A NEW BAYESIAN FRAMEWORK FOR THE INTERPRETATION OF GEOPHYSICAL DATA

Hadrien Michel

A thesis presented for the degree of

Doctor in Philosophy (PhD)

Faculty of Sciences, Department of Geology

Ghent University

and

Faculty of Applied Sciences, Urban and Environmental Engineering

University of Liège

Promoters:

Prof. Frédéric Nguyen, University of Liège

Prof. Thomas Hermans, Ghent University

Jury:

Prof. Alain Dassargues, University of Liège

Prof. David Dudal, KU Leuven & Ghent University

Prof. Stéphane Garambois, University Grenoble Alpes

Dr. James Irving, University of Lausanne

Dr. Eric Laloy, SCK-CEN

January 2023

Hadrien Michel is a fellow from F.R.S.-FNRS.

It just works!

Jensen Huang, 2018

ABSTRACT

Providing images of the subsurface from ground-based datasets is at the heart of the geophysicist's work. Multiple approaches have been applied to tackle this task. Most of the time, this task is performed in a deterministic framework, meaning that for a given dataset, a single model is provided to explain the data. However, those deterministic approaches lack the ability to provide reasonable uncertainty estimations, that take into account the non-unicity of the solution, noise in the data and modelling error. To provide precise and accurate models of the subsurface along with uncertainty, geophysicists use probabilistic approaches. Those approaches are able to sample the ensemble of a priori possible models (the prior) in order to extract models that can reasonably explain the datasets (the posterior). Such approaches, even though superior in terms of the reliability of their results, are rarely applied in practice due to their significant computational requirements.

In this manuscript, the aim is to propose a new Bayesian framework to interpret those geophysical datasets. This new framework, called Bayesian Evidential Learning, promises to enable a fast, precise and accurate estimation of the uncertainty. This framework is applied and adapted for 1D geophysical datasets (BEL1D). The new and adapted framework presents several advantages when compared to classical probabilistic approaches: from fast computations due to the limited number of forward runs needed, to providing insight about the experiment sensitivity and the validity of the prior. Moreover, it benefits from its construction as a Machine Learning algorithm, leading to quasi-instantaneous models of uncertainty.

RÉSUMÉ

Fournir des images du sous-sol à partir d'ensembles de données terrestres est au cœur du travail du géophysicien. De multiples approches ont été appliquées pour s'attaquer à cette tâche. La plupart du temps, cette tâche est réalisée dans un cadre déterministe, ce qui signifie que pour un ensemble de données déterminé, un modèle unique est fourni pour expliquer les données. Cependant, ces approches déterministes ne permettent pas de fournir des estimations raisonnables de l'incertitude, qui tiennent compte de la non-unicité de la solution, du bruit dans les données et des erreurs de modélisation. Pour fournir des modèles précis et exacts du sous-sol tout en tenant compte de l'incertitude, les géophysiciens utilisent des approches probabilistes. Ces approches sont capables d'échantillonner l'ensemble des modèles a priori possibles (le prior) afin d'extraire les modèles qui peuvent raisonnablement expliquer l'ensemble des données (le posterior). De telles approches, bien que supérieures en termes de fiabilité des résultats, sont rarement appliquées en pratique en raison de leurs importantes exigences en termes de temps de calcul.

Dans ce manuscrit, l'objectif est de proposer un nouveau processus bayésien pour interpréter ces données géophysiques. Ce nouveau système, appelé Bayesian Evidential Learning, promet de permettre une estimation rapide, précise et exacte de l'incertitude. Ce processus est appliqué et adapté aux jeux de données géophysiques 1D (BEL1D). Ce système présente plusieurs avantages par rapport aux approches probabilistes classiques : il permet des calculs rapides grâce au nombre limité d'exécutions nécessaires, et donne un aperçu de la sensibilité de l'expérience et de la validité de l'antériorité. De plus, il bénéficie de sa construction en tant qu'algorithme de Machine Learning, conduisant à la construction de

modèles d'incertitude quasi-instantanés.

SAMENVATTING

Het maken van beelden van de ondergrond op basis van grondgegevens is essentieel voor het werk van de geofysicus. Er zijn verschillende manieren om deze taak uit te voeren. Meestal wordt deze taak binnen een deterministisch kader uitgevoerd, wat betekent dat voor een bepaalde dataset één enkel model wordt verstrekt om de gegevens te verklaren. Deze deterministische benaderingen zijn echter niet geschikt om redelijke onzekerheidsanalyses te maken, die rekening houden met gegevensruis, modeleringsfouten, enz. Om precieze en nauwkeurige modellen van de ondergrond te berekenen, samen met onzekerheid, gebruiken geofysici probabilistische methoden. Die benaderingen mogen het ensemble van alle mogelijke modellen (de prior) bemonsteren om er modellen uit te halen die de datasets redelijkerwijs kunnen verklaren (de posterior). Deze benaderingen, hoewel superieur door de betrouwbaarheid van hun resultaten, worden in praktijk bijna nooit toegepast omdat ze veel computerwerk vergen.

In dit manuscript wordt een nieuw Bayesiaans raamwerk voorgesteld om die geofysische gegevenssets te interpreteren. Aan de hand van dit nieuwe systeem, genaamd Bayesian Evidential Learning, kan de onzekerheid snel, nauwkeurig en accuraat worden ingeschat. Dit kader wordt toegepast en aangepast voor 1D geofysische gegevenssets (BEL1D). Het nieuwe en aangepaste raamwerk biedt verschillende voordelen in vergelijking met klassieke probabilistische strategieën: van snelle berekeningen door het beperkte aantal forward runs die nodig zijn, tot inzicht in de sensitiviteit van het experiment en de validiteit van de prior. Bovendien profiteert het van zijn constructie als Machine Learning-algoritme, wat leidt tot quasi-onmiddellijk modellenren van onzekerheid.

ACKNOWLEDGEMENTS

During the four years that lead to this manuscript, many people helped me, directly or indirectly, to achieve this work. Among those, my promotor and co-promotor, Frédéric and Thomas, have been continuously there to discuss the advances of the work, possible new research areas, etc. They were the ones that gave me the opportunity to work on a PhD, and pushed me to go beyond what I would have produced.

My colleagues, even though they would mostly tell that they do not clearly understand what I do, were a crucial part to this work. They were the reason of the great atmosphere in which I was able to evolve. They were the ones that followed me on weird ideas leading to improbable results. Furthermore, they gladly participated to field works taking place mid-winter, with negative temperatures and only a fire to heat-up. I promise that I will one day do something with those datasets! Of all my colleagues, I will especially thank David, who had to put up with my presence in the office during those years.

Then, my family, for the continuous support, the unplanned vacations and week-ends, and their presence all around. I would especially thank my mom Patricia, for taking the time to proofread this manuscript.

Finally, I would like to thank my friends. They were always there for discussions about all and nothing, games nights, going out, ... and of course their support (all of those often taking place around a drink). They had to listen to me talking enthusiastically about my research in terms I can only imagine being closer to an incoherent flow of ideas than anything else. And for that, they largely deserve my praises!

SCIENTIFIC PRODUCTION

Publications

- **Michel, H.**, Nguyen, F., Kremer, Th., Elen, A. and Hermans, Th. (2020). 1D geological imaging of the subsurface from geophysical data with Bayesian Evidential Learning. Computers & Geosciences

- Kremer, Th., Müller-Petke, M., **Michel, H.**, Dlugosch R., Irons, T., Hermans, Th. and Nguyen, F. (2020). Improving the accuracy of 1D Surface Nuclear Magnetic resonance surveys using the multi-central-loop configuration: synthetic and fields examples. Journal of Applied Geophysics

- **Michel, H.**, Hermans, Th. and Nguyen, F. (2022). Iterative Prior Resampling and rejection sampling to improve 1D geophysical imaging based on Bayesian Evidential Learning (BEL1D). Geophysical Journal International

Conferences

- **Michel, H.**, Hermans, Th. and Nguyen, F., Application of BEL1D for sNMR data interpretation. Oral presentation at MRS2021, October 2021, Strasbourg (France)

- **Michel, H.**, Hermans, Th. and Nguyen, F., Improving the sNMR signal-to-noise ratio using cost-effective EM shielding. Poster presentation at the 48th IAH Congress, September 2021, Bruxelles (Belgium)

- **Michel, H.**, Hermans, Th. and Nguyen, F., Improving BEL1D accuracy for geophysical imaging of the subsurface. Oral presentation at SEG 2020, October 2020, Houston (Texas)

- **Michel, H.**, Nguyen, F. and Hermans, Th., Bayesian Evidential Learning for 1D geological imaging from geophysical data. Oral presentation at iEMSs 2020, September 2020, Bruxelles (Belgium)

- **Michel, H.**, Nguyen, F. and Hermans, Th., BEL1D: 1D imaging using geophysical data in the framework of Bayesian Evidential Learning. Oral presentation at EGU General Assembly 2020, May 2020, Vienna (Austria)

- **Michel, H.**, Nguyen, F. and Hermans, Th., Improving Bayesian Evidential Learning 1D imaging (BEL1D) accuracy through iterative prior resampling, Poster presented at AGU Fall meeting 2019, December 2019, San Francisco (California)

- **Michel, H.**, Hermans, Th. and Nguyen F., Uncertainty quantification of the subsurface water content through BEL1D interpretation of sNMR data, Poster presented at the Groundwater Quality conference, September 2019, Liège (Belgium)

- **Michel, H.**, Hermans, Th. and Nguyen F., 1D geological modeling of the subsurface from geophysical data with Bayesian Evidential Learning, Poster presented at the EGU General Assembly, April 2019, Vienna (Austria)

Seminaries

- **Michel, H.**, Hermans, Th. and Nguyen, F., Surface waves inversion in the framework of Bayesian Evidential Learning (BEL), Oral presentation at the

CERES conference, October 2020, Liège (Belgium)

- **Michel, H.**, Nguyen, F. and Hermans, Th., Geological imaging using geophysical data in the framework of Bayesian Evidential Learning (BEL), Oral presentation at the ISTerre Waves and Structures research group seminar, January 2020, Grenoble (France)

Computer programs

- **Michel, H.** (2022). pyBEL1D - A Python Implementation of BEL1D. *Github.*
- **Michel, H.** (2022). SardineReborn - New implementation of the Sardine Software for seismic refraction. *Github*
- **Michel, H.** (2020). BEL1D - Matlab toolboxes for the application of Bayesian Evidential Learning 1D imaging. *Github.*

Conferences as co-author

- Ahmed, A., Aignar, L. Deleersnyder, W., **Michel, H.**, Flores-Orozco, A. and Hermans, Th. (2022). Assessing the robustness of BEL1D for inverting TEM data. NSG2022 28th European Meeting of Environmental and Engineering Geophysics.
- Hermans, Th., **Michel, H.**, Lopez-Alvis, J. and Nguyen, F. (2021). Imaging the subsurface to inform hydrological models: a geophysicist's perspective. EGU General Assembly 2021.
- Ahmed, A., **Michel, H.**, Deleersnyder, W., Dudal, D. and Hermans, Th. (2021). Applying BEL1D for transient electromagnetic sounding inversion.

EGU General Assembly 2021.

- Ahmed, A., Deleersnyder, W., **Michel, H.**,Dudal, D. and Hermans, Th. (2021). Uncertainty quantification of salinity in coastal aquifer (Binah Thuan province, Vietnam) using BEL1D inversion of transient electromagnetic data. 48th IAH Congress.

- Caterina, D., Isunza Manrique, I., **Michel, H.**, Bobe, Ch., Lucas, H., Nguyen, F and Van De Vijver, E. (2020). Evaluating the resource recovery potential of fly ash deposits using electrical and electromagnetic methods. EGU General Assembly 2020.

Contents

Abstract **i**

Résumé **iii**

Samenvatting **v**

Acknowledgements **vii**

Scientific production **ix**

Contents **xiii**

List of Tables **xvii**

List of Figures **xviii**

Chapter 1:

Introduction **1**

1.1 Geophysical inversion 1

1.1.1 Deterministic inversion 3

1.1.2 Probabilistic inversion 6

1.2 The rise of Machine Learning in geophysics 8

1.3 Outline 10

Chapter 2:

An overview of 1D imaging in geophysics **15**

2.1 Seismic methods 16

2.1.1 Basic wave propagation in layered 1D media 17

2.1.2 Vertical seismic profile 18

2.1.3 Seismic refraction 20

2.1.4 Surface waves 22

2.2 Electromagnetic methods 23

2.2.1 Frequency domain electromagnetic 23

2.2.2 Time domain electromagnetic 24

2.2.3 Magnetotelluric 25

2.2.4 Surface Nuclear Magnetic Resonance 26

2.3 Electrical methods 27

2.4 Conclusion 29

Chapter 3:

Machine Learning: a geophysicist perspective **31**

3.1 Introduction 31

3.2 Machine Learning algorithms 33

3.2.1 Artificial Neural Networks & Deep Learning 34

3.2.2 Support Vector Machine 37

3.2.3 Decision Trees, Random Forests & Gradient Boosters 38

3.2.4 Others 39

3.3 Common use cases in geophysics 40

3.3.1 Classification 40

3.3.2 Surrogate modelling 41

3.3.3 Inversion 43

3.3.4 Signal processing 44
3.3.5 Petrophysical relationships 44
3.4 Uncertainty in Machine Learning 45
3.5 Conclusion 47

Chapter 4:

Bayesian Evidential Learning for 1D geophysical data: theory and development **51**

4.1 Theory and development 52
4.1.1 Introduction 52
4.1.2 BEL applied to 1D imaging: methodology 54
4.1.3 Numerical example 65
4.1.4 Field case study: Mont Rigi (Belgium) 73
4.1.5 Discussion 76
4.2 Application: sNMR multi-central configuration 81
4.2.1 The multi-central configuration 81
4.2.2 Application of BEL1D 83
4.3 Conclusion 86

Chapter 5:

Improving the posterior estimate through Iterative Prior Resampling **91**

5.1 Introduction 92
5.2 Methods 94
5.2.1 Iterative Prior Resampling 94
5.2.2 Rejection Sampling 97
5.2.3 Surface Waves Forward simulation 98

5.3 Synthetic benchmark 99
5.3.1 Results and McMC comparison 101
5.3.2 Parametric study of iterative prior resampling 108
5.4 Validation: application to Mirandola field dataset (InterPacific) . . 113
5.5 The case of sNMR 121
5.5.1 Results after one iteration 122
5.5.2 Results after applying IPR and rejection sampling 122
5.5.3 Improving the results from the Mont Rigi field data 125
5.6 Conclusion 127

Chapter 6:

General discussion and conclusion **131**

6.1 BEL for the interpretation of geophysical data 132
6.2 BEL advantages over other techniques 133
6.3 BEL generalization in geophysics 136
6.4 Further improvements of BEL 138
6.4.1 Defining new priors 138
6.4.2 Experimental design with BEL in geophysics 143
6.4.3 Exploring the impact of non-linear transforms 143

References **145**

Appendix A: BEL1D on a pendulum **173**

List of Tables

3.1 Summary of the tasks commonly performed by geophysicists classified by the type of learning 47

4.1 Characteristics of the synthetic model 65
4.2 Prior model space for the sNMR experiment 67
4.3 Prior model space for the Mont Rigi sNMR experiment 74
4.4 Description of the synthetic model used for the sensitivity demonstration along with the prior model 84

5.1 Benchmark model description 100
5.2 Description of the uniformly distributed prior model space for the benchmark. 100
5.3 Description of the benchmark model and associated uniform prior . 121

A.1 Prior model space for the pendulum model 175

List of Figures

1.1 Inversion and forward modelling. 2

1.2 L-Curve criterion for selecting an optimal λ 5

1.3 Classical approach versus BEL 11

2.1 Different types of common seismic waves 16

2.2 Illustration of the physical phenomena of seismic rays interacting with a plane interface. 17

2.3 Zero-offset VSP illustration 19

2.4 Illustration of seismic refraction for a 1D model 21

2.5 EM spectrum in geophysics 23

2.6 Illustration of the sNMR physical principle 26

2.7 Illustration of a vertical electrical sounding 28

3.1 Evolution of Machine Learning in geophysical publications 33

3.2 Illustration of a neural network. 34

3.3 Illustration of the support vector machine algorithm 37

3.4 Illustration of a decision tree. 38

4.1 Illustration of the BEL1D framework 56

4.2 Illustration of the Kernel Density Estimation process 60

4.3 Illustration of the posterior extraction in the CCA space 62

4.4 Noise propagation example 64

4.5 Simulated data for the sNMR experiment 66
4.6 DGSA for the synthetic benchmark 67
4.7 Data space versus model space in the canonically correlated space . 68
4.8 Combinations of parameters representing the canonically correlated model space dimensions 69
4.9 Results of BEL1D on the synthetic model 70
4.10 Results 1D profile visualization 71
4.11 Noise propagation example 72
4.12 sNMR at Mt. Rigi - CCA space 74
4.13 sNMR at Mt. Rigi - Posterior models distributions 75
4.14 sNMR at Mt. Rigi - Posterior models distributions with misfits . . . 77
4.15 Impact of the number of models sampled in the prior 78
4.16 Impact of the bandwidth of Kernel Density Estimation 80
4.17 Comparison of the classical coincident configuration and the multi-central configuration in sNMR 82
4.18 Sensitivity kernels for the multi-central configuration 83
4.19 Posterior model space on the peatland synthetic case 85

5.1 Illustration of the IPR principle 96
5.2 Visualization of the CCA space relationship at the first and last iterations 102
5.3 Results of BEL1D before and after IPR 104
5.4 Analysis of the convergence of IPR 105
5.5 Improvements of rejection sampling after IPR 107
5.6 Improvements of McMC after IPR 109
5.7 Parametric study of the number of models and the mixing ratio . . . 111
5.8 Impact of the number of layers 114

5.9 Mirandola - Models resulting from BEL1D, IPR and rejection sampling 118
5.10 Mirandola - Posterior model space resulting from BEL1D, IPR and rejection sampling 119
5.11 Mirandola - Estimation of the depth to the bedrock 120
5.12 Results of BEL1D on the synthetic sNMR case. 123
5.13 Results of BEL1D and IPR on the synthetic sNMR case. 124
5.14 Analysis of the impact of correlation on the estimated parameters for the total observed water content. 125
5.15 Results of IPR on the Mont Rigi field case 126

6.1 Testing 10 fixed thickness layers 140
6.2 Testing the effect of propagating the prior 142

A.1 Illustration of the pendulum experiment 173
A.2 Results of BEL modelling on the pendulum experiment 176
A.3 Probability estimation of the initial position of the bob 177
A.4 Pendulum experiment, data spaces 178

Chapter 1

Introduction

Geophysics aims at providing methods that inform about the subsurface without costly invasive investigations such as drilling. To perform this task, geophysicists investigate the earth through estimation of physical properties of soil, sediments and rocks. Those properties and their spatial or temporal variations inform the geophysicist on the geology and subsurface processes of a given site.

There is a large diversity of geophysical methods (see Chapter 2, and textbooks such as Schubert, 2015, M. Nabighian, 1987 or M. Nabighian, 1991), and each method is adapted to provide specific information about the subsurface, depending on the physical property (or properties) to which it is sensitive. However, interpreting the raw geophysical data is difficult and often requires to transform the data into geophysical models of the subsurface which is also known as geophysical inversion.

1.1 Geophysical inversion

Inversion is the opposite of forward modelling (Figure 1.1). Usually, the forward modelling is available as a closed-form expression, a matrix equation or a set of partial derivative equations. It can be solved either exactly or numerically (Equa-

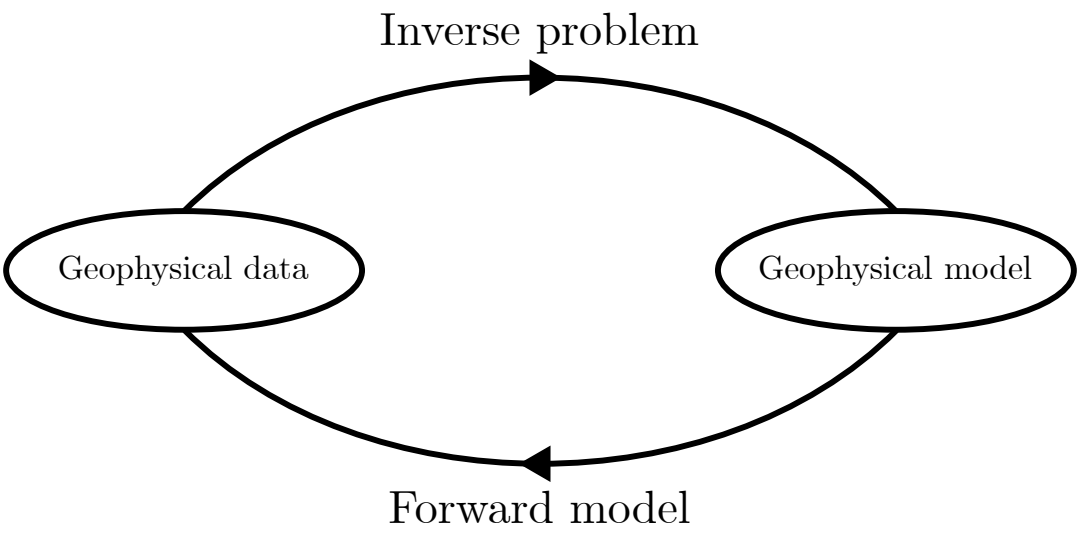

Figure 1.1: Inversion and forward modelling.

tion 1.1). Forward modelling can be expressed as:

$$g(\mathbf{m}) = \mathbf{d} \tag{1.1}$$

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces of finite or infinite dimensions. The non-linear mapping or forward operator $g() : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ contains the physics of the experiment and maps the model space to the data space. It acts as a direct link between the geophysical model ($\mathbf{m}$) and data ($\mathbf{d}$). In some cases, the forward operator can be formulated as a linear operator (e.g. for gravity data inversion), but it is most often non-linear (e.g. resistivity imaging or seismic imaging). Solving the inverse problem consists in finding a model ($\mathbf{m}$) or a set of models in a way that the data ($\mathbf{d}_{obs}$, originating from the field) is reasonably matched. Such processes are described in details in the literature (e.g.: Aster et al., 2013 or Kaipio and Somersalo, 2005).

1.1.1 Deterministic inversion

There are multiple approaches to solve the inverse problems. One can use a deterministic method, where the goal is to find the best model to explain the data, according to a misfit function (usually a p-norm, see Equation 1.2) formalizing the difference between the field data ($\mathbf{d}_{obs}$) and the modelled data issued from the forward model (Equation 1.1).

$$\psi_d = \|\mathbf{d} - \mathbf{d}_{obs}\|_p = \left(\sum_{i=1}^{n} |d_i - d_{obs,i}|^p \right)^{1/p}, \tag{1.2}$$

where n is the number of simulated data points and p is a real number greater than or equal to 1. The main issue with approaches that minimize the misfit of the data (ψ_d) is the resulting overfitting of the observed data, since the latter is affected by uncertainty and the model is affected by inaccuracies. This often results in a chaotic distribution of the model parameters, not realistically reflecting the geological structure or the studied process. Moreover, often, the problem is ill-posed (Lavrentev et al., 1986). This means that the solution to the geophysical inversion can be non-unique, instable or not existing.

A common way to avoid such issues, is to regularize the inversion. Doing so, the aim is to minimize an objective function $\psi(\mathbf{d}, \mathbf{m}, \lambda)$ which is the sum of the data misfit ($\psi_d(\mathbf{d}, \mathbf{m})$) and a model misfit ($\psi_m(\mathbf{m})$) which can for example be the distance to a reference model. The two terms are balanced using a regularization parameter (λ). The model misfit compares a given model to a reference model and specific constraints that may be present in the model given an a priori knowledge of the site. The definition of this misfit is thus itself a parameter of the inversion that must be explored to provide satisfactory images of the subsurface (e.g.: Acar and Vogel, 1994; Ajo-Franklin et al., 2007; Nguyen et al., 2016; Jordi et al., 2018).

This introduces bias in the inversion process in order to make the solution to the problem unique (Constable et al., 1987).

The regularization consists in tuning the impact of the model misfit using a regularization parameter (λ).

$$\psi(\mathbf{d}, \mathbf{m}, \lambda) = \psi_d + \lambda\psi_m \tag{1.3}$$

The value of λ is controlling the resulting model. Many approaches exist to optimize this parameter: L-curve, discrepancy principle, fixed value. All of those might lead to different resulting models.

The discrepancy principal aims at providing a data misfit that is reasonable given the data noise (Farquharson et al., 2003). Assuming a Gaussian independent noise for each n data point, one can estimate the expected maximum misfit (Equation 1.4).

$$\psi_d \approx n \tag{1.4}$$

This limits the values of the regularization parameter. However, attaining this value is complex at early iterations. This criterion is thus often combined with a cooling-schedule that progressively reduces λ by either a fixed factor or selects the optimal λ that satisfies the criterion. This requires at each iteration to perform a line-search to find this optimal value.

Another common approach is the L-curve criterion (Aster et al., 2013). Here, the idea is to optimize the trade-off that exists between the model and the data misfit. To do this, multiple values of the regularization parameter are tested, and the optimal choice should be the one that minimizes both misfits the most. Empirically, this optimum is the inflexion point in the so-called L-curve, linking the misfits for several values of λ (Figure 1.2).

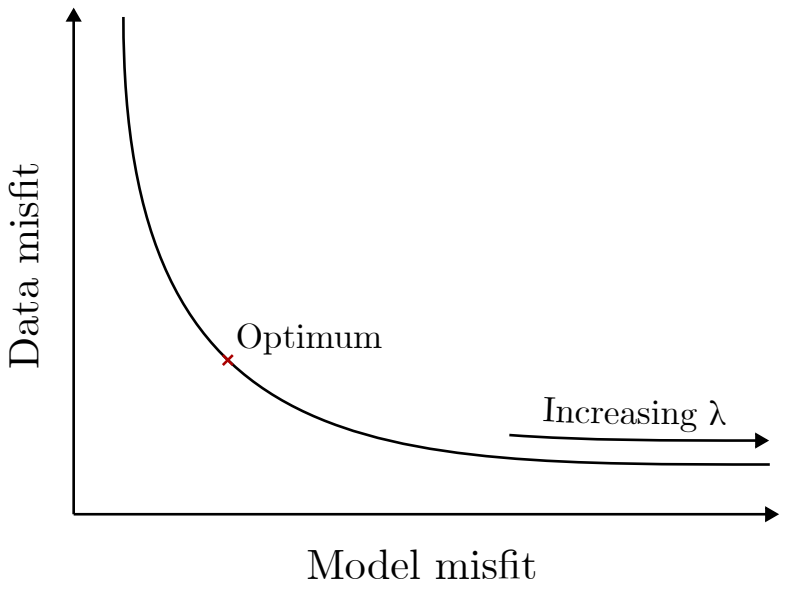

Figure 1.2: L-Curve criterion for selecting an optimal λ

In deterministic approaches, it is still possible to estimate the uncertainty on the model parameters. Error propagation in the linear forward model ($\mathbf{G}$) is among the most common approaches (Aster et al., 2013). The covariance matrix can be directly computed using the linear forward model and the data, assuming a Gaussian model for the error on the data with a known covariance matrix ($\text{cov}(\mathbf{d})$) (Equation 1.5). This model of the uncertainty on the parameters is performing adequately for linear well-posed problems, but lack the ability to provide reasonable uncertainty for non-linear and ill-posed problems which would then require the use of the Jacobian for a linearized error propagation or the use of Monte Carlo methods.

$$\text{cov}(\mathbf{m}) = \left(\mathbf{G^T G}\right)^{-1} \mathbf{G^T} \text{cov}(\mathbf{d}) \mathbf{G} \left(\mathbf{G^T G}\right)^{-1} \tag{1.5}$$

1.1.2 Probabilistic inversion

Instead of seeking a single optimum model along with its uncertainty, one may consider probabilistic methods. Here, the goal is to find a distribution of models with their corresponding probabilities. This is usually performed under Bayes' rule (Equation 1.6) where the solution of the inverse problem is given as a posterior probability distribution $p()$.

$$p\left(\mathbf{m}|\mathbf{d}\right) = \frac{p\left(\mathbf{d}|\mathbf{m}\right)p\left(\mathbf{m}\right)}{p\left(\mathbf{d}\right)}, \tag{1.6}$$

where $p\left(\mathbf{m}|\mathbf{d}\right)$ is the posterior probability of a model ($\mathbf{m}$), given the data ($\mathbf{d}$). To compute this value, one should know the prior probability of a given model $p(\mathbf{m})$, the probability of the data $p(\mathbf{d})$, and the probability of the data given the model $p\left(\mathbf{d}|\mathbf{m}\right)$ which we can link to the forward modelling. The probability of the data is a constant for a given set of data and prior ($p(\mathbf{d}) = \int p(\mathbf{d}|\mathbf{m})p(\mathbf{m})\,d\mathbf{m}$). The two critical terms are thus the prior probability of the model and the probability of the data given a model (Ulrych et al., 2001).

The probability of the dataset given a model is called the likelihood (written $L(\mathbf{m}|\mathbf{d})$). This likelihood reflects the difference between the data originating from the field measurements and the simulated data for the model $\mathbf{m}$ using the forward model (Equation 1.1). This likelihood function has multiple formulations, depending on the uncertainty distributions of the data and their correlations. However, under the hypothesis of Gaussian independent noise with standard deviation $\boldsymbol{\sigma}$, we can write Equation 1.7 (Lindsey, 1996).

$$L(\mathbf{m}|\mathbf{d_{obs}}) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)\exp\left(-\frac{1}{2}\left(\frac{d_{obs,i}-d_i}{\sigma_i}\right)^2\right) \tag{1.7}$$

Defining the prior probability will govern the posterior probability. In theory, in order to achieve the most unbiased inversion, the prior should be as large as possible (Tarantola, 2006). However, in practice, this is difficult as the resulting prior model space to explore would be very broad and would require unrealistic computational time. Therefore, the geophysicist usually defines a limited prior according to prior knowledge of the site and previous experiences. Careful attention must be attributed to this task as a prior that is too constrained will lead to biased estimations.

To sample the prior and extract the posterior, the most common and validated approach is to use Markov chain Monte Carlo simulation (McMC). McMC approaches sample models from a proposal distribution (or jumping distribution) in order to randomly perturb the current position of the chain. For each new proposed model, both the likelihood and the prior probability will be evaluated to assess the Metropolis acceptance probability (Equation 1.8).

$$P_{accept} = \min\left(1, \frac{L(\mathbf{m_i}|\mathbf{d_{obs}})p(\mathbf{m_i})q(\mathbf{m_j}|\mathbf{m_i})}{L(\mathbf{m_j}|\mathbf{d_{obs}})p(\mathbf{m_j})q(\mathbf{m_i}|\mathbf{m_j})}\right) \tag{1.8}$$

In Equation 1.8, the parameters are $\mathbf{m_j}$ the model in the current chain, $\mathbf{m_j}$ the perturbed model and d_{obs} the dataset; with $L()$ the likelihood, $p()$ the prior probability and $q()$ the proposal distribution.

Equation 1.8 can be simplified in case of symmetrical proposal distributions as in Equation 1.9, as is the case in DREAM (Vrugt, 2016).

$$P_{accept} = \min\left(1, \frac{L(\mathbf{m_i}|\mathbf{d_{obs}})p(\mathbf{m_i})}{L(\mathbf{m_j}|\mathbf{d_{obs}})p(\mathbf{m_j})}\right) \tag{1.9}$$

Sampling the posterior with a long enough chain, the geophysicist will obtain a stable set of posterior models that is representative of the actual posterior model space (Malinverno, 2002).

Another approach is the ensemble smoother technique (e.g.: Evensen, 1994; Y. Chen and Oliver, 2011; Emerick and Reynolds, 2013; Bobe et al., 2020). In this method the idea is to apply a transform iteratively to models from an ensemble according to the knowledge of observed data, under the assumption that the prior and the data error are both Gaussian and independent. This procedure is called data assimilation. For more details on those techniques and equations, the reader is directed towards van Leeuwen and Evenson (1996).

Nowadays, even though probabilistic inversion provide a more accurate quantification of uncertainty to account for in decision-making processes (Lawyer et al., 2001), they are often limited to scientific studies (e.g.: J. Chen et al., 2008; Trainor-Guitton and Hoversten, 2011; Linde et al., 2017; Andersen et al., 2018). The main obstacle is the computational requirements of such methods. An approach to improve on this aspect would be to enable more computations of multiple models to happen at the same time (parallelization). In the simplest Metropolis-Hasting McMC approach, the algorithm can not be parallelized. There are some versions of McMC that can be parallelized, but they are limited to the number of chains that are running in parallel (e.g. DREAM: Vrugt, 2016). Those aspects usually limit the applicability of probabilistic inversion to low dimensionality spaces and forward models that do not require prohibitive computational time.

1.2 The rise of Machine Learning in geophysics

Machine Learning is currently a popular research topic. From medical research (Rajkomar et al., 2019) to video games (Justesen et al., 2020), those algorithms are widely used in all areas of the society. Often, they are applied to solve problems that are either unsolvable using classical algorithms or that are too slow or not

scalable with other approaches. Chapter 3 is dedicated to the different algorithms and applications that arise from Machine Learning in geophysics.

These algorithms can improve the workflow of the geophysicist in multiple aspects (see Chapter 3). The main advantage of Machine Learning over traditional techniques originates from the core of the technique: it learns to perform a task from a given training set. Therefore, the learning phase results (which is usually computationally intensive) are often reusable.

The use of Machine Learning for geophysical data inversion is not new. In the 90s, the inverse problem was solved by a neural network for the first time for surface waves (R. Meier & Rix, 1993). Since then, the use of machine learning has risen for geophysical inversion (see Chapter 3). Multiple approaches have been tested. However, Machine Learning is mostly deterministic by design (one input provides a single output).

Even though the design of ML algorithms (especially neural networks) is not originally suited for uncertainty quantification, multiple approaches have been proposed to overcome this impediment (Bishop, 1995). In early researches, the idea was to learn the uncertainty from the training set (Devilee et al., 1999). The result is a set of discrete values that describe the posterior density function. With such a design, the training set must previously know the probability density functions of any given set $(\mathbf{m}, \mathbf{d})$. It is thus relatively impractical to apply with large priors.

More recently, other approaches, using the standard deviation from predictions of random forests (Zou et al., 2021) or adapted neural networks have been proposed (X. Zhang and Curtis, 2021; U. Meier et al., 2007). However, those methods have their own limitations:

- The posterior must be Gaussian or of known type (Zou et al., 2021).
- The posterior is unable to recover correlations between parameters efficiently

(U. Meier et al., 2007).

- The algorithm needs to be trained along with noisy data to be able to take it into account (X. Zhang & Curtis, 2021).

In this research manuscript, we aim at proposing a new alternative Machine Learning algorithm to solve the inverse problem in geophysical imaging in 1D and compute the associated uncertainty. This new algorithm is based on the Bayesian Evidential Learning (BEL) framework (Scheidt et al., 2018). This framework addresses multiple concerns of geophysicists in a Bayesian way. At first, for a given prior, it helps to acknowledge the data worth by means of sensitivity analyses. Doing this, we can assess if the experiment is a suitable design to attain the stated objective. Then, once the experiment validity has been confirmed, BEL may process prior model falsification. There, the technique consists in verifying the consistency of the field data with the prior corresponding data space. This step makes it easier to avoid complex and CPU intensive computations with an inadequate prior. Finally, BEL may be used to obtain a direct prediction of the parameter of interest, such as the wellhead protection area (Thibaut et al., 2021), on the basis of the field data. To do so, BEL learns from a set of models in the prior that are associated to data and predictions. This differs from classical approaches that first invert the data into a geophysical model and then, using petrophysics, converts those models in terms of the interset parameter (Figure 1.3).

1.3 Outline

After the current introduction, we will present different 1D geophysical methods (Chapter 2) relevant in today's world. This chapter will also show the diversity of methods and their related datasets.

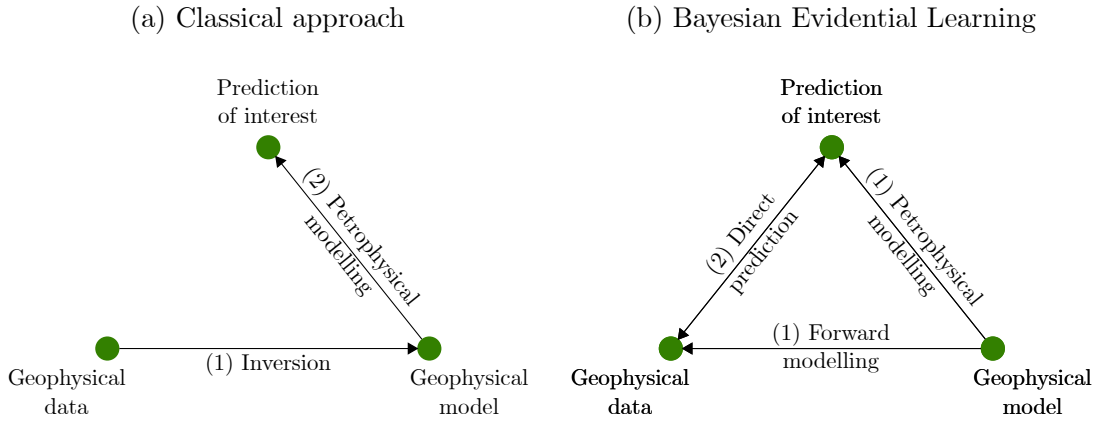

Figure 1.3: Classical approach versus BEL (modified from Scheidt et al., 2018). (a) The classical approach: first, the data is inverted, and then the prediction is forecasted. (b) The Bayesian Evidential Learning approach: first, a statistical relationship is learned from models data and their associated predictions, and then this relationship is exploited to provide a direct forecast of the prediction of interest given the data.

Then, we present the use of Machine Learning in geophysics (Chapter 3). This chapter will be the opportunity to present the reader with typical Machine Learning algorithms and their common applications in geophysics.

Next, in Chapter 4, we present the BEL1D algorithm. The BEL1D algorithm is the implementation of BEL for the 1D geophysical imaging of the subsurface based on geophysical data. In this chapter, we will present the difficulties that rose from the specificity of geophysical data and what was performed to improve the BEL algorithm. The algorithm is presented with a synthetic example using sNMR data and the results are compared to a classical McMC approach. We will then validate the algorithm on a field dataset against other geophysical measurements.

Along with this chapter, the MATLAB source code for the BEL1D algorithm is released under the BSD 2-clause licence (Michel, 2020). This MATLAB code is adapted for the application of BEL1D with sNMR and seismic surface waves dispersion curves interpretations using graphical user interfaces.

From the results of Chapter 4, we will show that the algorithm tends to overestimate the uncertainty when dealing with large prior uncertainty or poor resolution datasets. To overcome this impediment, we propose a new iterative algorithm in Chapter 5. This algorithm, called iterative prior resampling (IPR), uses insights learned at a given iteration to improve the knowledge of the probabilistic inversion. This time, we will show the algorithm working on seismic surface waves datasets. A first synthetic case is explored and compared to an McMC method. Then, we again validate the approach by using the algorithm on a published dataset from the InterPACIFIC project (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016 and Garofalo, Foti, Hollender, Bard, Cornou, Cox, Dechamp et al., 2016). Those examples empirically demonstrate that the algorithm is efficient to recover accurate and precise posterior model space estimations.

Along with this chapter, the fully open-source Python code (BSD 2-clause licence) for BEL1D and IPR is released (Michel, 2022) with applications for sNMR and seismic surface waves dispersion curves and easily adaptable for other methods. This code, contrary to the previously released MATLAB version is much more versatile and improves on the parallelization of the algorithm.

Finally, we will discuss on the results of this thesis and the insights learned from this work (Chapter 6). In this discussion, we will analyse key aspects that BEL applied to geophysical inversion should improve:

1. Is BEL effective to perform the inversion of geophysical data in a probabilistic framework?

2. What does BEL provide that other techniques do not?

3. Can BEL be generalized to any type of geophysical datasets? What are the possible caveats?

4. How can we further improve the performances of BEL?

Note that the main content of chapters 4 and 5 is published in Computers & Geosciences (Michel, Nguyen et al., 2020) and Geophysical Journal International (Michel et al., 2022) respectively.

Chapter 2

An overview of 1D imaging in geophysics

Geophysics aims at investigating the Earth structure and composition using non-intrusive and indirect approaches. The different methods are sensitive to bulk properties such as seismic velocities or electrical resistivity, which may inform on parameters of interest such as stiffness or water content, and image the subsurface structures and estimate its composition (e.g.: Schubert, 2015; Bagdassarov, 2021). Bulk properties of the Earth can be translated into information that are of direct concern to the end-user through petrophysical relationships whereas the estimation of bulk properties depend on the solution of inverse problems.

In this chapter, we will present a panel of near-surface geophysical approaches that can be interpreted as vertical 1D profiles. The aim of this section is to present the diversity of methods and the types of data that can be encountered by the geophysicist. For more details on the different techniques, the reader is directed towards references in the literature that cover in depth the different methods. We will here consider seismic methods, electrical and electromagnetic methods.

2.1 Seismic methods

Seismic methods are among the most widely used. They rely on the propagation of a mechanical wave across the subsurface. Different types of waves can be studied (Figure 2.1). Body waves travel through the entire subsurface. They are dissociated into two classes: P-waves and S-waves. The first is the compression wave, whose displacement is along the direction of the wave. The second one is the shear wave, whose propagation is transverse to the direction of the wave. Contrary to compression waves, shear waves are not transmitted by fluids.

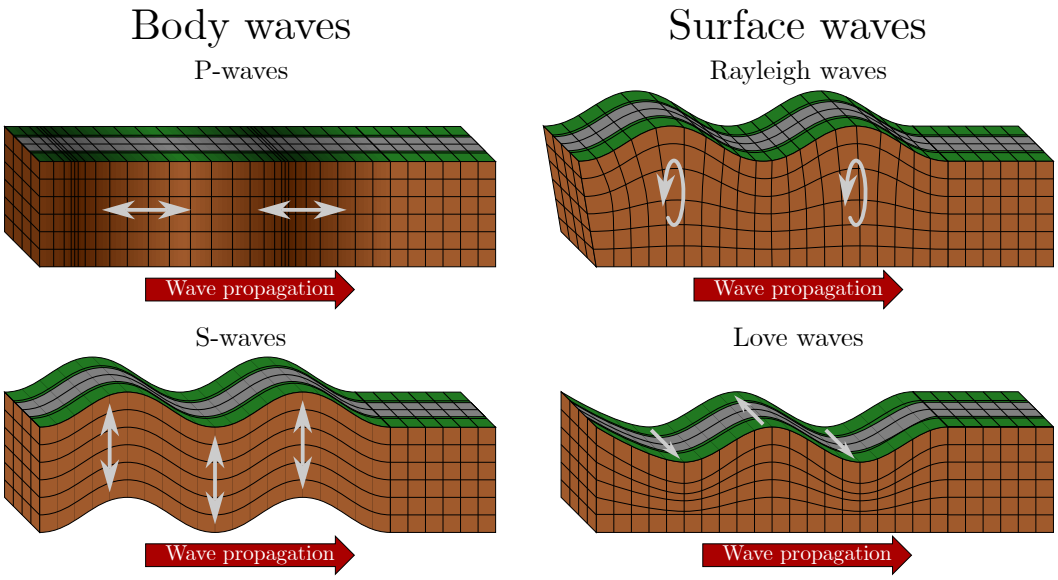

Figure 2.1: Different types of common seismic waves

Surface waves originate at the interface between air and the subsurface. The most common in geophysics are Rayleigh waves and Love waves. Rayleigh waves are characterized by a particle motion in an elliptical path in the vertical plane. For Love waves, the particle motion is horizontal and propagates perpendicularly to the wave direction. The particularity of such waves is that their velocity depends on

the wavelength.

The propagation of seismic waves in the subsurface depends on different rock properties. The main parameters of interest are the velocities V_P and V_S, the compression- and shear-waves velocities, which may be defined in terms of elastic properties and the density. However, those parameters depend themselves on numerous factors, from the mineralogical composition of the rocks to the fluid properties contained in the soil/sediments/rocks (Schmitt, 2015).

Seismic waves are typically recorded using geophones on terrestrial environments and hydrophones in marine environments. Those devices record the ground displacement as a function of time. Geophones are positioned with a specific orientation, which enables the distinction between different wave types. On the other hand, hydrophones are omnidirectional and pressure dependent.

2.1.1 Basic wave propagation in layered 1D media

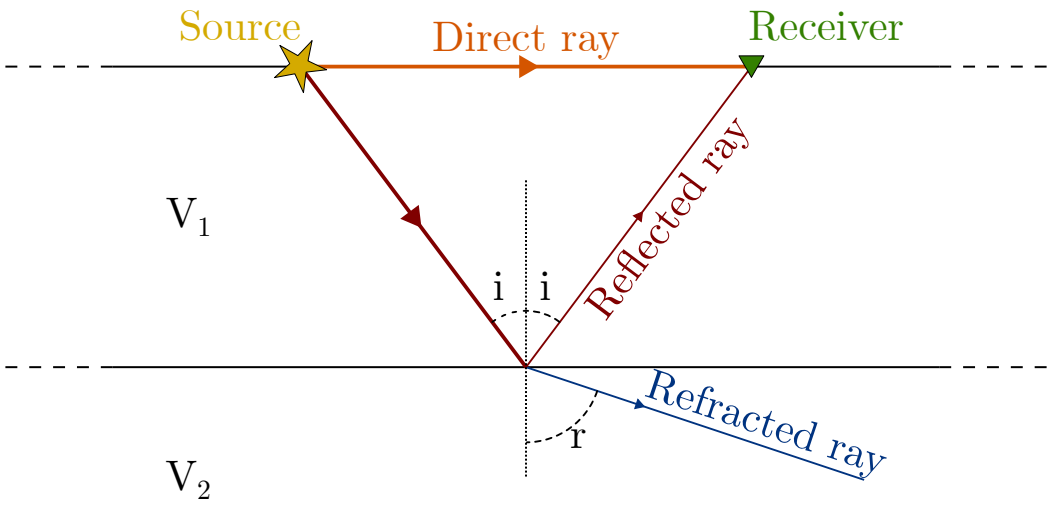

Figure 2.2: Illustration of the physical phenomena of seismic rays interacting with a plane interface.

From the surface, a seismic wave propagates through the subsurface with ve-

locities that depend on the impedance of the rocks. Typically, the impedance is directly related to the seismic velocity of the rocks and its density (Bagdassarov, 2021).

$$Z \approx \rho \times V \tag{2.1}$$

When there is a sharp contrast of velocities between two layers, the wave may be in part refracted deeper, and in part reflected to the surface (Figure 2.2) depending on the impedance contrast. By analysing the signal at the receivers on the surface and after data processing, the geophysicist can deduce the position of those interfaces that inform on the geological structures with different approaches (e.g.: D. R. Cox et al., 2020).

2.1.2 Vertical seismic profile

This technique requires the presence of a borehole. Geophones or hydrophones are placed inside the borehole at several depths and a source is positioned at the surface (Kennett et al., 1980). Usually, Vertical Seismic Profiles (VSP) use multiple sources at the surface. From the surface sources' configuration, the experiment can enable the recovery of a 1D profile (Zero-Offset VSP, see Figure 2.3), 2D transects or 3D models (Offset VSP, Walkaway VSP, etc.) (Campbell et al., 2005).

The physics for 1D layered systems is rather simple if we neglect the interaction with the borehole casing and diffraction. Two phenomenons are observed in a VSP after the source is triggered:

1. The wave is received directly on the different geophones/receivers (downgoing raypaths in Fig. 2.3).

2. The wave reflected on the different interfaces reaches the receivers (upgoing

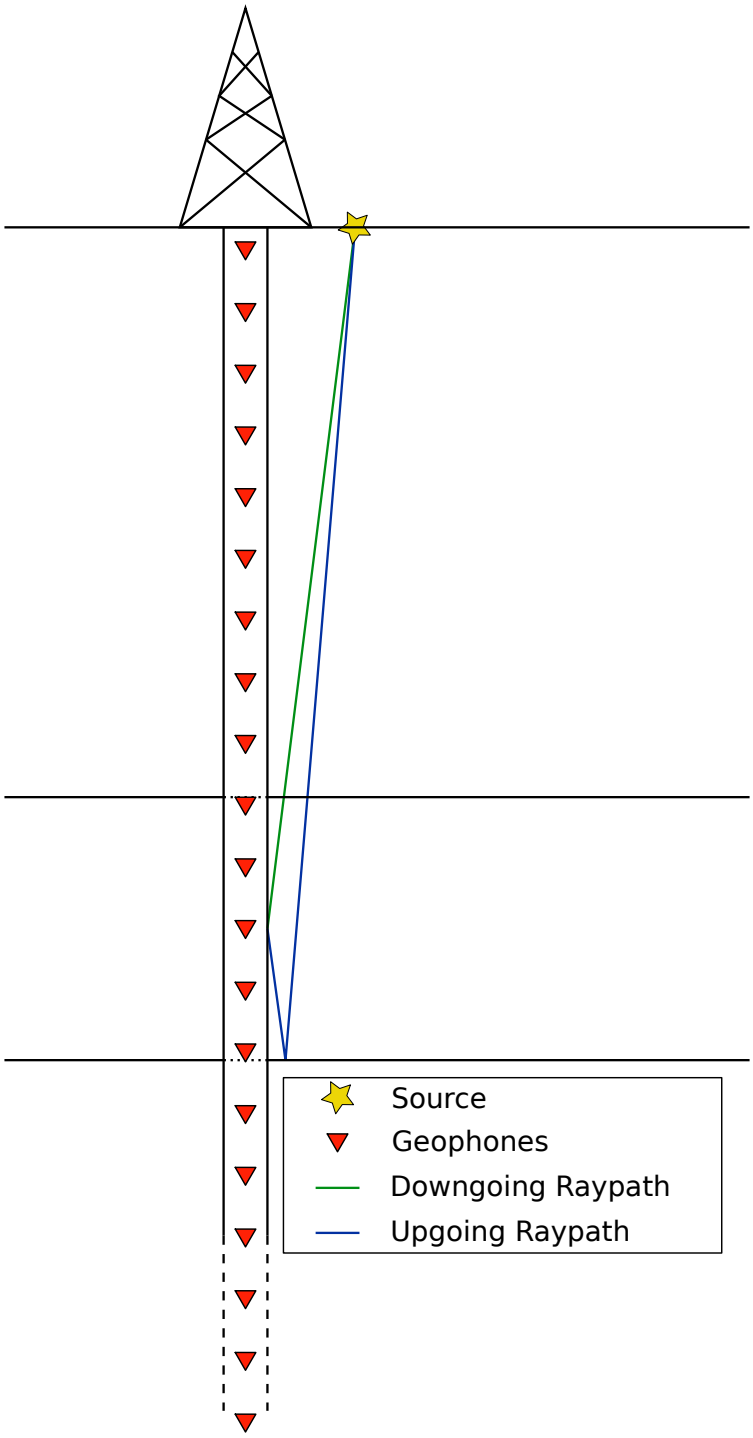

Figure 2.3: Zero-offset VSP illustration

raypaths in Fig. 2.3).

This means that the interpretation in terms of seismic velocities is straightforward until the last receiver. However, one of the advantages of such techniques is that they enable the detection of structures below the bottom of the borehole by means of the reflected waves/upgoing raypaths (Kennett et al., 1980).

Interpretation of such data is often performed while drilling to inform on the potential characteristics of the rocks below the drill head. Interpreting the direct wave immediately gives the seismic velocities for each layer between the geophones. Using more advanced techniques, it is also possible to retrieve the attenuation factor (Q) from those datasets (e.g.: Campbell et al., 2005; Zhao et al., 2014).

VSPs are commonly applied in oil and gas exploration (e.g.: Larki et al., 2021; Matsushima et al., 2016), where the knowledge of in-depth rock properties is crucial. VSP profiles are also commonly applied to calibrate models from other surface-based methods (e.g.: Hurich and Deemer, 2013).

2.1.3 Seismic refraction

In seismic refraction, the main hypothesis is that the waves are - as the name suggests - refracted at the interface between geological layers (see Figure 2.2). Using an array of geophones positioned on the surface, one can measure the first arrival of seismic waves (P- or S-waves depending on the equipment). Refraction of the waves at the interfaces occurs at any angle (Green, 1974). Given the incidence angle (θ_1) and the velocities of the waves in the 2 layers (V_1 and V_2), we can infer the angle of refraction (θ_2) in the second layer according to Snell's law:

$$\sin(\theta_1)/\sin(\theta_2) = V_1/V_2 \tag{2.2}$$

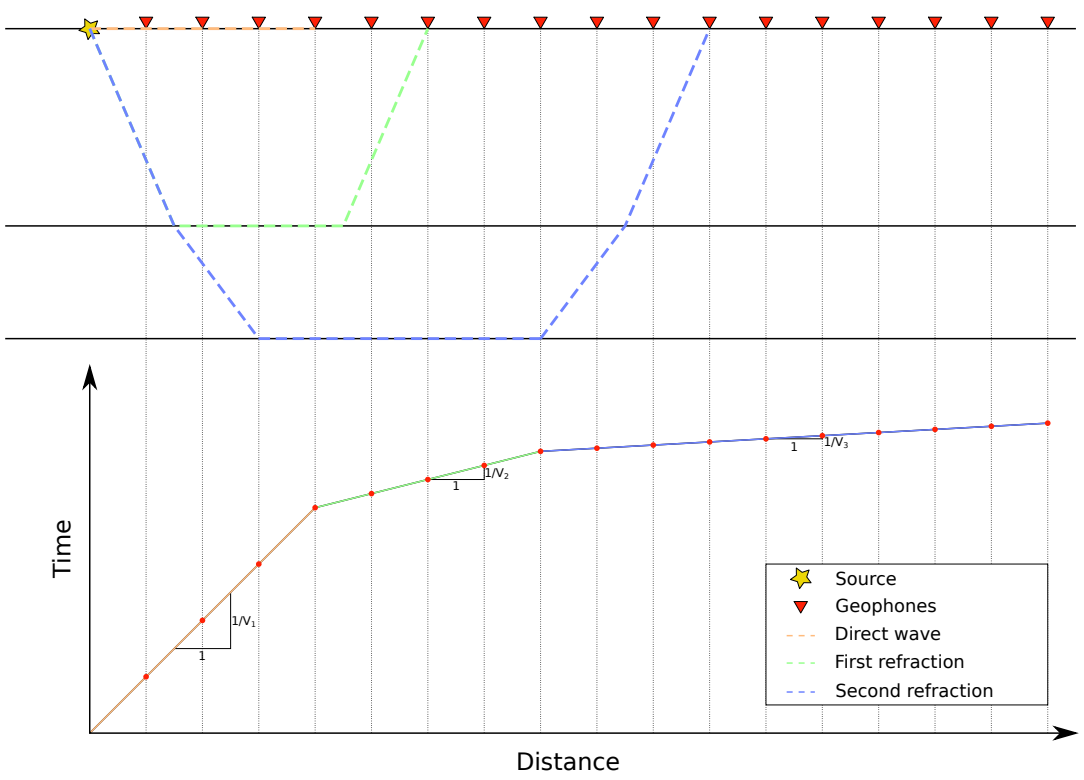

Figure 2.4: Illustration of seismic refraction for a 1D model

From this equation, we can infer the critical angle of incidence, at which the refraction occurs along the interface between the layers ($\theta_2 = 90°$). Given a model of the subsurface with purely horizontal layers, the first break of the seismograms will align along slopes that correspond to the inverse of the layers velocities (Figure 2.4; Green, 1974).

As can be seen in Figure 2.4, the wave propagation is indeed 2D but the geometry of the system is essentially 1D. With a single source and under the hypothesis of horizontal layers, only the thicknesses and velocities of the layers are impacting the model and thus recoverable. More complex modelling can take into account 2D features such as slopes of interfaces when there are multiple sources on the array (e.g.: Green, 1974; White, 1989).

Seismic refraction is now commonly applied using seismic refraction tomography,

where inversion of the 2D or 3D geological properties and structures is possible (e.g.: Flinchum et al., 2022). However, applications of seismic refraction to establish simple 1D (or slightly dipping) structures are still common, for example to recover the water table without the need for piezometers (e.g.: Clark and Page, 2011).

2.1.4 Surface waves

Surface waves are seismic waves whose propagation follows the Earth's surface (see Figure 2.1). Their propagation is dependent on the mechanical and geometric properties of the subsurface. Their amplitude is generally superior to body waves, making their acquisition much easier. Surface waves undergo the phenomena of dispersion where the wave velocity is frequency dependent and which can be observed on a computed dispersion curve (C. B. Park et al., 1999). In most interpretations, the focus is made on the fundamental mode dispersion curve (Socco et al., 2010). After inversion, the resulting models are expressed in terms of P-wave velocity (V_P), S-wave velocity (V_S) and density (ρ). However, surface waves are mostly sensitive to S-wave velocity (Xia et al., 1999; B. R. Cox and Teague, 2016). Often, a link between V_P and V_S is considered through the Poisson's ratio (σ) (Equation 2.3; e.g.: Wathelet, 2008):

$$V_S = V_P \times ((0.5 - \sigma)/(1 - \sigma))^{1/2} \tag{2.3}$$

Surface waves are mainly interpreted in terms of 1D models, with applications in several domains. They are commonly applied to recover the S-wave velocity profile of the first 30 meters of the subsurface, required for seismic design of buildings and infrastructure (Moss, 2008). In another case, Mreyen et al. (2021) used surface waves to infer the depth of the interface of a landslide. Surface waves are also capable of sounding depths that are not reachable with other methods, using passive arrays. The method is thus also common to model the Earth's crust (e.g.: Lebedev et al.,

2013).

2.2 Electromagnetic methods

In geophysics, electromagnetic fields are commonly used to sound electrical and electromagnetic properties of the subsurface (M. Nabighian, 1991). Those properties are the electrical conductivity (or resistivity), the dielectric permittivity and the magnetic permittivity (M. Nabighian, 1987; Zhdanov, 2009). Electromagnetic methods are often used for their capabilities to sound the subsurface without the need for a direct contact with the ground and thus their ability to rapidly cover large areas. The electromagnetic response of the subsurface can either be induced by a transmitter or naturally present.

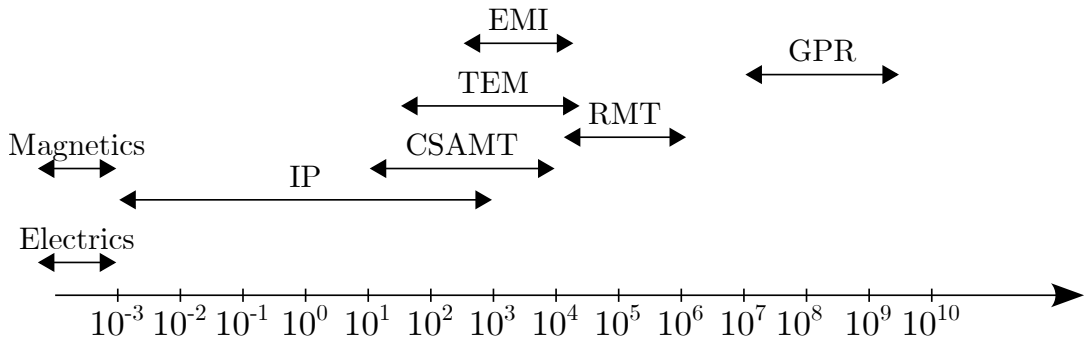

Figure 2.5: EM spectrum in geophysics. Modified from Van De Vijver (2017)

The spectrum of electromagnetic waves can be separated for the different geophysical techniques (see Figure 2.5, Van De Vijver, 2017).

2.2.1 Frequency domain electromagnetic

In frequency domain EM, the device produces a harmonic primary magnetic field, using an induction coil undergoing an alternative current. The generated magnetic

field propagates through the subsurface and if conductors are met, eddy currents are created which in turn generates a secondary magnetic field. The measuring coil then receives this secondary magnetic field, along with the primary field. Then, under the low induction number (LIN) hypothesis, the user can retrieve apparent conductivities (based on the quadrature) and the magnetic susceptibility (in-phase ratio) of the investigated volume (Van De Vijver, 2017; Spies and Frischknecht, 1991).

Those measures can then be translated into models of conductivity and magnetic susceptibility via inversion (e.g.: Farquharson et al., 2003). Often, those inversions are carried out in 1D profiles, but (pseudo-)2D and 3D inversions are becoming more and more common, through lateral constraints on juxtaposed 1D models (e.g.: Siemon et al., 2009) or full 3D models (e.g.: L. H. Cox et al., 2010).

This method is largely applied in agricultural soils investigation (e.g.: Brogi et al., 2019) or landfill characterization (e.g.: Van De Vijver, 2017). Frequency domain EM is also used in archaeology to map ancient human-made structures (e.g.: Smedt et al., 2014). All of those applications greatly benefit from the ease of use of frequency EM and its ability to quickly map large areas.

2.2.2 Time domain electromagnetic

Time domain electromagnetic (TDEM) or transient electromagnetic (TEM) measurements are using the transient signal originating from eddy currents produced in the subsurface after the rapid interruption of direct current in a loop (M. N. Nabighian & Macnae, 1991). Here, the signal measured is only resulting from the secondary field, making it much easier to set up on the field (Christiansen et al., 2008).

The resulting decay curves can then be inverted to retrieve models of the con-

ductivity of the subsurface. Those models are often 1D, since the computation of the forward model is rather complex and therefore inversion on larger dimensions requires much more powerful computing (e.g.: Heagy et al., 2017).

TEM has recently seen a rise in popularity due to the development of airborne TEM (Legault, 2015). This method has seen multiple applications in mineral exploration and hydrogeophysics (N. B. Christensen et al., 2009). Often, the inversion for those measurements are performed in 1D with lateral constraints to provide smooth 2D or 3D models (Auken et al., 2015).

2.2.3 Magnetotelluric

Magnetotelluric (MT) measurements use the naturally occurring EM sources, due either to the solar activity or lightning. The electric fields and magnetic fields that are resulting from those perturbations through the subsurface are measured (Vozoff, 1991). From those measurements, it is possible to retrieve 1D models of the layered Earth conductivity structure.

The main impediment of this approach is the weakness of the resulting signal. Therefore, controlled sources MT techniques have been developed: controlled source audio-frequency MT (CSAMT; Zonge and Hughes, 1991) and radio-frequency MT (RMT; Tezkan, 2008).

MT is often interpreted using 2D or 3D models (e.g.: Mackie and Madden, 1993). However, the method can also be interpreted into 1D models of the subsurface conductivity (e.g.: Kirkby et al., 2015).
This method is most commonly applied for oil and gas (e.g.: Patro, 2017), mineral (e.g.: Farquharson and Craven, 2009) or geothermal exploration (e.g.: Amatyakul et al., 2015).

2.2.4 Surface Nuclear Magnetic Resonance

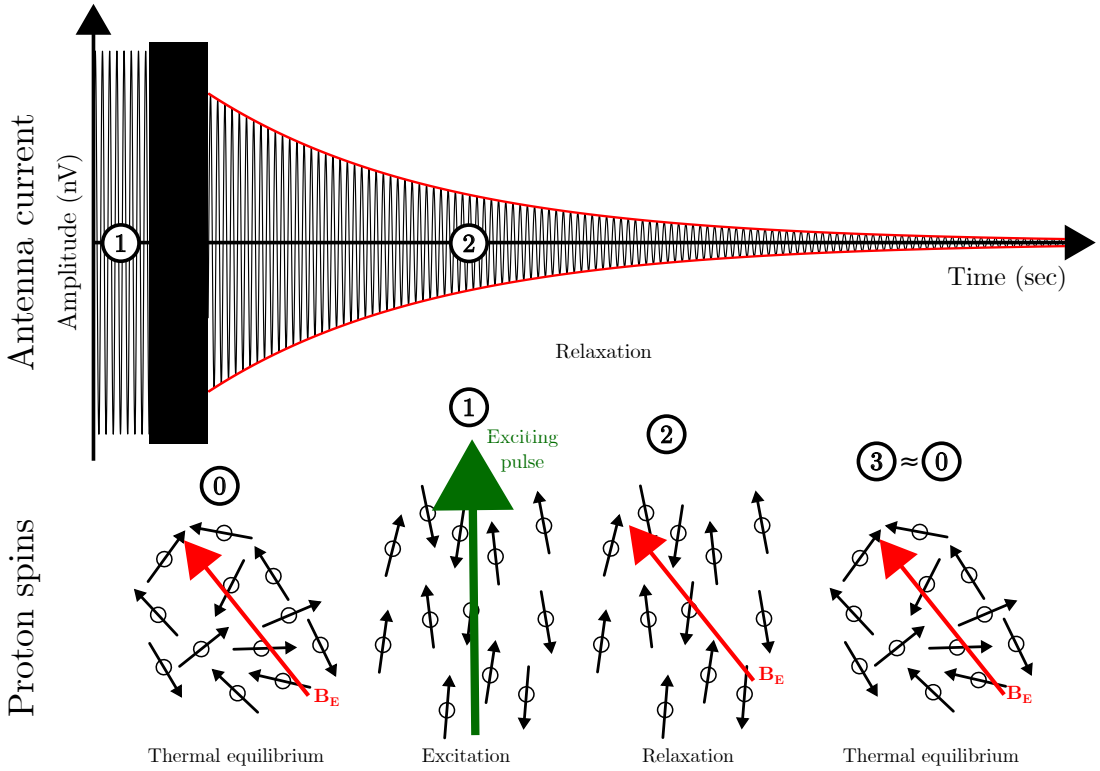

Figure 2.6: Illustration of the sNMR physical principle. At first, the alternating current is injected, creating a strong magnetic field (1), perturbing the initial thermal equilibrium of the protons spins (0). Then, when the current injection is stopped, the protons relax back (2) to the thermal equilibrium (3), producing a decaying alternative current in the receiving antenna.

Surface Nuclear Magnetic Resonance (sNMR) is a geophysical technique dedicated to the detection and characterization of aquifers, whose main advantage and particularity is that the measured signal (an exponential decay) is directly linked to the amount of water present in the subsurface. For more details about the nature and origin of the sNMR signal, we refer to Behroozmand et al. (2015). The main goal of sNMR experiments is to retrieve the distribution of the water content and the relaxation time with depth. The latter depends on the pore space geometry and can be related to hydrodynamic parameters such as the porosity or the per-

meability. The acquisition of sNMR data on the field consists in the injection of an alternative current in a transmission loop (typically circular or in an eight-shape) directly followed by the measurement of the induced current in the reception loop (often the same as the transmission loop). The injection is tuned at the resonance frequency of hydrogens protons contained in water (called the Larmor frequency), perturbing the quantum state of protons contained in water molecules. Once the injection is stopped, the protons are relaxed back to their original state, inducing a current with a decreasing amplitude with time in the receiver loop. This response is called the free-induction decay (FID) and constitutes the collected data (Figure 2.6). The injection is repeated with different pulse moment intensities, sounding different zones of the subsurface.

Although the method can be used for 2D or 3D applications (e.g.: Legchenko et al., 2012), most of the applications are focused on 1D soundings due to the high number of stacking required to achieve a reasonable signal-to-noise ratio. Physically, sNMR is limited to groundwater-related studies. It is thus extensively used for groundwater monitoring and exploration (e.g.: Grombacher et al., 2021; Legchenko et al., 2020).

2.3 Electrical methods

On the lower end of the electromagnetic spectrum, the magnetic component of the signal is lost and only the electrical signal remains. This results in direct current methods which require a contact with the subsurface (Tezkan, 1999). Such methods are very common for their robustness in the measurements of electrical properties of the subsurface (Binley & Slater, 2020). For electrical methods, an electrical current is injected using a dipole of electrodes (AB). This injected current

creates a potential field in the subsurface. The potential is then measured using another dipole (MN) (Binley & Kemna, 2005). Larger dipoles sound deeper in the subsurface (Kirsch and Yaramanci, 2008; Koefoed, 1979). Those measurements enable retrieving models of resistivity in the subsurface using inversion.

In electrical methods, the simplest approach is the vertical electrical sounding (VES, Figure 2.7; Binley and Kemna, 2005). The idea is to dispose a set of dipoles symmetrically around a point and to progressively use larger dipoles, hence sounding deeper geological structures. Multiple types of arrays can be applied to measure the resistivity. The most common for VES is the Schlumberger array, but the Wenner array (as in Figure 2.7) or Dipole-Dipole configurations are also possible (Kirsch & Yaramanci, 2008).

Nowadays, electrical methods are mostly applied using 2D or 3D arrays (Loke et al., 2003) which require multichannel equipment. The use of 1D electrical methods is still common for low budget applications such as groundwater exploration (e.g.: Jha et al., 2008; Clark and Page, 2011; Gaikwad et al., 2021).

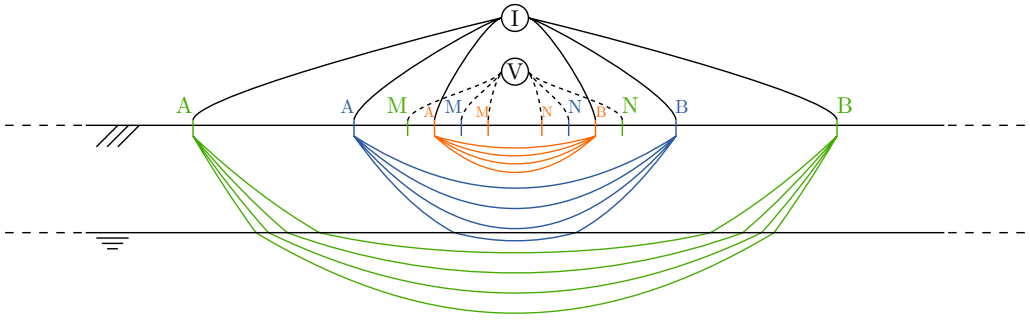

Figure 2.7: Illustration of a vertical electrical sounding (here with a Wenner array)

2.4 Conclusion

Among the different methods discussed in this chapter, all are related to processes that are happening in the actual 3 dimensional space. However, an interpretation in terms of a 1D system has several advantages. The first one is computational since we can rely either on closed-form expressions for the forward model or computationally efficient numerical schemes with a reduced number of parameters. This opens the possibility to perform interpretations at a low computational cost. This advantage is critical for probabilistic interpretations of the geology, as those require multiple thousands of runs of forward modelling. The second is that the corresponding field acquisition can be much faster, allowing to cover more grounds for mapping purposes. However, the 1D hypothesis is in fact scale dependent and may only be accurate if 1D+ effects could indeed be neglected. As an example, the sensitivity of the sNMR for a classical configuration shows that the subsurface outside the loop area is only weakly contributing to the signal making it in fact a 1D method at the scale of the loop. On the contrary, VES or seismic refraction data can be interpreted in terms of 1D models under strong hypothesis, but the sensitivity of those experiment is in nature 3D.

From this chapter, it is clear that 1D geophysical methods are relevant depending on the context (actual layered geology), or on the method (1D sensitivity). The relatively fast computation of the forward models for 1D problems allows us to consider probabilistic approaches that require to run thousands of forward models. Our current implementation of BEL (see chapters 4 and 5) is thus limited to those methods. Among the methods presented in this chapter, seismic surface waves and sNMR are most commonly interpreted in terms of 1D profiles. They are therefore selected for applications in the next chapters.

Chapter 3

Machine Learning: a geophysicist perspective

3.1 Introduction

Machine Learning (ML) describes a series of algorithms that can learn from a given dataset to perform a given task without being explicitly programmed to do so. One of the commonly used definition is written by Mitchell (1997):

> *A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance on T, as measured by P, improves with experience E.*

The different tasks can vary widely and depend on the available datasets. Such algorithms have been present in geosciences for a long time (Dramsch, 2020). Among Machine Learning, one can find a multitude of specific algorithms, with their own sub-variations. One of the earliest ML algorithm is a neural network. Those categories of algorithms are still among the most studied nowadays, with plenty of variations, from Artificial Neural Networks in their simplest form to Deep Learning algorithms that can contain dozens of layers and thus are able to perform more advanced learning (Goodfellow et al., 2016). Support Vector Machines, Ran-

dom Forests and K-means (among others) are ML algorithms commonly applied in geosciences.

Those algorithms can be implemented in two ways (Badillo et al., 2020):

- Supervised learning: the data are labelled, and the task consists in predicting the label of new data.
- Unsupervised learning: the data are unlabelled, and the task is to organize the data based on patterns.

Generally, ML will also aim one of the two tasks below:

- Learning a relationship: regression if supervised, dimensionality reduction if unsupervised.
- Separating the dataset into groups: classification if supervised, clustering if unsupervised.

With those distinctions, we can easily classify different ML approaches. Other approaches also exist (semi-supervised learning, reinforcement learning, etc.) but are more sparsely applied in geophysics. Some algorithms can be used for multiple purposes.

To introduce some common terms of ML, let us use the example of regression. It is a supervised ML algorithm whose task is to learn a relationship between the data variables (features) and specific values (labels). In the most simple cases, the relationship is an equation with parameters that are unknown. From this learning phase, one can then obtain the expected value of the target for any new sample (a set of features).

ML is a very active domain of research in geophysics. Figure 3.1 shows the evolution of the occurrence of the keyword '*Machine Learning*' in geophysics' scientific

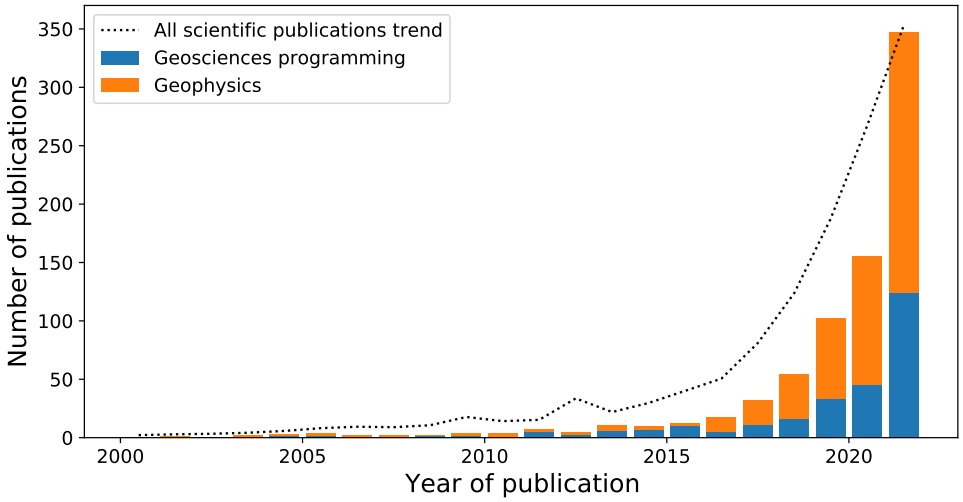

Figure 3.1: Evolution of Machine Learning in geophysical publications from 2000 to 2021. The dataset is separated into journals that have a deeper focus on the programming side of geosciences and journals that are focused on the geophysical aspect of science. Data gathered from *Dimensions* (Digital Science, 2018-).

publications. The graph shows that geophysicists have a slight delay when compared to the whole research community. However, the number of publications has rapidly increased in the last five years. Moreover, ML is not confined to programming focused journals but has also spread to more mainstream geophysical research journals.

In this chapter, we will briefly review ML algorithms and their usual applications in geophysics along with recent research.

3.2 Machine Learning algorithms

A plethora of ML algorithms have been developed in the past century. We will not define them all but rather present the main categories of algorithms. This section familiarizes the reader with ML terms.

3.2.1 Artificial Neural Networks & Deep Learning

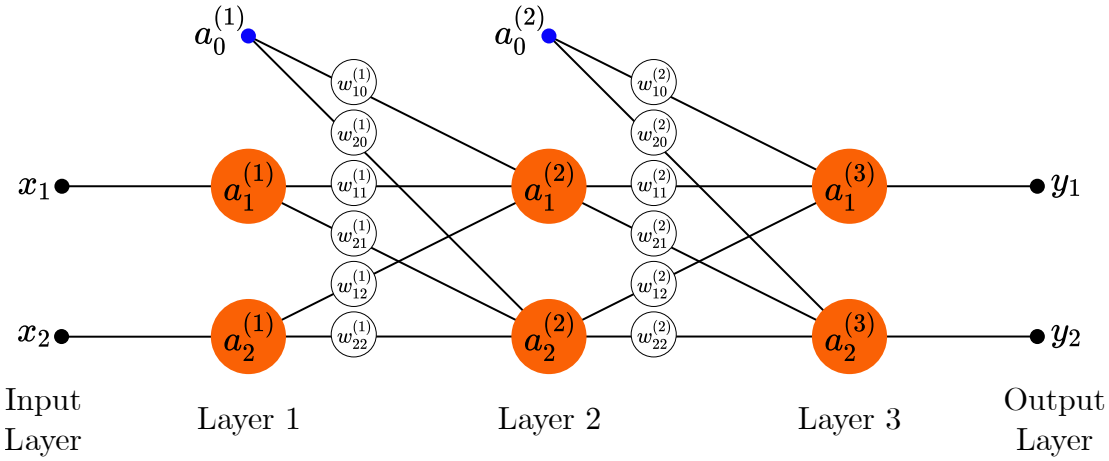

Figure 3.2: Illustration of a neural network.

Artificial Neural Networks are among the first forms of historic Machine Learning and are still widely applied in research and industrial applications. At their core, they consist of multiple neurons that are interconnected. In Figure 3.2, we illustrate a feedforward network. This neural network can be written in matrix form, enabling the simple conversion to computer code (Goodfellow et al., 2016). The network in Figure 3.2 consists of a first layer $a^{(1)}$ built from the inputs $x_1, x_2 \ldots$ and a bias term ($a_0^{(1)}$ with a value of $+1$). This vector is multiplied by a weight matrix $\mathbf{W}^{(1)}$ to provide the values fed into the activation functions g to provide the values of the activation layer $a^{(2)}$. The operation is repeated to propagate from the second to the third layer that corresponds to the output layer. In matrix form, this gives:

$$a^{(l+1)} = g\left(\mathbf{W}^{(l)} a^{(l)}\right), l = 1, 2 \tag{3.1}$$

where the matrix $\mathbf{W}^{(l)}$ contains all the weights $w_{ij}^{(l)}$:

$$\mathbf{W}^{(l)} = \begin{bmatrix} w_{10}^{(l)} & w_{11}^{(l)} & w_{12}^{(l)} \\ w_{20}^{(l)} & w_{21}^{(l)} & w_{22}^{(l)} \end{bmatrix}, \tag{3.2}$$

and the vector $a^{(l)}$ contains the values of the network at the nodes from the layer l:

$$a^{(l)} = \begin{bmatrix} a_0^{(l)} \\ a_1^{(l)} \\ a_2^{(l)} \end{bmatrix}, \tag{3.3}$$

with $a_0^{(l)}$ being the bias term added to each layer and equal to one.

The activation function $g()$ can take multiple forms. The most common are (Goodfellow et al., 2016):

- Sigmoids, for example:
  - The logistic function:
$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.4}$$
  - The hyperbolic tangent:
$$f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.5}$$
- ReLU (Rectified Linear Unit):
$$f(x) = \max(0, x) \tag{3.6}$$

- SoftMax, applied to a vector input ($\mathbf{z} = (z_1, \ldots, z_K)$):

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \text{ with } j \in \{1, \ldots, K\} \tag{3.7}$$

The choice of an activation function will depend on the type of neural network and the type of layer (Goodfellow et al., 2016). Hidden layers typically have the same type of activation function: Sigmoids or ReLU. The ReLU function has grown in popularity in the last decade, while sigmoids were more common in the 90s. For the output layer, there can be either no activation function (the input is the output), a sigmoid to bound the output between 0 and 1, or a SoftMax function (normalized logistic function) to transform the input into probabilities with a sum equal to 1.

Depending on the architecture of the neural network and the combination of activation functions, the algorithm can take different names. The first distinction is made between a low number of hidden layers (artificial neural network) and a large number of hidden layers (deep learning). The more the layers, the more complex the learning capabilities of the neural network. On the other hand, this addition of layers requires more models in the training set to achieve reasonable training and validation errors.

Neural networks can be trained to perform different types of tasks. Learning for a neural network is often supervised but can also be unsupervised as is the case for autoencoders, generative adversarial networks, self-organizing maps, ... (Dike et al., 2019). The different uses of those networks are discussed in the next section (3.3).

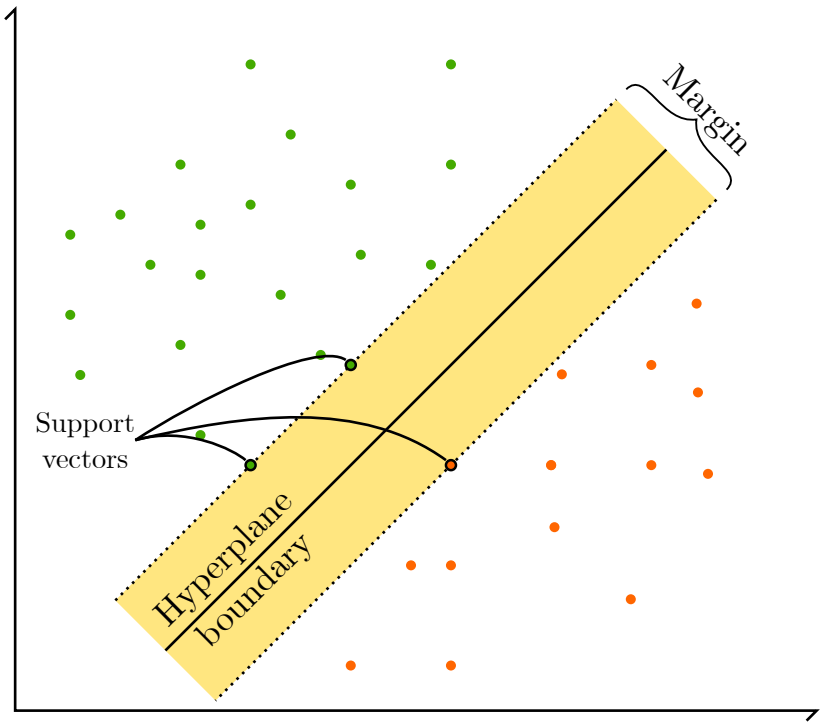

Figure 3.3: Illustration of the support vector machine algorithm for a linearly separable dataset with 2 classes (green and orange).

3.2.2 Support Vector Machine

Support Vector Machine (SVM) were first described by Boser et al. (1992). Their aim is to "*maximize the margin between the training patterns and the decision boundary*" (Boser et al., 1992). They are designed to provide the best decision boundary between categories of a dataset (different labels). Figure 3.3 illustrates the process for a linearly separable dataset. The decision boundary is a hyperplane whose dimension depends on the dimensions of the dataset itself. It is defined by the support vectors: points that are on the margins. The goal is thus to maximize this margin (Badillo et al., 2020; Noble, 2006; Hearst et al., 1998). By nature, SVM is limited to supervised learning and classification purposes.

In the case of non-linearly separable datasets, multiple approaches can be ap-

plied:

- Accept a given number of wrong classifications (called *slack variables*) but keep the boundary as simple as possible.
- Transform the dataset using a kernel to make it linearly separable (in the hyperspace)

Both approaches can be mixed to avoid an over-complex hyperplane boundary (Badillo et al., 2020).

3.2.3 Decision Trees, Random Forests & Gradient Boosters

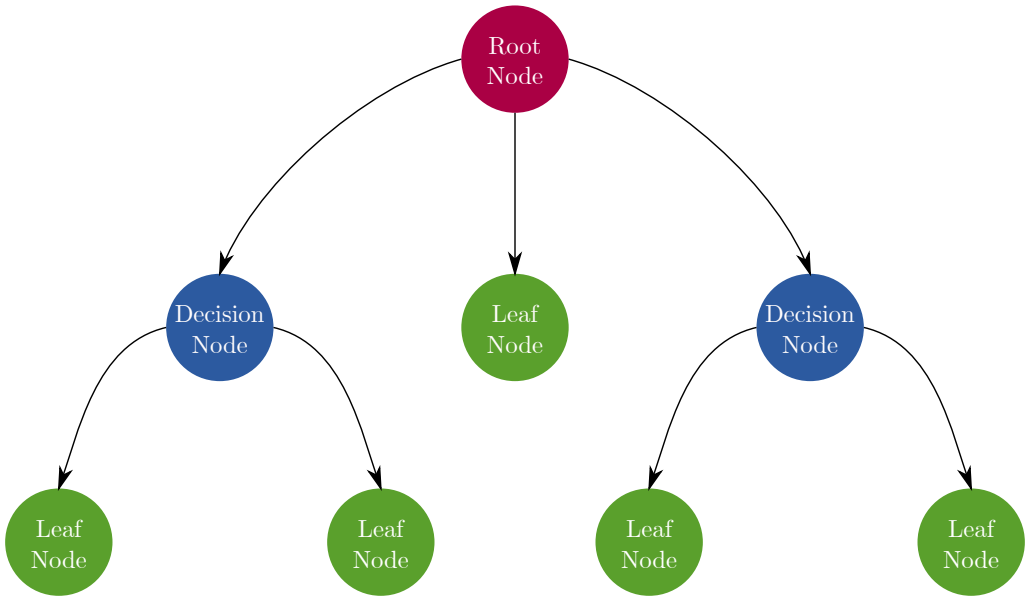

Figure 3.4: Illustration of a decision tree.

Decision trees are among the earliest forms of supervised ML. A decision tree is composed of multiple nodes (Figure 3.4). For each node, the dataset is split at a position in the features set. The split provides the tree with a better separation of

the dataset and thus a better ability to predict the class label (classification trees) or the value (regression trees) (Suthaharan, 2016).

A decision tree is defined by its number of layers. A deep decision tree is often able to reproduce exactly the training set labels. However, this can lead to an overfitting of the dataset (Badillo et al., 2020). To overcome this impediment, multiple approaches have been developed.

Random forests are composed of multiple deep decision trees and use their different outcomes to predict a value or class by a majority vote. Gradient boosters on the other hand use multiple shallow decision trees subsequently to improve on the prediction (Badillo et al., 2020).

3.2.4 Others

Even though the different ML algorithms described in the subsections above are constituting the majority of applications nowadays, other methods also exist and have been developed along with improving computing capabilities. Citing all the existing algorithms would be overcomplicated and not productive. We will nonetheless cite and briefly describe some popular options.

K-Nearest Neighbours (kNN) are one of the simplest classifiers that exist (Kramer, 2013). The goal is to provide a label for a new sample. The algorithm finds the k nearest samples in the training data and extracts their labels. The nearest samples are defined by a metric, usually the Minkowski metric of degree p (Eq. 3.8). The class for the new features set is deduced from the most prominent class in the neighbouring samples.

$$\|\mathbf{x}^{'} - \mathbf{x_j}\|^p = \left(\sum_{i=1}^{q} \left| (x_i)^{'} - (x_i)_j \right|^p \right)^{1/p} \tag{3.8}$$

K-Means is the pendant of kNN for clustering. The algorithm tries to minimize

a metric (usually the Minkowski metric with p=2, i.e. the Euclidean distance) between k centroids and the samples that are present in the k clusters (Likas et al., 2003).

Naive Bayes is a form of classification that uses a simplified version of the Bayes' rule to compute the expected probability of a sample being in a given class from its different features (H. Zhang, 2004).

$$p(c|E) = \frac{p(E|c)p(c)}{p(E)} \tag{3.9}$$

Logistic regression (D. R. Cox, 1958) is a classifier (and not a regression) algorithm commonly used in industrial applications. The idea is to fit a modified logistics sigmoid (Eq. 3.4) with a linear regression on the exponent of the exponential. This fit will provide with a probability of a sample being part of a given binary class. The method has been extended for multi-classes cases since its development (Böhning, 1992).

Principal Components Analysis (PCA) is also a common ML algorithm for dimensionality reduction. The algorithm seeks for a set of orthogonal axes in such a way that the first axis encompasses the maximum variation (Hotelling, 1933).

3.3 Common use cases in geophysics

Geophysicists use Machine Learning for multiple tasks. We will briefly discuss most of them and explain their advantages over other approaches.

3.3.1 Classification

Classifying rock properties based on geophysical data is one of the key challenges in geophysics. The idea is that the end-users are rarely interested in the raw geophys-

ical results but rather in a proxy that can be of interest to them (Aster et al., 2013). For example, in Isunza Manrique et al. (2019) and Isunza Manrique et al. (2022) studies, one of the goal was to transcribe the geophysical models into resources distribution models. Brogi et al. (2019) classified soils to enhance agricultural yield based on Electromagnetic Induction data. Building geological models from different geophysical data is also a common task (e.g.: Marzan et al., 2021; Grana et al., 2020; Bressan et al., 2020). ML algorithms can help to automate those tasks and provide unbiased interpretations.

Multiple algorithms can perform those classification tasks. The most common are Neural Networks (NN - e.g.: Di et al., 2019; Bressan et al., 2020) and Random Forests (RF - e.g.: C. W. Christensen et al., 2020). Other approaches are also utilized, such as Support Vector Machine (SVM), Decision Tree (DT) or Logistic Regression, etc. (Bressan et al., 2020). Usually, the training set is built from collocated ground-truth data and geophysical data. This means that the separation of features by ML algorithms is usually supervised (Osisanwo et al., 2017). However, some unsupervised clustering algorithms are also used to identify patterns into datasets (Wallet & Hardisty, 2019). Those patterns can later be identified by the expert as different lithologies or any other parameter of interest in the absence of ground-truth data (e.g.: Bauer et al., 2012). The applied unsupervised ML algorithms can vary. Among many others, Bauer et al. (2012) applied Self-Organizing Maps to cluster patterns from seismic and magnetotelluric models and Wallet and Hardisty (2019) applied Gaussian mixture models for facies recognition from seismic data.

3.3.2 Surrogate modelling

The forward modelling of geophysical processes is at the heart of geophysics. The operation consists in providing a simulation of a dataset based on a given geophysical

model (Aster et al., 2013). This operation can be CPU consuming and optimization is often required. ML is a suitable candidate to improve on this aspect as it is able to efficiently learn a direct relation between models and datasets from a large training set. Such an approach is efficient when (1) the forward modelling is difficult to compute or (2) when the number of models to simulate is large. In both cases, even thought training the algorithm is time-consuming, the end-goal is to provide the end-user with computationally efficient and fast proxy for the forward operations.

Wave-based methods are among the main methods to see a large application of such an approach. Especially, when dealing with full waveform datasets. In those cases, the full waveform forward requires heavy computations and CPU time that make the use of such approaches impractical. Simplifying the operations results in a significant gain for the inversion. Giannakis et al. (2019) used neural networks to approach the full-waveform ground penetrating radar datasets. In a similar approach, Conway et al. (2019) trained a neural network for the forward modelling of 3D magnetotelluric datasets. Those solvers are then used for faster inversion with classical approaches and could be applied for stochastic inversions as well.

ML based forward operations can also boost the performances of a probabilistic approach (Valentine & Sambridge, 2017). In this case, even though the forward operator is already relatively efficient, there is still a use case for faster computations, since the operations need rerunning several thousands of times. For example, Hansen and Cordua (2017) used a neural network based forward for GPR travel-time Monte Carlo sampling. Aleardi et al. (2022) used a neural network approximation of the electrical resistivity tomography forward operator to perform a Markov chain Monte Carlo and Ensemble Smoother inversions.

The main issue with those approaches is the need to fine-tune a network for every case. To circumvent this, Siahkoohi et al. (2019) proposed to apply transfer

learning to better model seismic data from previously trained neural networks.

3.3.3 Inversion

ML is often applied to replace the inversion procedure in geophysics (Kim & Nakata, 2018). The advantage of ML algorithms in geophysical inversion lies in the ability of such algorithms to model highly non-linear effects easily (Russell, 2019). The most common approach is to apply an artificial neural network (or any given variation) to a problem. Training the algorithm can be done either by gathering a large dataset of collocated geological models and geophysical data (e.g.: Joshi et al., 2021) or by building a training set from scratch (e.g.: Michel, Nguyen et al., 2020; Guo et al., 2021). The latter approach is more common, due to the need of large training sets of models and geophysical data that show features similar to the data samples to interpret. Even though geophysicists are used to deal with large datasets, the variations of contexts make it complex to reuse them for different sites. Some databases of complex synthetic models are beginning to see the light, with for example Noddy (Jessell et al., 2022).

Another approach that has been applied is to learn the gradient from a model to the optimal update during the inversion (Sun & Alkhalifah, 2020). Since this process requires the computation (or at least the approached version) of the inverse of the Hessian matrix, a ML approach can be beneficial if efficiently trained.

It is also possible to use ML to perform the inversion in the latent space of those neural networks (lower dimensionality space). Such an approach was applied by Lopez-Alvis et al. (2021), Lopez-Alvis et al. (2022) and Laloy et al. (2017).

Another possible help from ML in inversion is the use of generative models. Those algorithms can generate models that show similarities to models that already exist (training set). This approach can be beneficial to perform Bayesian inversion

(e.g.: Laloy et al., 2018; Scheiter et al., 2022). The generative models of predilection are either Generative Adversarial Networks (e.g.: Laloy et al., 2018; Mosser et al., 2020) or Variational Autoencoders (e.g.: Lopez-Alvis et al., 2022).

3.3.4 Signal processing

Processing the raw signal from geophysical equipment into usable data is key in geophysics, especially for noise prompt data such as sNMR and seismic traces. Various methods have been developed to overcome this issue. The case of sNMR is still very active in research today, as evidenced by the review of Kremer et al. (2022) on the subject. This has led to several attempts at this task with ML.

Yu et al. (2019) train a convolutional neural network (CNN) algorithm to filter out noise in seismic images through supervised learning. The idea is to produce noisy samples and to label them with the corresponding noise-free sample. The CNN will thus be able in the end to reproduce noise-free data from raw data. Such a concept has also been used for sNMR (Wei et al., 2021). Another method is to use autoencoders to process the noise out (Bhowmick et al., 2019; Vincent et al., 2010). In this approach, the idea is to benefit from the dimensionality reduction capabilities of autoencoders to remove noise. This comes from the fact that noise is usually gathered in higher dimensions as it is less prominent in the dataset. Therefore, a network trained to reproduce the data itself will reproduce the data without those higher dimensions and remove the noise (Vincent et al., 2010). Bhowmick et al. (2019) applied this approach to seismic data and well logging.

3.3.5 Petrophysical relationships

Using ML, it is also possible to learn petrophysical relationships. The advantage of ML in this application is the possibility to learn complex relationships between para-

meters when compared to more classical approaches limited to analytical formulas with few parameters to remain usable (e.g.: Moghadas et al., 2019; Gottschalk and Knight, 2022).

Those applications offer the advantage to avoid the need for the development of complex geostatistics models. Among others, Gottschalk and Knight (2022) used a Gradient-Boosted Decision Trees (GBDTs) to extend borehole descriptions to field-scale geophysical models. Moghadas et al. (2019) used an Artificial Neural Network (ANN) to model the soil moisture from geophysical data in a spatio-temporal approach.

3.4 Uncertainty in Machine Learning

By design, most ML algorithms provide a single output for a given input, and are mostly deterministic. However, several techniques exist to extend ML to probabilistic approaches.

Among those methods, classification algorithms are the most suitable methods for providing probabilities. Using Neural Networks with a SoftMax output layer, the user receives as output the probability of all the classes given the input (e.g.: Grana et al., 2020). Often, those are used to extract the most probable output and assign this class to the values. However, it is possible to use the output directly as an estimation of the probability of any class. Similarly, Random Forests are providing the user with multiple outputs, from which a majority vote is made to declare a single output. However, it is possible to use those multiple outputs as possible realizations of the classes and to deduce probabilities of classes from those (e.g.: Cracknell and Reading, 2013).

For inversion, it is also possible to use Neural Networks designed to provide un-

certainty on the parameters estimations. Multiple approaches have been developed to tackle this task (Bishop, 1995). In geophysics, the appraisal of uncertainty is a key parameter when dealing with inversion, and therefore, ML probabilistic inversion is an active research topic (e.g.: X. Zhang and Curtis, 2021; U. Meier et al., 2007; Aleardi et al., 2022). In Aleardi et al. (2022) approach, they use a ML forward operator to perform a probabilistic inversion, avoiding the cost of the forward operator. Devilee et al. (1999) were among the first to use neural networks to obtain posterior probability from geophysical data. They proposed to apply histogram and median networks, whose training is based on the known probability densities for the training datasets. More recently, scientific work has been focused on providing a better estimation of the posterior uncertainty, using more advanced descriptions of the posterior via Gaussian kernels in Mixture Density Networks (MDNs) (U. Meier et al., 2007) or sampling the posterior distributions using Invertible Neural Networks (INNs) (X. Zhang & Curtis, 2021).

U. Meier et al. (2007) used MDNs to perform the inversion for the Moho depth (single parameter) from seismic dispersion curves. There, the distributions are described by Gaussian kernels that are summed to provide the posterior. The algorithm is capable of learning the uncertainty on the bases of numerous (here 500000) forward models runs.

X. Zhang and Curtis (2021) used INNs to approximate the posterior model space for multiple seismic data types. In this method, the network maps models to datasets augmented by random variables representing the noise and vice-versa (Ardizzone et al., 2019). Doing so, X. Zhang and Curtis (2021) were able to produce posterior models for a given field dataset (augmented by sampled random variables), thus producing different outputs representing the posterior.

3.5 Conclusion

ML proposes a large variety of algorithms that can solve common problems in geophysics. In this short review, we showed that the diversity of ML algorithms can offer the geophysicist with several options when facing a challenge. Neural networks are the most versatile ML algorithms as they can take multiple forms and answer different questions, depending on the task at hand. They are well suited for supervised learning tasks, where the author already knows the labels that are associated with each sample in the training set and wants to label a new sample. However, for unsupervised tasks, neural networks are less frequent, due to the difficulty to train a network on an undefined (or ill-defined) objective. Many algorithms are then available to propose adequate handling of the defined tasks. The different tasks performed by geophysicists are often supervised, but unsupervised tasks are also possible (Table 3.1).

	Supervised Learning		Unsupervised Learning	
	Classification	Regression	Clustering	Dim. reduction
Classification	+		+	
Surrogate modelling		+		
Inversion		+		+
Signal processing		+		+
Petrophysics		+		

Table 3.1: Summary of the tasks commonly performed by geophysicists classified by the type of learning

ML algorithms can present significant advantages when compared to more classical approaches. In the case of classification, the tedious task of assigning rock properties to a given cell of a model, based on geophysical inversion results, can be efficiently assigned to ML algorithms in order to provide fast and accurate models of the geology. This approach can be performed either supervised or unsupervised.

In the first case, the groundtruth data must be present in sufficient quantity on the studied site, but the classification is effective at retrieving similar patterns across the site. In case of absent or insufficient groundtruth, the analysis can be performed unsupervised. Then, the algorithm will discover categories of the response that follow some similar patterns and will be able to locate those categories. The geophysicist is then responsible to label the clusters that have been discovered based on prior knowledge of the site and experience.

Then, ML can significantly reduce the CPU cost of the forward modelling operations of geophysical phenomenons. The non-linearity of such operations, making the use of costly finite difference or elements methods to solve the problem, can be learned by ML algorithms and will result in much faster operations. This is greatly helpful for Bayesian inversion of complex datasets where multiple thousands of forward runs are required.

ML can also lead to efficient ways to provide geophysical images on the basis of geophysical data. The ML algorithm can perform the full inversion operation, similarly to what a classification algorithm would provide in a discrete space, but with geophysical parameters on a continuous space. Other approaches consist in using Machine Learning to perform the inversion in a lower dimensionality space (the latent space) or to approach the updates given the dataset and current models to perform gradient descent. To top it all, ML provides efficient algorithms to sample priors defined from training images, which is beneficial since the definition of complex prior in 2D and/or 3D can be mathematically cumbersome and CPU consuming.

Finally, the geophysicist often faces noisy datasets. ML is able under specific circumstances to improve the signal-to-noise ratio of the datasets. The noise reduction can either be performed by training an algorithm to recognize the noise

and remove it from the signal or by using dimensionality reduction algorithms, that naturally tend to reduce the impact of the noise on the lower dimensionality space.

In conclusion, geophysicists greatly benefit from ML in their workflows. The main issues with those approaches are the training phases, where large databases are required and often lacking. Such databases are currently being built (e.g.: Jessell et al., 2022). Nonetheless, direct applications of ML in geophysics are currently available.

However, ML is still not yet suitable for a widespread application of inversion with uncertainty quantification at a reasonable training cost. The different approaches that exist have limitations.

- They can be limited to specific distributions types (e.g.: Maiti et al., 2013)
- They can be limited in the number of models variables due to the need for a large number of variables to describe non-specific distributions (e.g.: U. Meier et al., 2007)
- They require large training sets which is CPU intensive (e.g.: U. Meier et al., 2007; X. Zhang and Curtis, 2021)
- They need to be trained either on known posterior distributions given the dataset and the prior (e.g.: Devilee et al., 1999) or with a form of likelihood for a model and the dataset with noise (e.g.: X. Zhang and Curtis, 2021)

The algorithm discussed in this manuscript (BEL1D) can be seen as a hybrid solution that overcomes those limitations at the cost of a lower versatility. It can be seen as a supervised regression approach. We build a training set from synthetic datasets whose features are the different discrete data points, with their associated labels being the models they originate from. This training set is sampled from a

prior model space, meaning that the learning phase is valid only for models that are inside this space. From this training set, BEL1D is able to learn a relationship that maps a given set of features to multiple probable labels in a latent space. Doing so, we provide an efficient approach to stochastic geophysical data interpretation through ML.

The algorithm in question is detailed in the next chapters.

Chapter 4

Bayesian Evidential Learning for 1D geophysical data: theory and development

In this chapter, we will present the algorithm for Bayesian Evidential Learning (BEL), the Bayesian approach that is developed in this dissertation[1]. We will also apply the algorithm to surface Nuclear Magnetic Resonance (sNMR) geophysical data in different contexts. To validate the approach, we will compare the results to results obtained using a classical McMC approach.

[1]This chapter is largely based on the publication entitled: 1D geological imaging of the subsurface from geophysical data with Bayesian Evidential Learning (Michel, Nguyen et al., 2020). The different authors of this paper all contributed different aspects to this research.

- Code development: Hadrien Michel - advices: Thomas Hermans and Frédéric Nguyen
- Data acquisition: Thomas Kremer, Ann Elen and Hadrien Michel
- Data processing: Hadrien Michel, Thomas Kremer, Ann Elen - advices: Thomas Hermans and Frédéric Nguyen.
- Bayesian inversion: Hadrien Michel - advices: Frédéric Nguyen, Thomas Hermans and Thomas Kremer
- Parametric analysis: Hadrien Michel
- Writing: Hadrien Michel
- Proof-reading: all contributing co-authors

4.1 Theory and development

4.1.1 Introduction

In this chapter, we present Bayesian Evidential Learning (BEL) (Scheidt et al., 2018) as an innovative method for the unidimensional geophysical imaging of the subsurface. Bayesian Evidential Learning (BEL) is a general framework to handle data in geosciences, from the optimization of the data acquisition to the data interpretation, prior falsification, global sensitivity analysis, and the prediction of relevant model responses (Scheidt et al., 2018).

We will refer to the approach presented here as Bayesian Evidential Learning 1D imaging (BEL1D). It uses statistics-based relationships between forecast variables and data learned from realizations sampled from a prior distribution. As opposed to inversion methods, this approach does not require the time-consuming inversion of the dataset, but rather numerous (and parallelized) runs of the faster forward model. So far, BEL has been used to predict (among others) the dynamical response of subsurface models from well monitoring data (Satija & Caers, 2015), time-lapse geophysical experiments (Hermans et al., 2018), push-pull tests (Hermans et al., 2019), updating a geological model from new boreholes and predicting gas volume production (Yin et al., 2020), defining wellhead protection areas (Thibaut et al., 2021), i.e. model predictions mostly characterized by smooth variations in time and/or in space. BEL was also applied to show the non-uniqueness of the gravity inverse problem (Phelps et al., 2018).

In this contribution, we extend BEL to 1D geophysical imaging. Such 1D imaging arises in many geophysical methods (see Chapter 2) where forward models assume a succession of homogeneous horizontal layers such as surface nuclear mag-

netic resonance (sNMR e.g., Behroozmand et al., 2015), surface seismic waves (e.g., Socco et al., 2010), electromagnetic surveys (e.g., Hanssens et al., 2019; Li et al., 2018), or vertical electrical sounding (e.g., Jha et al., 2008). The main contributions in this chapter compared to previous works are:

1. The application of BEL1D for static geophysical imaging. Here, we predict directly the posterior distribution of spatially distributed geophysical model parameters

2. The estimation of the data-prediction relationship using kernel density estimation instead of Gaussian regression, because the former can be applied when the relationship deviates from linearity and Gaussian assumptions

3. The validation of the approach by comparing it with state-of-the-art McMC

4. The illustration of the similarity between the distance-based global sensitivity analysis and the canonical correlation analysis

5. The analysis of the number of samples in the prior

This chapter first introduces our implementation of BEL: Bayesian Evidential Learning 1D imaging (BEL1D). The method is then validated within a numerical example. To demonstrate the broad applicability of our approach, we also provide a validation within a non-geophysical related inverse problem (Appendix A: "Testing BEL1D on a synthetic mind experiment: oscillations of a pendulum"). Then, we present an example of BEL1D applied to field data, providing evidence that the method is already mature. Finally, we discuss elements governing the accuracy of BEL1D, such as the number of models sampled in the prior and the choice of the bandwidth for kernel density estimation.

4.1.2 BEL applied to 1D imaging: methodology

In this work, the focus is on computing geophysical models and assessing the uncertainty of the inferred models (Phelps et al., 2018). This interpretation part of BEL is also known as the prediction-focused approach (PFA), introduced by Scheidt et al. (2015). Contrary to deterministic approaches, this method does not rely on the stabilization of the ill-posed inverse problem, such as regularization for example, that imposes non-realistic constraints to the solution. It rather relies on the constitution of statistical relationships between targets (the set of parameters of interest to the end-user) and data (predictors). These relationships are originating from models and (numerical) simulations of the geophysical problem that reflect the available prior knowledge.

BEL has already been applied previously to geophysical data - time-lapse electrical resistivity - by Hermans et al. (2016). Hermans et al. (2016) demonstrated the applicability of BEL to predict variations of subsurface physical properties with time-lapse electrical resistivity data. More broadly, they also demonstrated that BEL was a possible tool for geophysicists.

However, Hermans et al. (2016) presented a scheme that required numerical simulations of groundwater flow and transport and petrophysical relationships specific to hydrogeophysical monitoring (i.e. time-lapse data), limiting the extent of the geophysical prior model, making it a hydrogeophysical coupled framework (Hermans et al., 2021). Solving the time-lapse problem in geophysics is easier because we can invert for changes in the model (e.g., Kemna et al., 2002; Nguyen et al., 2016), and that the petrophysical relationship is generally simpler (Linde et al., 2015). In Hermans et al. (2016), a petrophysical relationship and the background distribution of resistivity were taken as known, which simplified the inference of the

posterior distribution. Static data are also characterized by a higher noise level and thus more challenging to process (LaBrecque et al., 1996; Lesparre et al., 2017). This is also the first application where BEL is used to directly estimate spatially distributed model parameters. In our case, the forecast variables are static geophysical models, directly linked to the geophysical experiment, hence no petrophysical relationship is required. Such situations generally bear more prior uncertainty and wider variability. Moreover, our aim is to propose a fast generation of subsurface models along with uncertainties.

The adapted BEL1D method consists of six steps (Figure 4.1):

1. Definition of the prior uncertainty based on prior field knowledge, generation of prior models (yellow box in Figure 4.1) and associated data (blue box in Figure 4.1) by forward modelling

2. Reduction of the dimensionality of the data using principal component analysis (PCA)

3. Constitution of statistical relationships between the model parameters and the reduced data (green cloud in Figure 4.1), using canonical correlation analysis (CCA)

4. Generation of posterior distributions for the model parameters in reduced space by constraining the bivariate distributions to field data (red box in Figure 4.1) using kernel density estimators (contrary to previously used Gaussian process regression (Hermans et al., 2016))

5. Sampling of the constituted (non-Gaussian) distributions

6. Back-transformation of the samples into the original space, delivering a set of 1D models of the subsurface that are constrained to the knowledge of the

geophysical data (purple box in Figure 4.1).

Each step is explained in detail in the following subsections.

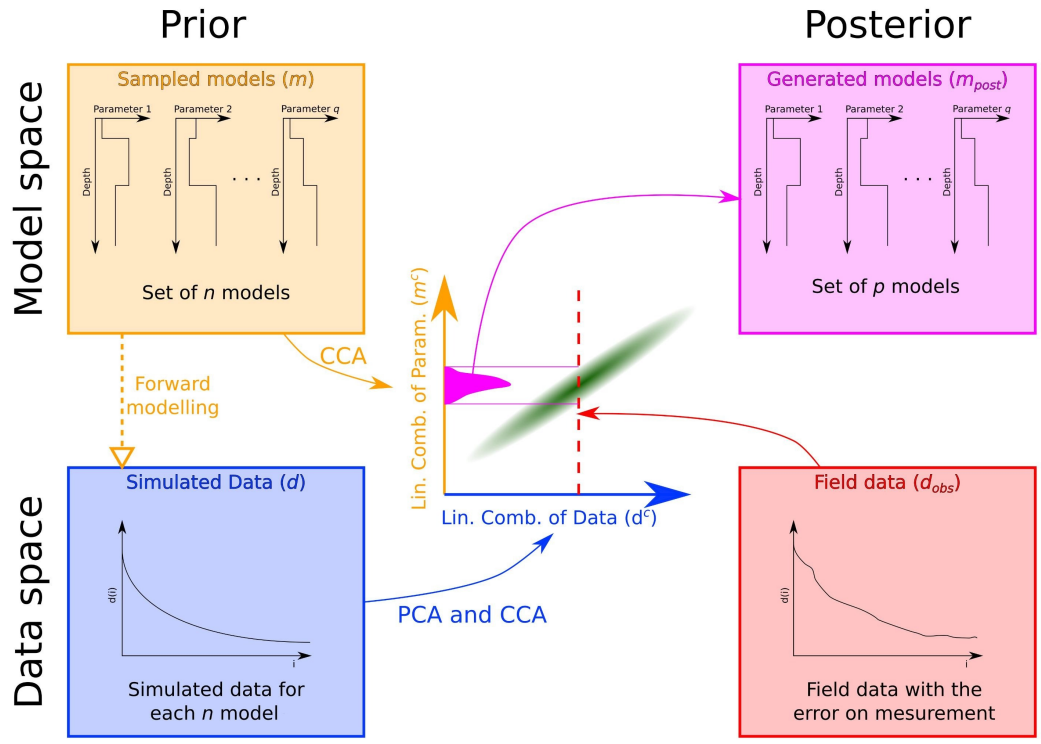

Figure 4.1: Illustration of the BEL1D framework. The prior is sampled, producing a set of models (yellow box), that yield a set of corresponding datasets (blue box) through forward modelling. The datasets' dimensionality is reduced using PCA, then a relationship linking models and datasets is deduced using CCA, producing a distribution of models in the CCA space (green ellipse). When the (noisy) field data is known (red box), it is projected into the CCA space, yielding to a posterior model space in this reduced space (pink distribution). Then, models tare sampled from those distributions and back-transformed into the original space (pink box), producing a set of samples from the posterior model space.

Although BEL1D can be applied to any 1D inverse problem with a limited number of parameters, for the clarity of this chapter, we choose to illustrate the method in using sNMR (see Chapter 2, Subsection 2.2.4).

4.1.2.1 Step 1: prior realizations and associated data

In BEL1D, the models are described in a classic manner with a finite number of layers N_L, the last one being a half-space. For each layer, the N_P physical properties related to the geophysical experiment and layer thickness are assigned. This means that a model can be described with a low number of parameters: $(N_L \times (N_P + 1)) - 1 = q$. Defining the prior model space $f_H(\mathbf{m})$, i.e. the space describing the prior knowledge of the subsurface physical properties before data acquisition, consists in assigning a distribution to the values of each of those q parameters. The shapes of those distributions are unconstrained and must best represent the prior knowledge of the survey site, either originating from previous experiments or from general geological and geophysical considerations. Once the prior model space is defined, we generate random models within the prior boundaries, hereafter called prior realization: $\mathbf{m}(q \times 1) \in f_H(\mathbf{m})$.

For example, the prior model space for sNMR experiments is defined using both the water content (in %) and relaxation time (in ms) distributions along depth. This leads to a finite number of layers, of unknown thickness, and of unknown water content and relaxation time, thus to the description of a model with only five parameters for a 2-layer model.

For all prior realizations, we compute their associated response using the same acquisition parameters as the experimental conditions. Doing so, we produce the synthetic data vector: $\mathbf{d}(l \times 1)$. To perform this operation, we use the (non-linear) forward operators (K) of the geophysical problem of interest such that (Equation 4.1)

$$\mathbf{d} = K(\mathbf{m}) \tag{4.1}$$

For more details about the forward operator associated with the sNMR method,

we refer to Hertrich et al. (2007). Often, the dimensionality of the data (l) is relatively large. For sNMR, the data space is composed of tens of time-decaying signals (one for each pulse moment) with sampling rates on the order of 100 Hz measured for about 1 second, thus resulting in several thousands of dimensions.

4.1.2.2 Step 2: reduction of the dimensionality of the data using principal component analysis

At this stage, we have generated the "model space" which contains the prior realizations, and the "data space" which contains the synthetic geophysical data sets associated with each model. To explore the statistical relationships that exist between these two spaces, one must first ensure that their dimensionality is sufficiently low (typically about 10 dimensions per space) to allow for reasonable computational costs. In the case of sNMR, as in most other cases if the prior models space remains simple, the model space will generally have a reasonable number of dimensions, thus no dimension reduction is required ($\mathbf{m^f} = \mathbf{m}(q \times 1)$). As stated previously, the data space is generally large and cannot be handled properly under reasonable computational costs. In this case, the Principal Component Analysis (PCA) technique can be used to transform the data space and reduce its dimensionality while preserving its original variability.

The PCA method seeks for linear combinations of variables that maximize the variability in the first dimensions using eigenvalue decomposition (Krzanowski, 2000). This operation results in PCA scores, the values of the transformed data. This way, the first dimensions are the most informative about the dataset, and the others only represent a small amount of the variability. Those latter can be discarded, to gain memory and ease following computations. In the sNMR context, the same process is applied to the original data space and performs very well since

the dimensions are often highly correlated. The obtained reduced data space has a dimensionality $k << l$ and is noted $\mathbf{d^f}(k \times 1)$. Note that in other geophysical problems, this PCA step may not be necessary if the dimensions of both the model and data space are low (among others: VES and TEM, see Chapter 2) or may also be necessary for the model along with the data spaces if their dimensions are too high (for example, if many thin correlated layers are used to describe the model). Noise analysis using Monte Carlo simulations (Hermans et al., 2016) provides co-variance between the different PCA components representing the data ($\mathbf{C^f}$). This translates in uncertainties on the PCA scores of the test data. The detailed procedure for noise propagation is presented in subsection 4.1.2.7.

4.1.2.3 Step 3: constitution of statistical relationships between the model parameters and the reduced data, using canonical correlation analysis

BEL1D relies on the constitution of statistics-based relationships between Earth models and simulated data. This step is the core of BEL1D. Several methods are suitable to derive such a relationship (see Chapter 3), but we choose canonical correlation analysis (CCA) for its simplicity. In essence, CCA transforms the (possibly PCA-reduced) model space ($\mathbf{m^f}$) and the PCA data space ($\mathbf{d^f}$), so that the canonical correlation (an approach formula of the correlation) between the resulting CCA spaces ($\mathbf{m^c}$ – Equation 4.2 – and $\mathbf{d^c}$ – Equation 4.3) is maximized (Krzanowski, 2000). The dimensions of the CCA spaces are linear combinations of the variables in their original space and are orthogonal to each other, meaning that there is no redundancy between the dimensions. We obtain:

$$\mathbf{d^c} = \mathbf{d^f A^T} \tag{4.2}$$

$$\mathbf{m^c} = \mathbf{m^f B^T} \tag{4.3}$$

where $\mathbf{A}$ is a matrix of dimensions $m \times k$ and $\mathbf{B}$ is a matrix of dimensions $m \times q$ (m is the minimum between k and q). Graphically, we can observe the statistical relationship between $\mathbf{d^c}$ and $\mathbf{m^c}$ as both spaces have the same dimensions $(1 \times m)$. As long as k is larger than q, it is possible to back-transform the models from the CCA model space to the original model space using the inverse of $\mathbf{B}$ (Hermans et al., 2016).

4.1.2.4 Step 4: generation of posterior distributions for the model parameters in CCA space using kernel density estimators

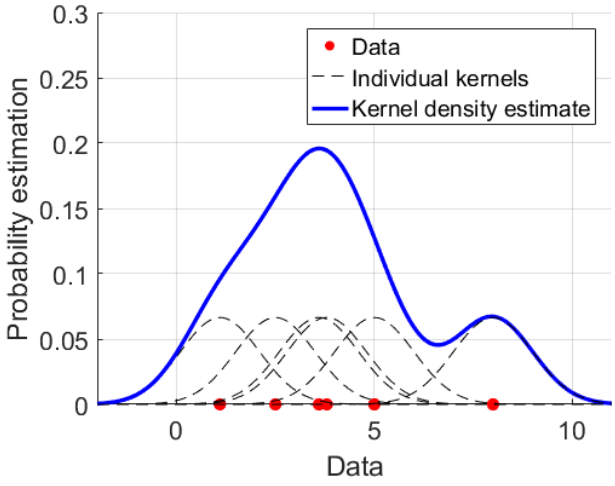

Figure 4.2: Illustration of the Kernel Density Estimation process

In canonically correlated space, the correlation between data and prediction variables within one dimension is maximized. Since the dimensions are orthogonal to each other, it is possible to produce a meaningful statistical description of each bi-variate distribution. This enables to determine a posterior distribution in the CCA model space constrained by the knowledge of field data. This posterior distribution represents the probability density function for the model parameters in

the CCA model space. To approximate the posterior distribution corresponding to the observed data $f_H(\mathbf{m^c}|\mathbf{d^c_{obs}})$, we use a kernel density estimator (KDE) with a Gaussian kernel (Wand & Jones, 1993) similar to Hermans et al. (2019) where they faced non-linearity issues when validating BEL for hydrogeology. Gaussian regression (Hermans et al., 2016; Satija & Caers, 2015) was tested; but in many cases, CCA did not yield perfectly linearly correlated relationships with Gaussian prior distributions, two necessary conditions to apply Gaussian regression. We thus implemented KDE for its robustness to estimate posterior distribution in a large variety of situations. KDE computes, for a given location in space, the sum of the contribution of each point from the CCA space as is illustrated in Figure 4.2. We use a multi-Gaussian kernel centred on the points with bandwidths chosen according to the point density (see subsection 4.1.5.2). The resulting distributions in the CCA model space are not constrained to any given distribution of known shape. The process is illustrated in Figure 4.3.

4.1.2.5 Step 5: sampling of the constituted distributions

The posterior distribution is not constrained to a known distribution. Therefore, sampling is done through the inverse transform sampling method (Devroy, 1986). This sampling procedure benefits from the properties of the cumulative distribution function that links a uniformly distributed variable to any distribution. The sampler thus transforms a value sampled in a uniform distribution to the corresponding value in the original distribution via the inverse of the cumulative distribution function. We can now easily generate a set of samples from the posterior distributions in the reduced space:

$$\mathbf{m^c_{post}} \in f_H(\mathbf{m^c}|\mathbf{d^c_{obs}}) \tag{4.4}$$

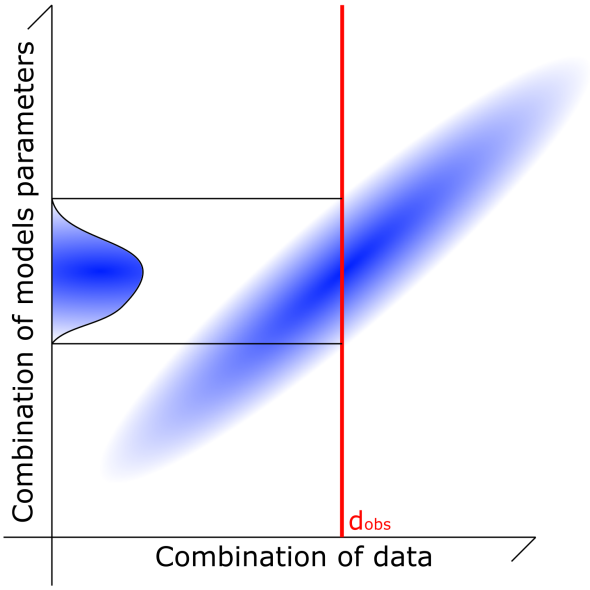

Figure 4.3: Illustration of the constitution of posterior distributions in CCA space. The cloud of points from the prior realizations and their associated data after KDE is represented by the blue ellipse, the red line represents the observed data in CCA data space and the blue distribution on the y-axis represents the distribution obtained through the kernel density estimator.

4.1.2.6 Step 6: back-transformation of the samples into the original space

The set of samples in CCA model space can be back-transformed into the original model space, using the inverse of the B matrix ($q \times q$). We thus have the posterior in the original physical space:

$$\mathbf{m_{post}} = \mathbf{m^f_{post}} = \mathbf{B^{-1}} \mathbf{m^c_{post}} \tag{4.5}$$

Here, in the sNMR context, since the model space was not reduced using principal component analysis, the models in PCA space are the same as the ones in original space. Because Gaussian regression or KDE with Gaussian kernel has no

limit on the value of sampled parameters, a few samples might occasionally be located outside the prior model space, in particular when the true model is at the extremity of the prior. Those models are removed from the sampled space as they do not correspond to the definition of the problem.

4.1.2.7 Noise propagation in BEL1D

Noise characterization is an essential step in any geophysical data acquisition. In order to investigate the effect of noise on BEL1D, one can analyse the variation of the PCA scores to the estimated field noise level (if the latter has been estimated). If the scores of the first k components are not varying significantly, then, one can assume that noise has no impact on the imaging process. On the other hand, if the noise impacts the scores significantly, BEL1D considers this impact by propagating uncertainty (represented by $\mathbf{C^f}$ in Equation 4.6) on the reduced observed data in the canonically correlated space, leading potentially to more complex computation of the posterior distributions (Hermans et al., 2016). This uncertainty is propagated in the CCA dimension according to linear error propagation from equation 4.2:

$$\mathbf{C^c} = \mathbf{A}\mathbf{C^f}\mathbf{A^T} \tag{4.6}$$

To estimate $\mathbf{C^f}$, we explore the PCA scores and their modifications when noise is present in the dataset. To do so, we take the data associated to each of the prior realizations as a basis and compute the PCA scores when one model is perturbed with a noise level similar to the data. The operation is repeated for a significant sub-sample (in the example, we used 50 randomly selected models) of the prior samples allowing to derive the PCA score covariance matrix $\mathbf{C^f}$.

Contrary to Gaussian regression where the addition of covariance to the observed reduced data is fully handled by matrix operations, and only slightly increases the

complexity of the formula for the sampling parameters (Hermans et al., 2016), the addition of uncertainty on the reduced observed data highly hinders performances of the kernel density estimation. Assigning uncertainty to the observed data (X in Figure 4.4) requires the computation of the kernel density estimation of the reduced posterior distributions for multiple sets of randomly sampled $\mathbf{d}^{\mathbf{c}}_{\mathbf{obs}} + \epsilon$, then processing to sampling and back transformation from those sets of distributions to constitute models of the posterior distributions.

This can be avoided by enlarging the bandwidth of the kernel density estimator (Bowman & Azzalini, 1997) in the X direction (in the CCA data space) according to $\mathbf{C}^{\mathbf{c}}$. This accounts for the uncertainty on the exact position of the data in the reduced space and therefore accounts for noise, without adding complexity to the computations, as is shown in Figure 4.4.

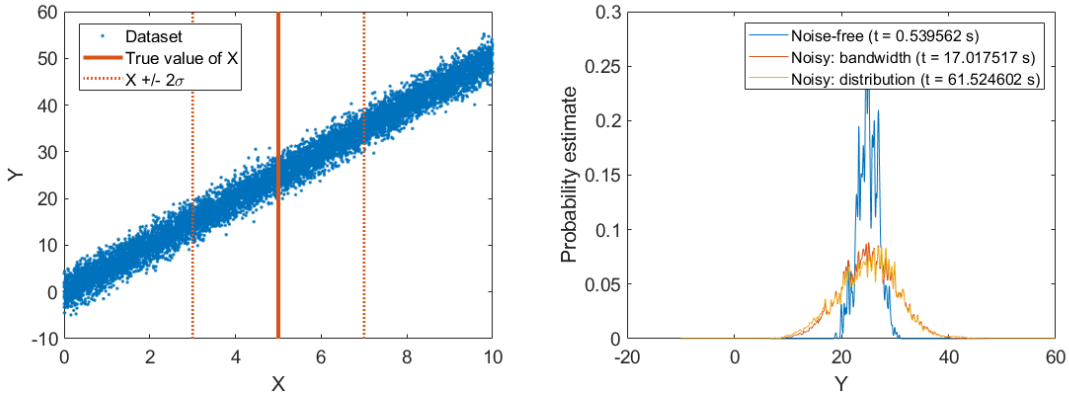

Figure 4.4: Example of the effect of noise on the computation of the kernel density estimator for the probability density functions. On the left, a random dataset that is highly correlated and the associated true value of X. On the right, the estimated distributions for the values of Y constrained to the known value of X and its associated error of 1. The computation time 't' is given for the different cases. The proposed *increased bandwidth* approach yields the same results as the exact sampling method.

4.1.3 Numerical example

In this section we will present results obtained by applying BEL1D to numerical examples of sNMR. The model that we will use corresponds to a classical context with two layers: a first unsaturated layer and a saturated half-space (Table 4.1). A field case is presented in section 4.1.4, and a non-geophysical related inverse problem is presented in Appendix A ("Testing BEL1D on a synthetic mind experiment: oscillations of a pendulum").

	Thickness (m)	Water content (%)	Relaxation time (ms)
Layer 1	5	5	100
Half-space	/	25	200

Table 4.1: Characteristics of the synthetic model

4.1.3.1 Computation of the model space and data space

The data are simulated for sNMR using the MRSMatlab toolbox (Müller-Petke et al., 2016). For the simulated experiment, the circular transmitter is the same as the receiver (the loops have a 50 m diameter with 2 turns). To simplify the computations, the Earth is set as resistive, thus neglecting the effect of electrical conductivity in signal attenuation and phase shift. Therefore, the signal is purely real. However, the process presented in this chapter is not limited to real-value data. sNMR is known to be impacted significantly by electromagnetic (EM) noise (e.g., Behroozmand et al., 2016). Different types of EM noise can affect the sNMR data. However, when most of these components have been filtered out, the sNMR signal will remain contaminated by a Gaussian noise distribution, whose magnitude cannot be reduced further. To represent this, we choose to add a Gaussian noise level to the dataset, corresponding to a mean signal-to-noise ratio of 10 (35 nV), which is a relatively high value in usual sNMR surveys.

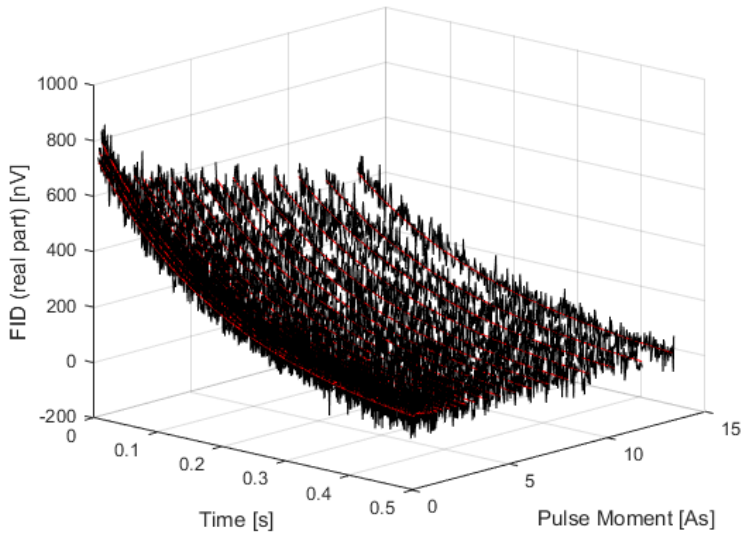

Figure 4.5: Simulated data for the sNMR experiment. The black curves represent the noisy dataset and the red curves the noise-free one.

As can be observed in Figure 4.5, the initial amplitude of the signal is around 500 nV, and a signal-to-noise ratio of 10 produces quite noisy signals, but the decaying behaviour remains observable.

The prior model space reflects the prior knowledge of the study site. In our case, we will assume that it is known that the subsurface is composed of one layer resting over an infinite half-space. Then, the rest of the properties will be defined as uniformly distributed variables varying in the intervals described in Table 4.2. The prior intervals are not necessarily centred on the actual value of the parameter. This is to better simulate real situations, where the user will only have rough information about the subsurface constitution. 5000 prior realizations are generated using a Latin-hypercube sampler (McKay et al., 1979). The forward model that was used to simulate the test data is reused here to generate synthetic data sets.

	Thickness [m]		Water content [%]		Relaxation time [ms]	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Layer 1	2.5	7.5	3.5	10	5	350
Half-space	/	/	10	30	5	350

Table 4.2: Prior model space for the sNMR experiment

4.1.3.2 Global sensitivity analysis and dimension reduction

We performed a distance-based global sensitivity analysis (J. Park et al., 2016) whose results are presented in Figure 4.6. The parameters corresponding to the first layer (W_1 and $T^*_{2,1}$) are not sensitive to the sNMR signal in this specific experimental configuration and should therefore lead to a poorly reduced uncertainty on those parameters.

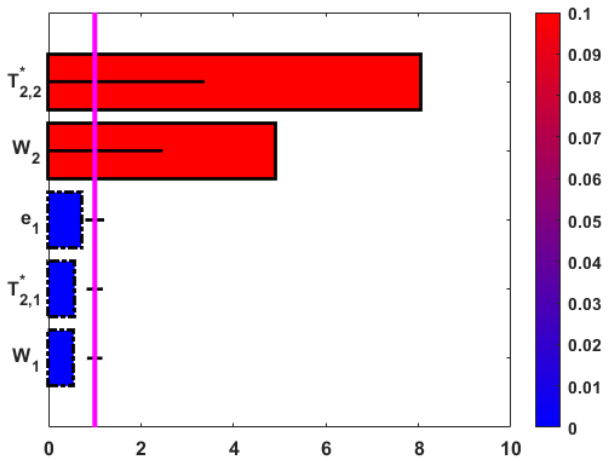

Figure 4.6: Multivariate sensitivity analysis performed on the simulated data. The red bars correspond to parameters that have an effect on the data and the blue ones to parameters that have a negligible effect.

From the initial 11904 dimensions in the dataspace, PCA enabled a reduction to only 5 dimensions, while managing to keep more than 90% of the variability. The canonical correlation analysis is shown in Figure 4.7. It is observed that (1) the correlation between the models and the data is larger for the first 3 CCA

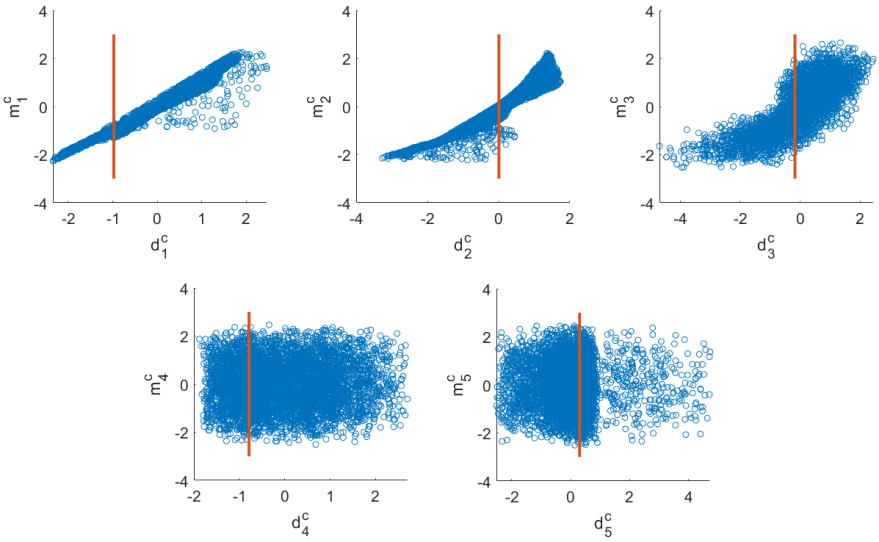

Figure 4.7: Data space versus model space in the canonically correlated space. The blue dots represent the prior models and data spaces and the orange line represents the exact value of the test data in the reduced space.

dimensions, (2) the correlation is not linear from the 2$^{\text{nd}}$ dimension and (3) from the 4$^{\text{th}}$ dimension, the points are scattered and not correlated. The fact that the last dimensions are scattered (note that each dimension is a linear combination of the model parameters) informs us on the inability of the data to efficiently predict all the parameters simultaneously. This is confirmed by the linear combinations that constitute the CCA model space dimensions (Figure 4.8): we observe that the three last reduced dimensions are mainly linear combinations of the insensitive values (Figure 4.6). Note that non-linear statistical techniques (e.g., Gorban et al., 2008; Lawrence, 2012) could potentially lead to better characterization of the data-model relationship.

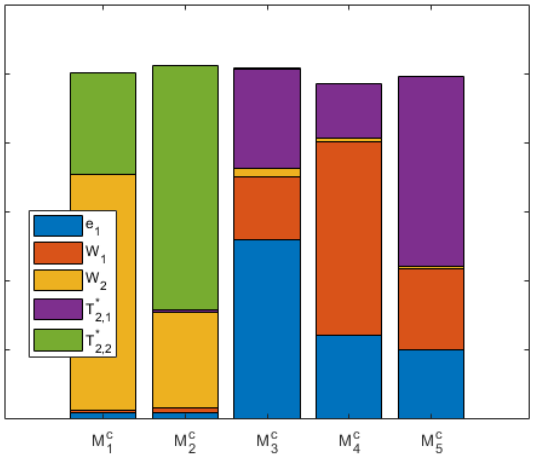

Figure 4.8: Combinations of parameters representing the canonically correlated model space dimensions. The values are scaled for visual purposes.

4.1.3.3 Results of BEL1D on the numerical example

BEL1D is applied to the simulated noisy dataset. Figure 4.9 shows the prior and posterior distributions of parameters obtained through BEL1D. We observe that the uncertainty on the half-space parameters is dramatically reduced through BEL1D. On the other hand, the insensitive parameters $T^*_{2,1}$ and W_1 are poorly reduced. This insensitivity to some parameters is also clearly identified in the pendulum example (Appendix A). As can be observed, the test value of the parameter is always contained in the posterior distribution. The multivariate analysis of the parameters shows that most of the parameters are independent. However, some correlations are observed for parameters that are related to the same layer ($T^*_{2,2}$ and W_2 for example). Figure 4.10 (Left) presents another representation of the posterior results, where the data RMS of posterior samples are colour-coded, and where each model is a set of linked parameters. The presented results show that a large part of the posterior models has a very low RMS error. The trend in the water content set through the prior model space is respected. In contrast, the trend for the relaxation

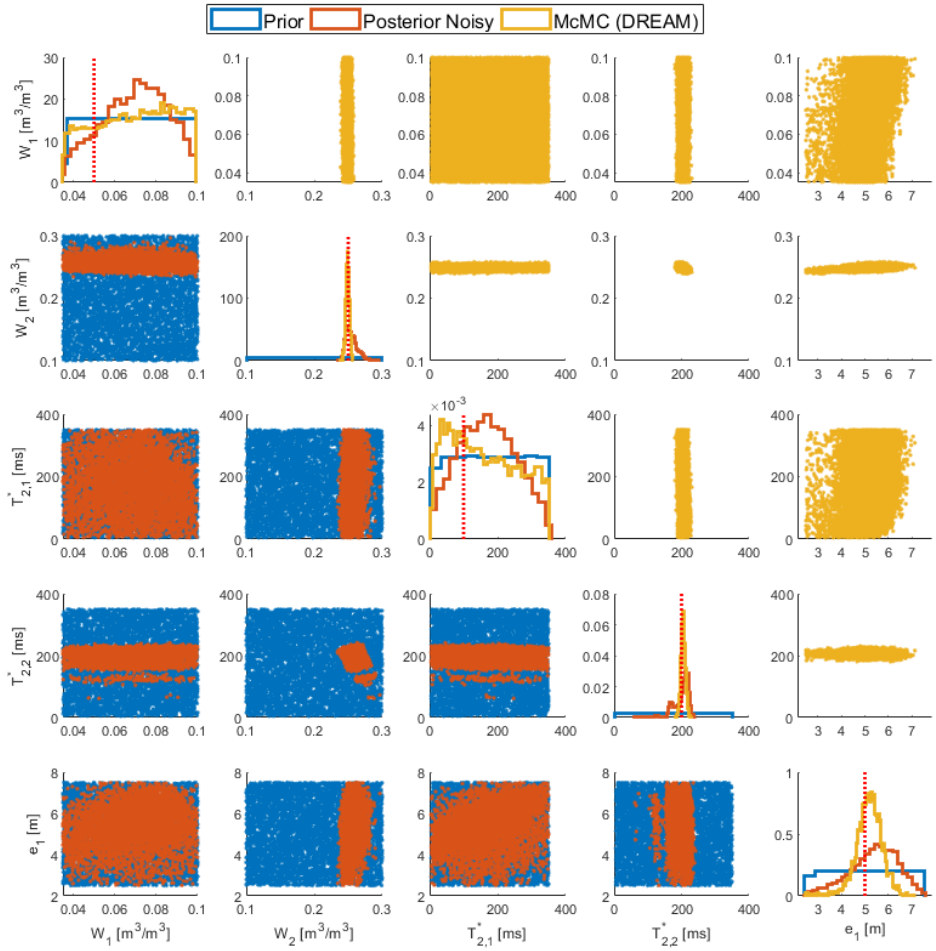

Figure 4.9: Results of the BEL1D process on the synthetic noise-free data from the 2 layers model. The obtained distributions are the posterior distributions estimated from the noisy dataset in red. The sampled prior model space is represented by the blue distributions. The test values are represented with dashed red lines in the histogram plots. Results of the McMC algorithm DREAM are presented in yellow.

time is not resolved by BEL1D. The analysis of the RMS and the comparison with DREAM (Figure 4.10, Right) show that this is related to the insensitivity of the data and not to the method itself.

Finally, we applied the DREAM McMC algorithm (Vrugt, 2016) to the same dataset. We applied a Gaussian likelihood with 15 generations and 10000 chains and a burn-in period of 5000 chains. However, to achieve convergence towards reasonable

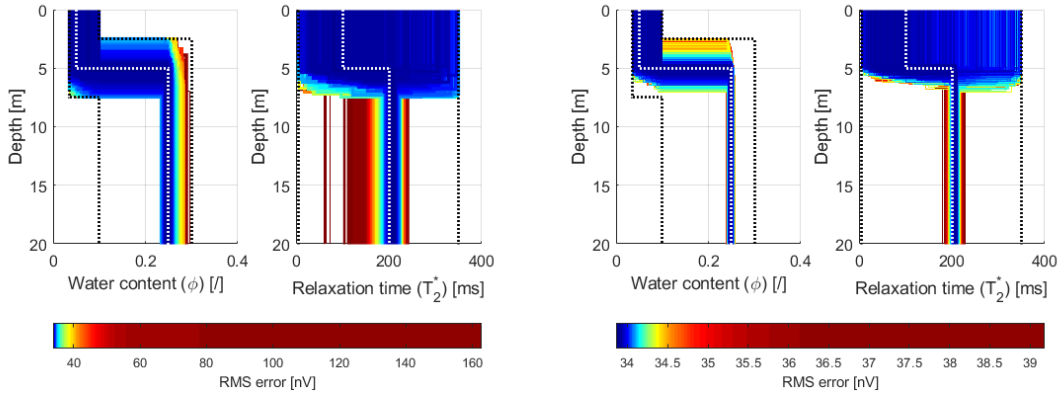

Figure 4.10: Left: results of the BEL1D process on the synthetic noisy data from the 2 layers model. Each displayed model is presented with its associated RMS error value. In order to enhance visualization, each colour is associated to the same number of models. The dashed white lines represent the test values and the dashed black lines represent the extent of the prior model space. The similar graph is also presented for DREAM results (Right)

uncertainties, changes to the definition of the dataset are required. We had to use gate integration, similarly to what is applied in MRSMatlab (Müller-Petke et al., 2016) in order to attribute higher loads to the first part of the decay curves, thus lowering the impact of noise on the estimation. For the presented results, we used 5 gates for each pulse moment. However, we must note that the results of DREAM are highly sensitive to the number of gates: with 1 or 2 gates, nearly no reduction of uncertainty is observed, whereas with a too large number of gates (to the limit with no gate integration), false posterior distributions are observed (the test values are outside the predicted posterior). The results are presented alongside the BEL1D results in Figure 4.9. They show that BEL1D tends to have a slightly lower performance than DREAM. This is related to the approximation of the problem in a lower dimension space where the actual data misfit is never computed. Applying a cut-off on the RMS values observed in BEL1D would result in the same distribution (Figure 4.10). In the next chapter (Chapter 5), we propose a solution to this issue by applying an iterative scheme with rejection at the final

iteration.

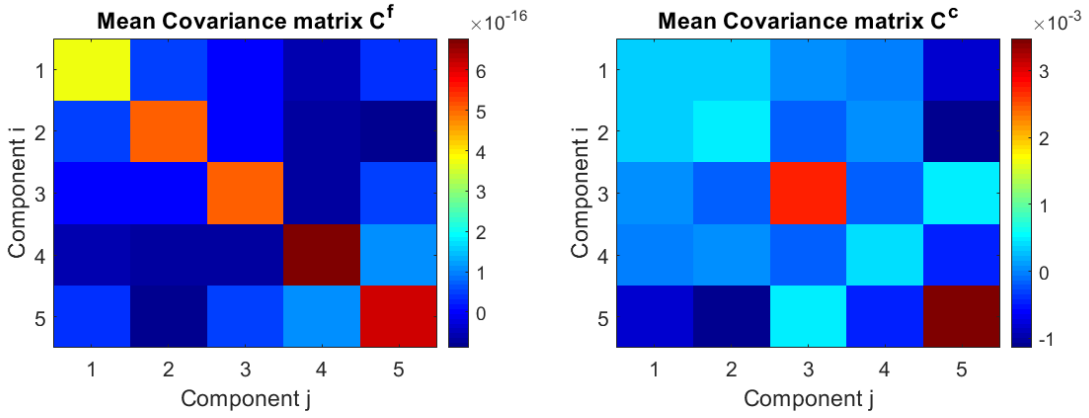

Figure 4.11: Impact of noise on the PCA scores (the first 5, kept after dimensionality reduction), covariance matrix for the PCA scores impacted by noise (left) and the corresponding covariance in CCA space (right). This first matrix is estimated through the mean of the covariance matrices observed between the noise-free scores and the noisy ones.

To estimate the posterior distributions, we needed to account for the impact of noise in the process. The matrix $\mathbf{C^f}$ presented in Figure 4.11 shows very low values for the covariance. However, once transformed into CCA space ($\mathbf{C^c}$, Equation 4.6), we observe that the effect of noise, even if limited, is still observable. From a default bandwidth of 0.01 in the X direction for the kernel density estimation, we computed bandwidths ranging from 0.017 to 0.059.

Those increased bandwidths explain the presence of some outliers in the data-space (Figure 4.10, Left). When observing the CCA space (Figure 4.7), it is noticeable that the second dimension (mainly representing the decay time of the second layer) shows models that are outside the main distribution but close to the position of the reduced test data. When using the bandwidth that takes into account the impact of noise, those models are included into the posterior. Nonetheless, those models correspond to large RMS error and are easily identified (Figure 4.10).

An example with more than two layers is presented in Section 4.2, where we

apply BEL1D to a new experimental configuration on a four-layer synthetic model.

4.1.4 Field case study: Mont Rigi (Belgium)

The Mont Rigi is located in the Belgian Fagnes region, in the Eastern part of the country. This site presents an ideal case study for sNMR as it is remote and far from any electromagnetic noise sources, due to its natural reserve classification.

Geologically, the site is characterized by a metric peat layer on top of a Cambrian bedrock (La Venne formation) known as an aquiclude with a very low water content (Gilson et al., 2017). In contrast, peat is known to present very high water contents with observed total porosities around 90% (Wastiaux, 2008). According to previous GPR exploration (Wastiaux & Schumacker, 2003), the peat layer at the experimental site should have a thickness between 2.5 and 4.5 metres. The sNMR response should therefore allow to easily distinguish the two layers with a properly designed experiment. The designed on-site experiment consisted of a classical coincident loop transmitter/receiver couple with a diameter of 20 metres.

The raw data have been preprocessed to lower the impact of noise with despiking, harmonic modelling, reference noise cancellation and despiking again (Müller-Petke et al., 2016). The signals have been further truncated to 0.2 seconds to decrease the impact of noise, mostly present at large timeframes. The resulting noise amplitude is quite low for sNMR (18 nV). Then, the signals have been inverted using the QT Inversion (Müller-Petke & Yaramanci, 2010), to constitute a benchmark to compare to the results of BEL1D. We used a smooth-mono model description (smooth models with a single relaxation time for a given depth) and a regularization parameter of 6000 for both water content and relaxation time using the L-curve criteria (Aster et al., 2013). Then, BEL1D was applied with the uniformly distributed prior model space described in Table 4.3. We sampled 5000 prior realizations and produced

5000 models for the posterior.

	Thickness [m]		Water content [%]		Relaxation time [ms]	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Layer 1	0	7.5	30	80	0	200
Half-space	/	/	0	15	100	400

Table 4.3: Prior model space for the Mont Rigi sNMR experiment

The CCA space analysis (Figure 4.12) shows that we could expect a significant reduction of uncertainty for some parameters, as the first three dimensions show high correlations. The first dimension, dominated by the water content of the half-space shows an especially narrow relationship between the reduced data and model space. On the other hand, the last dimension, mostly represented by the water content of the first layer, shows no specific correlation. We should therefore expect a much better reduction for the parameters of the half-space than for the first layer.

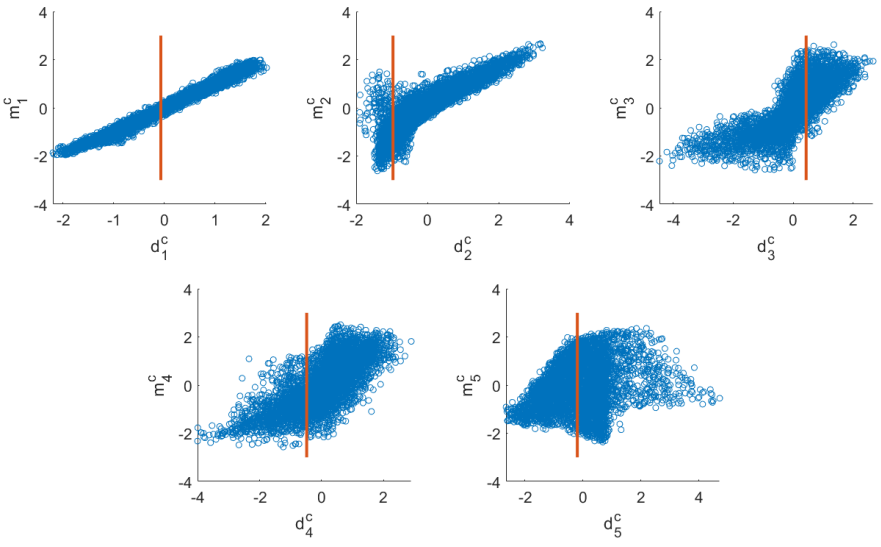

Figure 4.12: Mont Rigi – Data versus model space in the CCA space. The blue dots represent each prior realization and the orange line the position of the field data.

The noise level (after denoising) impact on BEL1D was assessed and resulted in significant changes in bandwidths for the kernel density estimation, between 0.04 and 0.12 instead of the default 0.01 value.

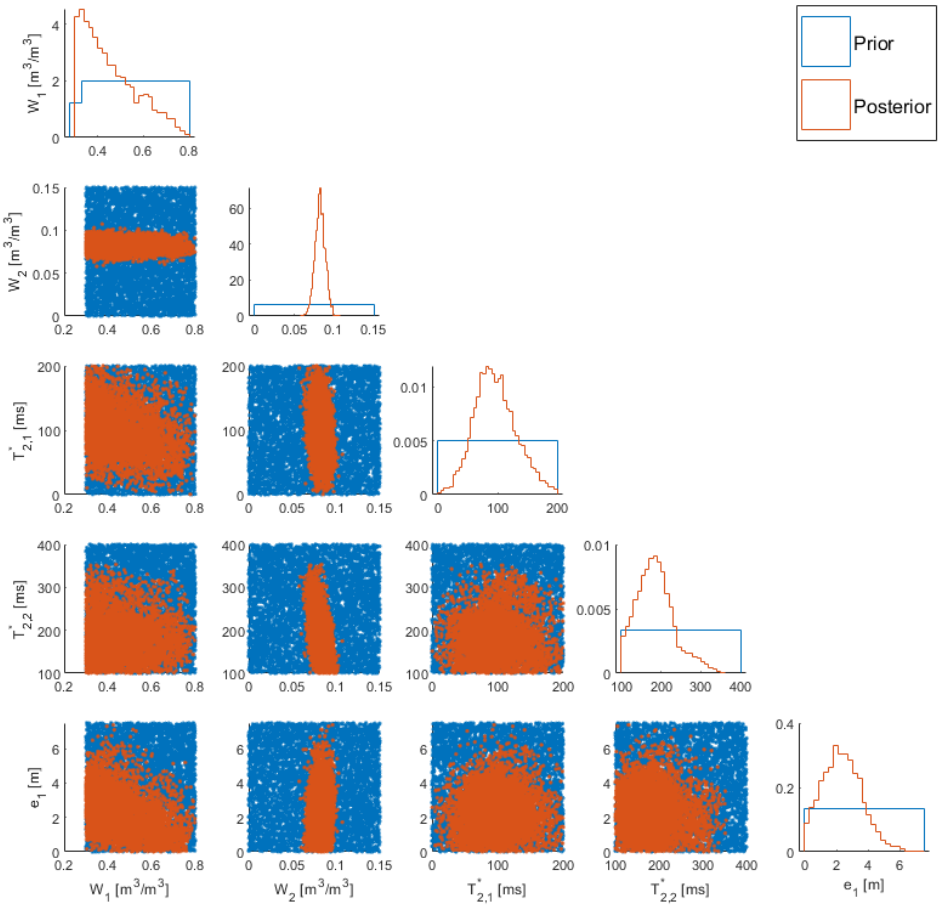

Figure 4.13: sNMR at Mont Rigi - Prior (blue) and posterior (orange) distributions of the parameters.

The analysis of the posterior parameters model space (Figure 4.13) shows that the reduction of uncertainty from the prior model space to the posterior is significant and produces consistent results, when compared to the geological context and the QT inversion solution (dashed grey lines in Figure 4.14). The thickness of the peat layer tends to be smaller than the maximum estimate (4.5 m) from the

GPR interpretation (Wastiaux & Schumacker, 2003). The use of a narrower prior taking into account the GPR information would prevent such a behaviour. This is illustrated by the decrease of the maximum value of the thickness for the first layer when the water content of the aforementioned layer increases. This behaviour arises from the non-uniqueness of the sNMR response stated above.

In the posterior models produced by BEL1D (Figure 4.14), a large majority shows a root-mean-square error lower than the noise level in the dataset, implying that the set of probable models provides an efficient estimation of the site parameters. The RMS errors of the posterior samples show that they all explain the data to a similar level; the uncertainty is thus intrinsic to the data set and the non-unicity of the solution. The non-uniqueness of the solution is easily observable, with two distinct behaviours: one with a median water content and a thick first layer and another with a very high water content but a thinner first layer. This confirms the realistic very high water content in the peat layer. Those two cases show similar RMS errors, hence are explaining the data to the same level. The probability of each posterior model is shown in the RMS error colour bar for which each colour represents the same number of models. Therefore, it is observed that the low RMS error models also correspond with the most probable ones.

4.1.5 Discussion

4.1.5.1 Impact of the number of sampled models in the prior model space

The optimum number of models in the prior model space sample is a trade-off between precision of the posterior models distributions and ease of computation: the larger the number of samples, the larger the memory requirement and the harder the computations. This parameter was not investigated in early applications

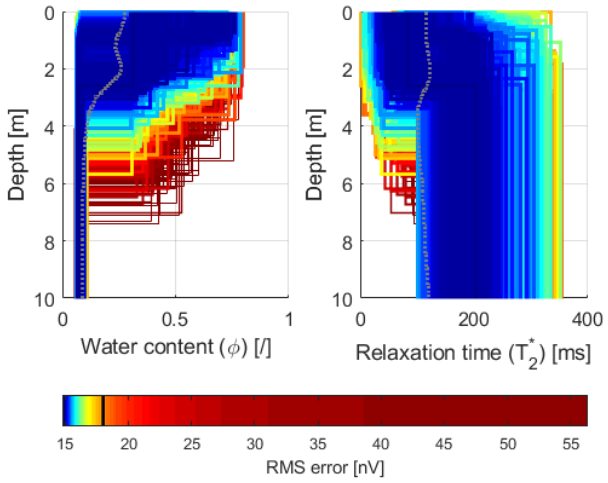

Figure 4.14: sNMR at Mont Rigi - Posterior models distributions with their associated misfit and results of the QT inversion (dashed grey line). The black line in the RMS error graph represents the level of noise in the original dataset.

of BEL, focussing on simpler predictions. However, when dealing with complex posterior predictions, the number of parameter begins to matter since more models are required to achieve convergence.

We ran BEL1D on the previous synthetic sNMR dataset with 10, 50, 100, 500, 1000, 5000, 10 000, 50 000 and 100 000 models in the prior samples. For each case, we ran BEL1D 10 times with different random samples and aggregated the results. It is observed that the RMS error distribution tends to stabilize above 1000 models, suggesting that this value is optimal in our case (Figure 4.15), especially for the minimum and 5^{th} percentile values. This observation is confirmed by the evolution of the parameter distribution characteristics (Figure 4.15). In those graphs, the mean values of the parameters are normalized with the test values, and the standard deviations are normalized with the standard deviation in the prior model space. We see that the parameter distributions are stabilized above 1000 models, even if the values are not minimal. This value is still relatively low, even though latter studies have found that fewer models were required for other applications (Thibaut et al.,

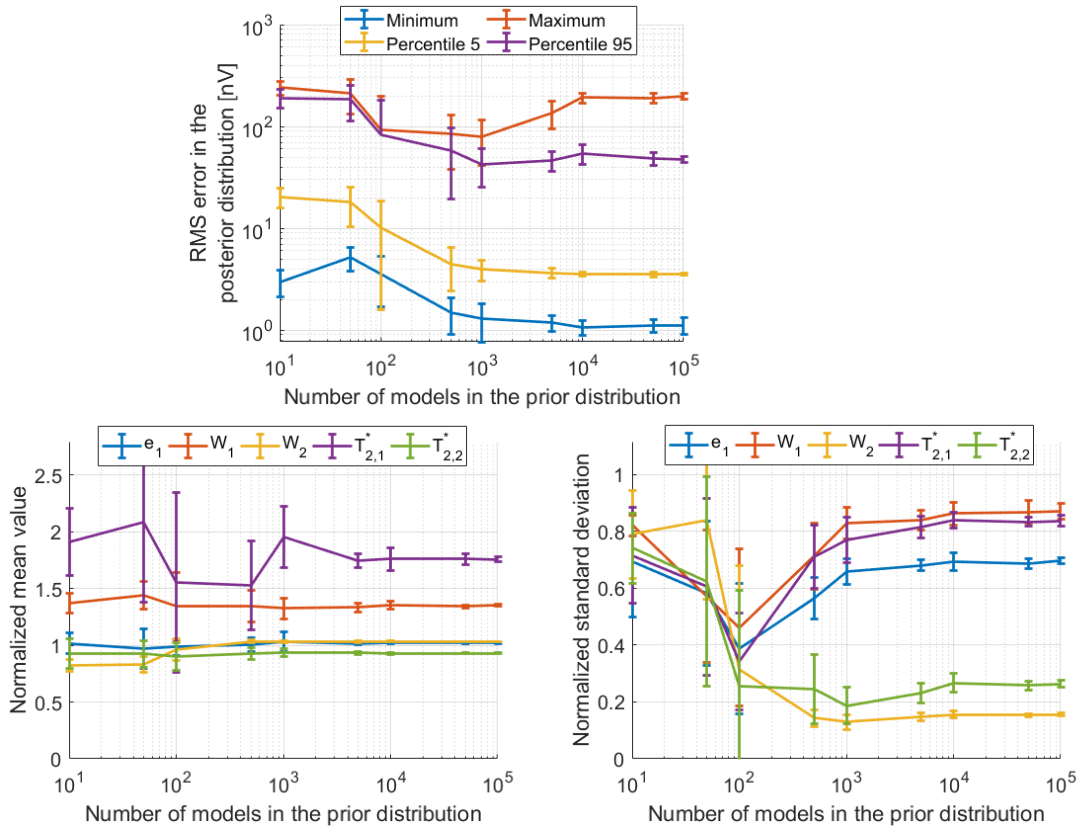

Figure 4.15: Impact of the number of prior samples on the posterior distribution. The top graph presents RMS evolution, while the two bottom graphs show (from left to right) the normalized mean value and the normalized standard deviation.

2021). Those results indicate that using a large number of models in the training set guarantees convergence, but that a fine-tuning of this parameter can lead to computational gains.

4.1.5.2 Impact of the chosen kernel density bandwidth on the posterior distributions

To assess the impact of the kernel density bandwidth on the posterior distribution, we ran 10 times 10 different values ranging from 10^{-5} to 1. In each case, 5000 models were sampled in the prior (Latin-hypercube sampler) and 5000 models were predicted, to ensure that no side effect is linked to other parameters. A similar analysis

to the one performed in the previous subsection is then performed (Figure 4.16). We observe that the bandwidth has a strong impact on the resulting distributions. Too low values result in erratic behaviours of the posterior distributions. On the other hand, too large bandwidths result in the smoothing of the posterior distributions, as is observed in the evolution of the standard deviation where this value increases sharply for the sensitive parameters. Therefore, the optimal bandwidth is a value located in-between the two extremes, where the reduction of uncertainty is significant for the sensitive parameters and the other parameters are stable. This occurs around 10^{-2} in our case. In our implementation of the kernel density estimator, the effect of a too small bandwidth is avoided by verifying that enough points (at least 1% of the sampled prior) are present in the direct surrounding of the reduced data ($d^c_{obs} \pm 3 \times bandwidth$). To perform the test presented in this bandwidth analysis, this safeguarding measure was disabled. In the case the default bandwidth is too small, then, the bandwidth is multiplied by 2 and the verification is made again until the criterion is satisfied and only then, the kernel density estimation is performed. We ran BEL1D with a bandwidth set to 10^{-5} and the safeguard algorithm enabled. The obtained result is here similar to the one obtained with a suitable bandwidth as is shown by the diamonds in the graph presenting the evolution of the normalized standard deviation in Figure 4.16.

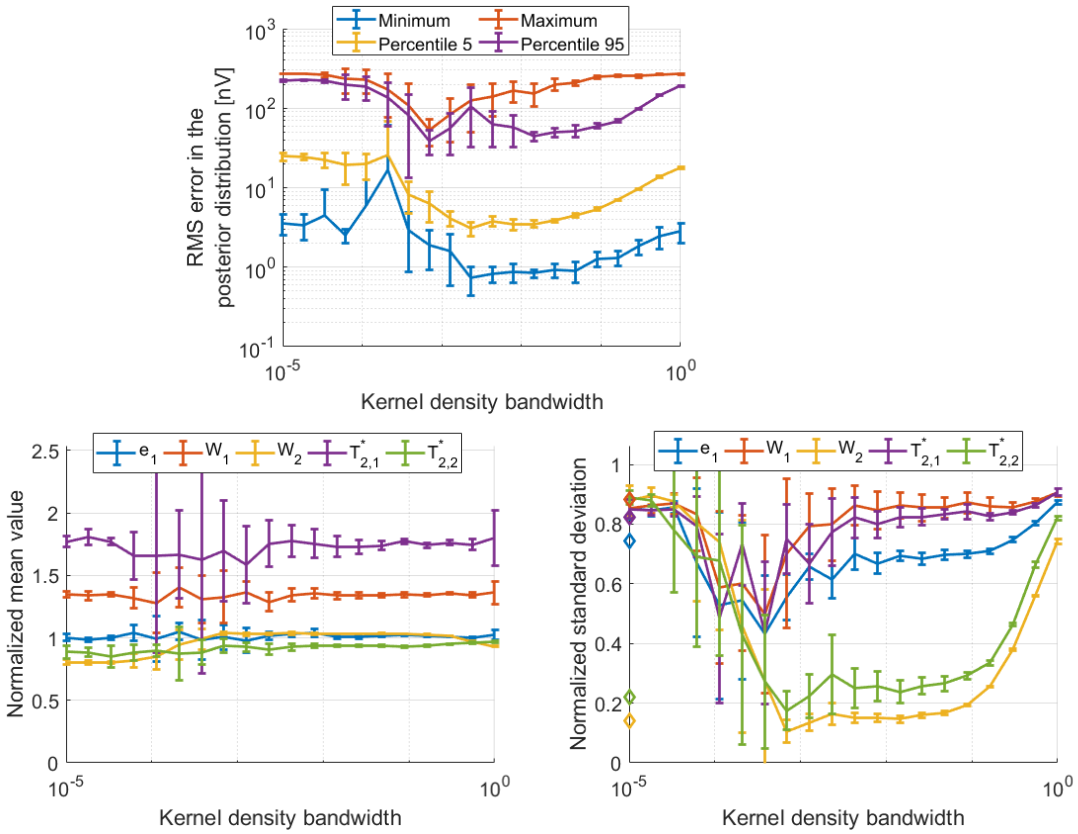

Figure 4.16: Impact of the bandwidth in the kernel density estimation on the posterior distribution. The top graph presents RMS evolution, while the two bottom graphs show (from left to right) the normalized mean value and the normalized standard deviation. This latter also presents results obtained with the safeguarding measure enabled (diamonds) for the most extreme case: a bandwidth of 1E-5.

4.2 Application: sNMR multi-central configuration

In this section, we will use the newly built algorithm to demonstrate the improved accuracy of 1D sNMR with the multi-central configuration.[2] We will use BEL1D on a synthetic dataset with both the classical configuration and the multi-central configuration with the idea to demonstrate the advantage in sensitivity gained from the multi-central configuration. This example also demonstrates the potential of BEL for experimental design purposes (Thibaut et al., 2021), although the experimental design framework would require to repeat the experiment for several true models (Thibaut et al., 2021).

4.2.1 The multi-central configuration

Classically, sNMR surveys are using coincident loops. However, even though those are very efficient to gather information at depth, those lack in sensitivity to the shallow subsurface. To circumvent this, Behroozmand et al. (2016) proposed the central-loop configuration: injecting on the outer loop and measuring the resulting signal on a smaller, central, antenna. This resulted in a higher sensitivity to shallow

[2]The results presented in this section have been partly published in Journal of Applied Geophysics (Kremer et al., 2020). The part presented here is mainly related to (and extends) my contribution to the paper. The different authors contributed to:

- Data acquisition: Thomas Kremer, Mike Mueller-Petke, Raphael Dlugosch and Hadrien Michel.
- Data processing: Thomas Kremer - advices: Trevor Irons, Mike Mueller-Petke and Raphael Dlugosch.
- Resolution analysis and inversion work: Thomas Kremer, Mike Mueller-Petke, Raphael Dlugosch and Hadrien Michel - advices: Frédéric Nguyen and Thomas Hermans
- Discussion: all contributing co-authors
- Writing: Thomas Kremer
- Proof-reading: all contributing co-authors

Coincident configuration Multi-central configuration

Figure 4.17: Comparison of the classical coincident configuration and the multi-central configuration in sNMR

layers. The multi-central sNMR configuration (Kremer et al., 2020) is an iteration over the central-loop (Figure 4.17). The idea is to benefit from both the coincident loop and the central-loop at the same time. This enables to gain sensitivity to the shallow subsurface, due to the presence of the central-loop. However, we do not lose sensitivity at greater depths as we still use the coincident loop. This also maximizes the data that is collected on the field during a single injection.

Figure 4.18 shows an example of the three sensitivity distributions associated with a multi-central loop configuration where the transmitter/receiver (Tx/Rx) loop is 50 m in diameter (1 turn), and the 2 smaller receiver loops are respectively 25 m in diameter (2 turns) and 10 m (4 turns) in diameter. The first graph corresponds to the classical 1D kernel shape observed for classical coincident loop configurations, whereas the others show the quite specific shapes of the central-loop configuration, where high sensitivity values are present close to the surface, also shown by Behroozmand et al. (2016).

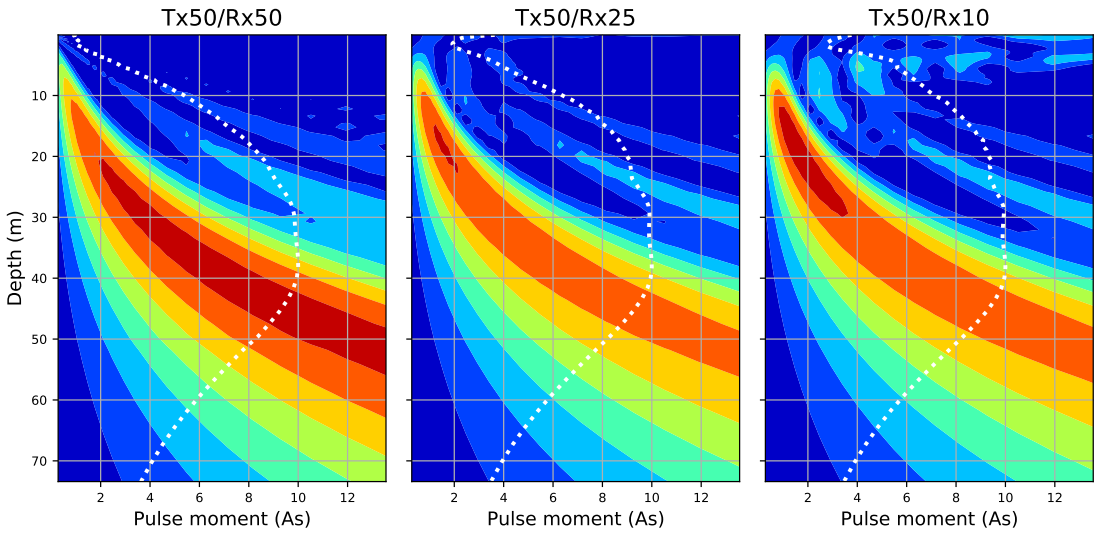

Figure 4.18: Sensitivity kernels values associated with the multi-central-loop configuration. All kernels have been computed with MRSMatlab, assuming a resistive Earth. The scale is normalized for every kernel. The total of the kernel for each depth is presented in dashed white lines.

4.2.2 Application of BEL1D

We applied the BEL1D method to investigate the interest of the multi-central-loop on a synthetic example. The model will represent a case typical of peatlands: a first layer (the peat, 2.5 meters) with a very high water content (25%), lying above a thick aquitard (1 to 3% water content), interrupted by an aquifer (10% water content) present at a depth between 50 and 70 meters. We use the multi-central configuration presented before. A 5 nV Gaussian noise was added across the board. The datasets were simulated with MRSMatlab, under the assumption of a resistive Earth.

In order to apply BEL1D, we must first define the extent of the prior model space. We will assume known that the model is composed of three layers, and use uniform distributions for the remaining parameters as described in Table 4.4. We sampled 10000 prior realizations and simulated their associated datasets according

to the sensitivity kernels, in order to replicate the "true" experiment. Then, we produced the posterior distributions of models by constraining the result of CCA to the knowledge of the benchmark dataset, enabling the production of models from the posterior model space constrained to the knowledge of the data.

	Thickness [m]			Water content [%]			Relaxation time [ms]		
	Min.	Bench.	Max.	Min.	Bench.	Max.	Min.	Bench.	Max.
Layer 1	0	5	10	10	25	40	0	250	400
Layer 2	20	45	60	0	3	10	0	70	400
Half-space	/	/	/	0	1	10	0	50	400

Table 4.4: Description of the synthetic model used for the sensitivity demonstration along with the prior model

The results of those computations are presented in Figure 4.19. There, we can clearly see that the resulting uncertainty is lower for the multi-central configuration, than it is for the classical coincident loops. This is explained by the higher sensitivity towards shallow layers, containing most of the water in this example.

The improvements in the uncertainty reductions are more pronounced for the water content and thickness of the first layer. This layer has a relatively low sensitivity overall using the coincident configuration (see Figure 4.18). Here, the central-loop is crucial for the added information about the peat layer. We can also note that the estimate is, at worst, similar for all the parameters. This indicates that the experiment does not reduce the sensitivity to some parameters.

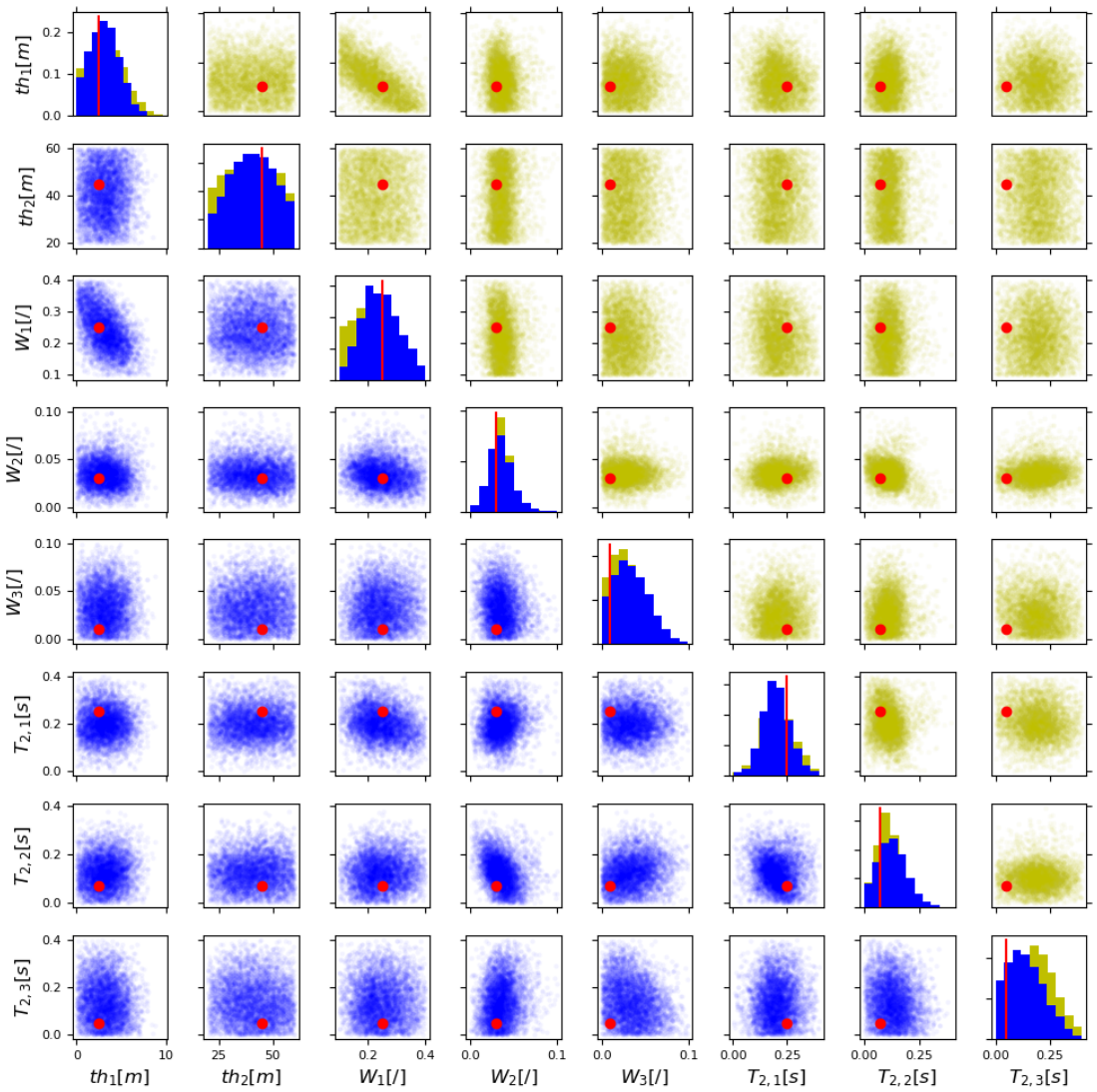

Figure 4.19: Posterior model space on the peatland synthetic case. The posterior obtained with the coincident configuration is in yellow and the multi-central configuration in blue.

4.3 Conclusion

We propose an adaptation of the Bayesian Evidential Learning framework as a new tool for the unidimensional interpretation of geophysical data called BEL1D. This approach offers an alternative to stochastic inversions, as it requires fewer and simpler computations, thus a lowered CPU cost. This approach is also presented as an alternative to deterministic inversions as its behaviour is superior both in terms of ease of implementation and completeness of the results (handling properly non-uniqueness, quantifying the uncertainty, etc.). We therefore developed this method theoretically, as a new general algorithm.

The framework does not require the use of regularization parameters or any other form of bias, but rather requires the proper definition of the prior model space. Contrary to other Bayesian methods, the use of a large prior model space (hence, introducing less bias in the posterior information) does not involve heavier computations to converge, as the algorithm is always building in one step the posterior model space that corresponds to the observed data (see, 4.1.2.4). This method can therefore be considered unbiased if the prior model space is sufficiently large. BEL1D is thus a computationally efficient alternative to stochastic inversions as it only requires n forward models corresponding to the prior samples. Moreover, in BEL1D noise is considered via an embedded theoretically based method, contrary to classically used high regularization parameters that lack a theoretical basis to be properly assessed and are therefore difficult to estimate.

When compared to McMC methods (such as DREAM), one of the key advantages of BEL1D is the traceability of all operations. When McMC classically applies random changes to the models and select them randomly, BEL1D constitutes a correlation between model parameters and the data. This means that one could seek

for the origin of all the posterior models in the prior model space. This enables, for example, to efficiently sample multi-modal distribution or to understand the origin of odd models in the posterior distribution (see the numerical example), impossible to track back in McMC algorithms. The other key advantage is the simpler tuning to convergence of BEL1D compared to DREAM (McMC). In DREAM, it was required to change the dataspace to enable convergence towards reasonable (or even correct) posterior distributions. In BEL1D, as long as the prior is correctly assumed, the only parameter that the user needs to take care of is the number of models to sample in this space. As shown in the discussion, this parameter can be selected by observing the stabilization of the posterior distributions, and can therefore be solved by sequentially increasing the number of sampled models. In DREAM, the application without changes in the dataset representation leads to either nearly no reduction of uncertainty (using the RMS as a proxy of the probability) or to a reduction towards level of uncertainty close to deterministic values but without the true model inside the posterior (using all data points as proxy of the probability). Therefore, applying this method to an unknown dataset might lead to erroneous interpretations, whereas BEL1D, even if producing slightly larger uncertainties, should always predict consistent posterior spaces.

We demonstrated that BEL1D is efficient for the interpretation of sNMR data. We presented results for noisy datasets originating from a numerical model and field. In both cases, we validate the results from BEL1D, either against the test values from the model and McMC methods - DREAM (Vrugt, 2016) - for the numerical example or other geophysical experiments and classical state-of-the-art inversion techniques for the field case. On the other hand, the careful analysis of the RMS error of the proposed models indicates that the range of uncertainty delivered by BEL1D is intrinsic to the non-unicity of the solution for geophysical

inverse problems: many models explain the data to the same level.

We also presented an application of BEL1D to the new sNMR multi-central configuration (Kremer et al., 2020). In this example, we show how this new configuration is superior to the classical design (coincident loops). This example also shows the ability of BEL1D to deal with more complex models (4 layers with 2 parameters each) and datasets. It also indicates that BEL (and BEL1D) is well suited for experimental design. Indeed, by comparing the results of BEL1D from multiple experiments and multiple probable models from the prior, it is possible to diagnose the best design for the objective at hand, similarly to what the work of Thibaut et al. (2021) explored.

In terms of performance, the computation cost of BEL1D is directly proportional to the efficiency of the forward model and the number of samples in the prior model space. It must be noted that the prior samples being independent, the simulation of synthetic data can be fully parallelized. The performance of BEL1D can thus be easily estimated from the available computing facilities. The cost is therefore much smaller than with other stochastic methods requiring ten to hundreds thousands of runs to converge towards a solution. Moreover, computation time could be dramatically reduced by pre-field forward modelling operations for large geophysical campaigns where the prior would be similar. Prior realizations and the associated datasets would then be reused (PCA and CCA operations also), leaving only the kernel density estimation, sampling and back-transformation operations to be applied after field acquisition, leading to extremely rapid imaging. This is an extreme advantage compared to other stochastic methods that rely on the data misfit (such as McMC) to sample the posterior.

Those applications, along with the pendulum case presented in Appendix A, demonstrate the versatility of BEL1D and its applicability to multiple cases with

very few adaptations. Provided that a forward model already exists, the adaptation of the codes to any other 1D geophysical method would be straightforward. For example, the forward operator from HV-Inv (García-Jerez et al., 2016) could be used for the interpretation of H/V data. System calls may also help to further extend the applicability of BEL1D. This may be the case for the use of all the 1D forward operators implemented in pyGIMLi (Rücker et al., 2017) - vertical electrical sounding, frequency- and time-domain electromagnetic and magnetotelluric.

In its present version, BEL1D is dedicated to the 1D inversion of geophysical data in layered Earth model where the contrast between layers is supposed to be sharp. In the future, we plan to extend the method to smoothly varying systems accounting for numerous thin, correlated layers.

Chapter 5

Improving the posterior estimate through Iterative Prior Resampling

Chapter 4 showed that BEL was able to identify correctly the posterior. However, we found that BEL1D tends to overestimate the uncertainty in some cases. In this chapter[1], we will present a way to improve the precision and accuracy of BEL1D through iterations on the prior (Iterative Prior Resampling - IPR).

[1]The results presented in this chapter are published in the paper entitled: Iterative Prior Resampling and rejection sampling to improve 1D geophysical imaging based on Bayesian Evidential Learning (BEL1D) (Michel et al., 2022). The contributions of the different authors are:

- Code development: Hadrien Michel - advices: Thomas Hermans and Frédéric Nguyen
- Bayesian inversion: Hadrien Michel - advices: Frédéric Nguyen and Thomas Hermans
- Parametric analysis: Hadrien Michel
- Writing: Hadrien Michel
- Proof-reading: all contributing co-authors

The results are also built upon results published in an extended abstract to SEG2020 entitled: Improving BEL1D accuracy for geophysical imaging of the subsurface (Michel, Hermans, Kremer & Nguyen, 2020). The contributions of the different authors are:

- Code development: Hadrien Michel - advices: Thomas Hermans and Frédéric Nguyen
- Bayesian inversion: Hadrien Michel - advices: Frédéric Nguyen, Thomas Hermans and Thomas Kremer
- Writing: Hadrien Michel
- Proof-reading: all contributing co-authors

Those improvements are required with large priors, and with poor correlations between the models and the data in this initial prior. Indeed, a large prior can mix models that have different data aspects and thus lead to a poor correlation between models and data. We will illustrate the results using seismic surface waves, that are usually prone to low data-model parameters correlation due to the presence of insensitive parameters in the prior (Xia et al., 1999). However, the code developed is not limited to this application, and a sNMR example will also be presented.

5.1 Introduction

In this chapter, we are using Bayesian Evidential Learning (BEL) (Scheidt et al., 2018) to solve the inverse problem of retrieving 1D shear waves velocity profiles from dispersion curves efficiently. This approach is at the crossing between Machine Learning and Bayesian inversion since it requires a learning phase, allowing to approximate the Bayesian problem in a reduced dimension space by developing a statistically significant relationship between data and predictions. This linearized statistical relationship between data and predictions is based on a canonical correlation analysis in reduced dimensions. It has seen several applications in geosciences from oil field production prediction (Satija et al., 2017) to hydrogeological prediction of wellhead protection areas (Thibaut et al., 2021) and hydrogeophysical modelling of heat transport (Hermans et al., 2018; Hermans et al., 2019). BEL was originally designed as a prediction-focused approach applied to time dependant processes, circumventing the need to estimate the posterior distribution of model parameters by inversion. In Chapter 4 (also, Michel, Nguyen et al., 2020), we introduced a new variation of BEL that enables instantaneous uncertainty quantification of static subsurface model parameters (1D imaging) together with a global sensit-

ivity analysis and validated the approach on surface Nuclear Magnetic Resonance. The approach has also been tested on time domain electromagnetic (TEM) datasets (Ahmed et al., 2021). In addition to the low computational cost associated with predicting 1D models, associated uncertainty, and the global sensitivity that BEL1D provides, it also enables tracking posterior models and identify unexpected models in the posterior.

A drawback of BEL1D lies in the overestimation of the uncertainty from the prior model space to the posterior model space for some low sensitivity parameters of the models (Michel, Hermans, Kremer & Nguyen, 2020). This often happens when the relationship obtained between the predictions and the data is not properly linearized for example when dealing with large prior uncertainty (Hermans et al., 2019; Michel, Hermans, Kremer and Nguyen, 2020; Ahmed et al., 2021). This uncertainty overestimation has been encountered previously in Bayesian frameworks (Jeong et al., 2017; Mariethoz et al., 2010; Dosne et al., 2016 or J. Park and Caers, 2020). To address this issue, Iterative Spatial Resampling (ISR) (Jeong et al., 2017; Mariethoz et al., 2010) proposes augmenting the prior with new information drawn from the posterior at the previous iteration. Sampling Importance Resampling (SIR) (Dosne et al., 2016) assign new loads to the prior (changes the statistical distribution that describes it) according to previously drawn models and their likelihood in McMC. These algorithms are often introduced to achieve better convergence of Bayesian approaches when dealing with large prior uncertainty, as is often the case in geophysics. More recently, J. Park and Caers (2020) proposed in the framework of Bayesian Evidential Learning to update the uncertainty of data and prediction variable, by using the obtained posterior models as prior models for the next iteration after running forward functions on the obtained posterior samples. The iterative procedure is stopped when the error in fit decreases below some specified

threshold.

In this contribution, we propose a resampling algorithm based on J. Park and Caers (2020), named Iterative Prior Resampling (IPR), to reduce the overestimation of the uncertainty. Here, IPR iteratively improves the posterior distribution by adding a sub-sample of posterior models from the previous iteration to obtain a "better informed" prior, improving the learning phase of the algorithm and leading to an improved quantification of the uncertainty. We show that the combination of BEL1D with IPR significantly improves the precision of BEL1D but is not sufficient to reach an uncertainty similar to McMC approaches. To further reduce the overestimation of the posterior distribution, we propose to apply a rejection sampling algorithm to the models obtained at the last iteration to reach a more accurate estimation. By comparison with McMC, we show that this approach is converging towards the posterior model space while still having reasonable computational times. In this chapter, we demonstrate the approach on the case of the dispersion curve from seismic surface waves and present the case for surface Nuclear Magnetic Resonance, although BEL1D can be applied to any 1D geophysical inversion scheme.

5.2 Methods

5.2.1 Iterative Prior Resampling

Iterative Prior Resampling (IPR) within BEL1D consists in adding a given number of the models sampled from the estimated posterior to the prior and rerunning all the BEL1D operations from this new augmented prior (Fig. 5.1). The number of models to be added is derived from a mixing ratio, which defines the proportion between the numbers of models added from the posterior compared to the number

of models in the initial prior. Since BEL1D approximates the Bayesian problem in a reduced space where learning operations occur, IPR is designed to improve the quality of the uncertainty estimation by enabling a better correlation between the model parameters and their corresponding datasets. IPR increases the number of models compared to a uniform distribution in the region of the prior space where the posterior distribution shows a higher probability density.

This algorithm is inspired by proven methods like ISR (Jeong et al., 2017; Mariethoz et al., 2010) and SIR (Dosne et al., 2016) which indicate that improving the prior knowledge iteratively does reach convergence towards the actual posterior. As is the case with BEL1D (Michel, Nguyen et al., 2020), we cannot prove that IPR converges towards the actual posterior model space, even though previous studies showed empirically that the results are comparable to McMC (Michel, Nguyen et al., 2020) and rejection sampling (Satija & Caers, 2015). We will thus empirically demonstrate that the method converges towards a posterior distribution similar to classical Markov chain Monte Carlo approaches (see section 5.3.1).

To stop the iterative process, the Kolmogorov-Smirnov distance (KS) (Thas, 2010) between the univariate model parameter distributions at the current and previous iterations is used by defining a threshold on the similarity between the posterior distributions at two successive iterations. If this is not reached after 100 iterations, the algorithm stops. The computations are conducted in a normalized space to avoid giving too much importance to some parameters due to their high relative magnitude. The threshold is defined through empirical testing on multiple distributions with a probability of falsely rejecting the null hypothesis (the two distributions are not equal) lower than 5%. The KS distance has the advantage to be computed on the cumulative distributions functions, hence is normalized (Thas, 2010). This ensures that the distance may be used even if the obtained distributions

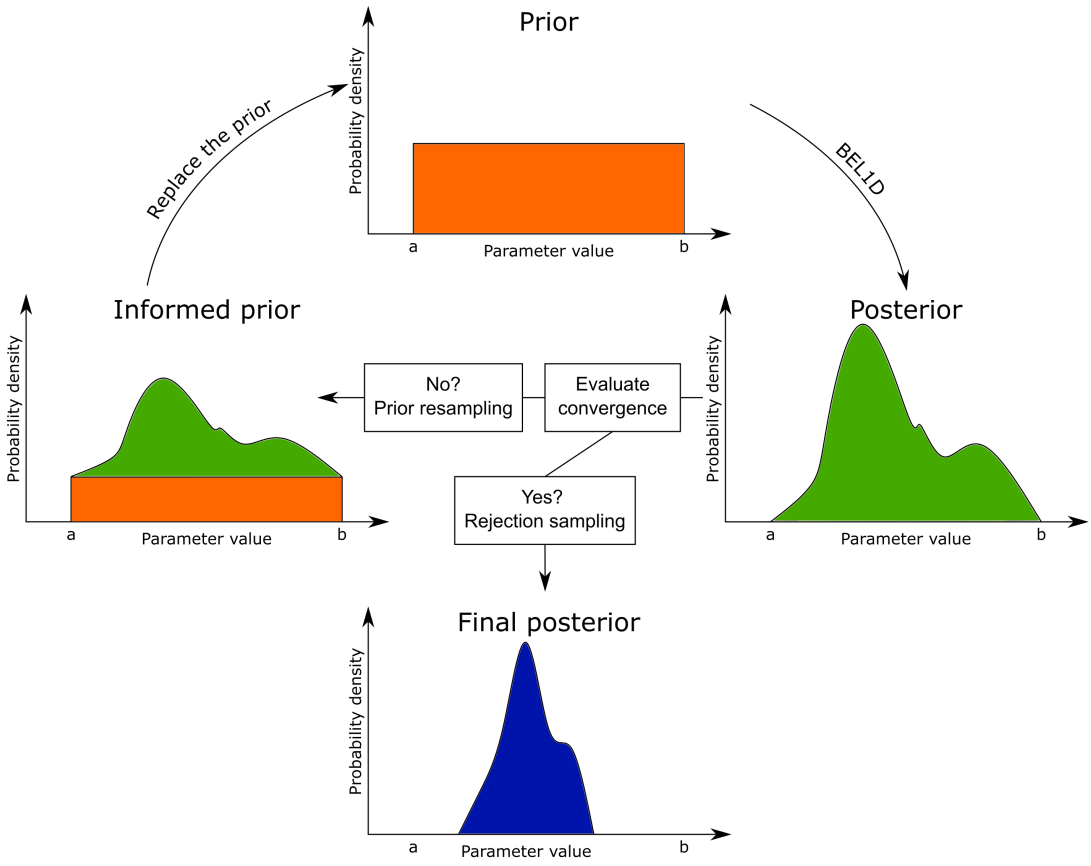

Figure 5.1: Illustration of the principle of Iterative Prior Resampling. The mixing ratio in this figure would be the ratio between the area in green and the area in orange in the "informed prior" graph. As a final step once the IPR algorithm has converged, we apply rejection sampling.

are not Gaussian. The threshold is thus only varying with the number of models in the compared distributions.

5.2.2 Rejection Sampling

Once IPR has converged, as a final step, a rejection sampling algorithm, accounting for data misfit, further improves the estimated posterior. This algorithm consists of a Metropolis Sampler applied to the models from the last iteration of IPR. This algorithm differs from an McMC approach fundamentally, as it re-samples models out of a set of already sampled models (the posterior distribution after IPR) contrary to generating new models from perturbations of previously accepted models.

1. For all models from the sampled posterior, compute the likelihood from the field data and the simulated data according to equation 5.1 (Vrugt, 2016).

$$L(\boldsymbol{m}|\boldsymbol{d_{obs}}, \boldsymbol{\sigma^2}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \times \exp\left(-1/2\left(\frac{d_{obs,i} - d_i(\boldsymbol{m})}{\sigma_i}\right)^2\right) \tag{5.1}$$

   In equation 5.1, $\boldsymbol{m}$ is the tested model, $\boldsymbol{d_{obs}}$ is the observed field data of dimensionality $(1 \times n)$, $\boldsymbol{d(m)}$ is the simulated data for the current model m and $\boldsymbol{\sigma^2}$ is the estimated Gaussian data variance (of dimensionality $(1 \times n)$) (Vrugt, 2016).

2. Initialize a counter for the number of rejected models to 0.

3. Select a random model (m_0) out of the samples with a likelihood λ_0. The first model is always accepted.

4. Select another random model (m_1) with a likelihood λ_1 and compute the ratio between the two likelihoods: $r = \lambda_1/\lambda_0$.

(a) If $r > U(0,1)$, accept the model m_1, replace λ_0 with λ_1 and reset the counter to 0.

(b) Else, reject the model m_1 and increase the counter of rejected models by 1. If the counter of rejected models is higher than 20, accept the model m_1, replace λ_0 with λ_1, and reset the counter to 0.

5. Select another random model m_1.

6. Repeat the operations of the 4th and 5th points until all the models have been analysed.

In practice, this step allows to remove the models of the posterior that are not fitting the data, i.e. which have a low likelihood. The order of the models is taken randomly to avoid introducing any bias in the computation. Assuming that the datasets corresponding to these models are already computed to calculate their Root Mean Square Error (RMSE), the computation of the likelihood and rejection sampling is straightforward.

5.2.3 Surface Waves Forward simulation

Seismic surface waves (see Chapter 2, Subsection 2.1.4) datasets are often reduced to dispersion curves. Here, we use a Python port of sdisp96 from surf96 (PySurf96) (Herrmann, 2013) to compute the dispersion curves (the data d) of 1D models, and we demonstrate our approach on the fundamental mode. The models m are expressed in terms of layer thicknesses, P-wave velocity (V_P), S-wave velocity (V_S) and density (ρ). Since surface waves are mostly sensitive to S-wave velocity (Xia et al., 1999; B. R. Cox and Teague, 2016), a link between V_P and V_S can be considered through the Poisson's ratio (σ) (Equation 5.2) (see e.g. Wathelet, 2008).

$$V_S = V_P \times \left(\frac{0.5 - \sigma}{1 - \sigma}\right)^{1/2} \tag{5.2}$$

In order for BEL1D to run accounting for potential uncertainties in the datasets, we use the frequency-dependent error model proposed by Boaga et al. (2011) :

$$\epsilon(V, f) = A \times V + B/f \tag{5.3}$$

With V the velocity, f the frequency, A [km/s] and B [s] two parameters that can be tuned according to the noise level. In this thesis, we use $A = 0.075km/s$ and $B = 20s$. Those values are empirically selected to accurately represent the uncertainty arising from the different experts interrogated during the InterPACIFIC project (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016; Garofalo, Foti, Hollender, Bard, Cornou, Cox, Dechamp et al., 2016) in the Mirandola case (see section 5.4 for more details). We will use the same model for the synthetic benchmark and the field validation (the previously mentioned Mirandola test case).

5.3 Synthetic benchmark

We first apply iterative prior resampling with rejection sampling to a synthetic surface wave's dataset generated from a model where only the shear-wave velocity profile is unknown. The other parameters (compression waves velocities and densities) are fixed mainly for visualization and comparison purposes. This simplifying assumption is later released. All the computations are performed on Windows 10 with an Intel Core i7-6800K running at stock with 64 GB of RAM at 2400MHz.

The benchmark model that we are selecting consists of two layers overlying a half-space (Table 5.1).

Table 5.1: Benchmark model description

	Thickness (th) $[m]$	S-wave velocity (V_s) $[m/s]$	P-wave velocity (V_p) $[m/s]$	Density (ρ) $[kg/m^3]$
Layer 1	10	120	300	1500
Layer 2	50	280	750	1900
Half-space	/	600	1500	2200

We are considering a noisy dataset with the noise distribution described in Equation 2. The frequency range that we use is between 1.25 Hz and 32Hz.

For the prior model space, only the thicknesses (th_i) and S-wave velocities ($V_{s,i}$) are varying. We use a uniformly distributed prior with ranges for V_s that are selected to obtain physically consistent Poisson's coefficient for all layers when considering the benchmark values for V_p. Therefore, we do not need to enforce specific conditions to ensure a valid range of Poisson's coefficient which is useful for McMC, where imposing relationships between parameters is complex. We define the prior on the five parameters with BEL1D in Table 5.2. The ranges for the uniformly distributed parameters are selected voluntarily to reflect a large prior uncertainty.

Table 5.2: Description of the uniformly distributed prior model space for the benchmark.

	Thickness (th) $[m]$		S-wave velocity (V_s) $[m/s]$	
	Minimum	Maximum	Minimum	Maximum
Layer 1	1	30	100	180
Layer 2	10	100	250	450
Half-space	/	/	500	900

This prior is automatically falsified in the BEL1D algorithm. When projecting the data into the CCA reduced space (d^C_{obs}, represented by the red lines in Figure 5.2 C to F), the position is compared to the kernel density estimation of the distributions $f(d^c_i, m^C_i)$. If the position of d^C_{obs} lies inside the probable parts of the distributions (between the 1% and 99% percentiles), the prior is accepted. Otherwise, the dataset is declared outside the prior and the algorithm stops automatically. Here, BEL1D

was initially run with 1000 models in the prior and progressively augmented with 1000 models at each IPR iteration (mixing ratio of 1). In the following benchmark study, we first present the improvements brought by the iterative resampling scheme and analyse its iterative behaviour in terms of uncertainty reduction and posterior estimation accuracy and compare it with the posterior estimated from McMC. We then show how the rejection sampler allows filtering the posterior by accounting for the data misfit. As an alternative, we show how the approach can also be used to define a pre-posterior before running McMC. Finally, we proceed with a parametric study of the algorithm.

5.3.1 Results and McMC comparison

In this section, we first show how the IPR approach allows to reduce the uncertainty of the posterior distribution compared to the original BEL1D algorithm. The improvement in the statistical correlation and the disentanglement of the model parameters in the reduced dimension brought by IPR is shown in Figure 5.2. Figure 5.3 shows the reduction of uncertainty in the posterior distribution over the prior space for the original BEL1D algorithm. We note that the estimated uncertainty remains large with a significant proportion of posterior models showing large RMSE. This modest uncertainty reduction is due to the poor relationship that we observe between the model parameters and the reduced data spaces (Figures 5.2 C and E) after one iteration. The higher dimensions, containing less sensitive parameters (Michel, Nguyen et al., 2020) display a large scattering, and hence a poor correlation. The first iteration therefore fails to resolve less sensitive parameters. In this case, as seen in Figure 5.2 (A), the different parameters are entangled in the different dimensions, indicating that all the parameters are similarly informed by the dataset given the obtained posterior.

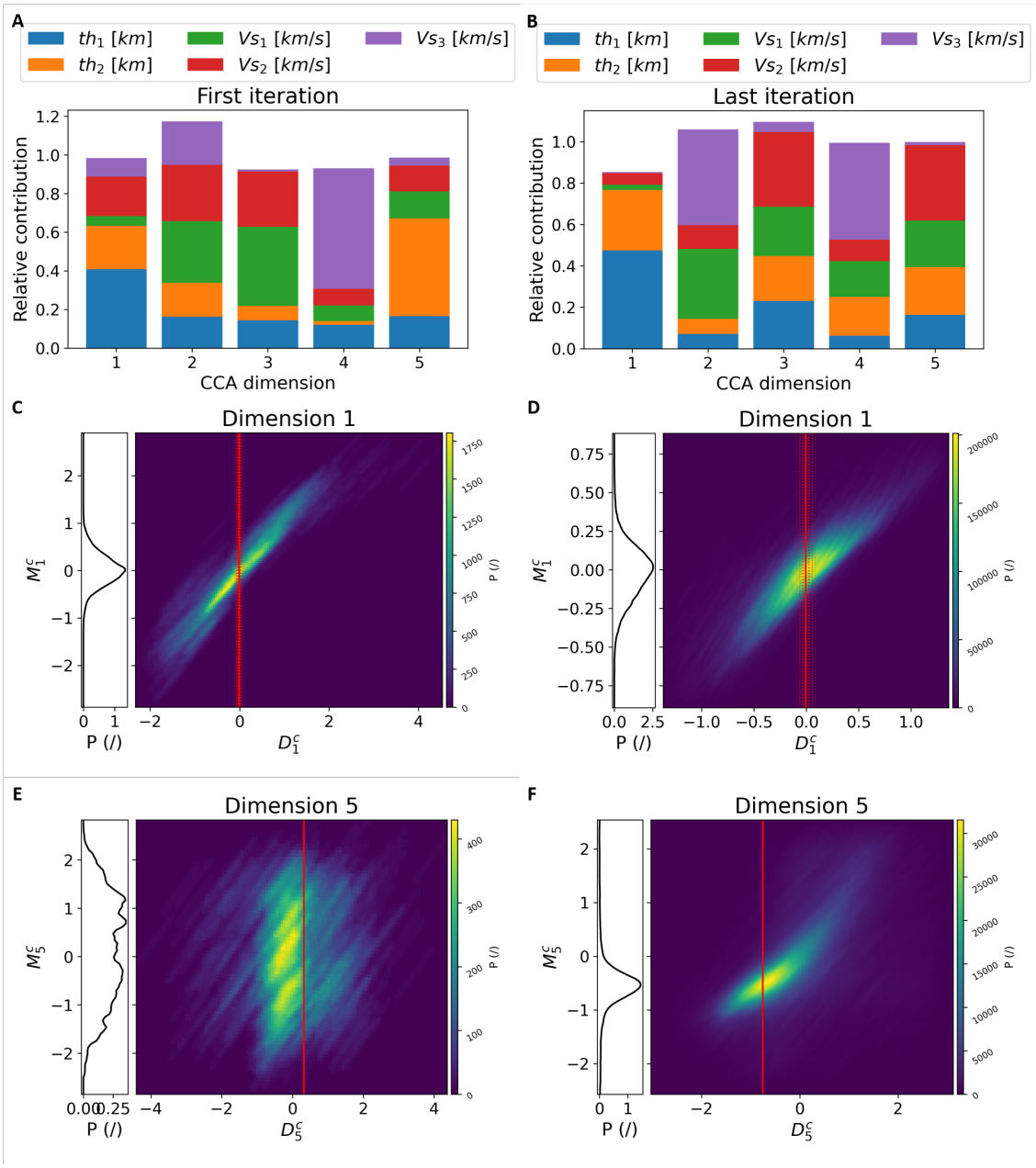

Figure 5.2: Visualization of the CCA space relationship between the models' parameters and the reduced data. A and B represent the relative contributions of all the models' parameters to the different CCA dimensions at (A) the first and (B) the last, 18th in this case, iterations. C and D represent the relationship for the first CCA dimension at (C) the first and (D) the last iterations. We present the same for the fifth dimensions in E and F for the first and last iterations respectively. In those figures, the vertical red bar represents the actual position of the data in the reduced data space (solid line) along with the noise propagated (dotted lines representing +/- 1, 2 and 3 standard deviations). The PDFs represented on the left are the kernel density estimated probability density functions of the reduced model's parameters given the dataset. We use those to sample back models from the estimated posterior. The narrower those distributions, the narrower the obtained posterior in the original model space.

The implemented iterative procedure allows a substantial reduction of uncertainty from the prior (yellow in Figure 5.3D) to the posterior at the last iteration (blue in Figure 5.3D) and a disentanglement of the model parameters in the reduced dimension (Figure 5.2 B). The improvement of the statistical relationship that is learnt through the procedure is evident from the comparison between Figures 5.2C and 5.2D (dimension 1, mostly sensitive to the thicknesses) and between Figures 5.2E and 5.2F (dimension 5 where all parameters are entangled but where the corresponding velocities dominate). Moreover, the posterior is more accurate with iterative resampling than without. The posterior models are shown as 1D profile in Figure 5.3 with their RMSE on the dataset as colour scale. As can be seen, most models of the posterior are fitting the data within their noise level, but some models with a high RMSE remain.

Figure 5.4 shows the evolution of the different parameters with the iterations. The mean value of the posterior for the first iterations are far away from the actual values and fall outside the likely part of the distribution for some parameters, with a corresponding too large uncertainty. IPR proves efficient in improving the accuracy of the posterior and in reducing its uncertainty due to the improved correlations between the data and the parameters (Figure 5.2), especially for the higher dimensions.

Since BEL1D (with or without IPR) does not theoretically ensure convergence towards the actual posterior, we benchmark our results with an McMC algorithm: a Metropolis sampler with on-the-fly adaptation of the scaling factor (Vrugt, 2016) (Figure 5.5). We can see that our implementation delivers an accurate estimation of the posterior mean but that the uncertainty is still overestimated. Note that BEL1D and IPR results were obtained in less than 5% of the time it took the McMC algorithm to converge and required only 3% of the number of models sampled by

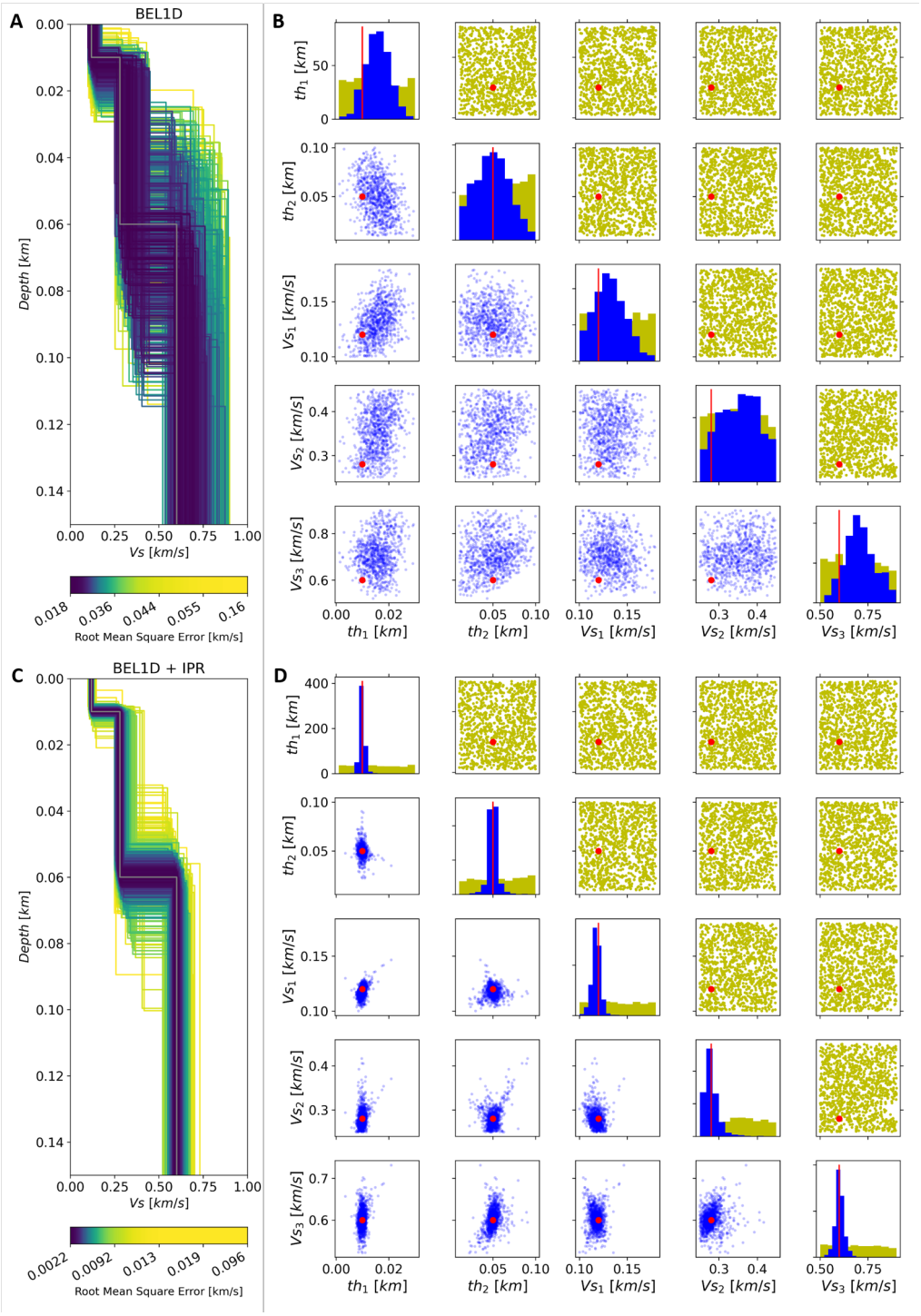

Figure 5.3: Results of BEL1D before (A and B) and after applying IPR (C and D) on the benchmark case. (A and C) Visualization of the obtained 1D profiles along with their associated RMSE and (B and D) comparison of the prior (yellow) and the posterior (blue) model space for the different variables of the models. The scale for the RMSE is linear, but we built the colour scale to represent percentiles, with each colour representing as many models as the others. The RMSE corresponding to the noise level is 0.026 km/s. Eighteen iterations were required to achieve convergence based on the KS distance in this specific case. They took under 20 seconds to run on the test computer.

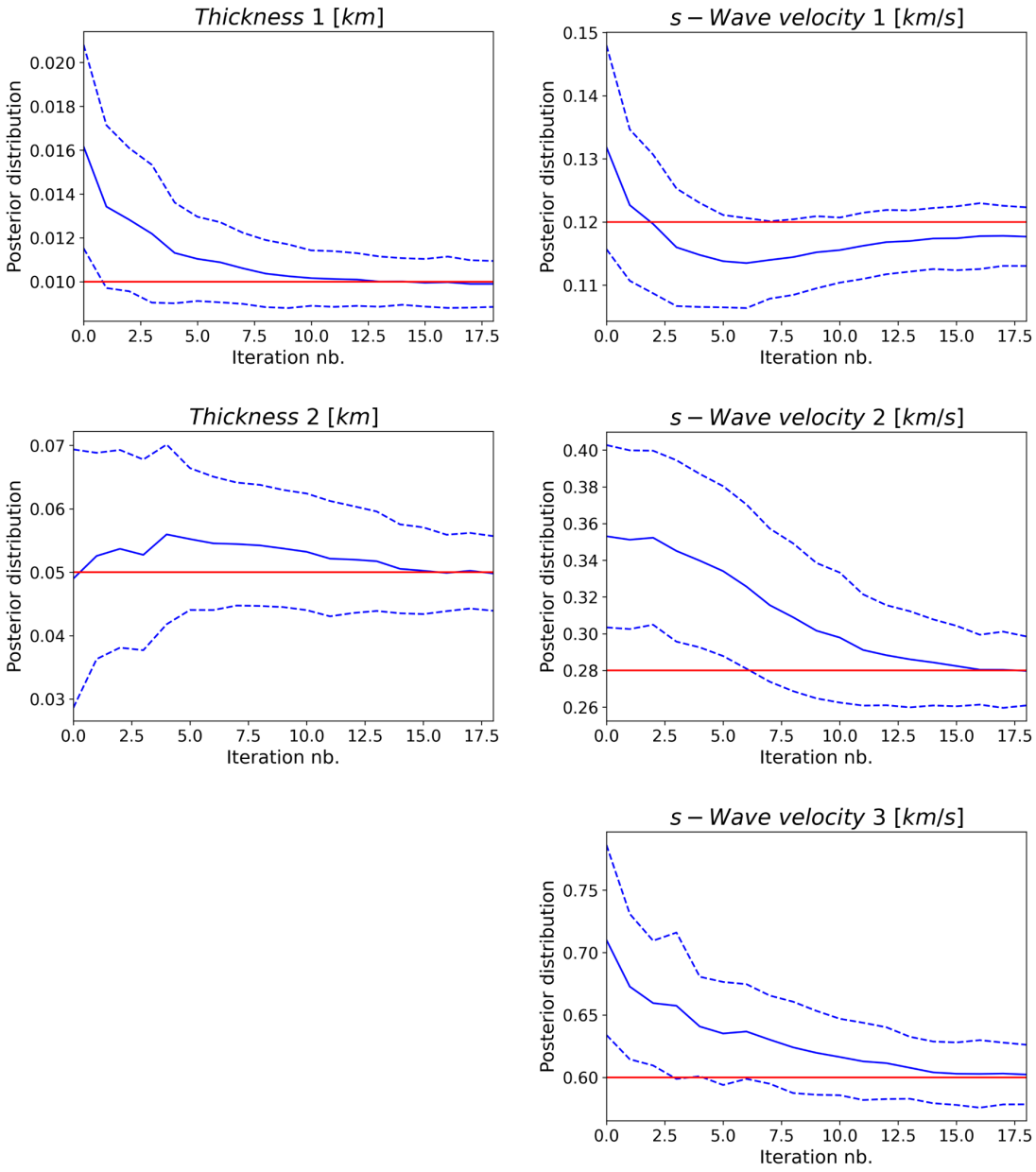

Figure 5.4: Analysis of the convergence of IPR. The solid blue lines represent the mean for the corresponding distributions, the dashed lines the mean +/- the standard deviation and the red lines represent the actual benchmark value that we are searching.

the McMC algorithm. In addition, McMC clearly identifies a strong correlation between parameters, $e_2/v_{S,2}$ being the most pronounced. The results obtained by using IPR on BEL1D show the same correlation, but it is partly hidden by outliers in the distribution. Since BEL1D does not require the data misfit or the likelihood of a model to determine if the sampled model from the posterior distribution is accepted or not as opposed to McMC, some models with high RMSE values are sampled. This can be observed in Figure 5.3C where the uncertainty range is impacted by sampled models having a high RMSE. It is thus inherent to BEL1D to find outliers that would be discarded in an approach that is based on the data misfit (or the likelihood) and is a result of the approximation of the inverse problem in the low dimensional space.

To account for this limitation, we propose to apply rejection sampling to the models that are present in the sampled posterior model space in Figure 5.5. Rejection sampling allows getting rid of highly unlikely models with high RMSE (or low likelihood) and to generate results much closer to the McMC benchmark with little computational efforts (1 sec in this case). After rejection sampling, the correlations that exist between the parameters are clearly identified.

Even though rejection sampling is very compelling as a quick approach to refine the obtained distributions, it does not provide the exact posterior model space. We therefore propose another approach to obtain the exact posterior distribution: we use the posterior obtained from BEL1D and IPR as the prior for a classical McMC approach. The results of this approach are presented in Figure 5.6. Using this approach, the retrieved posterior distribution is similar to the one obtained by McMC with the original prior. In terms of computation time, all proposed approaches significantly reduce computation time compared to McMC on the original prior (which takes 490 sec). BEL1D + IPR takes 20 sec (4% of the McMC) while BEL1D + IPR

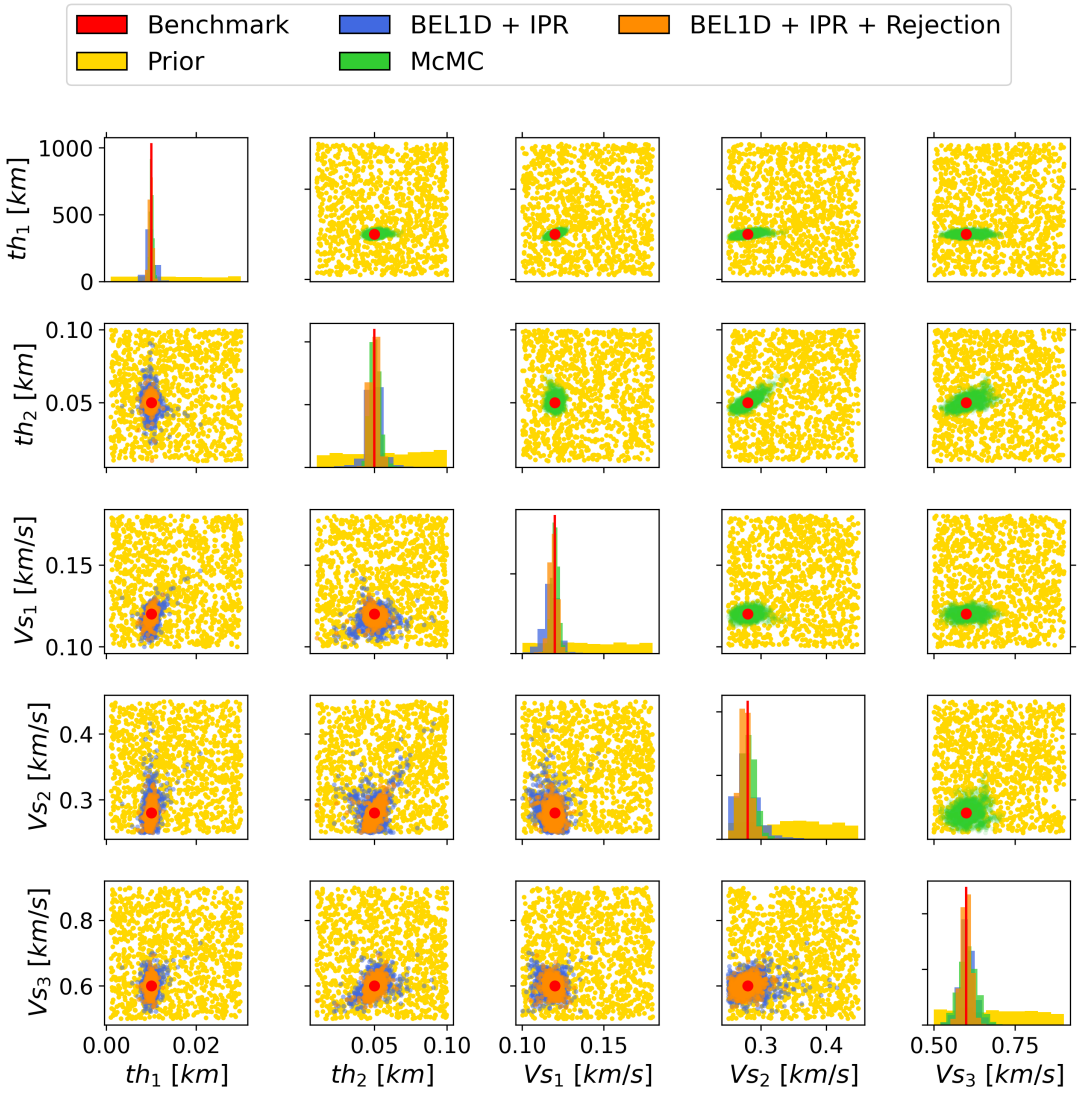

Figure 5.5: Improvements obtained by using rejection sampling after IPR. The results are compared to the results obtained by an McMC algorithm applied on the prior model space (green).

+ McMC takes 125 sec (25.5%). Nonetheless, applying rejection sampling provides the best compromise between accuracy and computation time as it only takes one second to compute after BEL1D and IPR for a total of 21 sec (less than 4.5% of the time for a standard McMC approach).

In summary, BEL1D when combined with IPR already provides a rapid first estimation of the posterior model space but overestimates the uncertainty (Figure 5.3). When combined with rejection sampling (obtained at a negligible CPU cost) the obtained distribution mimics the behaviours of the results from an McMC approach, but small discrepancies remain (Figure 5.5). Therefore, if one wants to improve even further the estimation of the posterior model space, it is possible to use the posterior obtained through the BEL1D + IPR workflow as a pre-posterior for an McMC approach that will provide an accurate estimation of the posterior (Figure 5.6) while still being faster to converge than the application of the McMC approach to the original prior.

5.3.2 Parametric study of iterative prior resampling

Here, we sequentially study the influence of key parameters in the proposed methodology. First, we discuss the impact of the number of models in the original prior and the mixing ratio $n_{posterior,add}/n_{prior,init}$ between the posterior to add to the prior and the initial prior. We will analyse their impact on key aspects of the posterior estimation and on the computation time, since it is crucial to remain under a reasonable threshold in order for the method to remain attractive. We are using different indicators to analyse the impact of the parameters. First, we use the normalized difference between the mean of the obtained posterior distribution and the actual benchmark value. Then, we use the reduction of the standard deviation from the prior model space to the posterior model space. Again, we normalize this reduc-

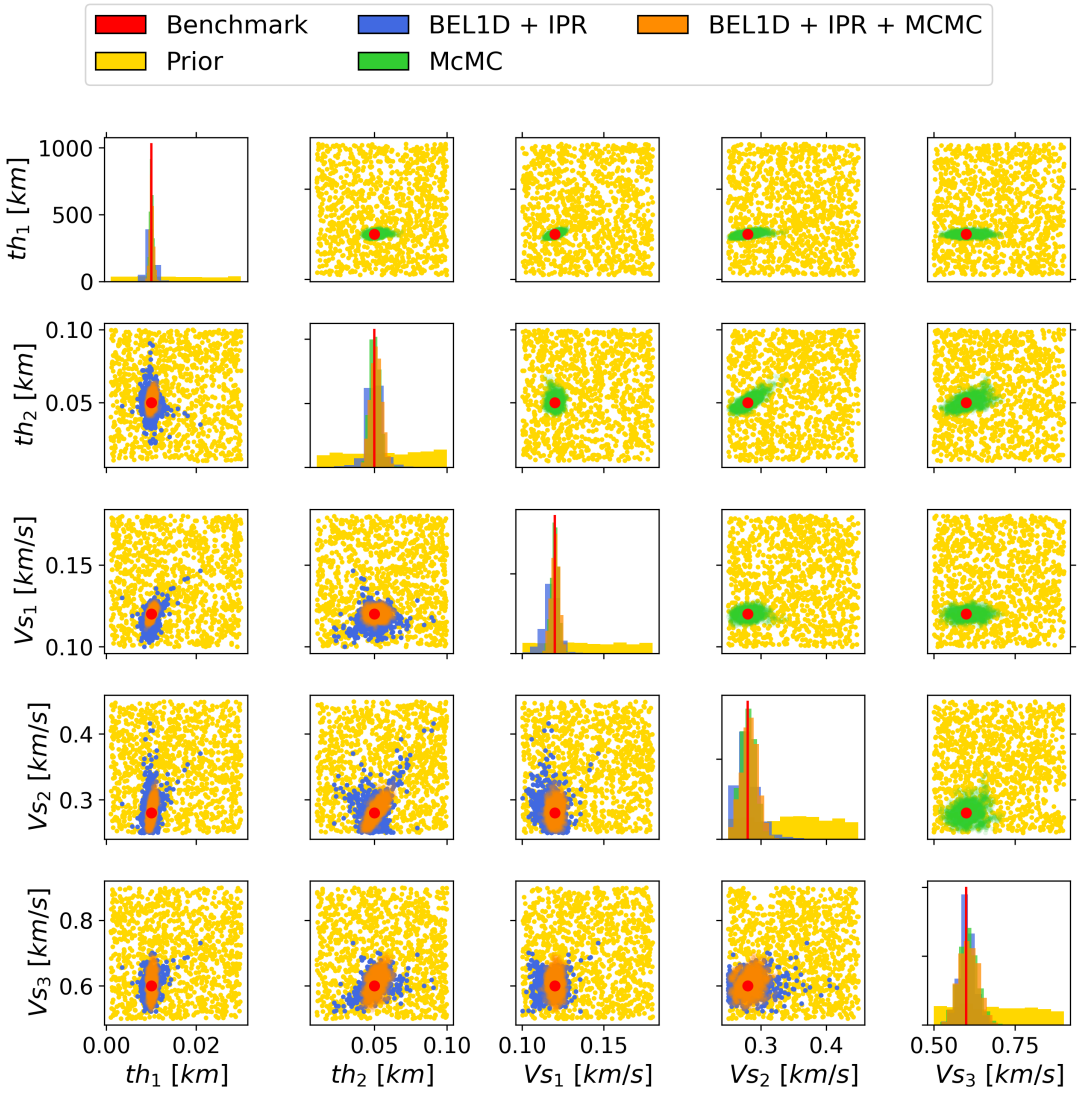

Figure 5.6: Improvements obtained by using McMC after IPR. The results are compared to the results obtained by an McMC algorithm applied on the prior model space (green).

tion for visual purposes. Finally, we are recording the needed computational time. We repeated each test 10 times to account for natural variability of the Bayesian process (standard deviation of the indicators presented with error bars).

By default, the number of models in the initial prior corresponds to the number of models to be sampled in the posterior which can be defined by the user, corresponding to a mixing ratio of 1. The number of sampled models can increase or decrease to reach the required mixing ratio that can also be defined by the user. However, the initial number of models sampled in the prior still largely governs the convergence of the algorithm, as it controls the accuracy of the algorithm in the early iterations (Michel, Nguyen et al., 2020).

The results of the analysis is presented in Figure 5.7 (first column). There is clearly an impact of the number of models in the prior on the convergence of the algorithm. For a low number of models sampled in the prior, we can see that the prediction is neither precise (Figure 5.7 – A), nor accurate (Figure 5.7 – C): the mean of the posterior is not properly estimated, and the uncertainty is overestimated. Above 500 models sampled, we see that the behaviour stabilizes itself towards a more precise and accurate prediction. This means that choosing any value above 500 for the number of models in the sampled prior model space is conservative. This behaviour is related to the requirement to sample enough models to properly learn the relationship in the low dimensional space. Otherwise, the first iteration will not identify properly the region of the prior that should be resampled, adding inappropriate models to refine the posterior and resulting in a prior not consistent with the data.

When analysing the impact on the computation time (Figure 5.7 – E), we note that the time is only increasing linearly, which indicates the order of complexity of the algorithm being approximately $O(n)$, n being the number of models sampled

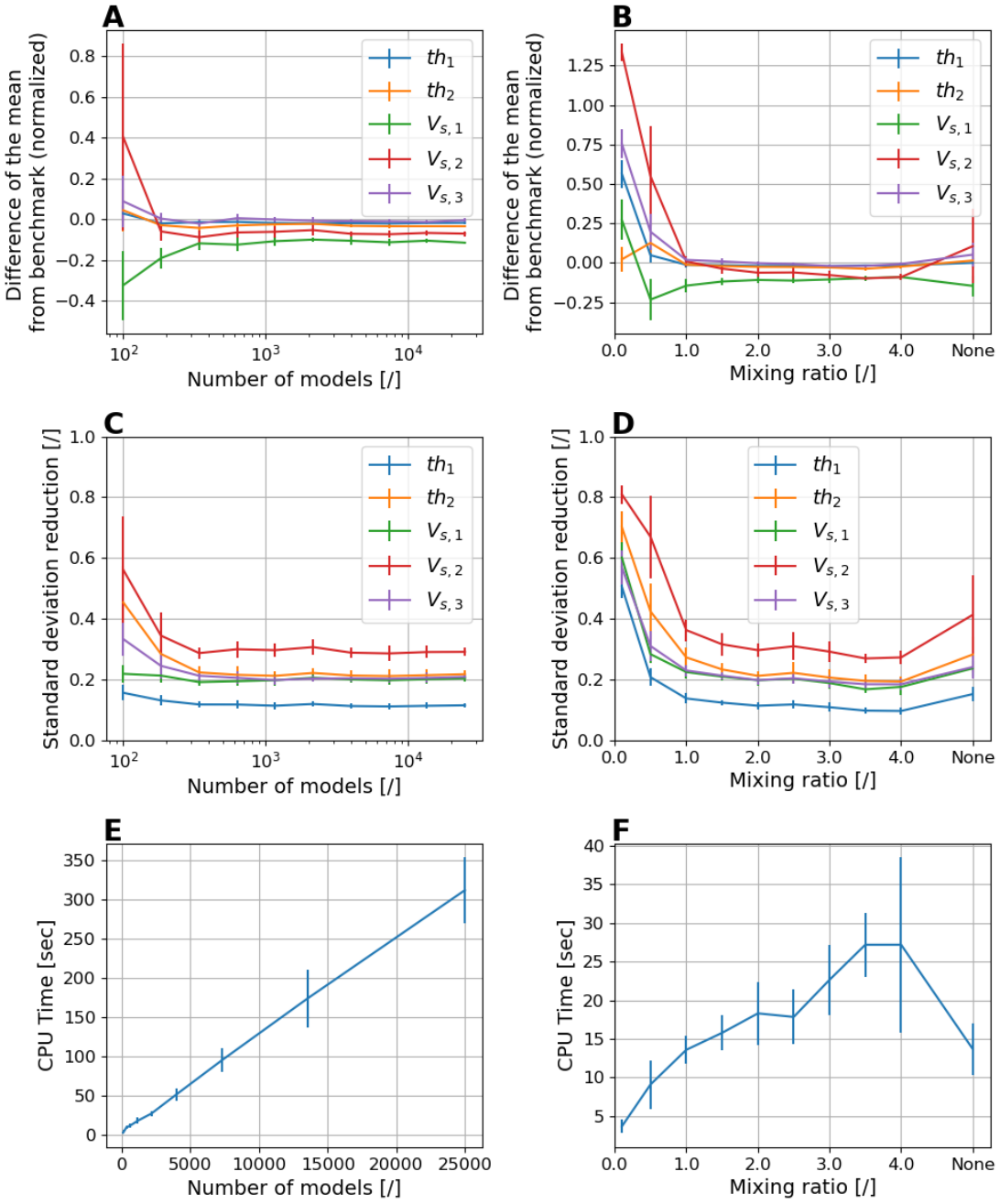

Figure 5.7: Impact of the number of models in the sampled prior (first column) and of the mixing ratio posterior/prior (second column). From top to bottom, (A and B) the accuracy of the obtained estimation (Euclidean distance between the mean of the obtained posterior distribution and the benchmark model, normalized by the initial prior standard deviation), (C and D) the precision of the obtained estimation (ratio of the obtained standard deviation for the posterior estimation and the initial prior standard deviation) and (E and F) the computation time. For each value presented on the graph, the computations where performed ten times to estimate the run-to-run standard deviations (error bars). None for the mixing ratio means adding everything that is sampled without taking care of the ratio.

in the prior. Therefore, we should stick to the lower values for this parameter. We introduced the mixing ratio in the model to avoid bias in the estimation of the posterior. This parameter (m) equilibrates the load between the prior and the posterior. The idea is that at a given moment, convergence will be achieved, not by reaching a stable posterior, but because perturbations in the prior from IPR become negligible. For low mixing ratios (Figure 5.7 – B and D), BEL1D converges towards a poorly resolved mean estimation of the posterior. The number of added samples is too small to really improve the learning process. From the benchmark, a mixing ratio around one seems appropriate, since adding more samples does not change significantly the mean estimation nor reduces the standard deviation, while remaining reasonably time-consuming. We also add the sampled models from the posterior at each iteration without taking care of the proportions between one-another (see 'None' in the figures). This leads to similar results globally, with the noticeable exception of the observed run-to-run standard deviation on the obtained distributions, which is larger. In terms of computation time (Figure 5.7 – F), we observe that the time increases with the mixing ratio. This increase in time is due to the increased number of models that need to be computed in order to satisfy the mixing ratio. It is also observable that the CPU time required is similar with a mixing ratio of "1" and no mixing ratio ("None"). Therefore, we advise to use a mixing ratio of one for an optimal behaviour control.

The parametrization of the prior model space has typically an important impact. We analyse the impact of having 3 (the above benchmark model), 4, 5, 6 and 7 layers defined in the prior. For each case, we use the same type of parametrization as in the benchmark. Hence, we only take into account the n-1 thicknesses and n s-wave velocities, n being the number of layers. For the algorithm to work properly with additional layers (thus more parameters to explore), we increased the number of

models sampled from the initial prior by multiplying it by 2 for each added layer, since the models are more complex to explore. This relationship was obtained through empirical testing. The results are presented in Figure 5.8. We observe that adding layers in the model slightly increases the uncertainty, which is logical since the prior uncertainty is larger. Nonetheless, we are still able to retrieve the main aspects from the model, namely the number of identified layers still remaining 3 for all the cases, showing that the method does not need a prior constrained to 3 layers to predict them. Some variations are observed inside the layers, but the estimation is always skewed at the boundaries between the layers and not in the middle, where the impact on the dataset is more important. This results in smoother transitions between layers and is inherent to the non-unicity of the solutions when the number of parameters increases.

Exploring a larger model space requires more time since more models are needed in the prior samples (multiplying by 2 for each added layer). In the test case presented above, the time was approximately doubled for each added layer. This is linked to the number of models required to cover the full model space and a similar increase applies to McMC approaches. Nonetheless, in case the number of layers in the model is uncertain, it can be beneficial to start with a large number of layers to identify how many layers are necessary to explain the data, and then to run BEL1D with a reduced number of layers.

5.4 Validation: application to Mirandola field dataset (InterPacific)

The InterPacific project (Intercomparison of methods for site parameter and velocity profile characterization) aimed at assessing the reliability of in-hole and surface

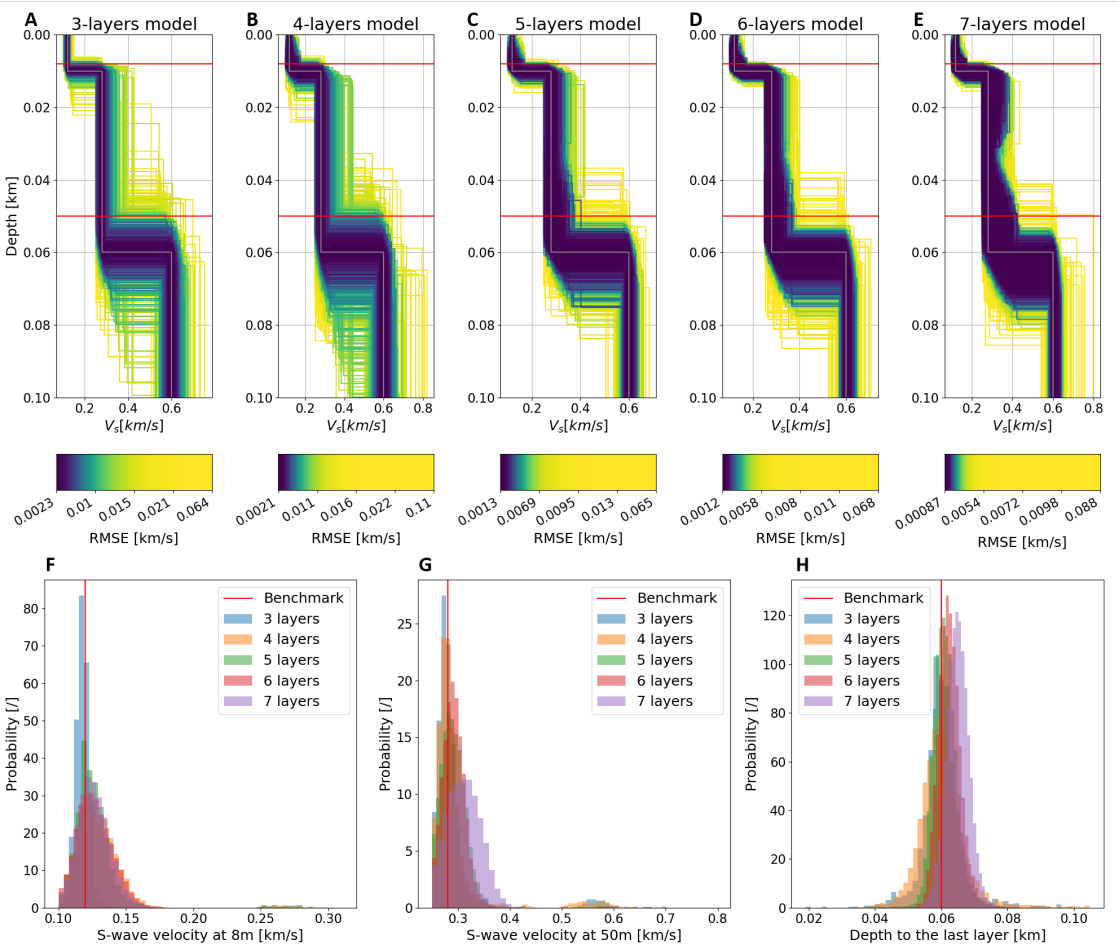

Figure 5.8: Impact of the number of layers on the obtained posterior model space. (Top) From A to E with respectively 3 to 7 layers with an increment of one. The colour scale is unique to every subplot as it represents the obtained probability for the represented case with each colour representing as many models in the subplot. The RMSE corresponding to the noise level is 0.026 km/s. (Bottom) Histograms of the obtained models at specific depths (8 and 50 meters for F and G respectively) and the estimated depth to the bedrock (H).

wave methods for the estimation of shear wave velocity by comparing the interpretations obtained from different experts (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016; Garofalo, Foti, Hollender, Bard, Cornou, Cox, Dechamp et al., 2016). In this section, we use the mean dispersion curve that arose from these experts' analyses of the raw data to demonstrate BEL1D applicability to retrieve the shear wave velocity profile. We will focus on the case proposed by Garofalo,

Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al. (2016) in the Mirandola test-site located in the Po plain. The site is characterized by soft sediments (sand, clays and gravels) overlaying a Pliocene bedrock (sand and marine pelite), found at a depth between 50 and 150 meters (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016). From this available information, we will assume that the terrain is composed of three layers: two soft layers, representing the alluvial plain and its heterogeneity and a third layer representing the bedrock. Above this, we will impose that the models must have Poisson's coefficient within the range 0.2 to 0.45 for each layer through added conditions on the prior. Including the Poisson's coefficient in the prior would be redundant with the inclusion of either V_s or V_p, as the knowledge of two out of the three parameters enables the computation of the third. We made the choice to use V_s and V_p in the prior and to exclude models that do not correspond to the acceptable range of Poisson's coefficient. Therefore, the prior is no longer uniformly distributed across the different dimensions, as some models are excluded through this added rule. The prior is composed of 11 parameters: 2 thicknesses (th_i), 3 S-waves velocities ($V_{s,i}$), 3 P-waves velocities ($V_{p,i}$) and 3 densities (ρ_i). Their uniformly distributed prior distributions are described below:

- $th_1 \in U[5, 50]m$
- $th_2 \in U[45, 145]m$
- $V_{s,1} \in U[100, 500]m/s$
- $V_{s,2} \in U[100, 800]m/s$
- $V_{s,3} \in U[300, 2500]m/s$
- $V_{p,i} \in U[200, 4000]m/s$ with $i = 1, 2$ or 3
- $\rho_i \in U[1500, 3500]kg/m^3$ with $i = 1, 2$ or 3

As is observed, the prior is very large given the context. This approach is used to show the capacity to reduce significantly the uncertainty when very few is known in advance from the site.

In this case, since there are more parameters in the model, we will use a larger set of models sampled from the prior. We settled on 10000 as being a good trade-off between fast computations and reliable estimations.

To sample the models, we used uniformly distributed variables with rejection in order to satisfy the Poisson's coefficient interval. From the original 60 dimensions of the data (the 60 couples frequency-phase velocity of the sampled dispersion curve), PCA reduced the dimensionality to 11, explaining more than 95% of the variability. In this case, the number of dimensions is constrained by the BEL1D algorithm and not the variability to explain as, to be able to perform further steps of BEL1D, the dimension of d^f (the PCA reduced dataset) must be larger or equal to the one of $m^f (= m)$ (the - potentially PCA reduced - model space).

We propagate the error on the data itself using the propagation process described in Chapter 4 with the noise model from equation 3. Once noise propagated, we can sample as many models from the posterior as wanted. Here, since we are iterating on the prior, we sample a number of models sufficient to satisfy the mixing ratio fixed to one, as for the previous benchmark case.

After 13 iterations (3min), we obtain the results presented in Figure 5.9 and Figure 5.10. They clearly show that we are encompassing the uncertainty observed by the different experts from the InterPacific project (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016). We can observe multiple interesting aspects. First, we see that we do not reduce the uncertainty for the densities. This is coherent with the well-known fact that surface waves have low sensitivities to this parameter (Xia et al., 1999). On the other hand, we observe that we have achieved

significant reduction for the parameters of the V_s profile and, to a lesser extent for the V_p profile. This latter observation is mostly due to the intrinsic link that exists between V_s and V_p through the Poisson's coefficient, as the measurement itself is not particularly sensitive to the compression waves velocities. This is especially visible in Figure 5.10 where we can see that the range of values for V_p is limited by the sampled values of V_s for a given layer. In Figure 5.9, we present the results of the algorithm at different steps. Figure 5.9A presents the results of BEL1D, where we easily see the large uncertainty in the estimated posterior. If we directly apply rejection sampling to the results of BEL1D (Figure 5.9B), we reach a better estimation, but we are still overestimating the uncertainty. Applying IPR (Figure 5.9C) results in a better overall estimation with some outliers still present. Finally, applying rejection sampling to BEL1D with IPR (Figure 5.9D) removes most of the high RMSE (low likelihood) models for a very reasonable cost (less than 1 second). The better estimation is also illustrated in the dataspace for the initial BEL1D run (Figure 5.9E) and the final BEL1D, IPR and rejection sampling results (Figure 5.9F).

We observe some discrepancies on the bedrock estimations compared to previous studies. We explain this behaviour by the use of the average dispersion curve as data, whereas there is a large discrepancy at low frequencies (long periods) for the dispersion curves used by the different experts (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016 and Figure 5.9).

We can further validate our approach by comparing the depth to the bedrock measured in drilling cores (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Dechamp et al., 2016) to the estimated depth from the proposed numerical process. This comparison is presented in Figure 5.11. There, we observe that our mean estimation is slightly underestimating the actual depth. Nonetheless, the measured depth is

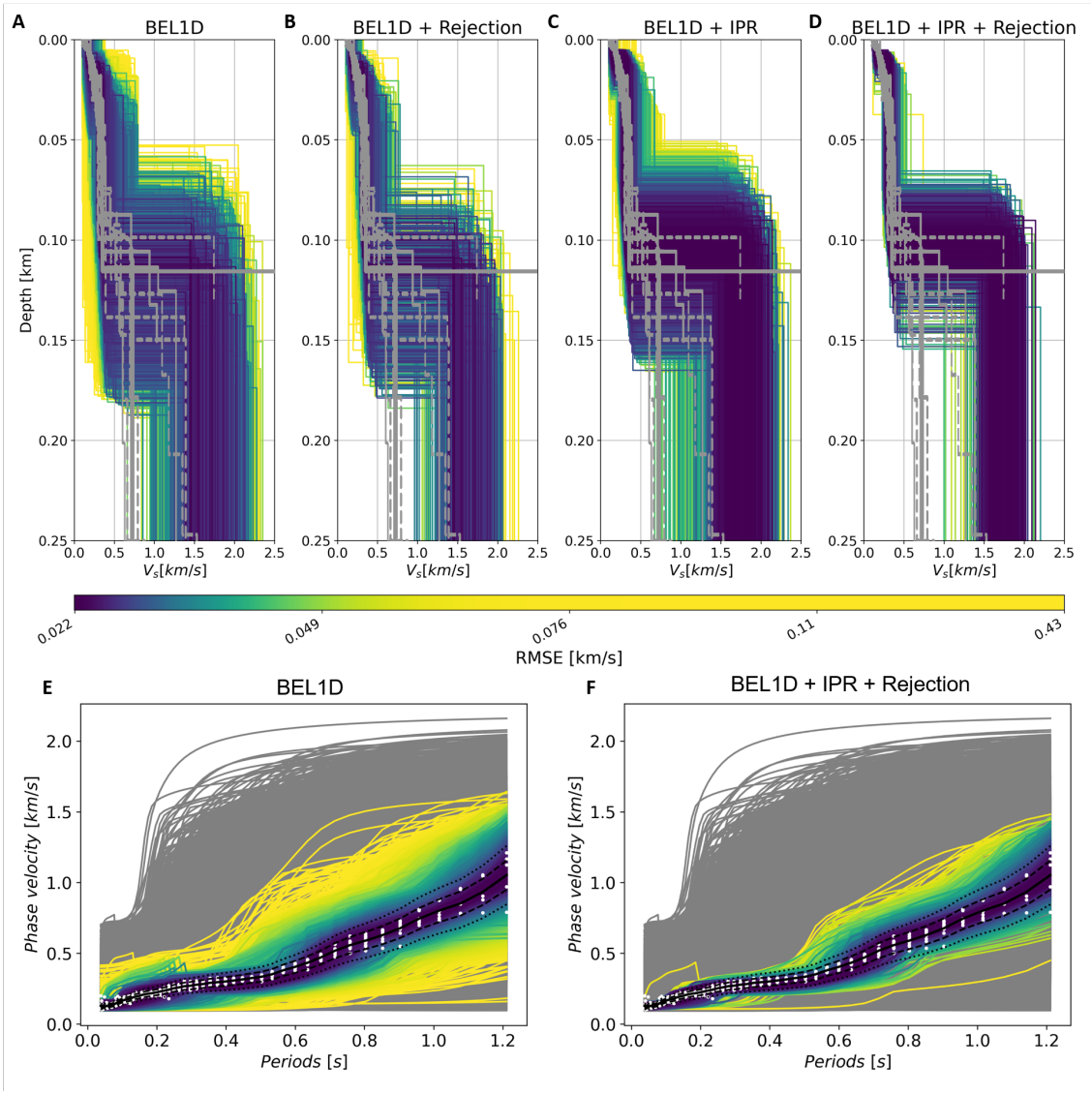

Figure 5.9: Results of BEL1D, IPR and rejection sampling applied to the Mirandola test case. (A to D) Models of the shear-wave velocity obtained from (A) BEL1D, (B) rejection sampling applied directly after BEL1D, (C) BEL1D and IPR and (D) rejection sampling applied after BEL1D and IPR with the colour scale representing the RMSE. Those figures present also the profiles obtained by the different experts (overlaid in grey) for the shear-wave velocities (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016). (E and F) Datasets associated to the different models (with their RMSE) after (E) BEL1D and (F) rejection sampling applied to the results of BEL1D and IPR. In white, we present the datasets as found by the different experts. The colour scale is the same for all the different graphs. The RMSE corresponding to the noise level is 0.038 km/s.

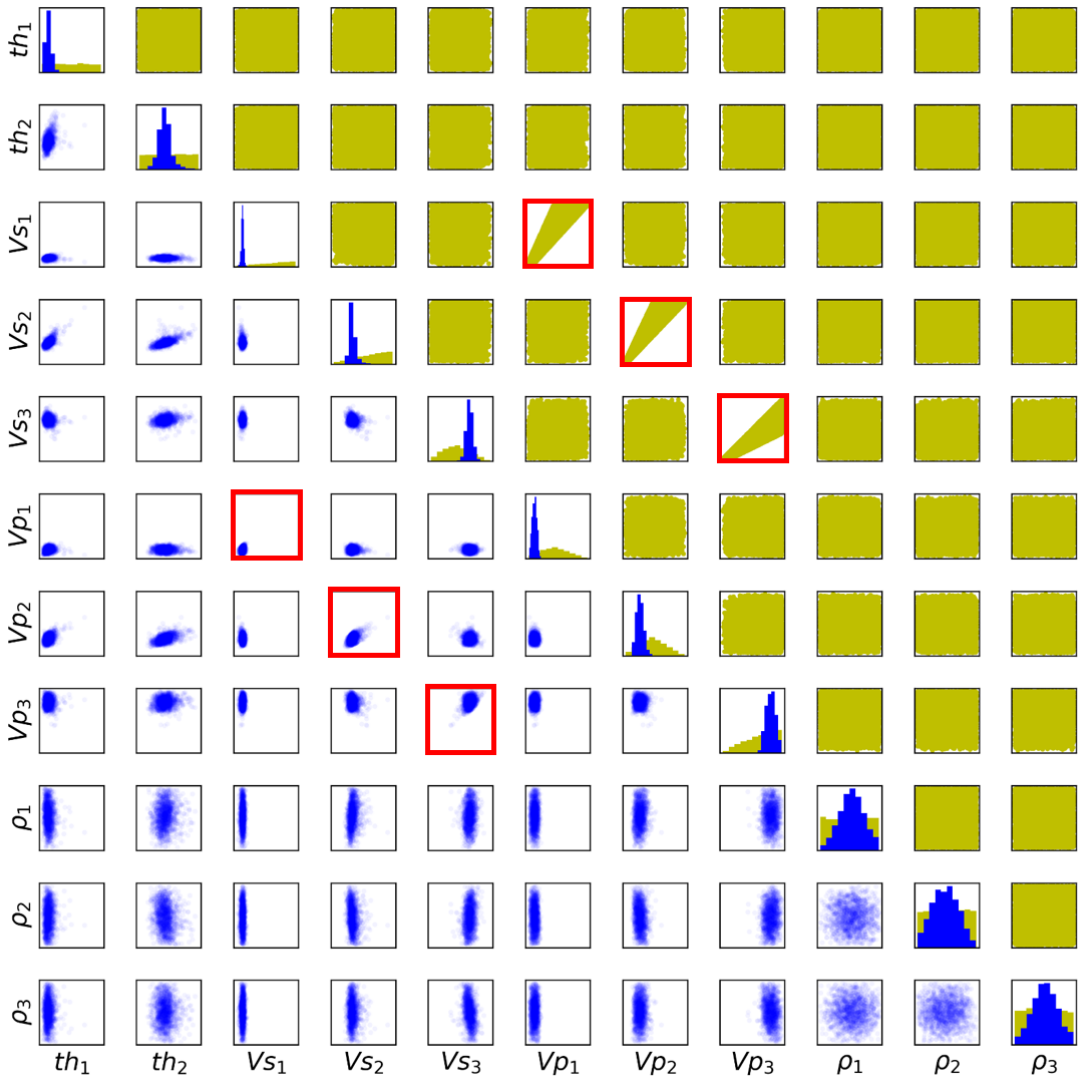

Figure 5.10: Results from BEL1D, IPR and rejection sampling applied to the Mirandola dataset. The prior models sampled from the prior model space are presented in yellow and the posterior models from the estimated posterior model space are shown in blue. The subplots presenting a light correlation between V_s and V_p are highlighted by red rectangles. This correlation is mainly due to the Poisson's coefficient that limits the ranges of possible values in the prior.

close to the mean and still in a high probability area of our estimated posterior. The relatively large uncertainty is related to the lack of constraints on the bedrock in the dispersion curve. In this figure, we illustrate once again the need for IPR to better constrain the posterior model space as only running BEL1D and rejection sampling results in a quasi negligible uncertainty reduction.

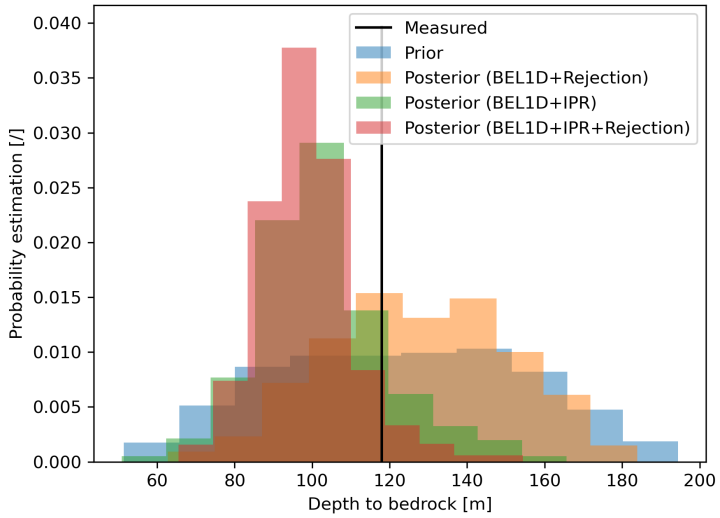

Figure 5.11: Estimation of the depth to the bedrock from BEL1D with IPR (and rejection). The actual measured depth (Garofalo, Foti, Hollender, Bard, Cornou, Cox, Dechamp et al., 2016) is displayed in black above the histograms.

5.5 The case of sNMR

Here, we will use the algorithm presented in the previous sections and apply it to sNMR data to demonstrate the applicability to other types of data[2]. We will apply BEL1D to a synthetic dataset created using MRSMatlab (Müller-Petke et al., 2016). On purpose, the prior will be large to force the demonstration that it is rather difficult to converge when very few is known in advance regarding the model.

Table 5.3: Description of the benchmark model and associated uniform prior

Layer number	th_i [m]			W_i [%]			$T^*_{2,i}$ [ms]		
	Min	True	Max	Min	True	Max	Min	True	Max
1	1	25	50	1	5	50	5	100	500
2	1	25	50	1	25	50	5	200	500
Half-space	/	Inf	/	1	10	50	5	50	500

The benchmark model along with the used prior is presented in Table 5.3. The dataset is simulated using a classical transmitter/receiver configuration with a 50 m diameter. The sampling frequency is 500Hz from 0.005 seconds to 0.5 seconds. The prior is barely informative, apart from the input knowledge that three layers can describe accurately the model. By design, the prior has zones where the dataset cannot be sensitive to the model; hence, it makes estimations of uncertainty even more complex.

[2]The results presented in this section were partially presented at the MRS2021 conference (Michel et al., 2021). The different authors contributed to:

- Code development: Hadrien Michel - advices: Thomas Hermans and Frédéric Nguyen
- Bayesian inversion: Hadrien Michel - advices: Frédéric Nguyen, Thomas Hermans and Thomas Kremer
- Writing: Hadrien Michel
- Proof-reading: all contributing co-authors

5.5.1 Results after one iteration

Let us first analyse the results obtained at the first iteration. Since the prior uncertainty is rather large, BEL1D is facing difficulties to retrieve an efficient correlation between the models and the simulated data. However, we are still able to reduce significantly the uncertainty on most of the parameters (Figure 5.12).

From the results at the first iteration, we already see that the experiment is mostly sensitive to the water contents. Then, the relaxation times and, finally, the layers thicknesses are the least sensitive parameters in this configuration. As is also expected, the first layer shows a higher uncertainty reduction at the first iteration for both the water content and relaxation time than the other parameters. This is due to the higher sensitivity of the experiment to this layer.

5.5.2 Results after applying IPR and rejection sampling

When applying IPR, we are using the information from the previous iterations to better constrain the prior. This leads to a more coherent reduction of uncertainty. Observing the results of the last iterations (Figure 5.13), we see that the obtained distributions tend towards a more accurate posterior.

If we analyse the correlation between the model's parameters (Figure 5.13), we observe that there is a significant correlation between the water content of the second layer and its thickness (Pearson's cross-correlation of -0.67). This result is corroborated by the distributions of the total water content (Figure 5.14, top). There, we observe that, even though the uncertainty on the water content and the thicknesses remains large, we reduce significantly the total water content.

In this figure, we aim to demonstrate the importance of sampling the whole ensemble and to take correlations between parameters into account. To do so, we

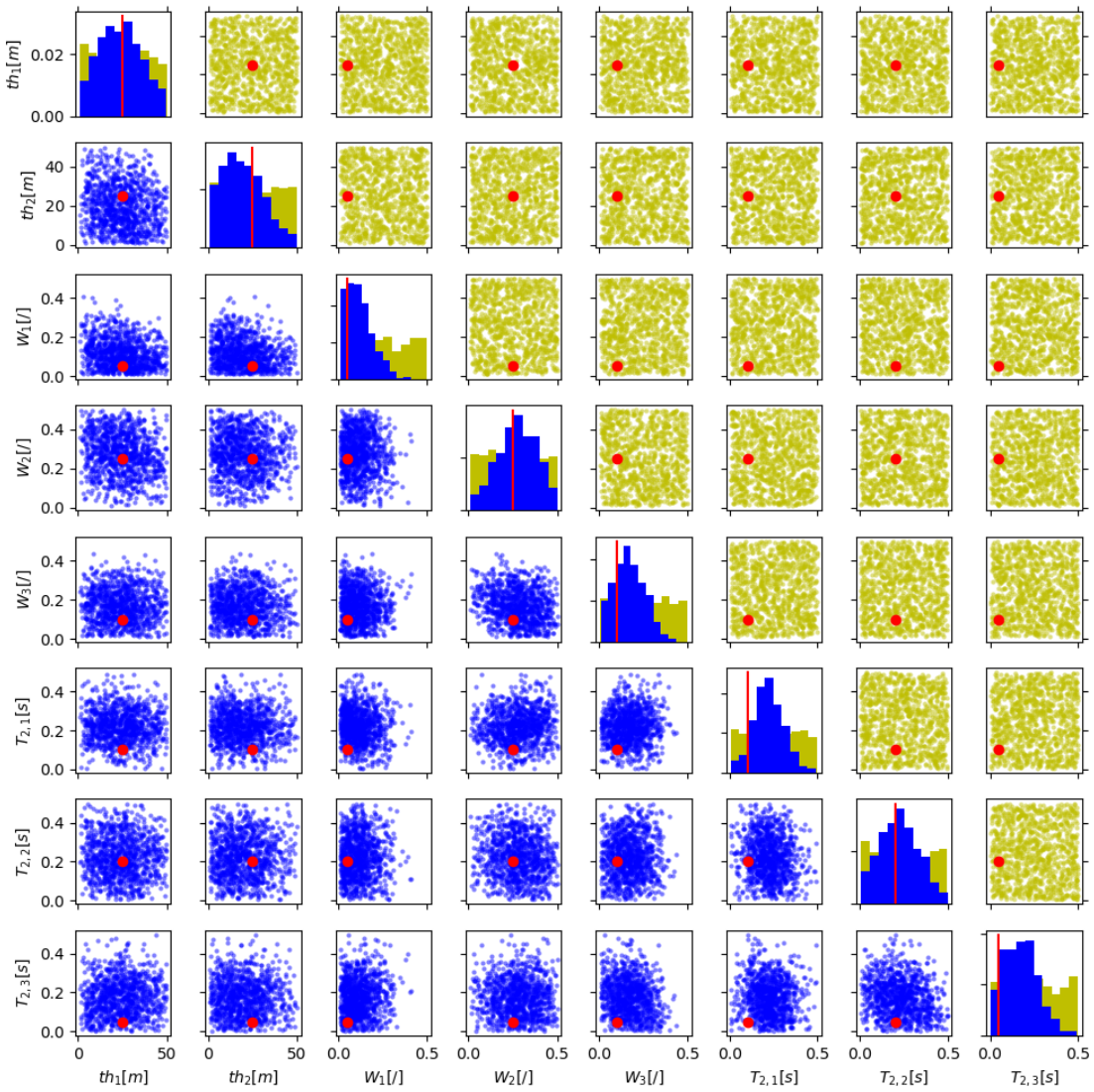

Figure 5.12: Results of BEL1D on the synthetic sNMR case.

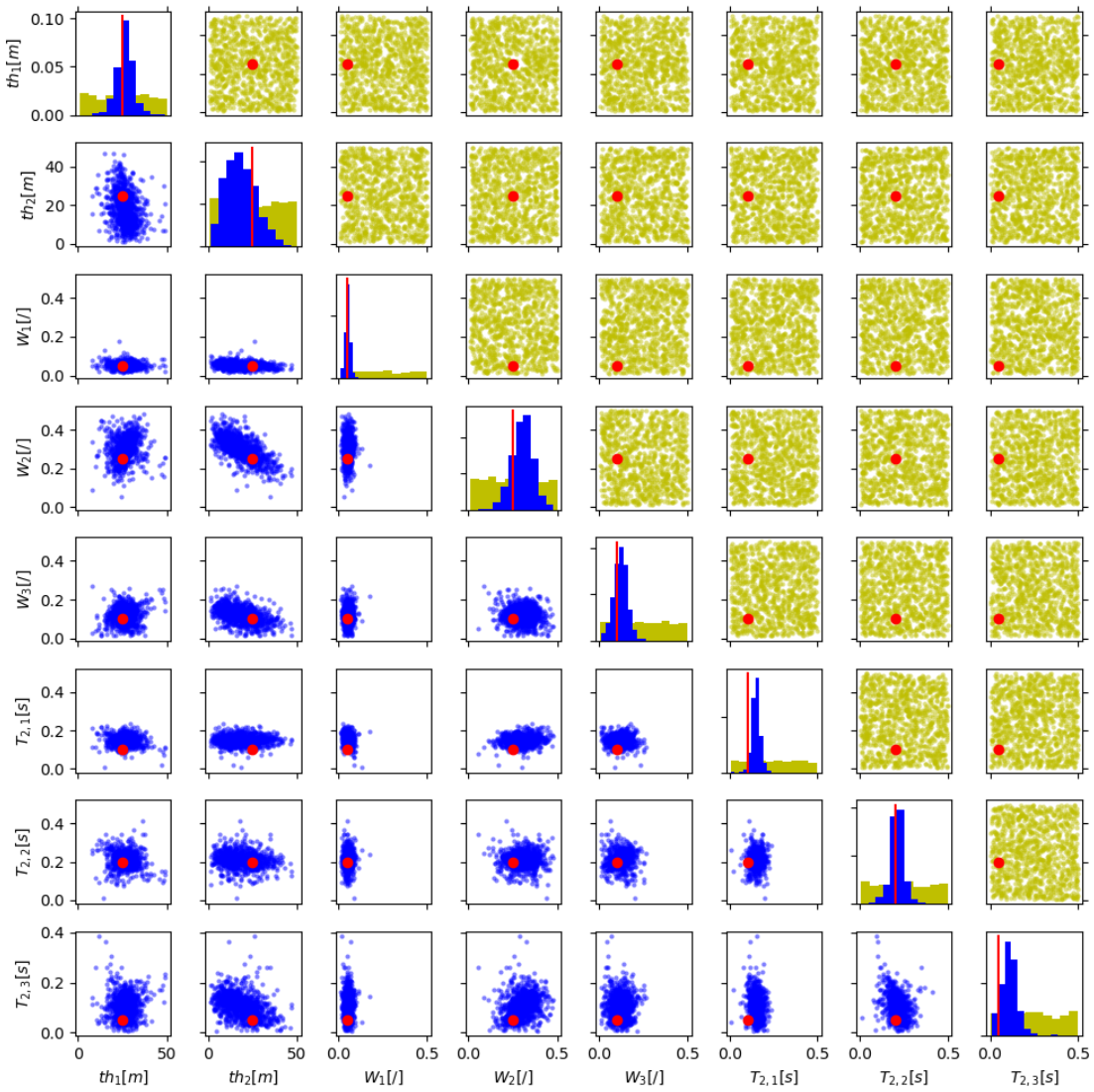

Figure 5.13: Results of BEL1D and IPR on the synthetic sNMR case.

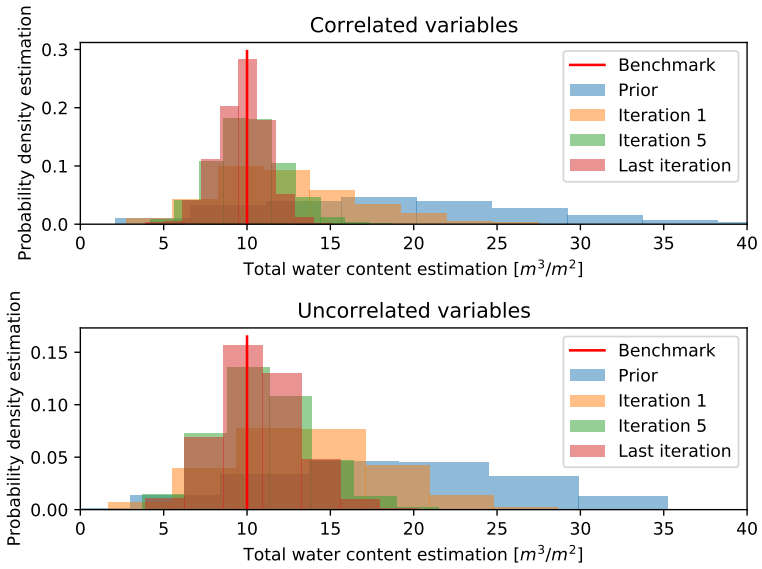

Figure 5.14: Analysis of the impact of correlation on the estimated parameters for the total observed water content.

compare the results from BEL1D to the same results, but where the link between the models variables is loosened, producing uncorrelated model parameters. We see that the correlation that exists inherently between the parameters is crucial to the model estimation.

Finally, we observe that the gain of the latest iterations is marginal. Knowing that the latest iterations are also the ones that are the longest to compute (more models in the informed prior), a user that is interested in a rapid but not especially precise estimation of the uncertainty could use fewer iterations to gain rapid insight on the uncertainty.

5.5.3 Improving the results from the Mont Rigi field data

In Chapter 4, we validated the BEL1D approach using a field dataset from the Mont Rigi natural reserve, in the Belgian Fagnes (Section 4.1.4). Those results were coherent with previous results from other geophysical methods' interpretation.

However, the uncertainty on the model parameters remained relatively large. Here, we apply IPR to the same field dataset and analyse the resulting model.

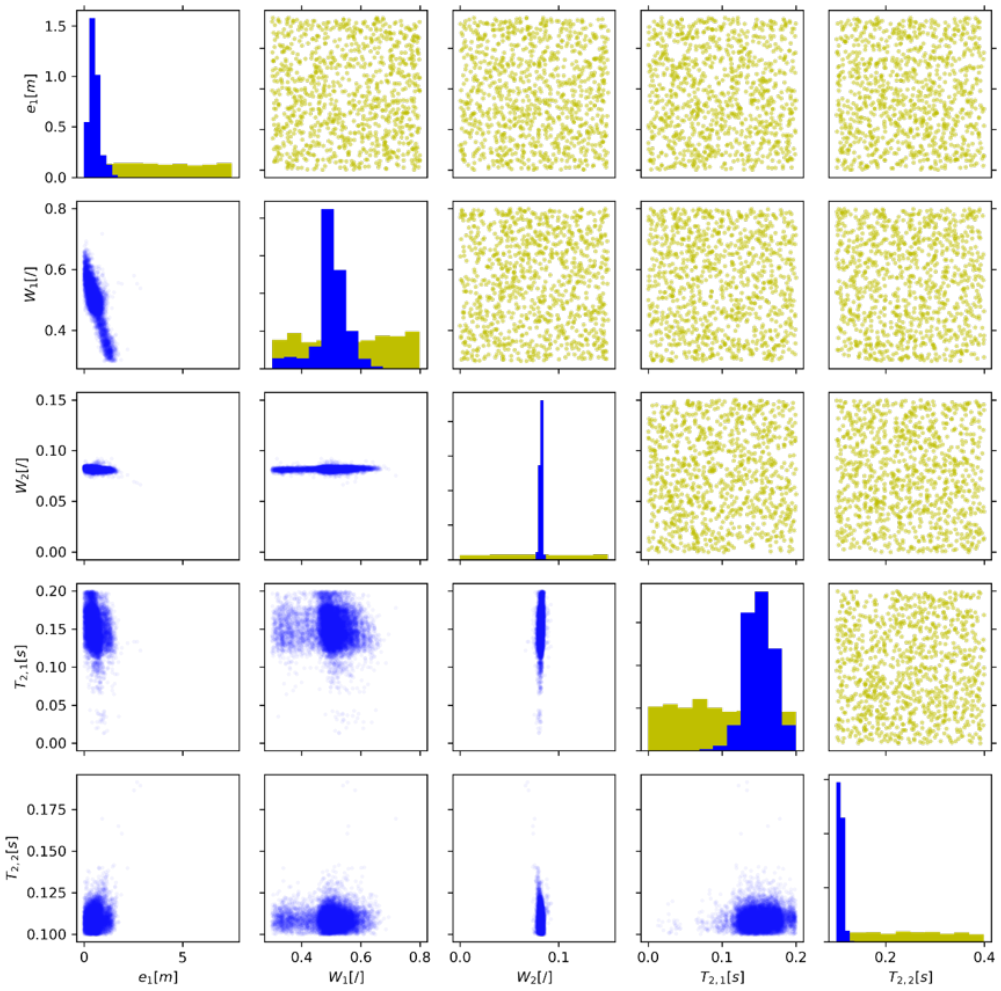

Figure 5.15: Results of IPR on the Mont Rigi field case. The different model parameters are well constrained and an even stronger correlation between the thickness of the peat layer and its water content is observed.

From the results in Figure 5.15, we see a strong correlation between the thickness of the first (peat) layer (e_1) and its water content (W_1). This result confirms the observation from the application of BEL1D alone (there is a strong non-unicity of the solution). However, instead of a limitation of the water content for a given thickness of the layer, we rather see a more coherent correlation between those two

parameters, resulting in a similar total water content in the explored system. We also observe that the parameters of the second layer are much more constrained. This is due to the very high sensitivity of the experiment to those deep layers, constituting most of the sNMR signal.

The distribution for the water content of the first layer (W_1) obtained through this algorithm also demonstrates the ability of the algorithm to recover multi-modal behaviours. Indeed, we observe that two types of models can explain the data, one being more likely than the other. This behaviour was already observed in the application from Chapter 4, but IPR preserves this multi-modal aspect.

5.6 Conclusion

In this chapter, we introduce an iterative process and rejection sampler to improve results from Bayesian Evidential Learning 1D imaging (BEL1D). This process is a combination of a first iterative step called Iterative Prior Resampling (IPR), aiming at establishing statistically representative relationships between model parameters and datasets, and a final rejection sampling test allowing to remove models not fitting the data from the posterior. Although using IPR deviates from the theoretical basis underlying Bayes rule, we show empirically that this addition to BEL1D leads to an improved quantification of the uncertainty. Our results indicate that we obtain posterior estimates from the modified IPR priors that are similar to ones obtained through McMC using the original prior.

We validate the method using dispersion curves obtained from surface waves analysis, both for a synthetic case and for an already documented field case (Mirandola, see Garofalo, Foti, Hollender, Bard, Cornou, Cox, Ohrnberger et al., 2016). Contrary to previous applications where the prior was relatively simple and inform-

ative (fewer parameters, low ranges of prior uncertainty and absence of insensitive parameters), here, the use of IPR was required to provide a satisfactory posterior estimation. This is most likely due to the low sensitivity to multiple parameters from the inverse problem (namely the compression wave velocities and the densities – Xia et al., 1999) that highly impair the ability to reduce the uncertainty through BEL1D.

The benchmark with synthetic models shows that the method evolves towards the posterior model space as estimated by an McMC algorithm. While still slightly overestimating the posterior uncertainty, the combination of BEL1D, IPR and rejection sampling approach seems the best compromise to reach the convergence at a limited computational cost.

The application to the field dataset validates the approach on a more complex case. In this example, we show that we are capable of retrieving a set of models that is coherent with results obtained by different experts and algorithms. We also show that our results are validated against in-situ field measurements for the depth to the bedrock and for prior distribution with related parameters. In both cases, the resulting distributions are not multimodal.

We also show results obtained by applying BEL1D and IPR to sNMR data with both a synthetic model with a large prior uncertainty and the field case study from the previous chapter (Chapter 4). Those show the applicability of the algorithm to other (much denser) data types. They also demonstrated the ability of BEL1D combined with IPR to obtain multimodal posterior distributions, confirming the findings on BEL by Hermans et al. (2019).

On the prior design for 1D imaging, we show that, even though using more layers is less time-efficient, priors with numerous layers converge to a simpler posterior model. This shows that even though prior uncertainty might still be large,

using many layers remains a feasible approach. Another approach that could be applied but is not explored here is coupling BEL1D with a regularized smoothed inversion, or a blocky inversion, to have a first idea of the subsurface configuration and gather the supposed number of layers from these results. This approach is less automated and therefore requires a deeper knowledge of the inversion algorithm as regularization can lead to very smooth models where layers might be difficult to extract.

In the proposed approach, the rejection sampling step is computationally very efficient since the likelihood is readily available, and allows retrieving posterior distributions similar to the ones from an McMC algorithm, with a 95% gain in computational time. This gain is explained by the limited number of models required to learn a meaningful statistical relationship between the models and the data and the full parallelization of the forward computation. This contrasts with even the most advanced McMC algorithms that are still limited to parallelizing at most the different chains (in the order of ten), since the knowledge of the previous model is required to calculate the next models to test. As an alternative, using the posterior from BEL1D combined with IPR as the prior for an McMC approach results in faster convergence of the McMC algorithm for similar results. This latter approach, even though more computationally intensive than the rejection sampling option, leads to the most precise and accurate approximation of the posterior with a significant computational gain (75%) compared to a standard McMC approach.

Chapter 6

General discussion and conclusion

In this manuscript, a new framework for the Bayesian inversion of geophysical data has been explored. This new framework uses Machine Learning to overcome the CPU intensive probabilistic inversion of the field data. It relies on the construction of a statistically significant relationship between synthetic models and their associated datasets.

In the introduction, we stated four questions that this manuscript aimed at answering:

1. Is BEL effective to perform the inversion of geophysical data in a probabilistic framework?

2. What does BEL provide that other techniques do not?

3. Can BEL be generalized to any type of geophysical datasets? What are the possible caveats?

4. How can we further improve the performances of BEL?

Each of those questions addresses a topic that is key for further developments of the proposed method.

6.1 BEL for the interpretation of geophysical data

In chapters 4 and 5, we explored our implementation of BEL aimed at the 1D geophysical imaging of the subsurface. This new implementation is called BEL1D.

In Chapter 4, the BEL1D algorithm is presented. It was applied to both a numerical and a field case study with sNMR datasets (see Subsection 2.2.4). In the first (numerical) case, the results of BEL1D are validated against an McMC algorithm. Our results empirically demonstrate that BEL1D provides similar posterior distributions as the ones obtained by McMC.
Then, the application to the Mont Rigi field case study demonstrates the ability of the newly built algorithm to work on real field data.

However, when BEL1D was applied to broad priors that are less informative about the experiment, the resulting posterior distributions are overestimating the uncertainty when compared to McMC approaches. We attribute this behaviour to the lack of correlation between prior models and their associated datasets in the canonically correlated space, due to the large diversity of possible behaviours that may be present in the original prior. To overcome this impediment, we implemented an iterative approach called Iterative Prior Resampling (IPR) that iteratively improves the posterior distributions until convergence by integrating models from the posterior of the previous iteration in a new, informed, prior (see Chapter 5). This approach was tested on both seismic surface waves dispersion curves (see Subsection 2.1.4) and sNMR data (see Subsection 2.2.4). The application of BEL1D combined with IPR on synthetic seismic surface waves dispersion curves demonstrates empirically that the iterative process converges towards posterior distributions obtained via an McMC approach. The other applications showed the ability of the algorithm to perform adequately on different geophysical datasets types, with even the ability

to recover multi-modal posterior distributions.

Overall, this manuscript empirically demonstrates the feasibility of using a Bayesian Evidential Learning framework for the interpretation of geophysical data into 1D geophysical images. We showed that the estimations of posterior model spaces obtained through BEL1D or BEL1D combined with IPR (and, to a lesser extent, rejection sampling) are similar to the ones that can be obtained through the careful tuning of McMC algorithms.

6.2 BEL advantages over other techniques

Through the different tests from chapters 4 and 5, multiple significant advantages of this approach were discovered. Among those advantages, the traceability of all the operation is key. When compared with other probabilistic approaches (classical or Machine Learning methods), the presence of any model in the posterior model space can easily be tracked down to the sampled models in the prior by analysing the canonically correlated space. This can help the user to understand the presence of odd models in the posterior, whose presence is difficult (or impossible) to track back with other techniques.

Then, BEL1D benefits from its construction as a Machine Learning approach. It enables rapid estimations of the posterior model space, assuming that the algorithm has already been trained on an appropriate prior. Moreover, the training phase can be very efficient, as the thousands forward computations can be fully parallelized. This contrasts with McMC algorithms, where even the most advanced version is limited to the number of chains being computed, since the knowledge of the likelihood at the previous iteration is required to build the new models of the chain (Vrugt, 2016).

Even though the combination of BEL1D with IPR (and, to a lesser extent, rejection sampling) is able to provide more precise and accurate posterior distributions, this comes at the cost of losing the advantage of the training phase not requiring the field data. Indeed, the initial training can still be performed without field data knowledge, but the iterations require them to build intermediate posteriors. This is a significant drawback of using an iterative approach. However, the iterative approach still requires few models to be computed (tens of thousands), which is to compare with other approaches that might require hundreds of thousands to millions of models to be able to converge. Also, the iterative process can still be heavily parallelized, as each iteration requires the computation of numerous independent forward model runs.

Another key advantage of the method is its ability to estimate any type of posterior distribution. In the Mont Rigi test case from Section 4.1.4 and Subsection 5.5.3, we showed that the algorithm was able to recover multi-modal posterior distributions. Also, in the Appendix A, the insensitive parameter M showed that BEL1D can recover uniform distributions. This means that no a priori knowledge of the posterior distributions statistical function is required. This contrasts with some approaches to uncertainty quantification in geophysics, where the shape of the posterior distributions is assumed (often a Gaussian), as is the case for most ensemble smoother techniques, or simple Machine Learning approaches providing moments of the posterior statistical distributions.

BEL1D also incorporates an easy to compute propagation of data noise. This means that it does not need specific training for a given noise to be able to take it into account, nor does it require a specific noise distribution to work. Thus, the training phase of BEL1D can take place before any field measurements, and any knowledge of the noise level on the field data.

The new framework also offers the advantage of being easy to tune to convergence, assuming the prior model space is coherent with the field data (falsification). Indeed, only two parameters are controlling the convergence: the number of models in the initial prior samples and the mixing ratio between the prior samples and the posterior samples in the iterative process. This heavily contrasts with other approaches, determinist or probabilist, where many parameters are used to tune the algorithms to convergence. For example, Neural Networks need a given number of layers with a given number of nodes each, with their activation function. Those parameters need to be fine-tuned to achieve a reasonable training and validation error (e.g. Leung et al., 2003). With the DREAM McMC implementation (Vrugt, 2016), there are twelve parameters (not related to the definition of the prior or the misfit computation) that can vary, leading to different results. Once a set of parameters is obtained, the method is very efficient, but tuning the algorithm can be cumbersome. Even worse, using an incorrect set of parameters, can lead in some cases to erroneous posterior estimations. In contrast, for BEL1D the number of models in the sampled prior can easily be optimized by testing increasing values until stabilization of the posterior. A similar approach can also work for the mixing ratio, governing the IPR behaviour.

Then, within the BEL1D training step, we are able to empirically assess the sensitivity of the experiment towards the model parameters. If the parameters of interest are mainly gathered in the first, highly correlated, reduced dimensions in the canonically correlated space, this indicates a strong sensitivity towards those parameters. On the contrary, when the parameters of interest are gathered in the higher dimensions (lower correlations), this indicates that the experiment design is not effective to retrieve information on those aspects. Since the training phase can be performed without the knowledge of the exact field dataset value, this can help to

design a suitable experiment. This idea is explored in the multi-central configuration versus the coincident configuration section 4.2, where two experimental designs are compared in terms of the recovered posterior model space for the same model. It is relevant to note that the BEL framework is designed to tackle experimental design, as is the case in Thibaut et al. (2021).

Finally, prior falsification is automated within BEL1D. Indeed, when projecting the field data onto the canonically correlated space, the program can, at a very low CPU cost, check that the prior is capable of explaining the dataset. This is a crucial point, since prior falsification is often a complex task that, if not taken into account, can lead to erroneous estimations as would be the case for McMC for example.

6.3 BEL generalization in geophysics

In this manuscript, we applied the new methodology to both sNMR and dispersion curves from surface waves datasets. However, the framework is theoretically not limited to those two methods. We also applied BEL1D to a synthetic case unrelated to geophysics (Appendix A). BEL has also seen applications to multiple types of data and predictions in research during the last years (e.g.: Hermans et al., 2019, Phelps et al., 2018, Thibaut et al., 2021, ...).

The codes for BEL1D are fully available open-source (Michel, 2022). This means that working upon the work provided here is easier for anyone and collaboration is highly encouraged. Building upon open-source forward operators provided by projects such as *pyGIMLi* (Rücker et al., 2017), *SimPEG* (Cockett et al., 2015), *Fatiando a Terra* (Uieda et al., 2013) and *emsig* (Werthmüller, 2017), should broaden the reach of BEL1D. This implementation for 1D methods should be relatively

straightforward.

This indicates that BEL1D could be adapted to any geophysical method. However, the main limitation to this adaptation in the definition of the prior. Indeed, contrary to more classical approaches whose building of the posterior model space is based on a misfit definition, BEL1D does not compute those misfits for the selection of posterior models. In BEL1D, the validity of a given dataset is related to its distance to other models in the canonically correlated space. This has the advantage of being a robust approach: BEL1D cannot provide a model if either the prior model space is not containing models explaining the dataset, or if the forward modelling is not accurate enough (a form of prior falsification). The definition of the prior model space is therefore of the utmost importance. At the current state, the code is written for blocky 1D profiles. This limits the usability of the algorithm to methods for which such models are coherent.

When the field dataset falls outside the canonically correlated space, it is linked to the prior definition and its adequation to the experiment. The most simple case can occur when the prior model space is too optimistic on the prior uncertainty, and the groundtruth is indeed not included in this prior. However, in cases of unbiased prior model spaces that are sufficiently large and assuming that the number of prior samples is sufficient to cover the full prior model space, this should not happen. The reason for the inability of BEL1D to provide a posterior model space in such cases is to find elsewhere.

The prior should be able to accurately represent the geophysical experiment. To do so, both the prior model space description and the applied forward model (also part of the prior hypothesis) must correspond to the experiment at hand. Using the current version of the code, the prior is limited to 1D blocky profiles. This limits applicability to physically blocky experiments. However, depending

on the scale of the experiment, blocky profiles can be an issue. There is thus a need for more versatile prior models definition. Also, the 1D aspect of the current code is a limitation in itself, since geophysical experiments (even if they can often be interpreted in terms of 1D model) are happening in the 3D world. The 1D hypothesis thus neglects or approximates the effects above the first dimension, which can sometimes lead to a misrepresentation of the actual field experiment. For such reasons, prior definition and adequation to the field data is a task that will require further work to enable the applicability of BEL (and BEL1D) more broadly.

6.4 Further improvements of BEL

6.4.1 Defining new priors

From the discussion on the adaptation of BEL1D for other geophysical methods, it is clear that implementing smoothly varying models could help in cases of higher uncertainty on the possible layering or for physically smoothly varying contexts (temperature gradients visualized by electrical imaging for example (Hermans et al., 2015)). An approach to achieve this is to use numerous fixed thickness layers with varying parameters, as is often found in deterministic approaches (e.g.: Müller-Petke et al., 2016). Ideally, the parameters could be correlated one-another, through some sort of geostatistical object like a variogram (e.g.: Omre, 1984). This would avoid the prior to be unrealistically large. However, introducing a prior correlation between the different layers could also result in the introduction of bias.

We test the application of this idea with a synthetic sNMR dataset (see Chapter 2) and fully independent prior parameters (see Figure 6.1). We use 10 layers of 10 meters each. In the tested case, the prior consisted of equivalent uniform distributions for each water contents and relaxation times, with ranges from 2.5% to 30%

and 5ms to 400ms respectively. Here, the benchmark model consists of a single deep aquifer with a smooth increase in both the water content and the relaxation time.

In Figure 6.1, we observe that the algorithm has difficulties to recover the main trend, even though the benchmark model is in the posterior model space. This lack of correlation between parameters in the posterior is due to the unrealistic prior, with water contents and relaxation times varying between their minimum and maximum values between layers.
Currently, this approach is still limiting due to the number of layers and their thicknesses, but this limitation is present because of the uncorrelated nature of the prior, which leads to an unreasonable prior with more layers. From this example, we see that implementing structural correlations in the prior definition is a key aspect to enable such models to work properly. Ideally, the model would have tens to hundreds of layers to be able to model adequately all the different phenomenon and such spaces would be nearly impossible to explore using uncorrelated variables.

Solving for 2D (or 3D) structures is another obvious evolution of the algorithm. To solve for those structures, there are several obstacles that need careful attention. In 1D blocky models, parameters in the prior could be fully uncorrelated, since each layer was representing completely different structures. This can still be the case with simplified views of the 2D/3D world. Drifting away from simplified models, building complex models of the geology is a challenge in itself. There are many techniques to define priors of the geophysical models (Caumon, 2018). In each technique describing the prior, a way to insert a geospatial constraint in the model is required to avoid unrealistic behaviours. This means that contrary to all the examples presented in this manuscript, the model parameters are going to be related one-another in a predefined way. Another possible approach is to use surface-based geophysical modelling (Galley et al., 2020). In this approach, instead of providing

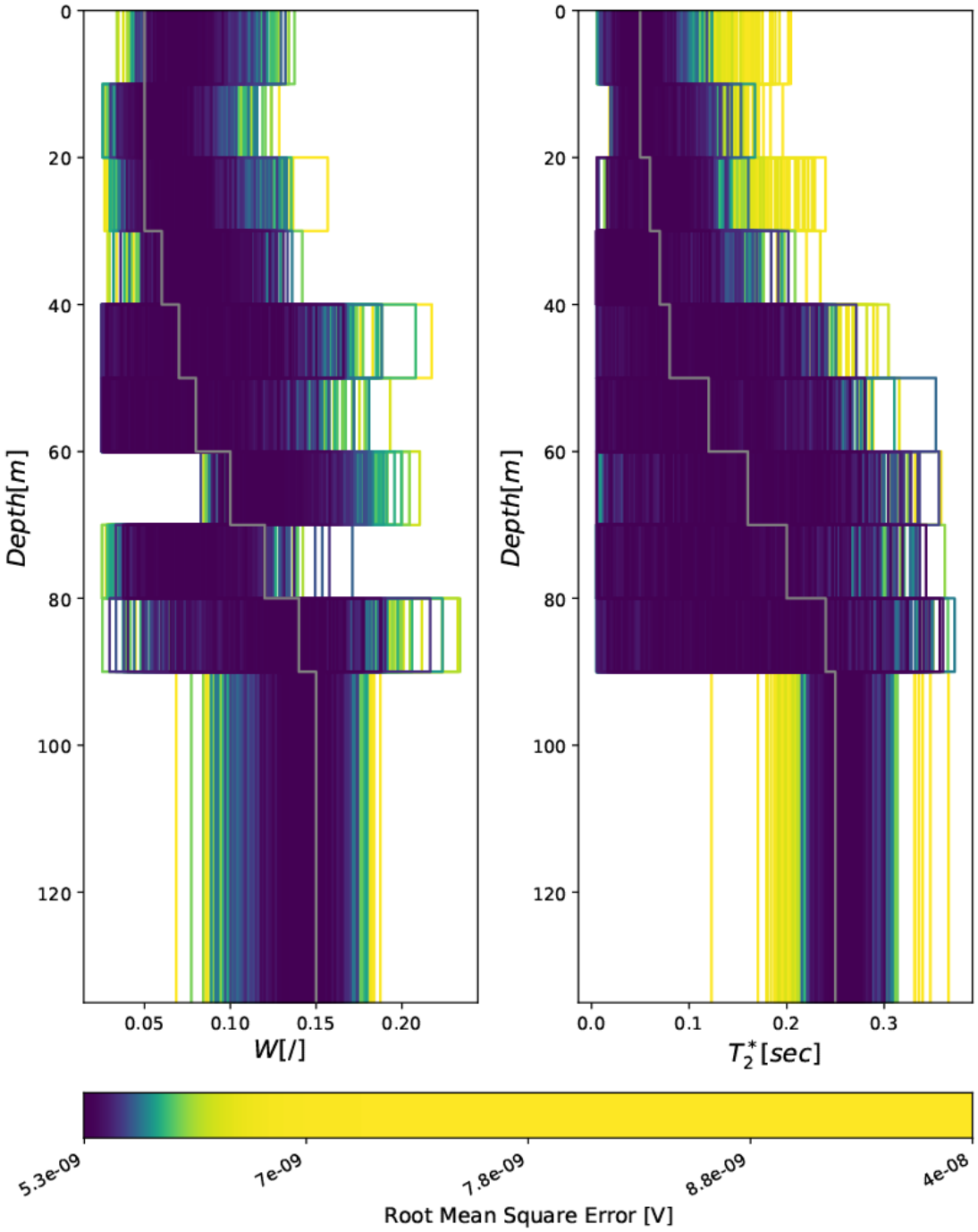

Figure 6.1: Testing 10 fixed thickness layers. The benchmark model is displayed in grey above the models obtained through BEL1D and IPR. The prior model space is represented in grey below the posterior model space. The colour scale reflects the RMSE distributions, with each colour representing the same number of models. The noise level on the dataset is Gaussian with $5.2nV$ in amplitude.

a single value for each cell of a mesh, the mesh is constituted of multiple structures with their own properties. This could help to reduce the number of parameters in the process.

Implementing for other methods in 2D or 3D would also bring more complex forward models to compute (e.g.: Adriaens et al., 2021). This could greatly benefit from the full parallelization of this operation in BEL1D training phase. Using super-computers with clusters of tens to thousands of CPU cores on those complex models could dramatically improve the training phase of the algorithm. Once the training phase is performed, this could lead to a rapid first estimate of the uncertainty on a low power field machine.

An intermediate approach towards 2D/3D models is to use information from close-by 1D profiles to infer a more informed prior. This is classical in geostatistical approaches, where propagating the information from a fixed set of points is required to build a full field of data (Chilès & Delfiner, 2012). Similarly, several inversion processes rely on the position of constraints between lateral models (e.g.: Auken et al., 2005).

Here, the idea would be to mix an initial prior with the posterior from a close-by point to infer an informed prior, similarly to the iterative process from Chapter 5. Doing so, the iterative process might be able to recover a reliable relationship with fewer iterations, leading to computational gains.

We tested this approach with 2 models that vary slightly (the depths to the layers are different) with seismic surface waves dispersion curves data. This situation is similar to what would happen with dipping layers and multiple surveys along a single line.

Results of this experiment are presented in Figure 6.2. The mix between the initial prior and the posterior of the first model is set to respectively 75% and 25%.

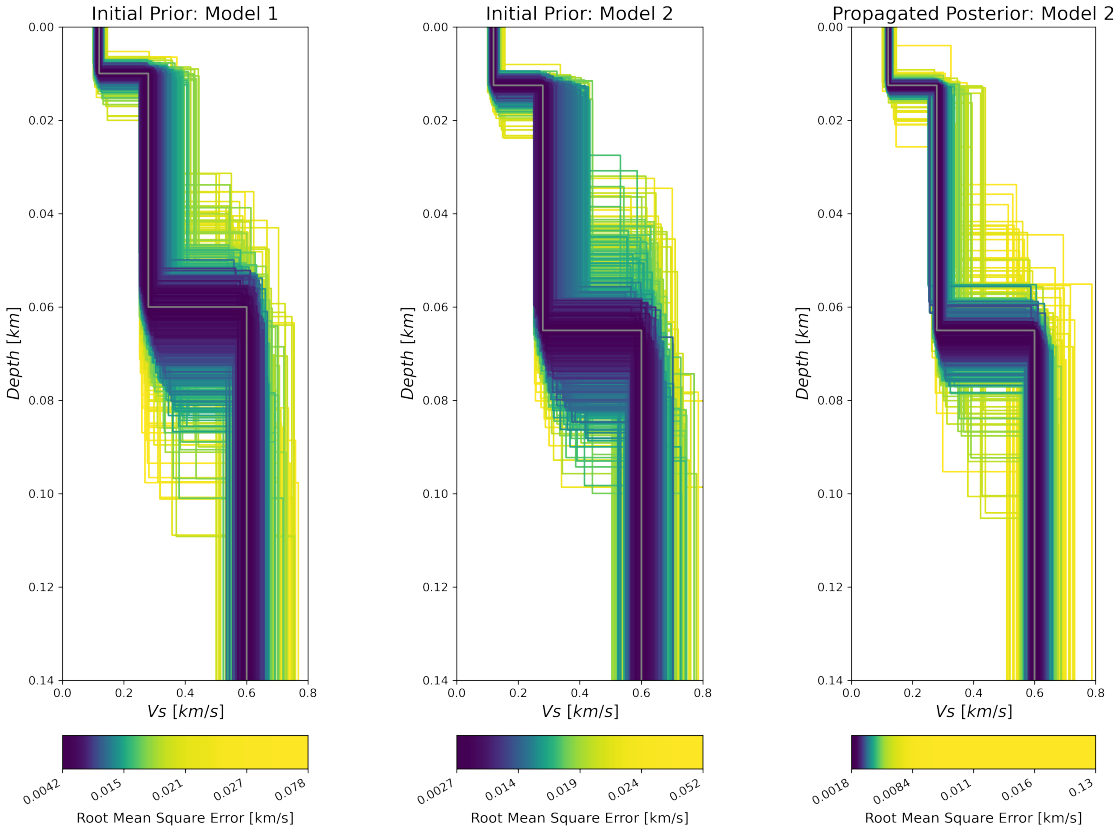

Figure 6.2: Testing the effect of propagating the prior. The initial prior is suitable for both cases. Using the posterior of the first case to inform the second, we reach a similar estimation, except for the fewer models presenting higher RMSE.

First, the obtained posterior models show a lower uncertainty for most of the parameters. However, the difference is mainly in the extreme values, that are highly unlikely. This is expected, since we are introducing biases in the problem by adding information from the first model to the location of the second model. This effect is similar to what is obtained using lateral constraints on 2D models, since the constraints avoid large variations.

This approach is not yet ready for standard use, as parameters governing the algorithm need to be investigated. Since the mixing initial prior/posterior of close-by model is introducing bias, it is important to be able to tune this parameter using an objective criterion (ideally based on the datasets themselves).

6.4.2 Experimental design with BEL in geophysics

Developing approaches for experimental design is also a task that BEL can improve, as is indicated by the multi-central test case (Section 4.2). Developing a way to automatically assess the validity of several experiments to achieve the goal set by the investigators, BEL could improve the field data acquisition itself.

To do so, similar approaches to what is proposed by Thibaut et al. (2021) need to be adapted for geophysical experiments. One could test the sensitivity of different configurations for a given prior model space (reflecting the prior knowledge of the site to experiment), using the correlation of the canonically correlated space of BEL1D as indicators (see Chapter 4) or another method such as the distance-based sensitivity analysis DGSA from J. Park et al. (2016). Then, the experiments can be compared in terms of sensitivity towards the different parameters (or only a selection of parameters), enabling the selection of the best experiment to perform a given task.

One of the advantages of this approach would be that prior samples for the selected experiments will already be drawn. Thus training of BEL1D is a low CPU cost to add.

6.4.3 Exploring the impact of non-linear transforms

Finally, one of the current limitation of BEL1D is its use of linear algorithms for dimensionality reduction and building a strong relationship between models and datasets. Implementing non-linear algorithms for the dimensionality reduction (Lee & Verleysen, 2007), such as kernel PCA or Neural Networks such as Autoencoders (Hinton & Salakhutdinov, 2006) could improve the performances in the case of high dimensionality data, or preserve more of the initial variability while remaining in a

low dimensionality space. The quantification of the interaction between the models and the datasets could also see a conversion to non-linear algorithms (Lai & Fyfe, 2000). Those improvements would result in more complex training phases of the algorithm, but should achieve better relationships between parameters and data points, leading to improved performances of the algorithm.

References

Acar, R., & Vogel, C. R. (1994). Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, *10*, 1217–1229. https://doi.org/10.1088/0266-5611/10/6/003

Adriaens, X., Henrotte, F., & Geuzaine, C. (2021). Adjoint state method for time-harmonic scattering problems with boundary perturbations. *Journal of Computational Physics*, *428*, 109981. https://doi.org/10.1016/J.JCP.2020.109981

Ahmed, A., Michel, H., Deleersnyder, W., Dudal, D., & Hermans, T. (2021). Applying BEL1D for transient electromagnetic sounding inversion. *EGU General Assembly*. https://doi.org/10.5194/egusphere-egu21-1131

Ajo-Franklin, J. B., Minsley, B. J., & Daley, T. M. (2007). Applying compactness constraints to differential traveltime tomography. *Geophysics*, *72*, R67–R75. https://doi.org/10.1190/1.2742496

Aleardi, M., Vinciguerra, A., Stucchi, E., & Hojat, A. (2022). Probabilistic inversions of electrical resistivity tomography data with a machine learning-based forward operator. *Geophysical Prospecting*, *70*, 938–957. https://doi.org/10.1111/1365-2478.13189

Amatyakul, P., Rung-Arunwan, T., & Siripunvaraporn, W. (2015). A pilot magnetotelluric survey for geothermal exploration in mae chan region, northern thailand. *Geothermics*, *55*, 31–38. https://doi.org/10.1016/J.GEOTHERMICS.2015.01.009

Andersen, K. R., Wan, L., Grombacher, D., Lin, T., & Auken, E. (2018). Studies of parameter correlations in surface NMR using the Markov chain Monte Carlo

method. *Near Surface Geophysics*, *16*, 206–217. https://doi.org/10.3997/1873-0604.2017064

Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., Maier-Hein, L., Rother, C., & Köthe, U. (2019). Analyzing inverse problems with invertible neural networks. *ICLR 2019*. http://arxiv.org/abs/1808.04730

Aster, R. C., Borchers, B., & Thurber, C. H. (2013). *Parameter estimation and inverse problems* (2nd). Elsevier Inc. https://doi.org/10.1016/C2009-0-61134-X

Auken, E., Christiansen, A. V., Jacobsen, B. H., Foged, N., & Sørensen, K. I. (2005). Piecewise 1D laterally constrained inversion of resistivity data. *Geophysical Prospecting*, *53*, 497–506. https://doi.org/10.1111/J.1365-2478.2005.00486.X

Auken, E., Christiansen, A. V., Kirkegaard, C., Fiandaca, G., Schamper, C., Behroozmand, A. A., Binley, A., Nielsen, E., Effersø, F., Christensen, N. B., Sørensen, K., Foged, N., & Vignoli, G. (2015). An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. *Exploration Geophysics*, *46*, 223–235. https://doi.org/10.1071/EG13097

Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An introduction to machine learning. *Clinical Pharmacology and Therapeutics*, *107*, 871–885. https://doi.org/10.1002/CPT.1796

Bagdassarov, N. (2021). *Fundamentals of rock physics*. Cambridge University Press. https://doi.org/10.1017/9781108380713

Bauer, K., Muñoz, G., & Moeck, I. (2012). Pattern recognition and lithological interpretation of collocated seismic and magnetotelluric models using self-organizing maps. *Geophysical Journal International*, *189*, 984–998. https://doi.org/10.1111/J.1365-246X.2012.05402.X

Behroozmand, A. A., Auken, E., Fiandaca, G., & Rejkjaer, S. (2016). Increasing the resolution and the signal-to-noise ratio of magnetic resonance sounding data using a central loop configuration. *Geophysical Journal International*, *205*, 243–256. https://doi.org/10.1093/GJI/GGW004

Behroozmand, A. A., Keating, K., & Auken, E. (2015). A review of the principles and applications of the NMR technique for near-surface characterization. *Surveys in Geophysics*, *36*, 27–85. https://doi.org/10.1007/S10712-014-9304-0

Bhowmick, D., Gupta, D. K., Maiti, S., & Shankar, U. (2019). Stacked autoencoders based machine learning for noise reduction and signal reconstruction in geophysical data. *arVix*. https://doi.org/10.48550/arXiv.1907.03278

Binley, A., & Kemna, A. (2005). DC resistivity and induced polarization methods. In Y. Rubin & S. Hubbard (Eds.), *Hydrogeophysics* (pp. 129–156, Vol. 50). Springer Netherlands. https://doi.org/10.1007/1-4020-3102-5_5

Binley, A., & Slater, L. (2020). *Resistivity and induced polarization : Theory and applications to the near-surface earth*. Cambridge University Press.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Clarendon Press.

Boaga, J., Vignoli, G., & Cassiani, G. (2011). Shear wave profiles from surface wave inversion: The impact of uncertainty on seismic site response analysis. *Journal of Geophysics and Engineering*, *8*, 162–174. https://doi.org/10.1088/1742-2132/8/2/004

Bobe, C., Vijver, E. V. D., Keller, J., Hanssens, D., Meirvenne, M. V., & Smedt, P. D. (2020). Probabilistic one-dimensional inversion of frequency-domain electromagnetic data using a Kalman ensemble generator. *IEEE Transactions on Geoscience and Remote Sensing, 58*, 3287–3297. https://doi.org/10.1109/TGRS.2019.2953004

Böhning, D. (1992). Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics 1992 44:1*, *44*, 197–200. https://doi.org/10.1007/BF00048682

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, 144–152. https://doi.org/10.1145/130385

Bowman, A. W., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis : The kernel approach with s-plus illustrations*. Clarendon Press; Oxford University Press.

Bressan, T. S., de Souza, M. K., Girelli, T. J., & Junior, F. C. (2020). Evaluation of machine learning methods for lithology classification using geophysical data. *Computers & Geosciences*, *139*, 104475. https://doi.org/10.1016/J.CAGEO.2020.104475

Brogi, C., Huisman, J. A., Pätzold, S., von Hebel, C., Weihermüller, L., Kaufmann, M. S., van der Kruk, J., & Vereecken, H. (2019). Large-scale soil mapping using multi-configuration EMI and supervised image classification. *Geoderma*, *335*, 133–148. https://doi.org/10.1016/J.GEODERMA.2018.08.001

Campbell, A., Fryer, A., & Wakeman, S. (2005). Vertical seismic profiles—more than just a corridor stack. *The Leading Edge*, *24*, 694–697. https://doi.org/10.1190/1.1993259

Caumon, G. (2018). Geological objects and physical parameter fields in the subsurface: A review. In B. Daya Sagar, Q. Cheng & F. Agterberg (Eds.), *Handbook of mathematical geosciences: Fifty years of iamg* (pp. 567–588). Springer International Publishing. https://doi.org/10.1007/978-3-319-78999-6_28

Chen, J., Kemna, A., & Hubbard, S. S. (2008). A comparison between Gauss-Newton and Markov-chain Monte Carlo - based methods for inverting spectral induced-polarization data for Cole-Cole parameters. *Geophysics*, *73*. https://doi.org/10.1190/1.2976115

Chen, Y., & Oliver, D. S. (2011). Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences 2011 44:1*, *44*, 1–26. https://doi.org/10.1007/S11004-011-9376-Z

Chilès, J.-P., & Delfiner, P. (2012). *Geostatistics : Modeling spatial uncertainty*. John Wiley & Sons.

Christensen, C. W., Pfaffhuber, A. A., Skurdal, G. H., Lysdahl, A. O. K., & Vöge, M. (2020). Large scale & efficient geotechnical soil investigations: Applying machine learning on airborne geophysical models to map sensitive glaciomarine clay. *Proceedings of the 6th International Conference on Geotechnical and Geophysical Site Characterization, Budapest, Hungary*, 7–11.

Christensen, N. B., Reid, J. E., & Halkjaer, M. (2009). Fast, laterally smooth inversion of airborne time-domain electromagnetic data. *Near Surface Geophysics*, 599–612.

Christiansen, A. V., Auken, E., & Sørensen, K. (2008). The transient electromagnetic method. In R. Kirsch (Ed.), *Groundwater geophysics* (pp. 179–226). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88405-7_6

Clark, J. A., & Page, R. (2011). Inexpensive geophysical instruments supporting groundwater exploration in developing nations. *Journal of Water Resource and Protection*, *03*, 768–780. https://doi.org/10.4236/jwarp.2011.310087

Cockett, R., Kang, S., Heagy, L. J., Pidlisecky, A., & Oldenburg, D. W. (2015). Simpeg: An open source framework for simulation and gradient based parameter estimation in geophysical applications. *Computers & Geosciences*, *85*, 142–154. https://doi.org/10.1016/J.CAGEO.2015.09.015

Constable, S. C., Parker, R. L., & Constable, C. G. (1987). Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *GEOPHYSICS*, *52*, 289–300.

Conway, D., Alexander, B., King, M., Heinson, G., & Kee, Y. (2019). Inverting magnetotelluric responses in a three-dimensional earth using fast forward approximations based on artificial neural networks. *Computers and Geosciences*, *127*, 44–52. https://doi.org/10.1016/j.cageo.2019.03.002

Cox, B. R., & Teague, D. P. (2016). Layering ratios: A systematic approach to the inversion of surface wave data in the absence of a priori information. *Geophysical Journal International*, *207*, 422–438. https://doi.org/10.1093/GJI/GGW282

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, *20*, 215–232. https://doi.org/10.1111/J.2517-6161.1958.TB00292.X

Cox, D. R., Newton, A. M., & Huuse, M. (2020). *An introduction to seismic reflection data: Acquisition, processing and interpretation* (N. Scarselli, J. Adam, D. Chiarella, D. G. Roberts & A. W. Bally, Eds.; 2nd). https://doi.org/10.1016/B978-0-444-64134-2.00020-1

Cox, L. H., Wilson, G. A., & Zhdanov, M. S. (2010). 3D inversion of airborne electromagnetic data using a moving footprint. *Exploration Geophysics*, *41*, 250–259. https://doi.org/10.1071/EG10003

Cracknell, M. J., & Reading, A. M. (2013). The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines. *Geophysics*, *78*, WB113–WB126. https://doi.org/10.1190/geo2012-0411.1

Devilee, R. J. R., Curtis, A., & Roy-Chowdhury, K. (1999). An efficient, probabilistic neural network approach to solving inverse problems' inverting surface wave velocities for eurasian crustal thickness. *JOURNAL OF GEOPHYSICAL RESEARCH*, *104*, 841–869. https://doi.org/10.1029/1999JB900273

Devroy, L. (1986). *Non-uniform random variate generation*. Springer-Verlag.

Di, H., Gao, D., & Alregib, G. (2019). Developing a seismic texture analysis neural network for machine-aided seismic pattern recognition and classification. *Geophysical Journal International*, *218*, 1262–1275. https://doi.org/10.1093/GJI/GGZ226

Digital Science. (2018-). *Dimensions* [Software]. https://app.dimensions.ai

Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2019). Unsupervised learning based on artificial neural network: A review. *2018 IEEE International Conference on Cyborg and Bionic Systems, CBS 2018*, 322–327. https://doi.org/10.1109/CBS.2018.8612259

Dosne, A.-G., Bergstrand, M., Harling, K., & Karlsson, M. O. (2016). Improving the estimation of parameter uncertainty distributions in nonlinear mixed effects models using sampling importance resampling. *Journal of Pharmacokinetics and Pharmacodynamics 2016 43:6*, *43*, 583–596. https://doi.org/10.1007/S10928-016-9487-8

Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics*, *61*, 1–55. https://doi.org/10.1016/BS.AGPH.2020.08.002

Emerick, A. A., & Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, *55*, 3–15. https://doi.org/10.1016/J.CAGEO.2012.03.011

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, *99*, 10143–10162. https://doi.org/10.1029/94JC00572

Farquharson, C. G., & Craven, J. A. (2009). Three-dimensional inversion of magnetotelluric data for mineral exploration: An example from the McArthur river uranium deposit, Saskatchewan, Canada. *Journal of Applied Geophysics*, *68*, 450–458. https://doi.org/10.1016/J.JAPPGEO.2008.02.002

Farquharson, C. G., Oldenburg, D. W., & Routh, P. S. (2003). Simultaneous 1D inversion of loop–loop electromagnetic data for magnetic susceptibility and electrical conductivity. *Geophysics*, *68*, 1857–1869. https://doi.org/10.1190/1.1635038

Flinchum, B. A., Holbrook, W. S., & Carr, B. J. (2022). What do P-Wave velocities tell us about the critical zone? *Frontiers in Water*, *3*, 187. https://doi.org/10.3389/FRWA.2021.772185

Gaikwad, S., Pawar, N. J., Bedse, P., Wagh, V., & Kadam, A. (2021). Delineation of groundwater potential zones using vertical electrical sounding (VES) in a complex bedrock geological setting of the west coast of India. *Modeling Earth Systems and Environment*, *8*, 2233–2247. https://doi.org/10.1007/S40808-021-01223-3

Galley, C. G., Lelièvre, P. G., & Farquharson, C. G. (2020). Geophysical inversion for 3D contact surface geometry. *https://doi.org/10.1190/geo2019-0614.1*, *85*, K27–K45. https://doi.org/10.1190/GEO2019-0614.1

García-Jerez, A., Piña-Flores, J., Sánchez-Sesma, F. J., Luzón, F., & Perton, M. (2016). A computer code for forward calculation and inversion of the H/V spectral ratio under the diffuse field assumption. *Computers and Geosciences*, *97*, 67–78. https://doi.org/10.1016/j.cageo.2016.06.016

Garofalo, F., Foti, S., Hollender, F., Bard, P. Y., Cornou, C., Cox, B. R., Dechamp, A., Ohrnberger, M., Perron, V., Sicilia, D., Teague, D., & Vergniault, C. (2016). InterPACIFIC project: Comparison of invasive and non-invasive methods for seismic site characterization. Part II: Inter-comparison between surface-wave and borehole methods. *Soil Dynamics and Earthquake Engineering*, *82*, 241–254. https://doi.org/10.1016/J.SOILDYN.2015.12.009

Garofalo, F., Foti, S., Hollender, F., Bard, P. Y., Cornou, C., Cox, B. R., Ohrnberger, M., Sicilia, D., Asten, M., Giulio, G. D., Forbriger, T., Guillier, B., Hayashi, K., Martin, A., Matsushima, S., Mercerat, D., Poggi, V., & Yamanaka, H. (2016). InterPACIFIC project: Comparison of invasive and non-invasive methods for seismic site characterization. Part I: Intra-comparison of surface wave methods. *Soil Dynamics and Earthquake Engineering*, *82*, 222–240. https://doi.org/10.1016/J.SOILDYN.2015.12.010

Giannakis, I., Giannopoulos, A., & Warren, C. (2019). A machine learning-based fast-forward solver for ground penetrating radar with application to full-waveform inversion. *IEEE Transactions on Geoscience and Remote Sensing*, *57*, 4417–4426. https://doi.org/10.1109/TGRS.2019.2891206

Gilson, M., Briers, P., Ruthy, I., & Dassargues, A. (2017). *Carte hydrogéologique de wallonie - Elsenborn- Langert - Dreiherrenwald (50/3-4; 50a/1)*. SPW.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gorban, A. N., Kégl, B., Wunsch, D. C., & Zinovyev, A. Y. (Eds.). (2008). *Principal manifolds for data visualization and dimension reduction* (Vol. 58). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73750-6

Gottschalk, I., & Knight, R. (2022). The development of a machine-learning approach to construct a field-scale rock-physics transform. *Geophysics*, *87*, MR35–MR48. https://doi.org/10.1190/GEO2020-0811.1

Grana, D., Azevedo, L., & Liu, M. (2020). A comparison of deep machine learning and Monte Carlo methods for facies classification from seismic data. *Geophysics*, *85*, WA41–WA52. https://doi.org/10.1190/GEO2019-0405.1

Green, R. (1974). The seismic refraction method—a review. *Geoexploration*, *12*, 259–284. https://doi.org/10.1016/0016-7142(74)90015-5

Grombacher, D., Liu, L., Griffiths, M. P., Vang, M., & Larsen, J. J. (2021). Steady-state surface NMR for mapping of groundwater. *Geophysical Research Letters*, *48*, e2021GL095381. https://doi.org/10.1029/2021GL095381

Guo, J., Li, Y., Jessell, M. W., Giraud, J., Li, C., Wu, L., Li, F., & Liu, S. (2021). 3D geological structure inversion from Noddy-generated magnetic data using deep learning methods. *Computers and Geosciences*, *149*. https://doi.org/10.1016/J.CAGEO.2021.104701

Hansen, T. M., & Cordua, K. S. (2017). Efficient Monte Carlo sampling of inverse problems using a neural network-based forward-applied to gpr crosshole traveltime inversion. *Geophysical Journal International*, *211*, 1524–1533. https://doi.org/10.1093/GJI/GGX380

Hanssens, D., Delefortrie, S., Bobe, C., Hermans, T., & Smedt, P. D. (2019). Improving the reliability of soil EC-mapping: Robust apparent electrical conductiv-

ity (rECa) estimation in ground-based frequency domain electromagnetics. *Geoderma*, *337*, 1155–1163. https://doi.org/10.1016/j.geoderma.2018.11.030

Heagy, L. J., Cockett, R., Kang, S., Rosenkjaer, G. K., & Oldenburg, D. W. (2017). A framework for simulation and inversion in electromagnetics. *Computers & Geosciences*, *107*, 1–19. https://doi.org/10.1016/J.CAGEO.2017.06.018

Hearst, M., Dumais, S., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18–28. https://doi.org/10.1109/5254.708428

Hermans, T., Compaire, N., Thibaut, R., & Lesparre, N. (2021). Bayesian evidential learning: An alternative to hydrogeophysical coupled inversion. *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*, *2021-September*, 3125–3129. https://doi.org/10.1190/segam2021-3580979.1

Hermans, T., Lesparre, N., Schepper, G. D., & Robert, T. (2019). Bayesian evidential learning: A field validation using push-pull tests. *Hydrogeology Journal*, *27*, 1661–1672. https://doi.org/10.1007/S10040-019-01962-9

Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A., & Caers, J. (2018). Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using Bayesian Evidential Learning. *Water Resources Research*, *54*, 2931–2948. https://doi.org/10.1002/2017WR022135

Hermans, T., Oware, E., & Caers, J. (2016). Direct prediction of spatially and temporally varying physical properties from time-lapse electrical resistance data. *Water Resources Research*, *52*, 7262–7283. https://doi.org/10.1002/2016WR019126

Hermans, T., Wildemeersch, S., Jamin, P., Orban, P., Brouyère, S., Dassargues, A., & Nguyen, F. (2015). Quantitative temperature monitoring of a heat

tracing experiment using cross-borehole ert. *Geothermics*, *53*, 14–26. https://doi.org/10.1016/J.GEOTHERMICS.2014.03.013

Herrmann, R. B. (2013). Computer programs in seismology: An evolving tool for instruction and research. *Seismological Research Letters*, *84*, 1081–1088. https://doi.org/10.1785/0220110096

Hertrich, M., Braun, M., Günther, T., Green, A. G., & Yaramanci, U. (2007). Surface nuclear magnetic resonance tomography. *IEEE Transactions on Geoscience and Remote Sensing*, *45*, 3752–3759. https://doi.org/10.1109/TGRS.2007.903829

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*, 504–507. https://doi.org/10.1126/SCIENCE.1127647

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*, 417–441. https://doi.org/10.1037/H0071325

Hurich, C., & Deemer, S. (2013). Combined surface and borehole seismic imaging in a hard rock terrain: A field test of seismic interferometry. *Geophysics*, *78*, B103–B110. https://doi.org/10.1190/GEO2012-0325.1

Isunza Manrique, I., Caterina, D., Dumont, M., Benoît, M., Antoine, M., Pauline, K., & Nguyen, F. (2022). Characterization of past metalurgical residues using geophysical imaging: A case study of Duferco site (Belgium). *6th Symposium on circular economy and urban mining (SUM2022)*.

Isunza Manrique, I., Caterina, D., Hermans, T., & Nguyen, F. (2019). Probabilistic joint interpretation of geoelectrical and passive source seismic data for landfill characterization. *AGU Fall Meeting*.

Jeong, C., Mukerji, T., & Mariethoz, G. (2017). A fast approximation for seismic inverse modeling: Adaptive spatial resampling. *Mathematical Geosciences 2017 49:7*, *49*, 845–869. https://doi.org/10.1007/S11004-017-9693-Y

Jessell, M., Guo, J., Li, Y., Lindsay, M., Scalzo, R., Giraud, J., Pirot, G., Cripps, E., & Ogarko, V. (2022). Into the Noddyverse: A massive data store of 3D geological models for machine learning and inversion applications. *Earth System Science Data*, *14*, 381–392. https://doi.org/10.5194/ESSD-14-381-2022

Jha, M. K., Kumar, S., & Chowdhury, A. (2008). Vertical electrical sounding survey and resistivity inversion using genetic algorithm optimization technique. *Journal of Hydrology*, *359*, 71–87. https://doi.org/10.1016/j.jhydrol.2008.06.018

Jordi, C., Doetsch, J., Günther, T., Schmelzbach, C., & Robertsson, J. O. (2018). Geostatistical regularization operators for geophysical inverse problems on irregular meshes. *Geophysical Journal International*, *213*, 1374–1386. https://doi.org/10.1093/GJI/GGY055

Joshi, D., Patidar, A. K., Mishra, A., Mishra, A., Agarwal, S., Pandey, A., Dewangan, B. K., & Choudhury, T. (2021). Prediction of sonic log and correlation of lithology by comparing geophysical well log data using machine learning principles. *GeoJournal*, 1–22. https://doi.org/10.1007/S10708-021-10502-6

Justesen, N., Bontrager, P., Togelius, J., & Risi, S. (2020). Deep learning for video game playing. *IEEE Transactions on Games*, *12*, 1–20. https://doi.org/10.1109/TG.2019.2896986

Kaipio, J. P., & Somersalo, E. (2005). *Statistical and computational inverse problems* (Vol. 160). Springer-Verlag. https://doi.org/10.1007/b138659

Kemna, A., Kulessa, B., & Vereecken, H. (2002). Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models. *Journal of Hydrology*, *267*, 125–146. https://doi.org/10.1016/S0022-1694(02)00145-2

Kennett, P., Ireson, R. L., & Conn, P. J. (1980). Vertical seismic profiles: Their applications in exploration geophysics. *Geophysical Prospecting*, *28*, 676–699. https://doi.org/10.1111/J.1365-2478.1980.TB01254.X

Kim, Y., & Nakata, N. (2018). Geophysical inversion versus machine learning in inverse problems. *The Leading Edge*, *37*, 894–901. https://doi.org/10.1190/TLE37120894.1

Kirkby, A., Heinson, G., Holford, S., & Thiel, S. (2015). Mapping fractures using 1D anisotropic modelling of magnetotelluric data: A case study from the otway basin, victoria, australia. *Geophysical Journal International*, *201*, 1961–1976. https://doi.org/10.1093/GJI/GGV116

Kirsch, R., & Yaramanci, U. (2008). Geoelectrical methods. In R. Kirsch (Ed.), *Groundwater geophysics* (pp. 85–117). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88405-7_3

Koefoed, O. (1979). *Geosounding principles, 1 : Resistivity sounding measurements*. Elsevier Scientifc Pub. Co.

Kramer, O. (2013). K-nearest neighbors. In *Dimensionality reduction with unsupervised nearest neighbours* (pp. 13–23). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38652-7_2

Kremer, T., Müller-Petke, M., Michel, H., Dlugosch, R., Irons, T., Hermans, T., & Nguyen, F. (2020). Improving the accuracy of 1D surface nuclear magnetic resonance surveys using the multi-central-loop configuration. *Journal of Applied Geophysics*, *177*. https://doi.org/10.1016/j.jappgeo.2020.104042

Kremer, T., Irons, T., Müller-Petke, M., & Larsen, J. J. (2022). Review of acquisition and signal processing methods for electromagnetic noise reduction and retrieval of surface nuclear magnetic resonance parameters. *Surveys in Geophysics*, *43*, 999–1053. https://doi.org/10.1007/S10712-022-09695-3

Krzanowski, W. (2000). *Principles of multivariate analysis* (2nd). Oxford University Press.

LaBrecque, D. J., Miletto, M., Daily, W., Ramirez, A., & Owen, E. (1996). The effects of noise on Occam's inversion of resistivity tomography data. *Geophysics*, *61*, 538–548. https://doi.org/10.1190/1.1443980

Lai, P. L., & Fyfe, C. (2000). Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, *10*, 365–377. https://doi.org/10.1142/S012906570000034X

Laloy, E., Hérault, R., Jacques, D., & Linde, N. (2018). Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, *54*, 381–406. https://doi.org/10.1002/2017WR022148

Laloy, E., Hérault, R., Lee, J., Jacques, D., & Linde, N. (2017). Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Advances in Water Resources*, *110*, 387–405. https://doi.org/10.1016/j.advwatres.2017.09.029

Larki, E., Tanha, A. A., Parizad, A., Soulgani, B. S., & Bagheri, H. (2021). Investigation of quality factor frequency content in vertical seismic profile for gas reservoirs. *Petroleum Research*, *6*, 57–65. https://doi.org/10.1016/J.PTLRS.2020.10.002

Lavrentev, M., Romanov, V., & Shishatski, S. (1986). *Ill-posed problems of mathematical physics and analysis* (Vol. 64). American Mathematical Society.

Lawrence, N. D. (2012). A unifying probabilistic perspective for spectral dimensionality reduction: Insights and new models. *Journal of Machine Learning Research*, *13*, 1609–1638.

Lawyer, L. C. (, Bates, C. C., & Rice, R. B. (2001). *Geophysics in the affairs of mankind: A personalized history of exploration geophysics*. Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560801788

Lebedev, S., Adam, J., & Meier, T. (2013). Mapping the Moho with seismic surface waves: A review, resolution analysis, and recommended inversion strategies. *Tectonophysics*, *609*, 377–394. https://doi.org/10.1016/J.TECTO.2012.12.030

Lee, J., & Verleysen, M. (2007). *Nonlinear dimensionality reduction*. Springer.

Legault, J. (2015). Airborne electromagnetic systems: State of the art and future directions. *CSEG Recorder*, 38–49.

Legchenko, A., Vincent, C., Baltassat, J. M., Garambois, S., Thibert, E., Gilbert, A., Descloitres, M., Girard, J. F., Gagliardini, O., & Guyard, H. (2012). Monitoring of water accumulation in the Tête Rousse glacier (French Alps) using 3D magnetic resonance imaging. *Near Surface Geoscience*, cp-306–00007. https://doi.org/10.3997/2214-4609.20143314

Legchenko, A., Baltassat, J. M., Duwig, C., Boucher, M., Girard, J. F., Soruco, A., Beauce, A., Mathieu, F., Legout, C., Descloitres, M., & Patricia, F. A. G. (2020). Time-lapse magnetic resonance sounding measurements for numerical modeling of water flow in variably saturated media. *Journal of Applied Geophysics*, *175*, 103984. https://doi.org/10.1016/J.JAPPGEO.2020.103984

Lesparre, N., Nguyen, F., Kemna, A., Robert, T., Hermans, T., Daoudi, M., & Flores-Orozco, A. (2017). A new approach for time-lapse data weighting in

electrical resistivity tomography. *Geophysics*, *82*, E325–E333. https://doi.org/10.1190/GEO2017-0024.1

Leung, F. H., Lam, H. K., Ling, S. H., & Tam, P. K. (2003). Tuning of the structure and parameters of a neural network using an improved genetic algorithm. *IEEE Transactions on Neural Networks*, *14*, 79–88. https://doi.org/10.1109/TNN.2002.804317

Li, J., Lu, X., Farquharson, C. G., & Hu, X. (2018). A finite-element time-domain forward solver for electromagnetic methods with complex-shaped loop sources. *Geophysics*, *83*, E117–E132. https://doi.org/10.1190/GEO2017-0216.1

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, *36*, 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2

Linde, N., Ginsbourger, D., Irving, J., Nobile, F., & Doucet, A. (2017). On uncertainty quantification in hydrogeology and hydrogeophysics. *Advances in Water Resources*, *110*, 166–181. https://doi.org/10.1016/J.ADVWATRES.2017.10.014

Linde, N., Renard, P., Mukerji, T., & Caers, J. (2015). Geological realism in hydrogeological and geophysical inverse modeling: A review. *Advances in Water Resources*, *86*, 86–101. https://doi.org/10.1016/j.advwatres.2015.09.019

Lindsey, J. K. (1996). *Parametric statistical inference*. Clarendon Press.

Loke, M., Acworth, I., & Dahlin, T. (2003). A comparison of smooth and blocky inversion methods in 2d electrical imaging surveys. *Exploration Geophysics*, *34*, 182–187. https://doi.org/10.1071/EG03182

Lopez-Alvis, J., Nguyen, F., Looms, M. C., & Hermans, T. (2022). Geophysical inversion using a variational autoencoder to model an assembled spatial prior

uncertainty. *Journal of Geophysical Research: Solid Earth*, *127*, e2021JB022581. https://doi.org/10.1029/2021JB022581

Lopez-Alvis, J., Laloy, E., Nguyen, F., & Hermans, T. (2021). Deep generative models in inversion: The impact of the generator's nonlinearity and development of a new approach based on a variational autoencoder. *Computers & Geosciences*, *152*, 104762. https://doi.org/10.1016/J.CAGEO.2021.104762

Mackie, R. L., & Madden, T. R. (1993). Three-dimensional magnetotelluric inversion using conjugate gradients. *Geophysical Journal International*, *115*, 215–229. https://doi.org/10.1111/J.1365-246X.1993.TB05600.X

Maiti, S., Erram, V. C., Gupta, G., Tiwari, R. K., Kulkarni, U. D., & Sangpal, R. R. (2013). Assessment of groundwater quality: A fusion of geochemical and geophysical information via Bayesian neural networks. *Environmental Monitoring and Assessment*, *185*, 3445–3465. https://doi.org/10.1007/s10661-012-2802-y

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, *151*, 675–688. https://doi.org/10.1046/J.1365-246X.2002.01847.X

Mariethoz, G., Renard, P., & Caers, J. (2010). Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research*, *46*, 11530. https://doi.org/10.1029/2010WR009274

Marzan, I., Martí, D., Lobo, A., Alcalde, J., Ruiz, M., Alvarez-Marron, J., & Carbonell, R. (2021). Joint interpretation of geophysical data: Applying machine learning to the modeling of an evaporitic sequence in Villar de Cañas (Spain). *Engineering Geology*, *288*, 106126. https://doi.org/10.1016/J.ENGGEO.2021.106126

Matsushima, J., Ali, M. Y., & Bouchaala, F. (2016). Seismic attenuation estimation from zero-offset VSP data using seismic interferometry. *Geophysical Journal International*, *204*, 1288–1307. https://doi.org/10.1093/GJI/GGV522

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*, 239. https://doi.org/10.2307/1268522

Meier, R., & Rix, G. (1993). An initial study of surface wave inversion using artificial neural networks. *Geotechnical Testing Journal*, *16*, 425–431. https://doi.org/10.1520/GTJ10282J

Meier, U., Curtis, A., & Trampert, J. (2007). Global crustal thickness from neural network inversion of surface wave data. *Geophysical Journal International*, *169*, 706–722. https://doi.org/10.1111/J.1365-246X.2007.03373.X

Michel, H. (2020). BEL1D: Matlab toolboxes for the application of Bayesian Evidential Learning 1D imaging. https://github.com/hadrienmichel/BEL1D

Michel, H. (2022). pyBEL1D: A Python implementation of BEL1D. https://doi.org/10.5281/ZENODO.6833249

Michel, H., Hermans, T., Kremer, T., & Nguyen, F. (2021). Application of BEL1D for sNMR data interpretation. *MRS2021 - 8th International Workshop on Magnetic Resonance*, 59–62.

Michel, H., Hermans, T., & Nguyen, F. (2022). Iterative Prior Resampling and rejection sampling to improve 1D geophysical imaging based on Bayesian Evidential Learning (BEL1D). *Geophysical Journal International*. https://doi.org/10.1093/GJI/GGAC372

Michel, H., Hermans, T., Kremer, T., & Nguyen, F. (2020). Improving BEL1D accuracy for geophysical imaging of the subsurface. *SEG Technical Program*

Expanded Abstracts, *2020-October*, 3562–3566. https://doi.org/10.1190/SEGAM2020-3426204.1

Michel, H., Nguyen, F., Kremer, T., Elen, A., & Hermans, T. (2020). 1D geological imaging of the subsurface from geophysical data with Bayesian Evidential Learning. *Computers & Geosciences*, *138*, 104456. https://doi.org/10.1016/j.cageo.2020.104456

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Moghadas, D., Jadoon, K. Z., & McCabe, M. F. (2019). Spatiotemporal monitoring of soil moisture from EMI data using dct-based bayesian inference and neural network. *Journal of Applied Geophysics*, *169*, 226–238. https://doi.org/10.1016/J.JAPPGEO.2019.07.004

Moss, R. (2008). Quantifying measurement uncertainty of thirty-meter shear-wave velocity. *Bulletin of the Seismological Society of America*, *98*, 1399–1411. https://doi.org/10.1785/0120070101

Mosser, L., Dubrule, O., & Blunt, M. J. (2020). Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *Mathematical Geosciences*, *52*, 53–79. https://doi.org/10.1007/s11004-019-09832-6

Mreyen, A. S., Cauchie, L., Micu, M., Onaca, A., & Havenith, H. B. (2021). Multiple geophysical investigations to characterize massive slope failure deposits: Application to the Balta rockslide, Carpathians. *Geophysical Journal International*, *225*, 1032–1047. https://doi.org/10.1093/GJI/GGAB028

Müller-Petke, M., Braun, M., Hertrich, M., Costabel, S., & Walbrecker, J. (2016). MRSmatlab — a software tool for processing, modeling, and inversion of magnetic resonance sounding data. *Geophysics*, *81*, WB9–WB21. https://doi.org/10.1190/GEO2015-0461.1

Müller-Petke, M., & Yaramanci, U. (2010). QT inversion - comprehensive use of the complete surface NMR data set. *Geophysics*, *75*. https://doi.org/10.1190/1.3471523

Nabighian, M. (1987). *Electromagnetic methods in applied geophysics: Volume 1, theory*. Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802631

Nabighian, M. (1991). *Electromagnetic methods in applied geophysics: Volume 2, application, parts a and b*. Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802686

Nabighian, M. N., & Macnae, J. C. (1991). Chapter 6. Time domain electromagnetic prospecting methods. In M. N. Nabighian (Ed.), *Electromagnetic methods in applied geophysics: Volume 2, application, parts a and b* (pp. 427–520, Vol. 2). Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802686.CH6

Nguyen, F., Kemna, A., Robert, T., & Hermans, T. (2016). Data-driven selection of the minimum-gradient support parameter in time-lapse focused electric imaging. *Geophysics*, *81*, A1–A5. https://doi.org/10.1190/GEO2015-0226.1

Noble, W. S. (2006). What is a support vector machine? *NATURE BIOTECHNOLOGY*, *24*. https://doi.org/10.1038/nbt1206-1565

Omre, H. (1984). The variogram and its estimation. In G. Verly, M. David, A. Journel & A. Marechal (Eds.), *Geostatistics for natural resources characterization* (pp. 107–125). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-3699-7_7

Osisanwo, F., Akinsola, J., Awodele, O., & Hinmikaiye, J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal*

of Computer Trends and Technology (IJCTT), *48*, 128–138. https://doi.org/10.14445/22312803/IJCTT-V48P126

Park, C. B., Miller, R. D., & Xia, J. (1999). Multichannel analysis of surface waves. *GEOPHYSICS*, *64*.

Park, J., & Caers, J. (2020). Direct forecasting of global and spatial model parameters from dynamic data. *Computers & Geosciences*, *143*. https://doi.org/10.1016/J.CAGEO.2020.104567

Park, J., Yang, G., Satija, A., Scheidt, C., & Caers, J. (2016). Dgsa: A matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments. *Computers and Geosciences*, *97*, 15–29. https://doi.org/10.1016/j.cageo.2016.08.021

Patro, P. K. (2017). Magnetotelluric studies for hydrocarbon and geothermal resources: Examples from the asian region. *Surveys in Geophysics*, *38*, 1005–1041. https://doi.org/10.1007/S10712-017-9439-X

Phelps, G., Scheidt, C., & Caers, J. (2018). Exploring viable geologic interpretations of gravity models using distance-based global sensitivity analysis and kernel methods. *Geophysics*, *83*, G79–G92. https://doi.org/10.1190/GEO2017-0742.1

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*, 1347–1358. https://doi.org/10.1056/NEJMRA1814259

Rücker, C., Günther, T., & Wagner, F. M. (2017). pyGIMLi: An open-source library for modelling and inversion in geophysics. *Computers and Geosciences*, *109*, 106–123. https://doi.org/10.1016/j.cageo.2017.07.011

Russell, B. (2019). Machine learning and geophysical inversion — a numerical study. *The Leading Edge*, *38*, 512–519. https://doi.org/10.1190/TLE38070512.1

Satija, A., & Caers, J. (2015). Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. *Advances in Water Resources*, *77*, 69–81. https://doi.org/10.1016/j.advwatres.2015.01.002

Satija, A., Scheidt, C., Li, L., & Caers, J. (2017). Direct forecasting of reservoir performance using production data without history matching. *Computational Geosciences 2017 21:2*, *21*, 315–333. https://doi.org/10.1007/S10596-017-9614-7

Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying uncertainty in subsurface systems* (Vol. 182). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119325888

Scheidt, C., Renard, P., & Caers, J. (2015). Prediction-focused subsurface modeling: Investigating the need for accuracy in flow-based inverse modeling. *Mathematical Geosciences*, *47*, 173–191. https://doi.org/10.1007/S11004-014-9521-6

Scheiter, M., Valentine, A., & Sambridge, M. (2022). Upscaling and downscaling Monte Carlo ensembles with generative models. *Geophysical Journal International*, *230*, 916–931. https://doi.org/10.1093/GJI/GGAC100

Schmitt, D. R. (2015). Geophysical properties of the near surface earth: Seismic properties. In G. Schubert (Ed.), *Treatise on geophysics* (Second Edition, pp. 43–87, Vol. 11). Elsevier. https://doi.org/10.1016/B978-0-444-53802-4.00190-1

Schubert, G. (2015). *Treatise on geophysics* (Second Edition). Elsevier.

Siahkoohi, A., Louboutin, M., & Herrmann, F. J. (2019). The importance of transfer learning in seismic modeling and imaging. *Geophysics*, *84*, A47–A52. https://doi.org/10.1190/GEO2019-0056.1

Siemon, B., Auken, E., & Christiansen, A. V. (2009). Laterally constrained inversion of helicopter-borne frequency-domain electromagnetic data. *Journal of*

Applied Geophysics, *67*, 259–268. https://doi.org/10.1016/J.JAPPGEO.2007.11.003

Smedt, P. D., Meirvenne, M. V., Saey, T., Baldwin, E., Gaffney, C., & Gaffney, V. (2014). Unveiling the prehistoric landscape at Stonehenge through multi-receiver EMI. *Journal of Archaeological Science*, *50*, 16–23. https://doi.org/10.1016/J.JAS.2014.06.020

Socco, L. V., Foti, S., & Boiero, D. (2010). Surface-wave analysis for building near-surface velocity models - established approaches and new perspectives. *Geophysics*, *75*. https://doi.org/10.1190/1.3479491

Spies, B. R., & Frischknecht, F. C. (1991). Chapter 5. Electromagnetic sounding. In M. N. Nabighian (Ed.), *Electromagnetic methods in applied geophysics: Volume 2, application, parts a and b* (pp. 285–425, Vol. 2). Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802686.CH5

Sun, B., & Alkhalifah, T. (2020). ML-descent: An optimization algorithm for full-waveform inversion using machine learning. *Geophysics*, *85*, R477–R492. https://doi.org/10.1190/GEO2019-0641.1

Suthaharan, S. (2016). Decision tree learning. In *Machine learning models and algorithms for big data classification* (pp. 237–269). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7641-3_10

Tarantola, A. (2006). Popper, bayes and the inverse problem. *Nature Physics 2006 2:8*, *2*, 492–494. https://doi.org/10.1038/nphys375

Tezkan, B. (1999). A review of environmental applications of quasi-stationary electromagnetic techniques. *Surveys in Geophysics 1999 20:3*, *20*, 279–308. https://doi.org/10.1023/A:1006669218545

Tezkan, B. (2008). Radiomagnetotellurics. In R. Kirsch (Ed.), *Groundwater geophysics* (pp. 295–317). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88405-7_10

Thas, O. (2010). Comparing distributions. https://doi.org/10.1007/978-0-387-92710-7

Thibaut, R., Laloy, E., & Hermans, T. (2021). A new framework for experimental design using Bayesian Evidential Learning: The case of wellhead protection area. *Journal of Hydrology*, *603*, 126903. https://doi.org/10.1016/J.JHYDROL.2021.126903

Trainor-Guitton, W., & Hoversten, G. M. (2011). Stochastic inversion for electromagnetic geophysics: Practical challenges and improving convergence efficiency. *Geophysics*, *76*. https://doi.org/10.1190/GEO2010-0223.1

Uieda, L., Jr, V. C. O., & Barbosa, V. C. F. (2013). Modeling the Earth with Fatiando a Terra. In S. van der Walt, J. Millman & K. Huff (Eds.), *Proceedings of the 12th Python in Science Conference* (pp. 92–98). https://doi.org/10.25080/Majora-8b375195-010

Ulrych, T. J., Sacchi, M. D., & Woodbury, A. (2001). A Bayes tour of inversion: A tutorial. *GEOPHYSICS*, *66*, 55–69. https://doi.org/10.1190/1.1444923

Valentine, A., & Sambridge, M. (2017). Using machine learning to accelerate sampling-based inversion. *AGU Fall Meeting 2017*.

Van De Vijver, E. (2017). *Proximal soil sensing in the context of urban (re)development : An evaluation of multi-receiver electromagnetic induction and stepped-frequency ground penetrating radar at landfills and industrial sites*.

van Leeuwen, J., & Evenson, G. (1996). Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, *124*, 2898–2913.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, *11*, 3371–3408.

Vozoff, K. (1991). Chapter 8: The magnetotelluric method. In M. N. Nabighian (Ed.), *Electromagnetic methods in applied geophysics: Volume 2, application, parts a and b* (pp. 641–712, Vol. 2). Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802686.CH8

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/J.ENVSOFT.2015.08.013

Wallet, B. C., & Hardisty, R. (2019). Unsupervised seismic facies using gaussian mixture models. *Interpretation*, *7*, SE93–SE111. https://doi.org/10.1190/INT-2018-0119.1

Wand, M. P., & Jones, M. C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association*, *88*, 520. https://doi.org/10.2307/2290332

Wastiaux, C., & Schumacker, R. (2003). *Topographie de surface et de subsurface des zones tourbeuses des réserves naturelles domaniales des Hautes-Fagnes (No. C60/1-2-3)*. Université de Liège, Service de phytosociologie, flore et végétations des Hautes-Fagnes.

Wastiaux, C. (2008). Les tourbières sont-elles des éponges régularisant l'écoulement ? *Bulletin de la Société géographique de Liège*, *50*, 57–66.

Wathelet, M. (2008). An improved neighborhood algorithm: Parameter conditions and dynamic scaling. *Geophysical Research Letters*, *35*, 9301. https://doi.org/10.1029/2008GL033256

Wei, M., Zhang, Y., Fan, T., & Lin, T. (2021). Noise attenuation of magnetic resonancesounding signal using fully convolutional networks. *8th International Workshop on Magnetic Resonance*, 81–83.

Werthmüller, D. (2017). An open-source full 3D electromagnetic modeler for 1D VTI media in Python: Empymod. *GEOPHYSICS*, *82*, WB9–WB19. https://doi.org/10.1190/geo2016-0626.1

White, D. J. (1989). Two-dimensional seismic refraction tomography. *Geophysical Journal International*, *97*, 223–245. https://doi.org/10.1111/J.1365-246X.1989.TB00498.X

Xia, J., Miller, R. D., & Park, C. B. (1999). Estimation of near-surface shear-wave velocity by inversion of rayleigh waves. *https://doi.org/10.1190/1.1444578*, *64*. https://doi.org/10.1190/1.1444578

Yin, Z., Strebelle, S., & Caers, J. (2020). Automated Monte Carlo-based quantification and updating of geological uncertainty with borehole data (autobel v1.0). *Geoscientific Model Development*, *13*, 651–672. https://doi.org/10.5194/GMD-13-651-2020

Yu, S., Ma, J., & Wang, W. (2019). Deep learning for denoising. *Geophysics*, *84*, V333–V350. https://doi.org/10.1190/GEO2018-0668.1

Zhang, H. (2004). The optimality of naive bayes. In V. Barr & Z. Markov (Eds.), *Proceedings of the seventeenth international florida artificial intelligence research society conference* (pp. 562–567). AAAI Press.

Zhang, X., & Curtis, A. (2021). Bayesian geophysical inversion using invertible neural networks. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1029/2021JB022320

Zhao, J., Gao, J., Wang, D., & Zhang, M. (2014). Q-factor and velocity inversion from zero-offset VSP data. *Journal of Applied Geophysics*, *101*, 51–67. https://doi.org/10.1016/J.JAPPGEO.2013.11.005

Zhdanov, M. S. (2009). Electromagnetic properties of rocks and minerals. In *Methods in geochemistry and geophysics* (pp. 395–447, Vol. 43). Elsevier. https://doi.org/10.1016/S0076-6895(08)00210-2

Zonge, K. L., & Hughes, L. J. (1991). Chapter 9: Controlled source audio-frequency magnetotellurics. In M. N. Nabighian (Ed.), *Electromagnetic methods in applied geophysics: Volume 2, application, parts a and b* (pp. 713–809, Vol. 2). Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802686.ch9

Zou, C., Zhao, L., Xu, M., Chen, Y., & Geng, J. (2021). Porosity prediction with uncertainty quantification from multiple seismic attributes using random forest. *Journal of Geophysical Research: Solid Earth*, *126*. https://doi.org/10.1029/2021JB021826

Appendix A

Testing BEL1D on a synthetic mind experiment: oscillations of a pendulum

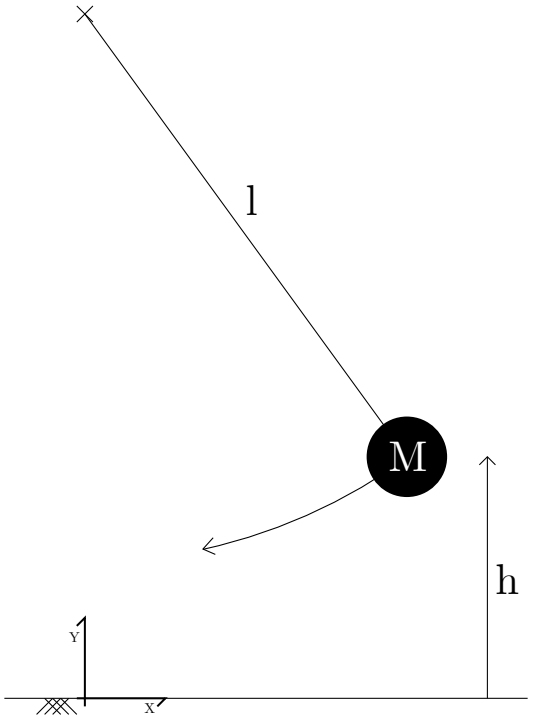

Figure A.1: Illustration of the pendulum experiment

In this purely theoretical case, we will demonstrate the ubiquity of BEL1D to solve inverse problems with a limited number of unknown parameters as encountered

in 1D geophysical data. We use a case not related to geosciences but a classical example of college physics. This is a rather easy way to show the strength of the BEL1D, as most of the readers will be familiar with the problem.

Imagine a pendulum, of unknown point mass (M) and unknown length (l), that has been launched from an unknown height (h) with a null velocity (Figure A.1). The height of the pendulum centre (H) is 10 m. Our experiment consists of taking pictures of the pendulum to determine the bob's Y position every second for 1min50sec, beginning 10 seconds after the launch of the bob (at null velocity). We will assume that the image processing induces a Gaussian uncertainty on the measured position with a standard deviation equal to 0.1 m.

The forward model is described by the system of differential equations presented in Equation A.1 where $\theta(t)$ is the angular position of the pendulum, $\theta_t(t)$ the velocity of the bob and $\omega^2 = \sqrt{g/l}$, the natural frequency of the pendulum. The different parameters of interest for the problem are all related to the value of $\theta(t)$, at the exception of the mass (M) which does not appear in the equations.

$$\begin{cases} \frac{\partial}{\partial t}\theta(t) = \theta_t(t) \\ \frac{\partial}{\partial t}\theta_t(t) = -\omega^2 \sin\left(\theta(t)\right) \end{cases} \tag{A.1}$$

For the presented example, the values of the parameters are:

- $l = 3m$
- $h = 7.5m$
- $M = 40kg$

The prior model space is defined by uniformly distributed variables that are within the ranges described in Table A.1, with the additional constraint that the sampled models should correspond to the criteria stated in Equation A.2.

$$l + h >= H \tag{A.2}$$

This latter condition guarantees that the pendulum is physically consistent.

	l[m]	h[m]	M[kg]
Minimum	1	1	0
Maximum	9	9	50

Table A.1: Prior model space for the pendulum model

BEL1D was applied with 10000 prior realizations in order to produce 10000 models in the posterior model space. From the 111 dimensions in the original data space (one point every second for 110 seconds), PCA enabled a reduction to 37 dimensions (while keeping at least 90% of the variance explained). In CCA model space, the first two dimensions were mainly representing the length (l) and the height of launch (h). The third dimension showed no correlation and represented mainly the mass (M). To constitute the posterior, we ignored the effect of noise.

The results of BEL1D on this particular experiment showed that the values of the length of the pendulum as well as the height of launch were reasonably solved (Figure A.2). However, the distributions in mass for the prior and posterior model spaces are similar. This is due to the intrinsic insensitivity of the experiment to the mass. It is also observed that the height of launch (h) and the length of the pendulum (l) are highly correlated in the posterior model space. Moreover, even if the real value of the length is relatively small, hence less likely than larger values in the prior model space, the BEL1D process is still able to produce an accurate posterior distribution for this parameter. Some processing of these results enables to retrieve an estimation of the probability distribution function of the launch position in the (X, Y) plane (Figure A.3, right). This estimation is very accurate, given the real initial position is located in the zone of the highest probability. Analysing the

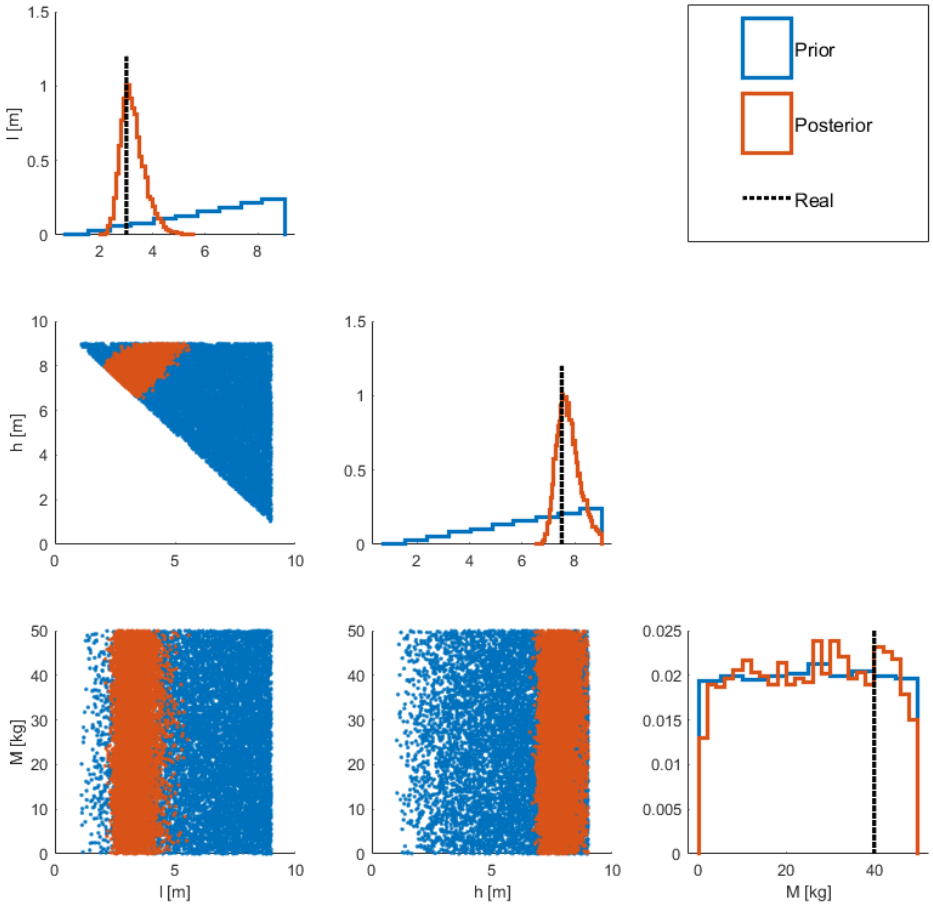

Figure A.2: Results of BEL modelling on the pendulum experiment. The prior space is presented in blue and the posterior in orange. The benchmark values are represented by the dashed black lines in the histogram plots.

RMS error on the simulated data from the posterior (Figure A.3, left) showed that the lowest RMS are observed closer to the real initial position, however with a larger scattering.

The data distributions showed the reduction from the prior model space to the posterior compared to the benchmark data (Figure A.4). It is observed that the reduction is substantial and that the posterior data are centred on the true data. In order to better reduce the uncertainty on the models, the experiment could

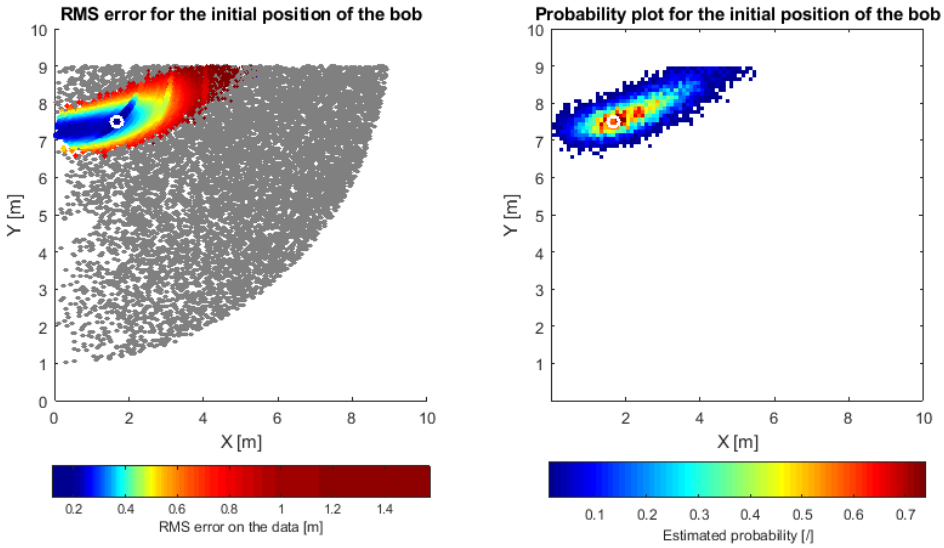

Figure A.3: Results of the BEL modelling on the pendulum experiment: (left) RMS error as a function of the initial position and (right) estimation of the probability distribution for the initial position of the bob. The true initial position is represented by the white circle.

be improved, either by considering the X component of the position, refining the time span, or reducing the noise on the data (better image processing). This is not performed here, since the example already demonstrates the ubiquity of the proposed algorithm which was the main objective.

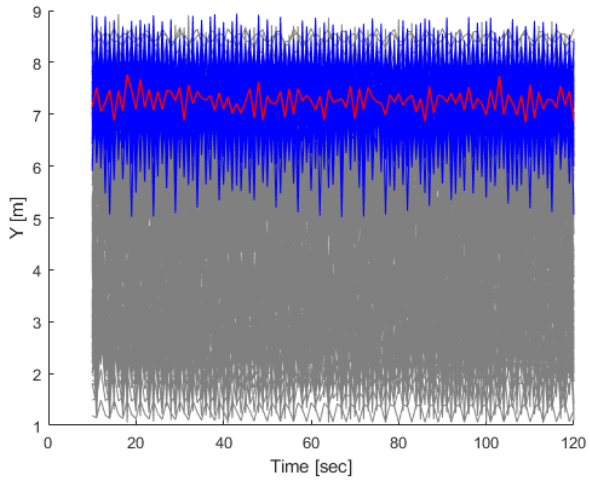

Figure A.4: Pendulum experiment, prior (grey) and posterior (blue) data spaces compared to the true values (red).