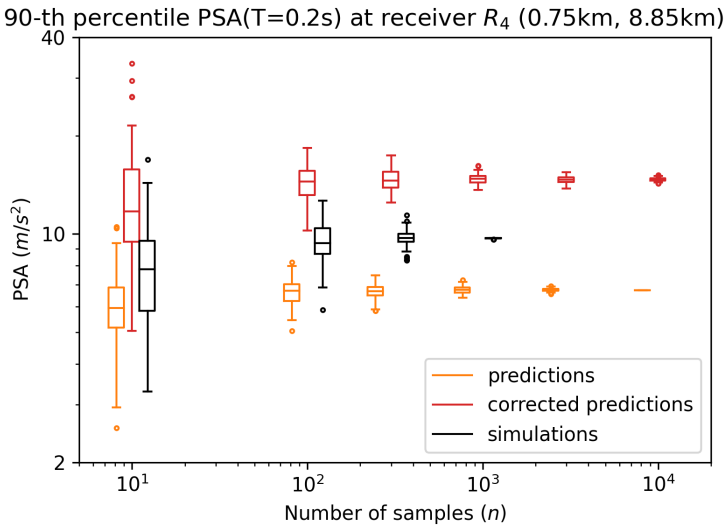


Figure A.35: Same as Fig. 4.14 for period T=0.2 s