UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Sciences de la Terre, de l'Univers et de l'Environnement**

Arrêté ministérial : 7 Août 2006

Présentée par

Paul John WELLINGTON

Thèse dirigée par **Jean VIRIEUX, et Romain BROSSIER**
et codirigée par **Stéphane GARAMBOIS**

préparée au sein **Institut des Sciences de la Terre**
et de **Terre Univers Environnement**

Efficient 1D, 2D and 3D Geostatistical constraints and their application to Full Waveform Inversion

Date de la soutenance: 22nd September, 2016 ,
devant le jury composé de :

Satish Singh
Professeur Institut de Physique du Globe de Paris, France, Président
Thibaut Allemand
Chercheur, CGG, France, Examinateur
Thomas Bohlen
Professeur Karlsruher Institut für Technologie , Karlsruhe, Allemagne, Rapporteur
Romain Brossier
Maître de Conférence au ISTerre, Université Grenoble Alpes, France, Directeur de thèse
Nobuaki Fuji
Maître de conférences,Institut de Physique du Globe de Paris, France, Examinateur
Stéphane Garambois
Maître de Conférence au ISTerre, Université Grenoble Alpes, France, Directeur de thèse
Jean Virieux
Professeur à l'ISTerre, Université Grenoble Alpes, France, Directeur de thèse

Acknowledgements

First of all I would like to thank my supervisors: Jean Virieux, Romain Brossier and Stéphane Garambois. I came to ISTerre looking for PhD opportunities after a long gap from being at University (more than 5 years). Of all the universities I spoke to, the people in the SEISCOPE group at ISTerre were the most optimistic and encouraging for me to pursue further academic study. The fact that they took a chance with me and offered me a position started what is best described as an intense ride. Throughout the process when things weren't working (which was often) or when I was lost (also quite often) they provided the perfect motivation for me. They wouldn't "spoon-feed" me the total solution, giving me tips, but at the same time, never saying I was incapable and that I just needed to go and think or look at problems from another angle. After 3 years this interaction has significantly improved my technical confidence and understanding, with the added bonus we made what appears to be a nice breakthrough, just in time, right near the end of my PhD.

I would like to thank the reviewers of my PhD manuscript (Thomas Bohlen, Satish Singh, Thibaut Allemand and Nobuaki Fuji) for taking the time to read my manuscript and providing insightful comments, questions and ideas for future work. I would like to thank the technical contributions from others who were not directly my supervisors. Firstly, I will give mention to the other senior researchers, Ludovic Metiviér, Stephane Operto and Michel Dietrich who provided technical support and administrative assistance along the way. Amir Asnaashari who helped me understand his work with significant patience and gave me an introduction to FWI at the start of my PhD. Okba Hamitou who proved to be a great mate, "office buddy" and collaborator in the second half of my PhD. When supervisors were unavailable Okba helped by taking an interest in understanding my work as it unfolded. He proved to be a useful "second-eye" for problems. Phuong-Thu Trinh, who started her PhD near the end of mine was always great to talk to and later became one of the strongest collaborators of my career. She went through the methodology of my PhD with a fine tooth-comb and explained the bits I put down down to "magic" with analytical mathematics. In the end she got, what I hope is an effective filter for her applications and I got some invaluable mathematical insights in to the "magic" of the filters described in this manuscript. A general thank you to all other members of the SEISCOPE consortium for providing ideas and discussions at meetings and technical events.

It would be remiss for me not to mention my good friends at ISTerre. There was many, too numerous to mention. I appreciate your friendship and patience when I behaved like a "mood-ring" being worn by my PhD topic. Mood swings were not uncommon from me (mainly related to whether things were working or not) and I appreciate the patience of my friends and other times their iron fist telling me to get over whatever was wearing me down. I thank my friends and family from back home who had to deal with the same thing, but, via teleconference. A special mention goes out to those who lived with me or put up with me on a more extended basis (Linda, Etienne, Natalie, Andrea, Gosia, Sergio and Ashley). I would also like to acknowledge my ski buddies: Natalie, Eric, Frans, Elle, René, Louis, Alain, Isa, Anne, Katya, Rita, Maor and Igor (among others). Enjoying the mountains around Grenoble in winter is something that will be one of my most treasured memories in the future. Hitting the slopes was also the perfect remedy to a clouded head.

Finally I would like to thank the sponsors of the SEISCOPE consortium, giving special mention to CGG (Gilles Lambaré, Diego Carotti, Ronan, Sablon and Thibaut Allemand) who provided data and assistance during the PhD. The administrative staff at ISTerre and ISSO for helping with the practicalities of living and studying in France and nurturing my still slowly blossoming French. The whole PhD experience has been a ride and one I am glad I had the courage to undertake.

I sincerely apologize if I have forgotten anyone, but there was many helpful people who made this PhD a reality.

Résumé

L'inversion de forme d'onde complète (FWI) est un processus non-linéaire et mal posé d'ajustement de données, dans notre cas, issues d'acquisitions sismiques. Cette technique cherche à reconstruire, à partir d'un modèle initial obtenu à faible nombre d'onde (faible résolution), des paramètres constitutifs contrôlant la propagation des ondes à grands nombres d'ondes (forte résolution). Durant ce processus itératif, certains artéfacts peuvent altérer la qualité du modèle reconstruit. Afin de diminuer ces artéfacts et d'assurer une reconstruction des paramètres qui soit cohérente d'un point de vue géologique, différentes techniques de pré-conditionnement ou de régularisation peuvent être proposées.

Cette thèse se focalise sur le potentiel de nouveaux filtres multi-dimensionnels construits dans l'espace des nombres d'ondes et orientés suivant les structures géologiques. Une stratégie de pré-conditionnement a été mise au point à l'aide de ces filtres et a été appliquée avec succès à la problématique FWI. La formulation analytique 1D de l'opérateur inverse de covariance laplacienne (Tarantola, 2005) constitue la base de la formulation d'opérateurs de dimension supérieure qui sont validés ici en les comparants avec l'opérateur analytique de covariance laplacienne 1D. Nous avons utilisé cette fonction analytique inverse 1D comme la base de filtrage de dimension supérieure, via l'addition de multiples fonctions inverses orientées orthogonalement. Ces fonctions laplaciennes inverses additionnelles (AIL) sont obtenues pour des configurations 2D et 3D après discrétisation par des techniques de différences finies. Nous montrons que l'on peut calculer un filtre en nombre d'onde de manière rapide et robuste en résolvant le système linéaire associé à ces opérateurs inverses. Lorsque des pentes sont inclues à l'étape de discrétisation par différences finies, il est alors possible d'utiliser ces opérateurs comme des filtres en nombre d'ondes orientés vers les structures géologiques, ceci avec une grande efficacité.

Ce filtre (AIL) montre des propriétés rapides de convergence et des performances indépendantes du vecteur à filtrer. Nous montrons notamment comment ce filtre peut être utilisé comme un opérateur utile pour le gradient associé à la FWI. Le pré-conditionnement du gradient peut atténuer les effets du problème mal-posé qui vont s'étendre dans l'espace des modèles. Deux exemples synthétiques (Valhall et Marmousi) calculés dans l'espace des fréquences sont proposés dans cette thèse. Le pré-conditionnement AIL s'avère efficace pour atténuer d'une part la signature mal-posée provenant de la présence de bruit ambient dans les données observées et d'autre part d'artéfacts liés aux effets de repliement

spatial liés aux conditions d'imagerie par FWI. La possibilité d'inclure des pentes permet de filtrer de manière préférentielle en considérant des pendages géologiques. Cette stratégie de filtrage permet l'atténuation d'artéfacts, tout en préservant le contenu en nombre d'ondes de la stratigraphie orthogonale au pendage.

Un cas réel d'inversion 2D FWI est finalement abordé permettant tout d'abord d'illustrer la sensibilité des résultats d'inversion au modèle initial. Celui-ci est d'importance majeure, particulièrement dans les régions profondes dépassant la pénétration maximale des ondes transmises. L'application de la technique FWI à cette acquisition sismique a permis d'améliorer de manière significative la cohérence sur une image migrée par renversement du temps (RTM). Nous montrons également que le pré-conditionneur AIL permet une décroissance significative du nombre de tirs requis à modéliser dans la boucle d'inversion, sans pour autant dégrader le contenu en nombre d'onde des structures géologiques principales dans les résultats finaux obtenus après inversion.

Abstract

Full waveform inversion (FWI) is a non-linear, ill-posed, local data fitting technique. FWI looks to moves from an initial, low-wavenumber representation of the earth parameters to a broadband representation. During this iterative process a number of undesirable artifacts can map into our model parameter reconstruction. To mitigate these artifacts and to ensure a geologically consistent model parameter reconstruction, various preconditioning and/or regularization strategies have been proposed.

This thesis details the construction of new, efficient, multi-dimensional, structurally-orientated wavenumber filters. A preconditioning strategy has been devised using these filters that we have successfully applied to FWI. The 1D analytical inverse Laplacian covariance operator of Tarantola (2005) forms the basis of higher dimensional operators and is initially validated by comparing to the 1D analytical Laplacian covariance operator. We use this analytical 1D inverse function as a basis for higher dimensional filtering via the addition of multiple, orthogonally orientated inverse functions. These additive inverse Laplacian functions (AIL) are shown in 2D and 3D configurations and are discretized using finite-difference techniques. We show that one can calculate, a rapid and robust wavenumber filter, by solving the linear system associated with these inverse operators. When dip is included at the finite difference discretization stage, it is possible to use these operators as highly efficient, structurally orientated wavenumber filters.

The AIL filter is shown to be rapid to converge and its performance is independent of the vector to be filtered. We show, that the filter can be a useful preconditioning operator for the FWI gradient. Preconditioning the gradient can mitigate against ill-posed effects mapping into the model-space. Two synthetic (Valhall and Marmousi) frequency domain FWI example are shown in this thesis. The AIL preconditioner has success at mitigating the ill-posed imprint coming from ambient noise in the observed data and also artifacts from spatial aliasing effects in the FWI imaging condition. The ability to include dip, allows one to preferentially filter along geological dip with anisotropic correlation operators. This filtering strategy allows the mitigation of artifacts, while simultaneously preserving the stratigraphic based wavenumber content that is orthogonal to dip.

A 2D, real data FWI case-study is also shown and we highlight the sensitivity of

the inversion result to the initial model. The initial model is of key importance, this is especially true in the areas deeper than the maximum penetration of transmitted waves. The application of FWI on this line is able to significantly improve gather alignment on a RTM image. We also see that the AIL preconditioner can allows us to significantly decrease the number of shot records we are required to model in our inversion workflow without degrading the key geological wavenumber content in the final inversion result.

Contents

Wave imaging techniques aim to gain insight about the properties of a medium whether it be a small rock sample, the womb of a pregnant mother or the structure of the earth. The "data" is gathered in a non-invasive way, via an array of receivers and represents the interaction of mechanical waves propagating within the media, from active or passive source as they travel towards the receivers. In the context of this PhD I will be focusing on the use of seismic waves to illuminate oil and gas reservoirs.

The importance of the seismic method for the oil and gas sector:
The acquisition of seismic measurements to help make decisions in the oil and gas sector has been around for many years. An article in Geophysics just after the USA enter World War II entitled *"How can Geophysicists best serve"* (Peacock, 1942) implored petroleum geophysicists to not go off to fight the war on the battlefield. Peacock instead suggested that those trained in geophysics were of key importance for US military dominance and the eventual allied victory, due to the dire need for oil to fuel the war machine. The war machine of world war two already relied on oil and was built not just from soldiers, but ships, trucks, tanks and air-crafts. With the war increasing the demand for oil, he reasoned, that if supply issues were encountered, it could have disastrous effects for the chances of victory. Statistics from the time highlighted the importance of geophysics on the successful placement of "Wild cat" exploration wells (Table 1). Explorers that used geophysics to look for structural traps, had a higher rate of success than those based on purely geology or on non-technical methods of positioning the well. He highlighted the key geophysical method of importance stating:
"Of all the proved methods the reflection seismograph offers the greatest benefit from detailed study and this is one direction in which effort should be made".

		Geology		Geophysics		G&G		No Technical	
		Dry	Prod.	Dry	Prod.	Dry	Prod.	Dry	Prod.
1938	Totals	1041	192	252	78	69	31	535	44
	Ratios	5.4		3.2		2.2		12.2	
1939	Totals	1006	135	400	69	40	13	666	43
	Ratios	7.5		5.8		3.1		15.5	
1940	Totals	1204	198	446	100	81	22	797	35
	Ratios	6.1		4.5		3.7		22.8	
1941	Totals	1380	300	437	143	111	28	771	30
	Ratios	4.6		3.1		4.0		25.7	
1938-1941	**Ratios**	**5.6**		**3.9**		**3.2**		**18.2**	

Table 1: The success (Prod.) to failure (dry) statistics for Wild cat wells in the USA 1938-1941. Wells constrained by geophysics had the best success rates, outperforming those that were positioned using no technical expertise by 400%.

To this day, the seismic method remains the key method for characterizing the sub-

surface structure. This is especially the case in offshore environments and away from well control. The UK North Sea provides a compelling historical example of the role of seismic technologies. When international exploration commenced in the late 1950s and early 1960s, the commercial exploration effort was highly challenged due to the level of uncertainty in the subsurface. From this period to the present day millions of kilometres of seismic data was acquired. The seismic data proved to be a key tool in understanding the complex geology that host the commercial oil and gas fields in the region(Glennie, 1998). The seismic acquisition methods themselves also developed from analogue to digital; single fold to 1000's of fold; and from sparse 2D to dense, azimuth-rich 3D.

The every increasing volume of information required an advancement in the methods of processing of seismic data, so that interpretations and inferences of the geology could be made feasible. The dense 3D acquisitions of the 1970s allowed detailed images of the subsurface to be constructed from the back scattered waves. The exploding-reflector concept was used to drive a family imaging techniques, that provide high wave-number images of the subsurface (Claerbout, 1971, 1976). These "imaging" workflows consist of a familiar two step process. Firstly, the construction of a kinematic macro-model, followed by amplitude projection, using a variety of different migration algorithms (Claerbout and Doherty, 1972; Gazdag, 1978; Stolt, 1978; Baysal et al., 1983; Yilmaz, 2001; Biondi and Symes, 2004). One of the key limitations of such workflows were their insensitivity to the intermediate wavelengths (Jannane et al., 1989). This insensitivity meant that there was a gap in the earth wavenumber reconstruction where the macro-model provided the low wavenumbers and the reflectivity images the higher wavenumbers (Figure 1). Such velocity analysis and migration techniques could efficiently produce good subsurface images and were sufficient for simple geology with limited heterogeneity. In more complex settings such as sub-salt, sub-basalt, below overburden anomalies and in thrust/faulted geological settings, the macro-model construction was quite troublesome. Although more robust approaches exist (Woodward et al., 2008) for attempting to update the low wavenumber velocity macro-model (Docherty et al., 2003) they still have limited sensitivity to the low and intermediate wave-numbers of the velocity structure present in the subsurface.

The importance of FWI:
The full waveform inversion approach (Lailly, 1983; Tarantola, 1984) was introduced as a feasible alternative to provide a wider wave-number coverage. The migration imaging principle (Claerbout, 1971, 1976) was recast as a local optimization problem, where the goal is to minimize the misfit between recorded and modelled seismic data. This non-linear, ill posed problem could be linearized using the Born approximation (Born, 1926). If the incident and back propagated wavefields were cross-correlated it was shown this yielded the first derivative of the misfit function, known as the FWI gradient. This gradient, provided the direction the subsurface parameters need to be updated so that the information contained in the modelled shot records were consistent with the true data. First and second order local descent based optimization strategies (see Nocedal, 1980,
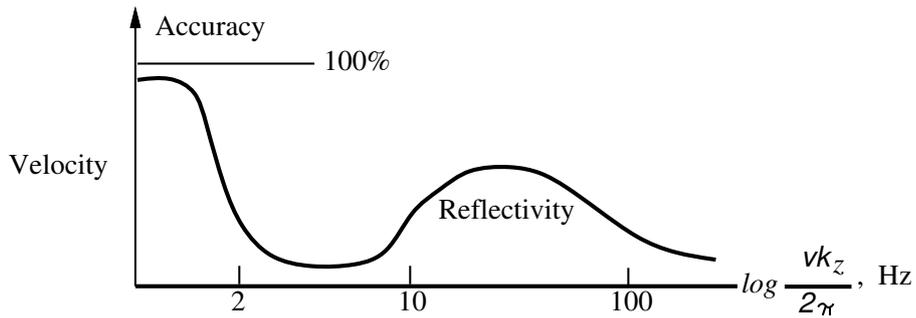
Figure 1: The different wave-number reconstruction of the subsurface from the velocity macro-model (low wavenumbers) vs reflectivity information (high wavenumbers). There is a gap in the intermediate wave-number region. *Taken from (Claerbout, 1985).*

for a review) have been applied where the gradient is also combined with an efficient approximation of the second derivative of the misfit function (Hessian) to develop an iterative inversion strategy. An example of the match between true and modelled shot records before and after FWI taken from the real data case study presented in this thesis is shown in Figure 2.

The immediate adoption of FWI in the 1990s and early 2000s, did not come to pass due a number of reasons. Two of the most dominant factors were the high cost of the forward modelling techniques at the time (Gauthier et al., 1986) and limitations in the acquisition configurations FWI was utilized in. The initial work on FWI was typically focused on the short spread reflections which were in-sensitive the intermediate wavelengths. The application of FWI in the presence of long offset acquisitions allowed waves more sensitive to the low and intermediate wavenumbers such as diving waves and supercritical reflections to contribute to the data fitting. This improved the optimization process and significantly improved results and the prospects of FWI (Mora, 1987, 1988; Pratt and Worthington, 1990; Pratt, 1999). The successful application of FWI to real 3D applications (Warner et al., 2007; Vigh and Starr, 2008; Ben Hadj Ali et al., 2008; Plessix, 2009; Sirgue et al., 2010; Prieux et al., 2011; Operto et al., 2015) coincided with a significant increase in the interest of the Exploration Geophysics community with FWI. This was marked by a large spike in the number of submissions to key journals in this topic (Figure 3).

The role of prior information:
Since the significant spike in interest FWI, the method has has been primarily promoted as solely a data fitting technique. Initially the level of prior information involved was just minor smoothing of the computed gradient or approximate inverse hessian to prevent numerical instabilities. The classical Tikhonov regularization (Tikhonov and Arsenin, 1977) was also applied in an attempt to mitigate against the ill-posed nature of the inversion. More recently the importance of more specific regularization and pre-conditioning approaches, such as those that preserve sharp edges (Guitton, 2012) or use a-priori dip information (Guitton et al., 2012) have showed the value of not considering
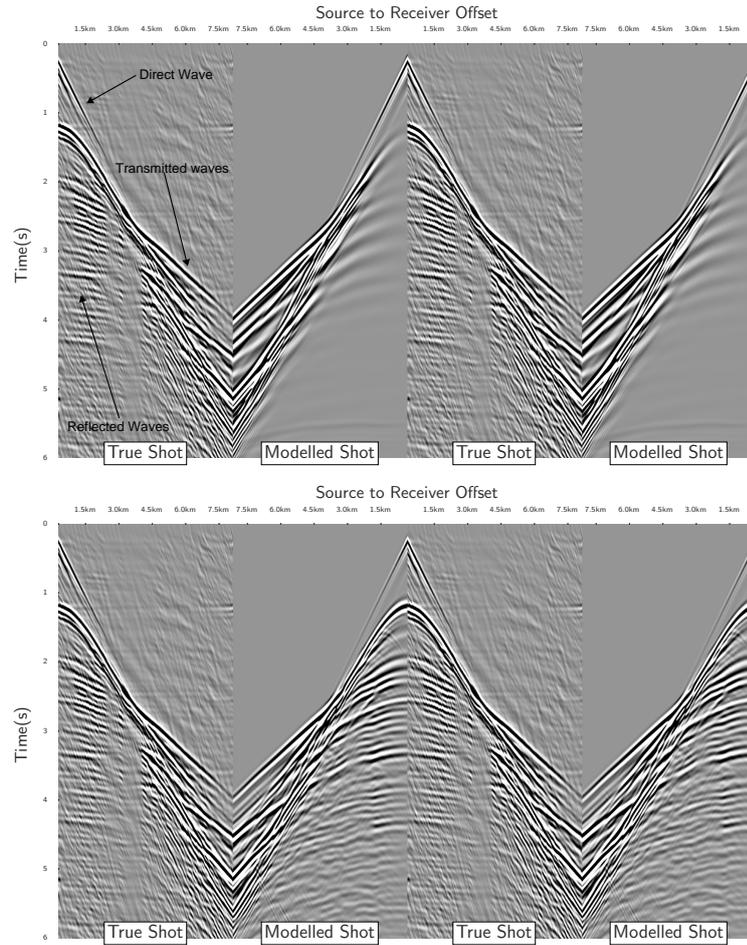
Figure 2: A true shot record prior to FWI compared to the modelled shot record *TOP:* Before FWI *BOTTOM:* After FWI. The shot records are mirrored so that the match between the near and far offsets can be observed. The key "families" of waves (direct, reflected and transmitted) are highlighted, all of which are involved in the fitting process in FWI.

FWI as a purely data fitting approach. The use of imperfect or incomplete information obtained from well data (Asnaashari et al., 2013) and/or pre-FWI imaging approaches, such as migration means we can recast the FWI problem as a damped least squares problem (Menke, 1984). The pitfalls and difficulty involved in including prior information relate to the ambiguity in building effective and correct constraints and the need for the prior based regularization/preconditioning to be computationally efficient to apply in the framework of FWI. In this thesis I present a flexible, simple to parametrize and efficient constraints for 1D, 2D and 3D applications.

Chapter 1: Commences with an introduction to inverse problems giving particular reference to the non-linear, ill-posed FWI problem and the local descent optimization

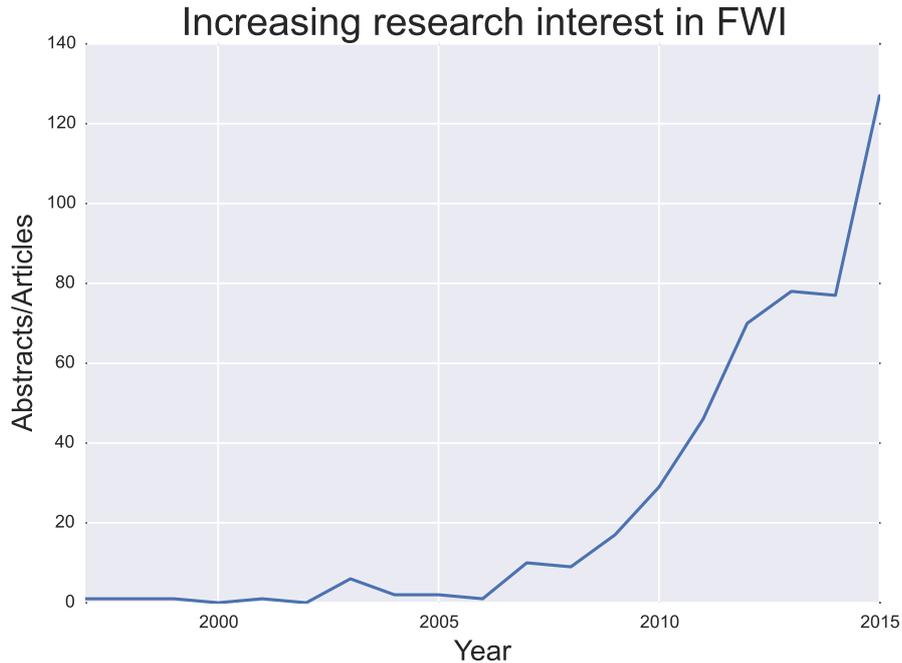Figure 3: The number of accepted submissions to the SEG for each year with the words "FWI" or "Full Waveform Inversion". *Taken from searching on http://library.seg.org/*

approaches we use to solve it. The second part of the chapter will look at explaining what work has been done by other others in order to regularize and precondition FWI with prior information in the hopes of limiting the ill-posed nature on the quality of the inversion result. I finalize the chapter by introducing the geology and acquisition configuration of the FWI case studies we will investigate in this dissertation.

Chapter 2: Briefly looks at geostatistics and introduce the idea of the spatial correlation of model parameters giving real and relevant analogies from the field of geography and geology. The Laplacian covariance and correlation operators are introduced and I highlight why both the action of both the operator it and its inverse are of interest, but typically illusive in "big data" situations where we have millions to billions of model parameters. The 1D inverse laplacian correlation function highlighted by Tarantola (2005) will be discussed and we will show in the rest of the chapter our attempts at making this inverse operator the framework for an efficient wavenumber filter in both 2D and 3D. We refer to this extension to the 1D inverse laplacian as the additive inverse laplacian (AIL).

Chapter 3: Discusses how the AIL operator described in Chapter 2 can be used to precondition the FWI problem. Two synthetic case studies are investigated, the Valhall synthetic model and the Marmousi model. A frequency domain FWI strategy is utilized to study the role the additive inverse Laplacian can have in reducing the ill-posed na-

ture provided by additive noise. Comparisons will be made as to how the AIL operator is simpler to parametrize than Tikhonov regularization while also being able to handle variable dip.

Chapter 4: Will investigate a 2D real dataset, provided by CGG that is shot over the Australian North Western Continental shelf margin. The initial model construction (which is of key importance for FWI) is detailed, prior to looking at how full waveform inversion results can improve the imaging in at the deeper gas bearing target level. The AIL framework will be used to mitigate some of the undesired artifacts that map into our model space reconstruction while allowing us to decrease the cost of FWI via shot decimation.

Chapter 1

Introduction to Inverse Problems and the role of a priori in FWI

Contents

Chapter overview: This chapter provides a general introduction to inverse problems (See Tarantola, 2005; Menke, 1984; Scales and Smith, 1994, for a more detailed description). We will initially discuss the general concepts of forward and inverse problems, before focusing on FWI. Key properties of FWI, such as the ill-posed nature and non-linearity will be discussed, while we introduce the iterative first and second order local descent techniques that are commonly used as the optimization strategies of choice in FWI (Nocedal and Wright, 2006; Bonnans et al., 2006). How prior information has been included by other authors as part this local optimization strategy will also

be briefly discussed. The chapter will finish by introducing the geology and acquisition configuration of the FWI examples that will be discussed in Chapters 3 and 4

1.1 Introduction to forward and inverse problems

The key goal of inverse theory is to attempt to transform measurement data into knowledge about the physical world. When we discuss such transformations in the context of seismic imaging and more specifically Full Waveform inversion, our goal involves attempting to transform measurements of the seismic wavefield $\mathbf{d}_{obs}$ into the parameters, $\mathbf{m}$, that define and control the behaviour of the wavefield. For such a transformation to be possible and reliable we need a physical link to exist between the model parameters and the observed data. The wave equation (which will be discussed both in the time and frequency domain formulation in later chapters) provides a mathematical connection between the model parameters and the seismic wavefield and as such we can express the modelled wavefield data as

$$\mathbf{d}_{mod} = \mathcal{F}(\mathbf{m}). \tag{1.1}$$

The operator $\mathcal{F}$, can be understood as the process of solving the wave equation for the parameters defined by the model parameters $\mathbf{m}$ and the extraction of this wavefield at the observation points (location of seismic receivers). The wave equation can include the full set of parameters that exactly describe wave propagation in complex media (Toksöz and Johnston, 1981), or alternatively a case-dependent pragmatic subset of the parameters of 1st order significance. In the cases we will be considering in this thesis the parameters we will use to describe wave propagation will be the compressional velocity, density and in some cases the VTI parameters epsilon and delta.

This application of $\mathcal{F}$ is akin to a mapping operation allowing one to move from the parameter or model space, to the data space. The application of $\mathcal{F}$ to the model parameters is often referred to as solving the "forward problem" (Figure 1.1). In wave propagation the forward problem is typically deterministic, meaning that if we have model parameters and we make an assumption of the physics of the wave equation, we will achieve a unique solution for the modelled data. The same cannot typically be said for the transformation from the data space to the model space. This transformation, referred to as the inverse problem often suffers from a number of issues. One such issue
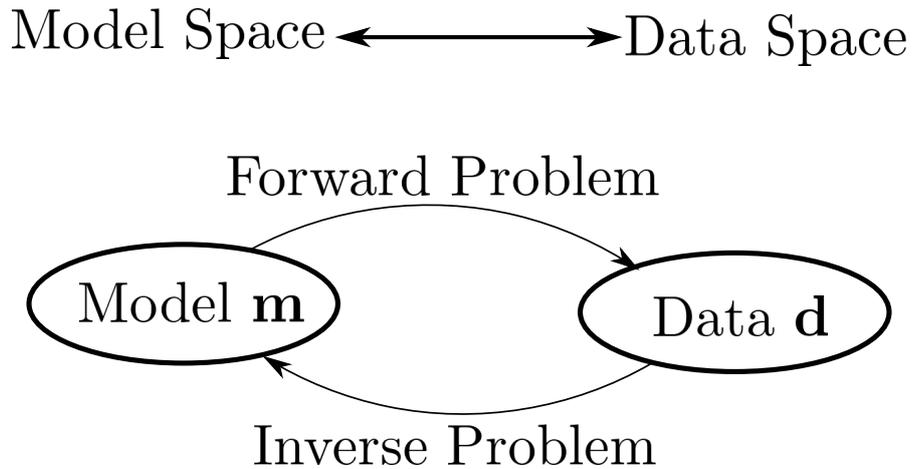
Figure 1.1: The transformation from the model space to the data space occurs from solving the often deterministic forward problem. Moving from the data space to the model space is referred to as the inverse problem

.

is that the solution to inverse problems are often ill-posed. The ill-posed nature means that the ability to find a solution, may not be guaranteed and even if the solution does exist, it will not be unique. A second complication is that, the link provided by between the data and model parameters is often non-linear. Both these issues are present in the application of FWI we will investigate and mean that the problem we look to resolve is, non-linear and ill-posed. I will further discuss the properties of these particular kinds of problems before moving to how we can look to solve them with iterative methods.

1.1.1 Ill-posed problems

Ill-posed problems can be defined as problems that don't adhere to the conditions described by Hadamard (1902). For a problem to be well-posed we require that, a solution exists to the problem, the solution is unique and the solution behaviour changes continuously and smoothly with the initial conditions. To understand ill-posedness, it is interesting to consider a simple well-posed problem. The heat equation with specified initial conditions represents such a well-posed problem. The heat equation,

$$\frac{\partial \mu}{\partial t} - \alpha \nabla^2 \mu = 0 \tag{1.2}$$

19

Heated region

Metal Plate

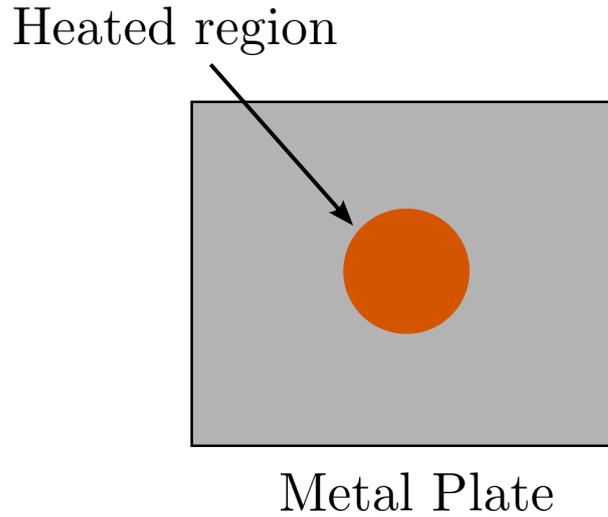Figure 1.2: A heat diffusivity problem. A round section of a square metal plate is heated by a blow torch. After heating stops, the heat equation can describe how the temperature will equilibrate over time

describes how an initial temperature distribution μ changes over time ($\frac{\partial \mu}{\partial t}$) where α is the thermal diffusivity and ∇^2 refers to the Laplacian operator. We can consider a simple thought experiment, where we have a square steel plate that is heated in the center by a blow-torch (Figure 1.2). If we solve the heat equation it will tell us how the edges of the heated circle will start to cool first, after-which the cooling effect will slowly moving towards the centre of the circle. Conversely, just outside the heated area, the temperature will gradually increases until the plate reaches a thermal equilibrium. This problem is well-posed in that one can continuously predict how the temperature will vary with time. However, in contrast if we wished to reconstruct an earlier temperature distribution from a later temperature distribution using the inverse heat equation the reconstruction will be ill-posed as the solution will be highly sensitive to the accuracy of the measured temperature distribution. This sensitivity to the initial conditions mean that ill-posed problem often require addition assumptions to ensure their stability. A commonly used example is to ensure that the resultant solution conforms to an expected smoothness (Tikhonov and Arsenin, 1977). The different strategies for stabilizing this ill-posed FWI will be discussed in more detail at the end of the chapter.

Other examples of the ill-posed nature of the inversion involve the additional simplifications that we often use in FWI. These include the fact the physics we involve in our forward modelling operator, $\mathcal{F}$ is often not able to completely explain all the information that appears in our observed data $\mathbf{d}_{obs}$. The reasons for the limitations of $\mathcal{F}$ are often

due to the high cost of fully modelling all the relevant physics in addition to the fact there often may be additional information such as noise that cannot be simulated by the wave equation. These factors mean that instead of trying to find the model, $\mathbf{m}$ that perfectly describes the observed data $\mathbf{d}_{obs}$, we try to minimize

$$C(\mathbf{m}) = ||\mathbf{d}_{obs} - \underbrace{\mathcal{F}(\mathbf{m})}_{\mathbf{d}_{mod}}||. \tag{1.3}$$

The goal is to select the model parameters, $\mathbf{m}$ that minimize the objective function $C(\mathbf{m})$. We typically look for the minimum of $C(\mathbf{m})$ using the description of the Euclidean length (Menke, 1984).This is also know as the least squares (ℓ_2) objective function.

$$C(\mathbf{m}) = ||\mathbf{d}_{obs} - \mathcal{F}(\mathbf{m})||^2. \tag{1.4}$$

Alternative norms do exist and have shown some advantages, such as increased robustness to noise and amplitude outliers, i.e. the ℓ_1 (Brossier et al., 2009a) or norms that provide a compromise between ℓ_1 and ℓ_2 (Amundsen, 1991; Crase et al., 1990; Guitton and Symes, 2003; Bube and Nemeth, 2007; Ha et al., 2009). In the presented studies, we will however restrict our discussion to the classical ℓ_2 norm.

1.1.2 Non-linear problems

As described previously, the link between the observed data $\mathbf{d}_{obs}$ and the model parameters $\mathbf{m}$ is non-linear. One of the key FWI relevant issues of non-linear problems occur when the objective function that we look to minimize in expression 1.4 begins to become multi-modal. We can visualize this for a two parameter case (Figure 1.3). The 2D contours of the objective function show that there are multiple low points where the data misfit appears low. Ideally one would like to ensure that our inversion reaches the lowest of these misfits known as the global minimum as opposed to only reaching one of the less optimal local minima. Our optimization strategy and approach needs to be tailored to achieve these goals.

A theoretically feasible method for finding the global minimum is via global optimization methods. Such methods allow one to attempt to completely probe the model space in search of the minimum (see Sambridge and Mosegaard, 2002, for an extensive review). The advantage of these approaches is that they can feasibly deal with the non-linear objective function. The Simulated annealing approach (Sen and Stoffa,
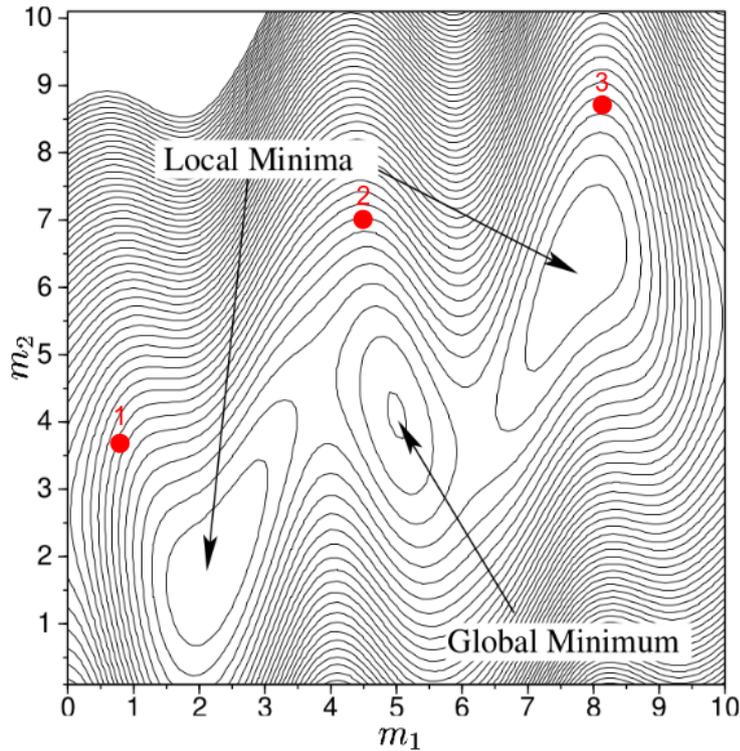
Figure 1.3: A two parameter case between parameters m_1 and m_2. The red dots highlight theoretical inversion starting positions. The further these dots are away from the global minimal the more sensitive they can be to falling into a local minimum. *Reproduced from Sirgue 2003.*

1995; Kirkpatrick et al., 1983), the Neighborhood approach (Sambridge, 1999a,b) and genetic algorithms (Holland, 1975) are examples that have been commonly applied in many geophysical application. Although such algorithms are utilized to tackle a range of topics in geophysics (Sambridge and Mosegaard, 2002) their application in FWI has been quite limited. Attempts have included the construction of initial velocity (Gao et al., 2014) or anisotropic (Debens et al., 2015) models. Both FWI examples are designed to produce low detail, low wavenumber models. In general, global approaches remain too computationally expensive to apply due to the need to the apply the operator $\mathcal{F}$ many times, to effectively probe the feasible extents of the model space. As global optimization techniques are largely seen to be infeasible, local, linearized optimization has become the preferred strategy in FWI. Linearisation of the problem is performed within the framework of the Born approximation (Born, 1926), where we require an accurate, initial background model, $\mathbf{m}_0$. The assumption is that an improved model $\mathbf{m}_1$ can be

described by

$$\mathbf{m}_1 = \mathbf{m}_0 + \Delta\mathbf{m}, \tag{1.5}$$

where $\Delta\mathbf{m}$ is a perturbation vector. By linearising the problem we are able to use efficient, iterative, local descent approaches to attempt to converge towards the global minimum. Our ability to reach this goal will be highly dependent on the initial starting model, $\mathbf{m}_0$. The simple two parameters example objective function (Figure 1.3) showed 3 potential initial model positions. As we assume the objective function to be linear and locally quadratic, it is likely only starting position 2 will reach the global minima and the convergence behaviour will depend on how we locally understand the shape of the objective function. Cases 1 and 3 will likely converge towards and become trapped in a sub-optimal local minima. The shape of the objective function in FWI is typically multi-modal with many local minima. This comes from the fact that FWI looks to match the oscillatory and band-limited vectors $\mathbf{d}_{mod}$ and $\mathbf{d}_{obs}$. The phenomena of "cycle-skipping" (Figure 1.4) is a direct result of this objective function "topography". Our initial model, $\mathbf{m}_0$ needs to be able to predict the kinematics of the observed data within a time shift of half the wave-field period, T (Beydoun and Tarantola, 1988). This requirement in the classical length based wave-field misfit approach has led many to attempt to decrease the effect of the non-linearity by a number of strategies. One approach is to initially focus on the inversion on the lower frequencies to decrease the non-linearity by increasing the size of $\frac{T}{2}$ relative to $\mathbf{m}_0$ kinematic errors (Pratt and Worthington, 1990; Bunks et al., 1995). Effort has been made to decrease the lowest acquired frequency in support of this goal. Other authors focus on the construction of accurate initial models using other ray based methods (see Woodward et al., 2008, for a review). A further and more recently applied approach is to try to transform the wave-field misfit objective function, to one that is less sensitive to the cycle-skipping phenomena. The alternative approaches include deconvolution based misfit (Luo and Sava, 2011; Warner and Guasch, 2014), those based on optimal transport (Métivier et al., 2015), phase unwrapping (Alkhalifah and Choi, 2012; Shah, 2014) or cross correlation (Lou and Schuster, 1991; Hale, 2013; Wang et al., 2015).

1.1.3 Local Optimization

Local optimization methods allow one to converge from our given starting point $\mathbf{m}_0$, using an indirect, guided exploration of the model space based. We are able to rewrite
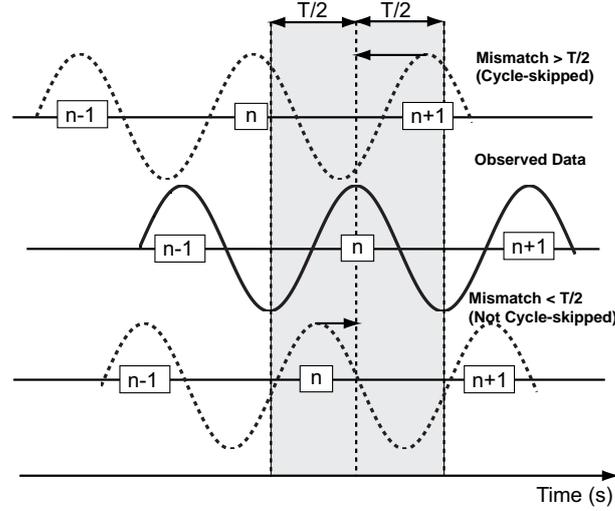
Figure 1.4: The "cycle skipping phenomena". Two cases for the modelled data, $\mathbf{d}_{mod}$ are shown. The top trace shows the case where the distance between the nth peak of the modelled data and the observed data $\mathbf{d}_{obs}$ is greater than half a period $\left(\frac{T}{2}\right)$. The data in this case is cycle skipped. The bottom trace shows a difference in timing of less than $\frac{T}{2}$. FWI will be able to account for this kinematic mismatch. *Modified from (Virieux and Operto, 2009).*

expression 1.5 as

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{p}_k \tag{1.6}$$

where k, represents the iteration number of the inversion, $\mathbf{p}_k$ is a descent direction and α_k is referred to as the step length. In order to perform the step length calculation also known as the line search it is typical to assume the objective function to be locally quadratic. Two additional forward problem are solved for steps in the direction of $\mathbf{p}_k$. This means that the objective function is known at the current point, $\mathbf{m}_k$ and two additional points. With these three points one can fit a parabola and extract the optimal value of α at the minima of the parabola (Figure 1.5). In order for the value of α_k to be deemed appropriate, it must satisfy the Wolfe conditions (Wolfe, 1969). This is composed of two individual conditions. The first requires that for a given size of step length, the objective function should decrease by a specified amount, while, the second specifies that the slope of the objective function at $\mathbf{m}_{k+1}$ should be less steep than at $\mathbf{m}_k$.

The optimization strategies applied in this thesis are based on the SEISCOPE Optimization Toolbox (Métivier and Brossier, 2016). The toolbox provides different options for calculation of the descent direction $\mathbf{p}_k$ while using a consistent approach for calculation of α_k based on the line search described above. The key different optimization
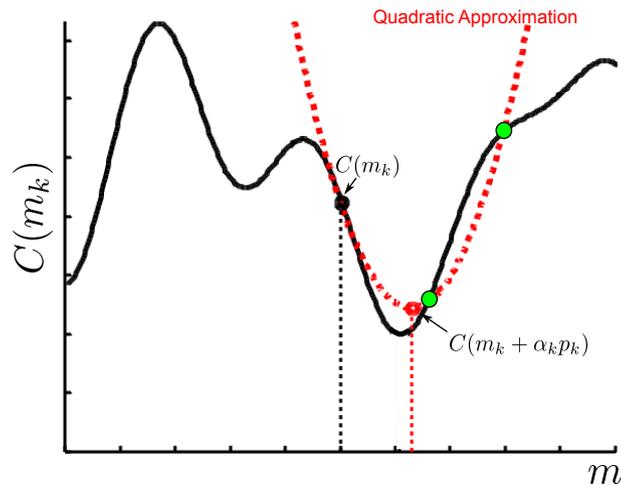
Figure 1.5: Illustration of the iterative decrease of the objective function by indirect local optimization. From the starting point m_k two additional forward modelling steps are applied for to calculate the cost function at two more positions (shown in green) along the current descent direction p_k. The updated model $m_{k+1} = m_k + \alpha_k p_k$ comes from the minimum of the quadratic function (red dot) that joins $C(m_k)$ and the two green points. This location must satisfy Wolfe's conditions (Reproduced from Menke, 1984; Vigh et al., 2009)

options are the steepest descent approach, non-linear conjugate gradient and the quasi-newton approach ℓ-BFGS and truncated newton approach.

Steepest Descent method

The steepest descent approach looks to follow the first derivative of the objective function

$$\mathbf{p}_k = -\nabla C(\mathbf{m}_k). \tag{1.7}$$

This approach can be quite effective if a small enough step length is used, but has proved to be extremely slow on difficult problems (Nocedal and Wright, 2006). A second issue can come from the sensitivity of the algorithm to poorly scaled problems. This is often the case in FWI when we are looking to invert more than one parameters (i.e. velocity and density).

Non-linear Conjugate Gradient method

The conjugate gradient algorithm approach is an often used linear solver that is effective for symmetric positive definite linear systems. The typical approach is to calculate the

descent direction as a combination of the the opposite of the gradient at the current iteration (the descent direction used in steepest descent) while also taking into account the descent direction from the previous iteration. The non-linear conjugate gradient approach in the SEISCOPE optimization toolbox uses the approach of Dai and Yuan (1999)

$$\mathbf{p}_k = \frac{||\nabla C(\mathbf{m}_k)||^2}{\left[\nabla C(\mathbf{m}_k) - \nabla C(\mathbf{m}_{k-1})\right]^T \nabla C(\mathbf{m}_{k-1})\mathbf{p}_{k-1}}. \tag{1.8}$$

The reduction in the number of iterations required to reach convergence with non-linear conjugate gradient as opposed to the steepest descent approach can be significant, however the acceleration is case dependent.

Newton Method

The steepest descent and conjugate gradient methods both represent 1st order optimization approaches. They rely, solely on the first derivative of the cost function. A first order Taylor expansion of the objective function (see Virieux and Operto, 2009, pp. 4 for the expansion) shows that

$$\mathbf{p}_k = -\left[\nabla^2 C(\mathbf{m}_k)\right]^{-1} \nabla C(\mathbf{m}_k), \tag{1.9}$$

where $\nabla^2 C(\mathbf{m}_k)$ represents the second derivative of cost function also known as the Hessian. The application of this inverse Hessian can have significant influence in improving the rate of convergence and is especially powerful in multi-parameter cases where scaling between the parameters can be troublesome. Typically we never compute the Hessian or it's inverse due to the significant cost. There has been some success applying this strategy to FWI (Métivier et al., 2013, 2014; Castellanos et al., 2015) by using an inexact truncated solution of the linear system associated with the Newton equation (Nash, 2000)

$$\nabla^2 C(\mathbf{m}_k)\mathbf{p}_k = -\nabla C(\mathbf{m}_k). \tag{1.10}$$

This requires the calculation of the Hessian operator times $\Delta\mathbf{m}_k$ which can efficiently be calculated using a matrix-free conjugate gradient algorithm. Although the approach has shown some promise it will not be considered in our applications.

Quasi-Newton ℓ-BFGS method

The ℓ-BFGS algorithm allows some 2nd order influence to be included in the descent direction $\mathbf{p}_k$ via application of an approximate inverse Hessian operator $\mathbf{H}_a^{-1}$ (Nocedal, 1980). A limited number (ℓ) of previous gradients are stored in memory and used to compute

$$- \mathbf{H}_a^{-1} \nabla C(\mathbf{m}_k) \tag{1.11}$$

following a two loop recursion strategy (Nocedal and Wright, 2006). The number of previous gradients tends to be typically small (i.e. 5 to 20) and the application of the approach to FWI has shown to improve the inversion results. This has been especially true in multi-parameter cases where the approximate inverse Hessian helps with parameter scaling. The ℓ-BFGS method is also advantageous as it adds minimal additional computational requirements (Brossier et al., 2009a).

1.2 Role of prior information in FWI

The optimization strategy discussed previously gives an efficient means for decreasing the misfit between the observed and modelled data. There can however be good reason to augment the purely data fitting goal with prior information. Reasons for this approach may be to mitigate against features that are not related to wave propagation, such as the presence of noise, to try to ensure that the FWI extracted parameters are consistent with information from other sources, such as well information or information sourced from seismic migration. I will give some examples of the current "state of the art" with regards how this prior information fits in to the standard FWI optimization strategy. To be consistent in nomenclature with many of the previous authors I will describe the first derivative of the objective function,

$$\mathbf{G} = \nabla C(\mathbf{m}) \tag{1.12}$$

as the FWI gradient $\mathbf{G}$. The efficient computation of this vector will be discussed in the later FWI application chapters.

The approaches I will detail can all be explained as forms of regularization or preconditioning. The nomenclature I will use to describe the different approaches of adding additional information will be consistent with Fomel and Claerbout (2003); Asnaashari

et al. (2013).

1.2.1 Model-space Regularization

Model based regularization involves the addition of a penalty term to the objective function to emphasize some desired feature in the model-space. Such a modification is often referred to as damped least squares (Menke, 1984).

$$C(\mathbf{m})_{\text{mod-reg}} = C(\mathbf{m})_{\text{data}} + \frac{\lambda}{2} C(\mathbf{m})_{\text{model penalty}}. \tag{1.13}$$

where $C(\mathbf{m})_{data}$ is our objective function from expression 1.4. The model space penalty is $C(\mathbf{m})_{\text{model penalty}}$ represents a secondary goal and the weight between this and the data-fitting goal is controlled by the hyper parameter λ. The model-space penalty term typically has the form

$$C(\mathbf{m})_{\text{model penalty}} = \mathbf{m}^T \mathbf{D} \mathbf{m}, \tag{1.14}$$

where the operator $\mathbf{D}$ is typically a roughening designed to emphasize whichever feature is desired. The most common uses are for it to emphasize smoothness (Tikhonov and Arsenin, 1977; Press et al., 1986) or the total variation (sparsity) (Vogel and Oman, 1996; Vogel, 2002; Askan and Bielak, 2008; Guitton, 2012) in the model space.

The addition of this penalty term also requires one to modify the gradient we use in FWI

$$\mathbf{G}_{\text{model-reg}} = \mathbf{G}_{\text{data}} + \lambda \mathbf{D} \mathbf{m}, \tag{1.15}$$

where $\mathbf{G}_{\text{data}}$ is expression 1.12.

1.2.2 Prior-model Regularization

Prior model regularization is designed not just to emphasize a given feature in the model space but to provide a penalty, when a given model parameters is not within an a priori, probabilistic range of values. In such a configuration our objective function and gradient become,

$$C(\mathbf{m})_{\text{prior-mod-reg}} = C(\mathbf{m})_{\text{data}} + \frac{\lambda}{2} \underbrace{(\mathbf{m} - \mathbf{m}_p)^T \mathbf{C}^{-1} (\mathbf{m} - \mathbf{m}_p)}_{C(\mathbf{m})_{\text{prior model penalty}}} \tag{1.16}$$

and

$$\mathbf{G}_{\text{prior-mod-reg}} = \mathbf{G}_{\text{data}} + \lambda \mathbf{C}^{-1}(\mathbf{m} - \mathbf{m}_p) \qquad (1.17)$$

respectively, where $\mathbf{m}_p$ is a prior model vector and $\mathbf{C}^{-1}$ is the inverse covariance operator that describes the uncertainty associated with $\mathbf{m}_p$. Asnaashari et al. (2013) applied this to FWI using the Marmousi synthetic (Figure 1.6). A linear interpolation between two pseudo-wells provided a velocity prior model, $\mathbf{m}_p$ and were used to contain the inversion. At the wells the velocity value was not permitted to vary significantly from the a-priori model. However as the location of a model parameter had a larger offset from the well-bore the inversion was more able to depart from $\mathbf{m}_p$. The degree to which such departures from $\mathbf{m}_p$ were permitted was controlled by the diagonal inverse covariance operator $\mathbf{C}^{-1}$. The distance from the well measurements was combined with a vertical weighting scheme to produce $\mathbf{C}^{-1}$. With the prior-model regularization defined, the inversion was able to both attempt to decrease the misfit associated with $C(\mathbf{m})_{\text{data}}$, while satisfying also satisfying the prior model constraint. This application of a priori model inversion represents a damped least square inversion in the Bayesian framework. Although such approaches are quite novel and suggested as good way of dealing with non-uniqueness of inversion problems (Scales and Snieder, 1997; Tarantola, 2006) the construction of both $\mathbf{m}_p$ and $\mathbf{C}^{-1}$ require considerable thought and strong deterministic choices (Clapp et al., 2004). We will not follow this strategy as part of this thesis.

1.2.3 Model-space Preconditioning

The final form of prior influence we will consider is referred to as model-space regularization (Fomel and Claerbout, 2003) or preconditioning. Unlike the regularization described, precondition approaches can be seen to directly remove unwanted a-prior features. Harlan (1995) suggested that instead of applying a penalty term that discourages complexity through roughening operators, we can perform a change of variable to a model-space that builds model simplicity directly. This was referred to as model-space preconditioning and the idea is that instead of looking to update the model parameter, $\mathbf{m}$, we use a change of variables to instead look to update a smart parameter $\mathbf{q}$ such that

$$\mathbf{m} = \mathbf{Sq}, \qquad (1.18)$$

where $\mathbf{S}$ is some preconditioning operator. Guitton et al. (2012) use this approach in the context of FWI, utilizing local dip filters to ensure model updates were aligned with
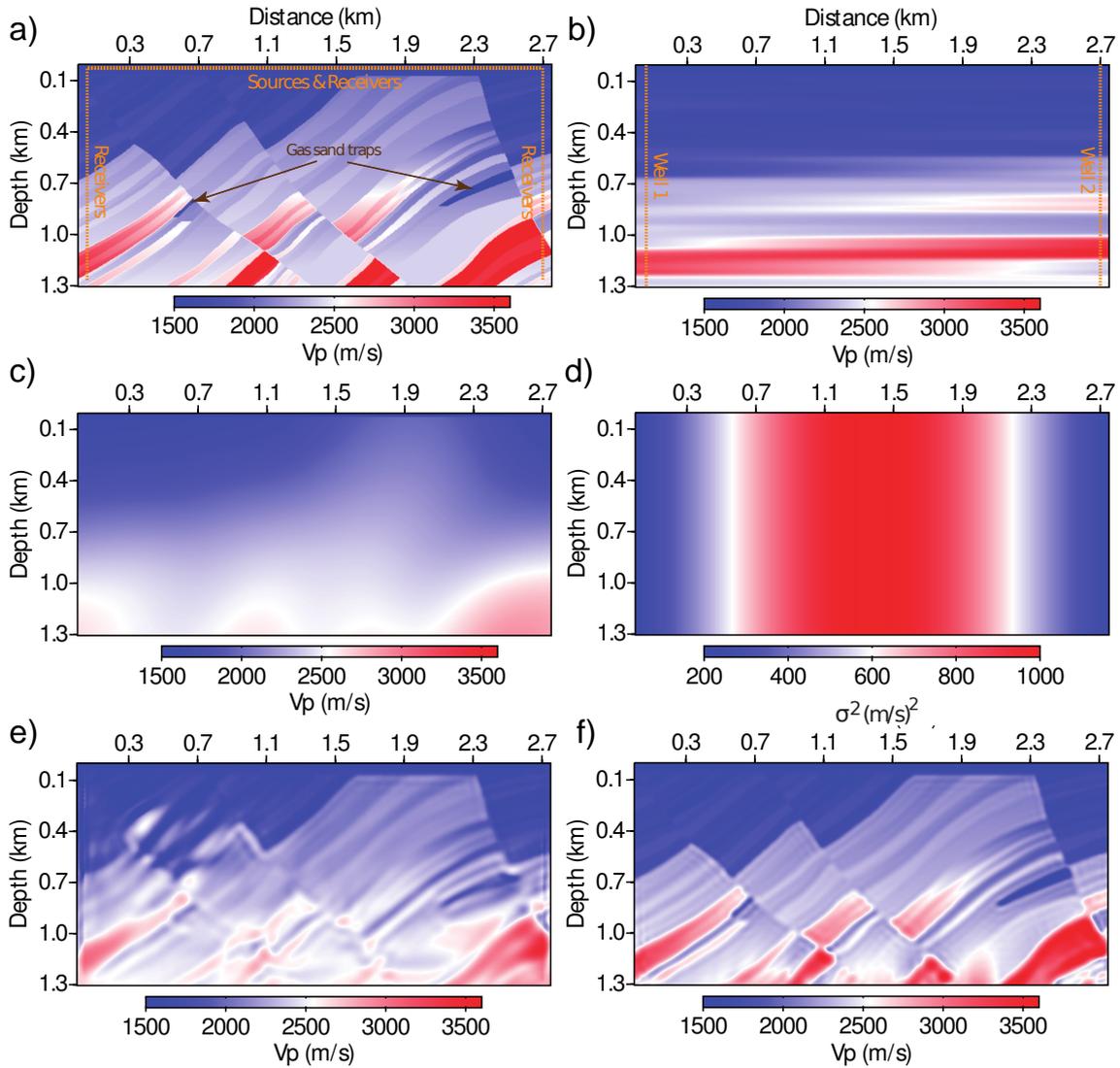
Figure 1.6: Application of prior-model regularization a) The true Marmousi velocity model, c) The initial model for FWI, b) The prior model vector $\mathbf{m}_p$ build from interpolation of the two wells d) The uncertainty of $\mathbf{m}_p$ e) The inversion result without using prior, f) the inversion result using prior model regularization (*Recreated from Asnaashari et al., 2013*).

geological dip. Such dip based operators are often referred to as steering filters and some key examples are the Wave-kill filter (Claerbout, 1992), the Plane wave destruction filter (Fomel, 2002) and filters based on directional Laplacians (Hale, 2007). The approach of Guitton used the directional Laplacian filters, to highlight how they could be effective in allowing FWI to converge even with a sparse acquisition of shots (Figure 1.7). The geological dip, estimated from RTM image constructed using the initial velocity model

$\mathbf{m}_0$ allowed the dip direction to be used to define inverse dip filters that re-projected the gradient ($\Delta C(\mathbf{m}_n)$) to be aligned with the geological dip ($\Delta C(\mathbf{q}_n)$). Application of this approach improved the reconstruction of the final inversion result making it comparable with the optimization result obtained using a much denser acquisition of shots.

An alternative preconditioning approach is also proposed by Alkhalifah (2015). This approach looks to precondition the FWI gradient so it is insensitive to the low-scattering angles at the early iterations. An efficient scattering angle filter is based on the representing the gradient in the time-lag normalized domain (Sava and Fomel, 2006; Khalil et al., 2013) and muting all but the large scattering angles. By focusing on the higher scattering angles it is suggested that the gradient will be richer in more stable low wavenumber information.

1.3 Investigated FWI case studies in this thesis

I will consider a number of FWI examples to investigate how prior information can be useful in mitigating some undesired features mapping in our model parameter reconstruction. Two synthetic datasets are considered (Chapter 3), the 2D Valhall synthetic model built by BP and the 2D Marmousi model. For these examples, FWI will be applied using the 2D frequency domain, visco-acoustic full waveform inversion code, TOY2DAC. A 2D real dataset is also consider (Chapter 4) that crosses the North Western Australian Continental Shelf margin. This dataset provided by CGG is a Broadseis marine streamer dataset and is inverted using the VTI acoustic FWI code TOYxDAC_TIME.

1.3.1 Valhall case study

The first synthetic dataset considered is based on the Valhall oil and gas field. The field located in the Norwegian North Sea (Figure 1.8) was discovered in 1975 and exists in relatively shallow water (< 100 m). The reservoir structure comprises a NNW-SSE trending anticline, occurring at approximately 2km depth. The oil is housed within the over-pressured Upper-Cretaceous Tor and Hod chalk reservoirs (Figure 1.9) which exhibit high porosity (>50pc) but low permeability. Oil is interpreted to have been sourced from the Upper Jurassic Mandal Formation with oil migration occurring from the Early Miocene through to present times (Barkved et al., 2003). The reservoir is
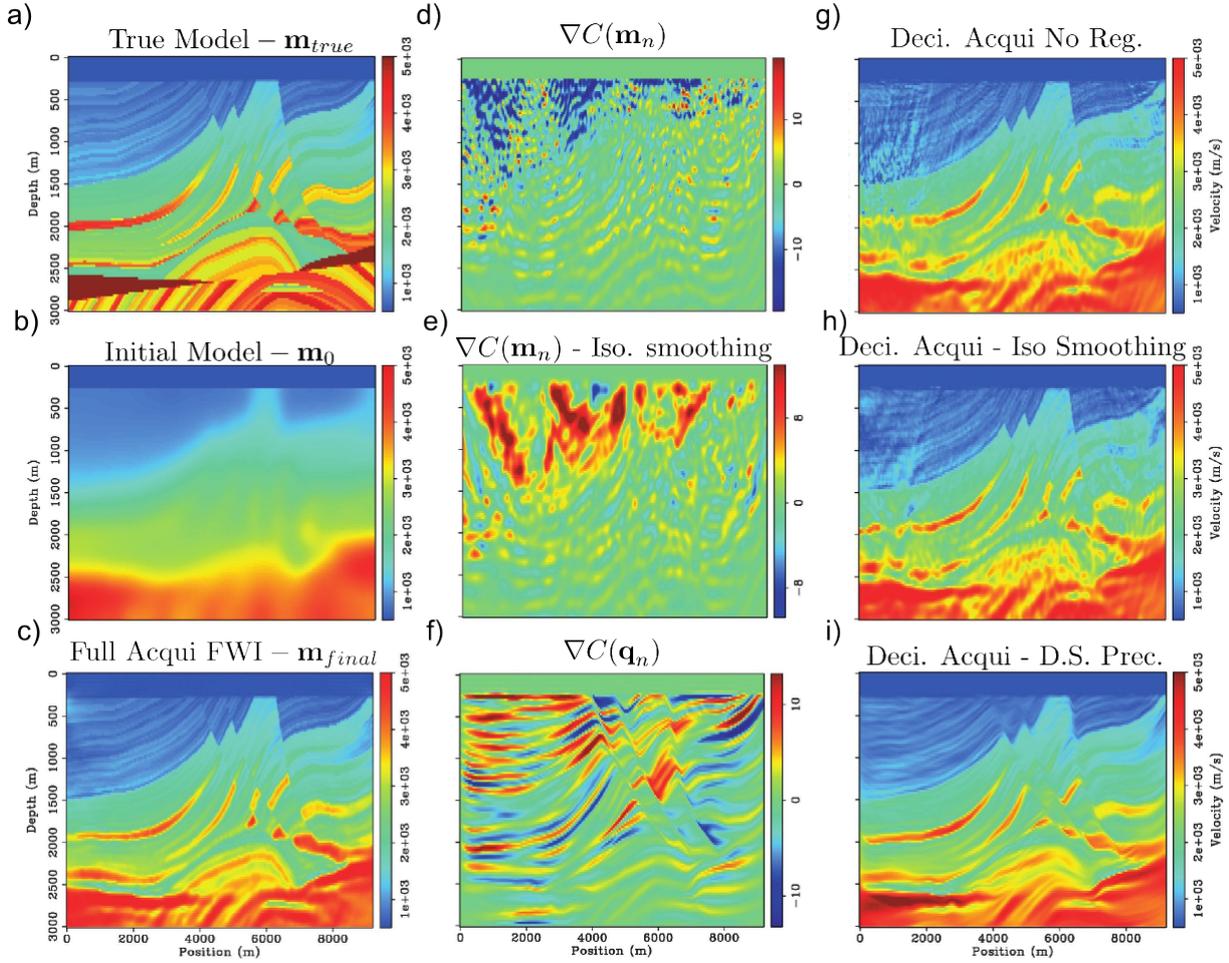
Figure 1.7: FWI applied to the Marmousi 2 model with a surface acquisition. a) True
model, b) Initial Model, c) Final inversion result using $\Delta S = 50$ m, d)Gradient achieved
using decimated shots $\Delta S = 2,500$ m e)$\Delta S = 2,500$ m gradient with isotropic smoothing
f), $\Delta S = 2,500$ m gradient using data-space preconditioning. g,h,i and correspond to
inversion results obtained from the gradients d,e,f respectively (*Reproduced from Guitton
et al., 2012*).

overlain by a 1000m thick Tertiary section comprised of claystones, silts and shales of
Paleocene, Eocene and Miocene age. Micro-fractures that occur in the chalk, allow gas
to migrate from the reservoir to diamateous and low-density sections in the Miocene.
This gas saturation produces a low velocity gas-cloud over the central part of the Valhall
field.

The Valhall oil field has been a widely presented example of the limitations of us-
ing conventional P-wave imaging techniques in the presence of a overburden gas cloud.
Prior to the end of the 1990s it was not possible to create a good image of the crest of
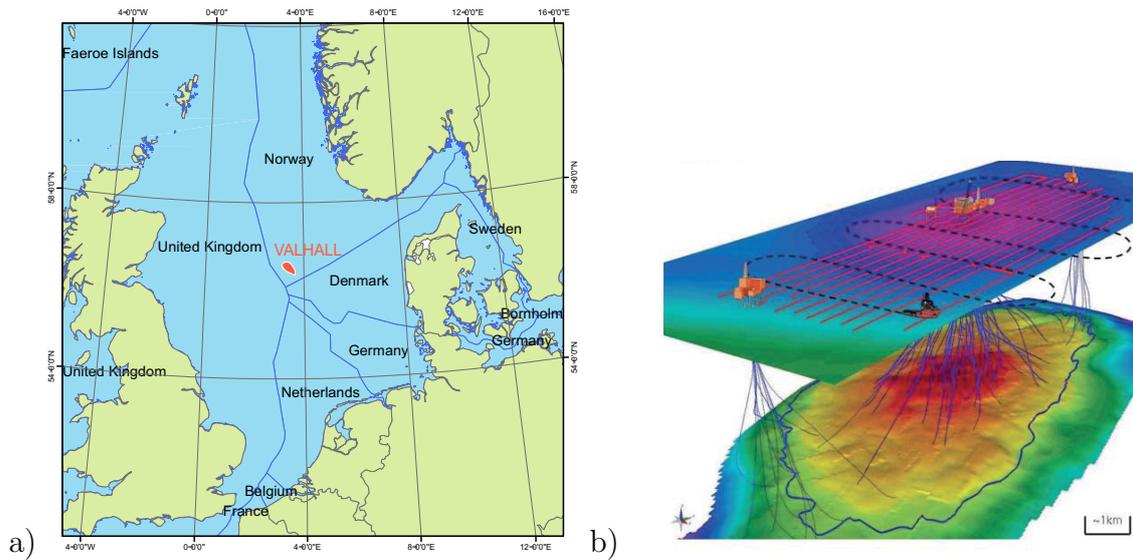
Figure 1.8: a) A map showing the location of the Valhall oil field, b) A schematic showing the OBC acquisition, wells and top reservoir structure.
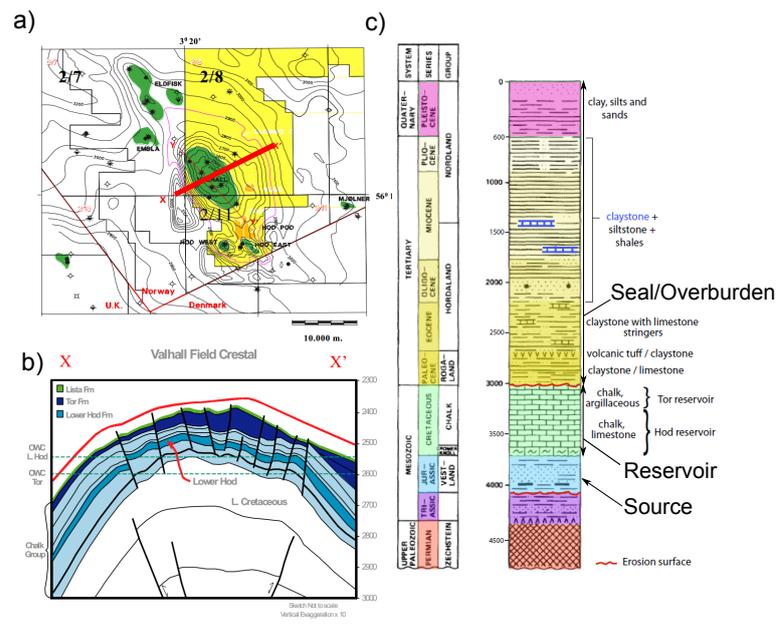

Figure 1.9: a) A contour map across the oil field with a WSW-ENE cross-section high-lighted in red, b)The cross section shows the faulted structure of the Valhall oil field reservoir while the stratigraphy is described in the stratigraphic column c). *Reproduced from* Barkved et al. (2003).

the structure in the vicinity of the gas cloud which resulted in a "no data zone". The acquisition of one of the worlds first 3D 4C surveys (Rosland et al., 1999; Barkved et al.,

33

1999) allowed effective imaging of this no data zone using shear wave imaging (Thomsen et al., 1997). In 2003 a simple, cost effective installation of permanent 3D, 4C ocean bottom cables occurred. The cost efficiencies of such an installation were afforded due to the constant water depth. The system was hoped to function for the life of the field (LoFS) to help support decision making regarding the optimal water-flooding of the oil field (Barkved et al., 2010). Frequent re-acquisition (11 re-shoots by 2010 (Barkved et al., 2010)) of 3D seismic over the permanent ocean bottom cable were driven by a number of factors including the strong 4D effect that could be expected due to pressure depletion and compaction effects induced by in the Chalk reservoir. The wide-azimuth, long offset data has been the subject of a number of studies to highlight the role 3D FWI can provide in building high resolution velocity models (Sirgue et al., 2010; Prieux et al., 2011; Operto et al., 2015).

Our analysis will utilize the Valhall synthetic model provided by BP. This synthetic model (Figure 1.10) contains many of the key features common in the real dataset. The model is 16.3 km long in the horizontal and 3.9 km in the vertical with a grid cell size of $h = 10$ m. We will look to recover this model from a smoothed initial model using FWI. A frequency domain FWI strategy will be applied by using an OBC style acquisition where we consider sources every 250 m and receivers every 50 m. It is important to note in our studies we consider a constant density model and we only consider the P-wave pressure component wavefield.

1.3.2 Marmousi Synthetic

The second model (Figure 1.11) we consider is a subset of the commonly used Marmousi model (Bourgeois et al., 1991). Build by l'Institut français du pétrole in 1988 it is based on a geological section taken through the North Quenguela trough in the Cuanza basin (Angola) (Verrier and Branco, 1972). The subset of the model we have taken is 3.9 km long in the horizontal and 1.3 km in the vertical direction, discretized with $h = 10$. In this model one of the key differences as opposed to the Valhall synthetic is the increased geological dip. A constant density model is employed and we look to reconstruct the true model from a 500 m isotropic smoothing of the initial model. A fixed-spread surface acquisition is employed with receivers every 10 m and sources every 50 m. One of the key features that we look to reconstruct is the two gas sands present in the fault blocks.
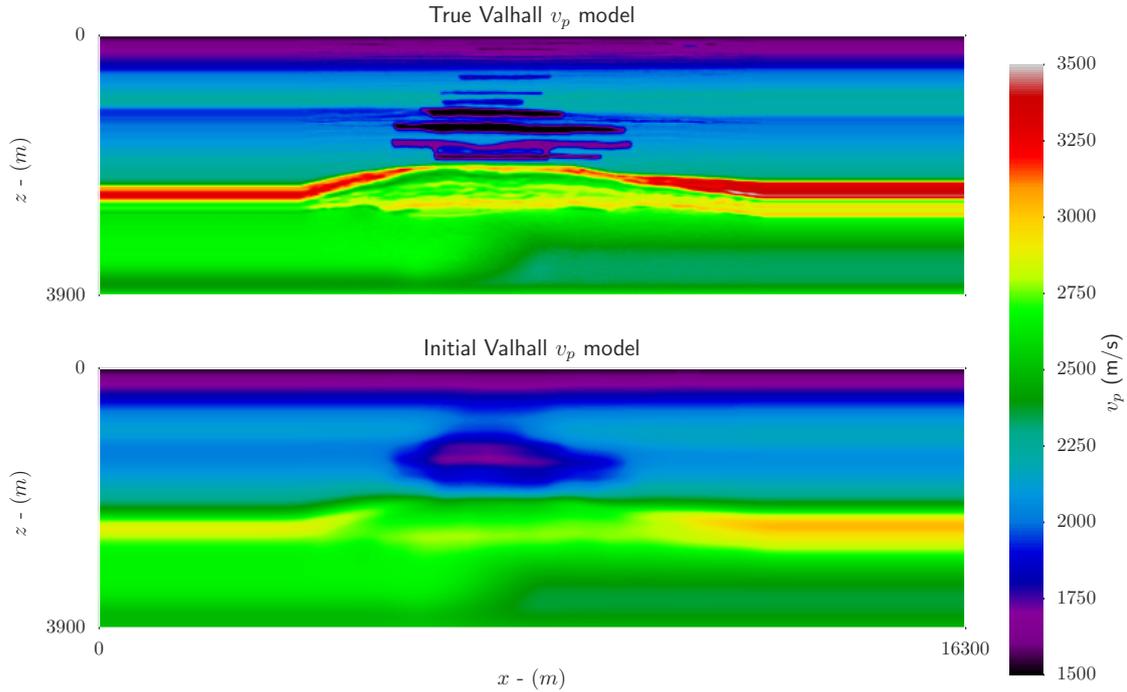
Figure 1.10: a) The true Valhall synthetic velocity model. b) 375 m Gaussian smoothed version of the true model that serves as the starting point for FWI.

1.3.3 NWA-006 Broadseis Dataset

A real dataset was generously provided by CGG. This dataset crosses the North Western Continental shelf of Australia (Figure 1.12). The line is approximately 35 km long and has been acquired both with a conventional, 7 m constant depth streamer and a broadband, "Broadseis" variable depth streamer (Cable depth 8-57.5 m). The eastern section of the line is acquired on the continental shelf where the water depth is less than 100 m. To the WNW along the line the water depth increases to greater than 1000 m as we move towards the basin depocentre. Within 100 km of the line there are some of the largest producing gas reservoirs in the world (Gorgon, Wheatstone, Pluto and the North West Shelf Venture). These giant gas fields typically have their reservoir based in the Triassic, Mungaroo formation. The Broadseis line was acquired between two previously drilled wells (Chrysaor 1 and West Tryal Rocks 4) that lie on the Broadseis line. These wells successfully targeted the same, highly productive, Triassic Mungaroo formation, discovering and appraising the Chrysaor and West Tryal Rocks gas fields respectively.

The stratigraphic and structural evolution of the region is of key importance to ap-
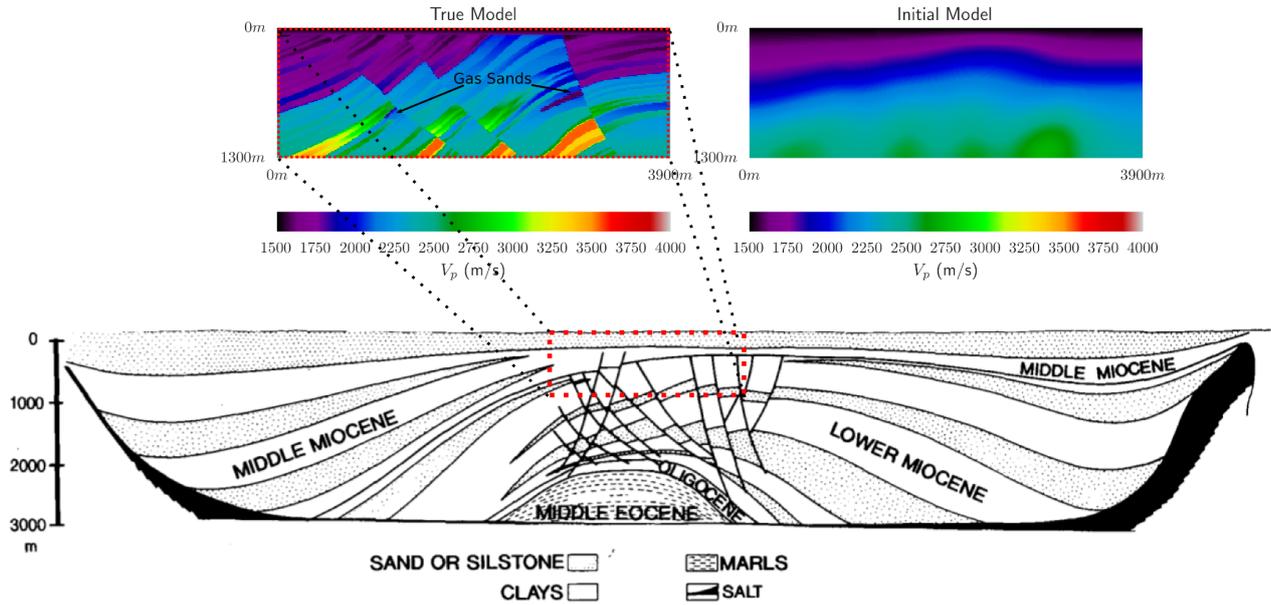
Figure 1.11: *TOP:* The Marmousi true and initial model that will be the subject of
the FWI study *BOTTOM:* The geological cross-section the Marmousi model is based on
(Verrier and Branco, 1972)

preciate and understand the FWI results. Detailed descriptions of the petroleum geology
and exploration history can be found in Barber (1988); Hocking (1988); Tindale et al.
(1998); Longley et al. (2002). I instead present a summary based on Walton (2008) in
which I will describe the key stratigraphic intervals, the structural evolution and some
key seismic horizon picks. The stratigraphic composite column that accompanies this
description is shown in Figure 1.13 where an effort has been made to make the key hori-
zons interpreted in a stacked image of the line. The two wells are highlighted on the
seismic section and red dash along the well path shows the "well pick" for each respective
horizon which is quite good for all but H1. The seismic section is from a time to depth
stretched PSTM stack provided by CGG, where the velocity model used corresponds to
our FWI initial model (shown and discussed in Chapter 4).

Starting from the deeper section (>3 km) of the NWA-006 line we encounter the
prospective, Triassic Mungaroo formation. This reservoir interval is composed of a thick
package of fluvial sandstones and non-marine brackish siltstones. In the uppermost 300
to 400 m however, the formation exhibits typically more marine influence and is com-
posed of coastal sandstones and claystones expected to have been deposited within a
broad, low relief, rapidly subsiding coastal plain. Gas accumulations have been discov-
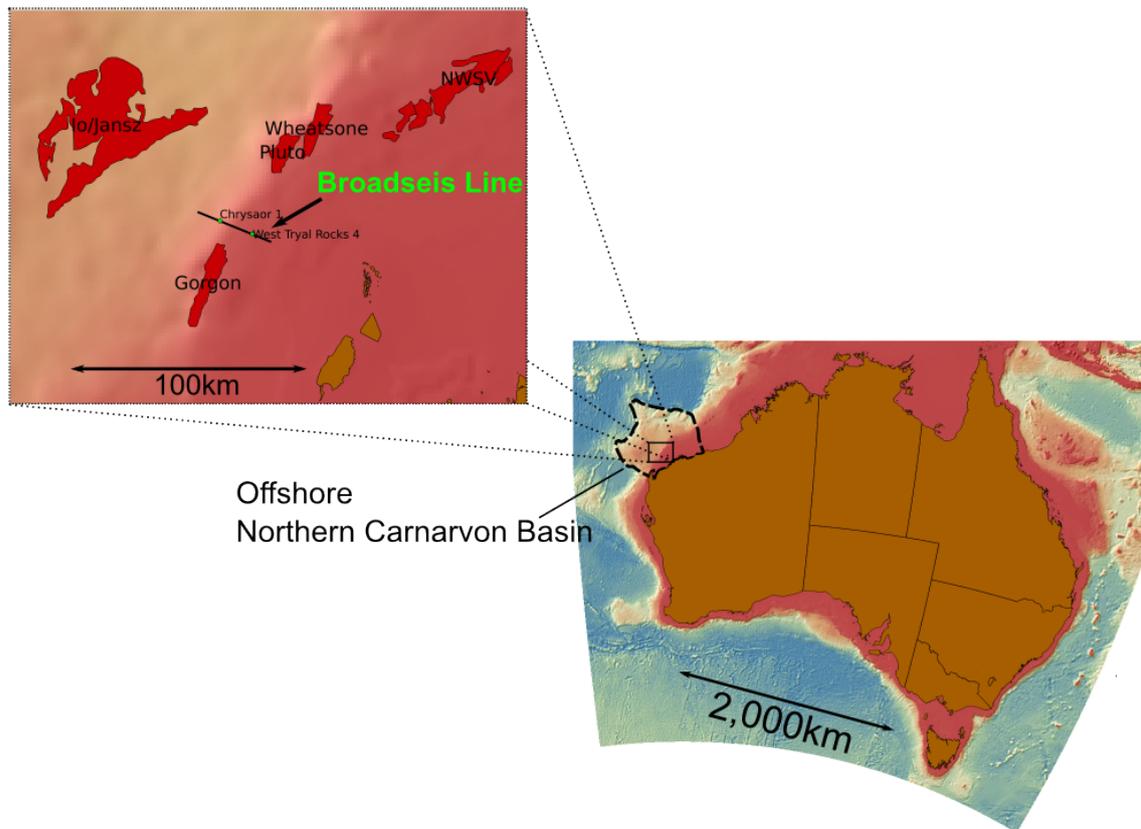
Figure 1.12: A location map showing the sea-floor bathymetry around the Australian Island continent. The Broadseis line crosses the Australian continental shell margin in the Northern Carnarvon basin and two wells. The line sits within 100 km of some of the largest natural gas fields in the world.

ered in the Mungaroo formation, both in the deeper fluvial section and in the shallower, shallow-marine reservoir. This gas charge is interpreted to be potentially coming from the deeper Locker Shale acting as a source rock. Within the Mungaroo formation, there are a number of seismically resolvable sequences, that based on the available well data appear to come form the impedance contrast that occurs from alternating sand-rich and clay-rich depositional units. The first horizon we identify (H4) is the Top Mungaroo formation.

Deposited atop of the Mungaroo formation is the Brigadier Formation and the Murat siltstone. The continental break up of the North Western Australian margin in the middle to late Jurassic caused significant rifting which can be seen with the rotated horst blocks and grabens offset by the faults interpreted in red. These horst structures form

the hydrocarbon bearing traps, that constitute much of the gas reserves in the offshore
Northern Carnarvon basin. Post-rifting, a period of lower sea-level exposed many of the
topographically higher horsts to sub-aerial erosion in the early Jurassic. This eroded sed-
iment provides the fill in some of the lower grabens where Jurassic sediments have been
eroded from exposed horsts and re-deposited in the grabens. Deposition continued dur-
ing the Jurassic until a significant rise in the sea-level produced an angular unconformity
surface known as the Intra-Jurassic Unconformity (IJU). This unconformity surface is
approximately coincident with the Top Mungaroo seismic reflection in the higher horsts
where erosion occurred and the overlying reflectors are approximately conformable. A
gap in the sediment record occurs in these regions where there is an abrupt shift from
Triassic sediments to Cretaceous aged sediments. In the deeper horsts however there is
potentially the presence of Jurassic, Murat and potentially Athol formations. In these
regions the depth separation between the Top Mungaroo and IJU reflector is more sig-
nificant and the angular unconformity is clearly visible.

In the Cretaceous there was deposition of thick sections of Berriasian to Valanginian,
Barrow group deltaic shales. These shales are formed by two broad units known as the
Barrow Group 1 and Barrow Group 2 delta which pro-grade from the continental shelf
basin-ward to the North West. A marine transgression starting in the Valanginisan,
resulted in the deposition of the argillaceous marine sediments of the Muderong shale.
The commencement of regressive sedimentation in the Aptian deposited a regionally ex-
tensive, silty radiolarian claystone. The contrast between this and the Top Muderong
provides one of the key reflections we interpret (the H3 horizon). Above this we find the
Lower and Upper Gearle siltstone. In the late Cretaceous the relative tectonic stability
and the decreasing supply of terrigenous sediments, resulted in the commencement of
deposition of widespread shelf carbonates. The Toolonga Calcilucite reflector (H2) is
approximately coincident with this change of lithofacies and is a carbonate rich section
overlain by the Miria Marl. Marls are a mix of clay and lime mud and are typically quite
hard and fast.

During or soon after the deposition of the Toolonga Calcilucite and Miria Marl subsi-
dence in the Exmouth Plateau to the North West and uplift in the Barrow and Dampier
Sub-basins to the South East resulted in shelf-edge instability and slumping of these
formations. This can be seen in the seismic by the relatively chaotic seismic structure in
the line. Although 3D seismic wasn't available I was able to to take screen-shot from the

report of a seismic section that goes from North to South directly crossing the NWA-006 line at the Chrysaor-1 location (Figure 1.14). This "tie-line" shows some significant variability in the Tertiary section and it is not unlikely that the 2D Broadseis line imaging effects will suffer some limitations due "out-of the plane" effects. The top of the slope failure section is marked by the Top Cape Range formation (H1) above which the seismic character is much more laterally continuous although there is some additional chaotic character on the shelfal section of the line. The horizon H0 is the interpreted waterbottom pick.
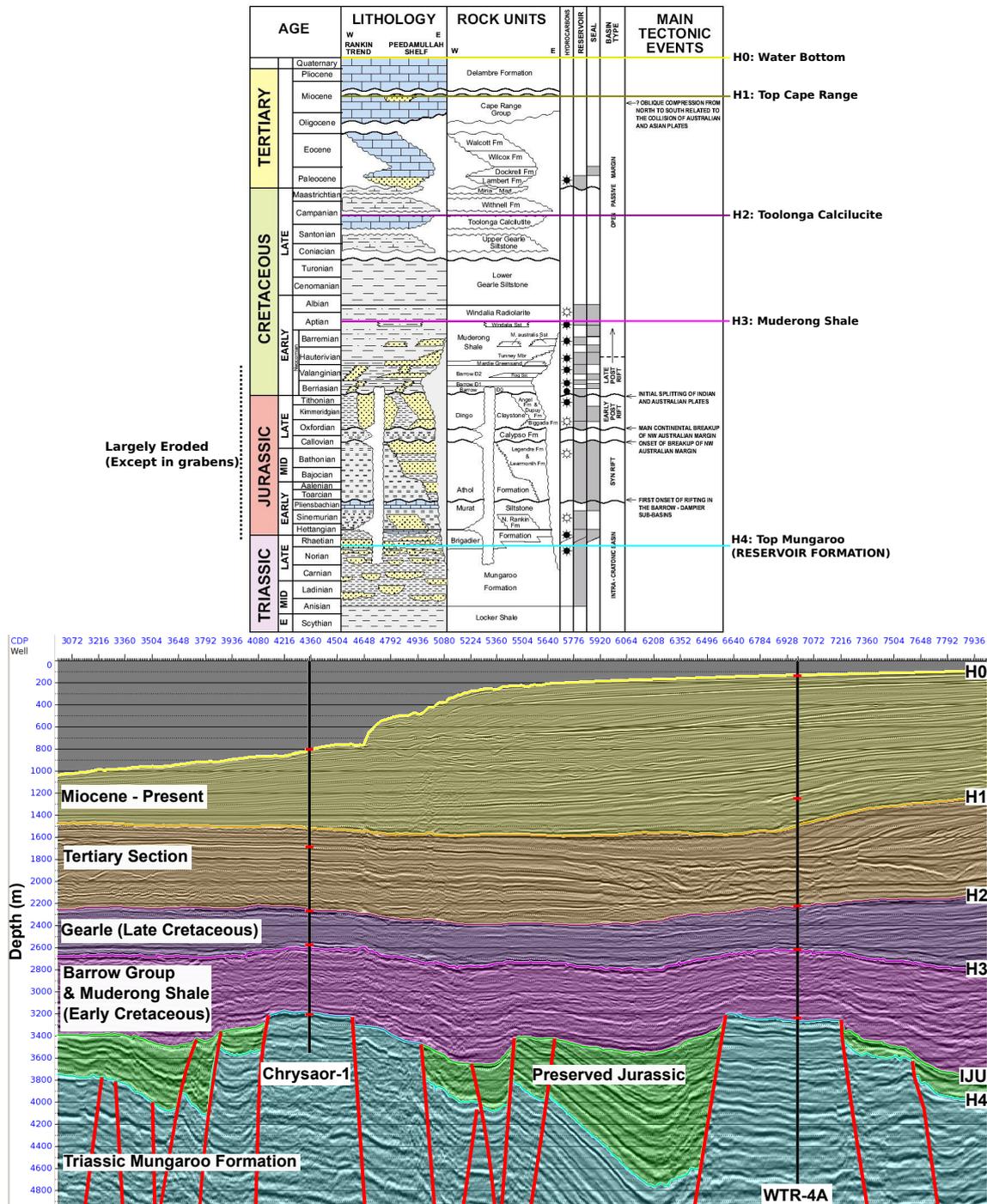
Figure 1.13: *TOP*: Stratigraphic column with key features and seismic horizons that are coincident with NWA-006 line. *Reproduced from (Walton, 2008) BOTTOM:*Depth stretched Conventional PSTM full stack. The key horizons are interpreted and the stratigraphic intervals are highlighted. An effort has been made to make the colour of the horizons consistent with Figure 1.14.

Figure 1.14: An image *taken from Walton (2008)* showing some of the chaotic seismic character that may cause "out of plane" effects on the NWA-006 line.

Chapter 2

Geostatistical Constraints

Contents

Chapter overview: Different approaches have been used in the past to attempt to regularize and/or precondition the ill-posed, full waveform inversion problem. In this chapter we are going to discuss the application of correlation operators to modify the wavenumber spectrum of a vector. This chapter will focus on the efficient

application of the matrix-vector product of these correlation **Corr** or covariance **Cov** operators to a vector, **u** (**Corr**(**u**) and **Cov**(**u**)). We will also look at efficient ways to calculate the inverse of these operators(**Corr**$^{-1}$**u** and **Cov**$^{-1}$**u**). For future reference our nomenclature describes a correlation operator as equivalent to a covariance operator with a variance $\sigma^2 = 1$. We use the convention of using **u** as the notation of vector as opposed to the more conventional **v**. Our descriptions of **u** in this chapter will deal with 1D, 2D and 3D spatial vectors of equal length n in each direction. These vectors will have a total number of elements of $N = n$ for 1D, $N = n^2$ for 2D and $N = n^3$ for 3D.

Initially, we will discuss why the concept of correlation is important in nature and more specifically, sedimentary geology. The discussion will then turn to analytical co-variance and correlation functions in 1D, 2D and 3D. These operators will be constructed to handle correlation lengths with respect to the Cartesian coordinates system as well as correlation lengths based on a rotated coordinates system. Although these operators will be discussed, their limitations in terms of the CPU requirement for calculating **Corr**(**u**) and **Cov**(**u**) for larger 2D and 3D vectors (Figure 2.1) in addition to the memory and CPU requirements for their inversion in order to calculate **Corr**$^{-1}$(**u**) and **Cov**$^{-1}$(**u**) (Figure 2.2) prevent them from being attractive to apply them in the framework of FWI.

Our focus will then shift to the curious, 1D analytical inverse Laplacian covariance operator (Tarantola, 2005, pp. 308-311). I will show that the presented analytical inverse covariance function can be extended to higher dimensions and allows a means to approximate **Corr**(**u**), **Cov**(**u**), **Corr**$^{-1}$(**u**) and **Cov**$^{-1}$(**u**) with a fraction of the memory and CPU requirements of using the true analytical formulation. The application of these operators in the framework of FWI will be discussed first of all on synthetic data (Chapter 3) followed by real data (Chapter 4).

2.1   Geological clustering and spatial dependence

As FWI is an ill-posed process, it benefits from some form of regularization and/or preconditioning as opposed to using just a data fitting goal. The potential role of these auxiliary constraints were discussed in Chapter 1. Although such measures can be useful, care must be taken in applying regularization and preconditioning as there can be significant peril with applying constraints to the FWI problem without adequate justification. It is well known that FWI is a non-linear problem, that is highly dependent on the initial
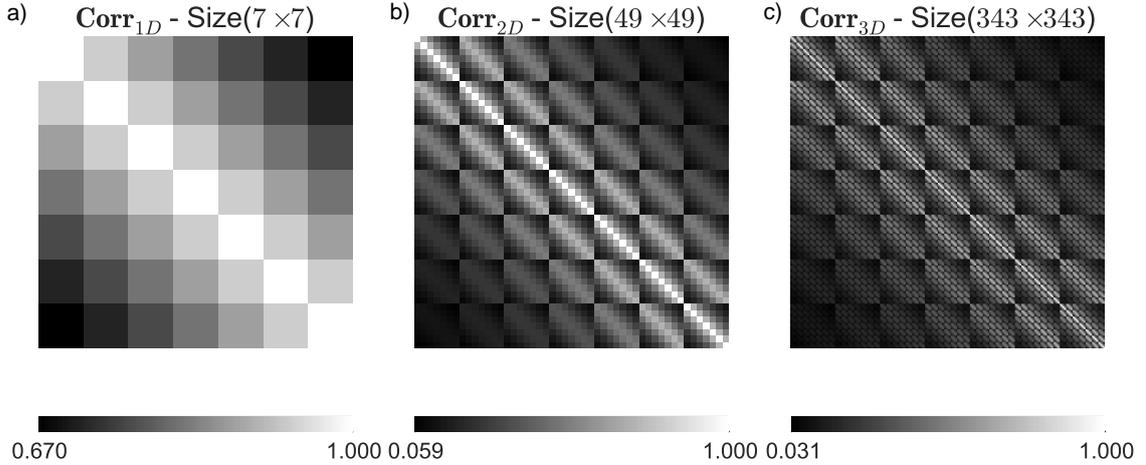
a) **Corr**$_{1D}$ - Size(7×7) b) **Corr**$_{2D}$ - Size(49×49) c) **Corr**$_{3D}$ - Size(343×343)

0.670   1.000 0.059   1.000 0.031   1.000

Figure 2.1: The correlation matrix describes how each point in the vector **u** is related to every other point. a) The number of elements N in the correlation matrix for a 1D vector of length $n = 7$ ($N = n^2$). b) The correlation matrix for a 2D vector with the dimensions 7×7 ($N = n^4$). c) The correlation matrix for a 3D vector with the dimensions $7 \times 7 \times 7$ ($N = n^6$).

model. If features are in the initial model, that do not kinematically match with what is in the true earth, then the inversion will likely not converge to the true solution. This behaviour can be particularly troublesome in sub-salt applications of FWI where the location and morphology of the salt is difficult to constrain. The incorrect application of regularization/preconditioning constraints that look to introduce or remove features that exist in the true earth, will prevent the inversion from succeeding. A simple example would be imposing a strong smoothness constraint on a model that has sharp and rapid variations. If the true earth contains high-wavenumber features, but the inversion is constrained to only focus on the low wavenumbers to maintain stability, then those high wavenumbers will not be recovered even if their imprint is in the data (Menke, 1984). I will propose to implement meaningful constraints based on the expected patterns and clustering of parameter values one can expect.

As a thought experiment we consider the special case of when we know the true earth velocity. We see in this case that much of the missing information comes from intermediate wavenumber contribution of the Valhall gas sands and stratigraphic layers. As the geology is relatively flat, most of the differences in the wavenumber spectrum are associated with intermediate to high vertical wavenumbers. The patterns we see in the $||\mathbf{v}_{p_{true}} - \mathbf{v}_{p_{initial}}||$ vector are also usually dominant in real earth models and relevant for
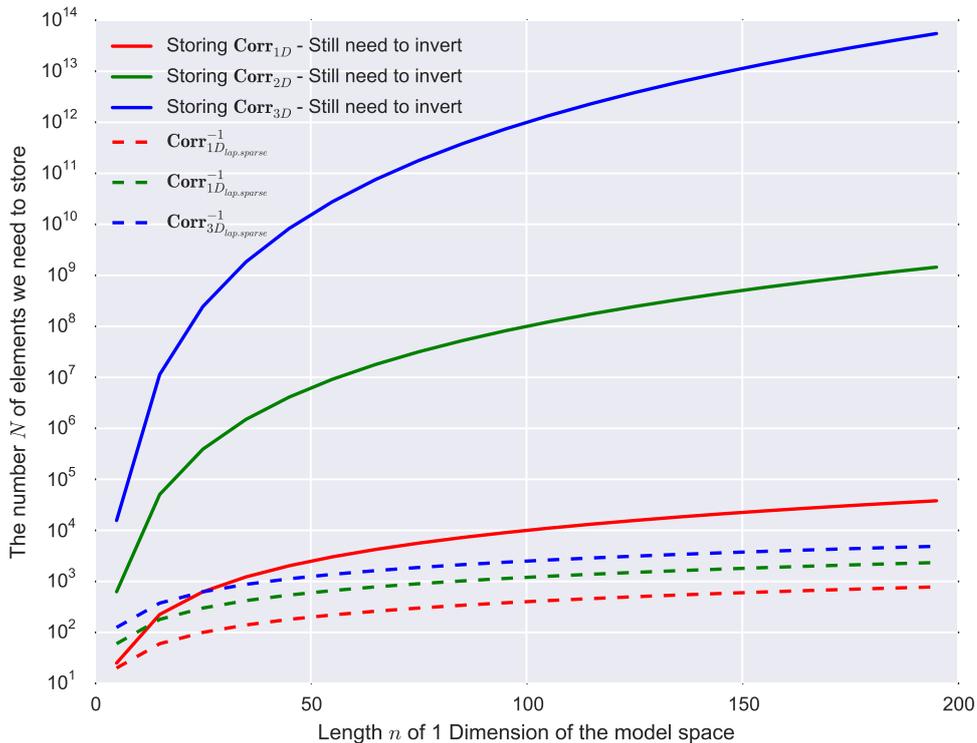
Figure 2.2: The memory required to store the correlation matrix **Corr**. We consider a spatial vector **u** of length n in 1D, length $n \times n$ in 2D and $n \times n \times n$ in 3D. If we want to store the correlation matrix (perhaps to calculate $\mathbf{Corr}^{-1}$), we see that as n increases the memory requirements increases $N = n^2$ for $\mathbf{Corr}_{1D}$, $N = n^4$ for $\mathbf{Corr}_{2D}$ and $N = n^6$ for $\mathbf{Corr}_{3D}$. This increasing memory complexity is shown by the solid lines. Our sparse inverse Laplacian memory requirements (dashed lines) require significantly less memory. Moreover, they do not need to be inverted for estimating the effect of the operator $\mathbf{Corr}^{-1}$.

many FWI applications. When we precondition or regularize our FWI we would like to emphasize this preferred wavenumber orientation in the hope it can help to mitigate the ill-posed nature of the inversion. To design the correct operators, we need to understand the natural, spatial clustering (also known as spatial dependence) that typically occurs in nature and how this relates to the addition of higher wavenumber components as we look do decrease $||\mathbf{v}_{p_{true}} - \mathbf{v}_{p_{final}}||$. It is important to understand why and how, these operators can and should be related to the wavenumber spectrum. To highlight some key concepts we look at the field of geography and geology. The natural, spectral behaviour of properties such as $\mathbf{v}_p$ from well measurements are of importance also.

To highlight some features of geological clustering we will consider the Valhall synthetic model previously introduced in Chapter 1 (Figure 2.3). One of the key promises of FWI is the theoretical ability to go from a kinematically accurate low wavenumber representation of the earth ($\mathbf{v}_{p_{initial}}$) to the true earth model ($\mathbf{v}_{p_{true}}$). The inversion proceeds, by adding progressively higher wavenumbers. In the examples we discuss, we expect that the low-wavenumber initial model ($\mathbf{v}_{p_{initial}}$) is sufficiently accurate that we do not suffer from cycle skipping. Starting from a kinematically accurate initial model, FWI minimizes the measured data misfit $||\mathbf{d}_{obs} - \mathbf{d}_{mod}||$ in the hope that our final inversion result ($\mathbf{v}_{p_{final}}$) minimizes $||\mathbf{v}_{p_{true}} - \mathbf{v}_{p_{final}}||$. Due to issues with the ill-posed nature of the inversion, the link between these goals is not always perfect. Artifacts from outside influences such as, noise and the limitations of the imaging condition can be introduced into our model parameter reconstruction.

2.1.1 Spatial dependence and spatial autocorrelation

When looking at the clustering of points, we start by looking at two key observations in the field of geography (Longley et al., 2001). It is interesting to examine the field of geography as it is a highly visual, yet statistical discipline that looks at the spatial distribution of attributes (Haining, 2009). Two observations are of particular relevance in relation to the principles of correlation and the goal of minimizing $||\mathbf{v}_{p_{true}} - \mathbf{v}_{p_{initial}}||$.

The first quote observation made by Walter Tobler (Tobler, 1970) is often referred to as the first law of geography. It states that:
"Everything is related to everything else, but near things are more related than distant things" - (Tobler's First law of geography) (Tobler, 1970)

Examples of this phenomena are numerous, such as the spatial distribution of rainfall or of property prices. Tobler (1970) made his observation while discussing the patterns of urban growth in the city of Detroit, USA. Although the exact mechanisms that cause this 'clustering' may or may not be understood, the observation that clustering exists has important implications for extrapolating away from given data points (i.e. away from rainfall measurement locations). The spatial distribution of property prices is another interesting example of this phenomena (Figure 2.4). Spatial trends often exist in property prices, for instance, the increase in cost per square meter in and around Paris, The French Riviera and the south west coast. The reason why these specific regions are expensive are often well understood (good weather, high wages), while the points around
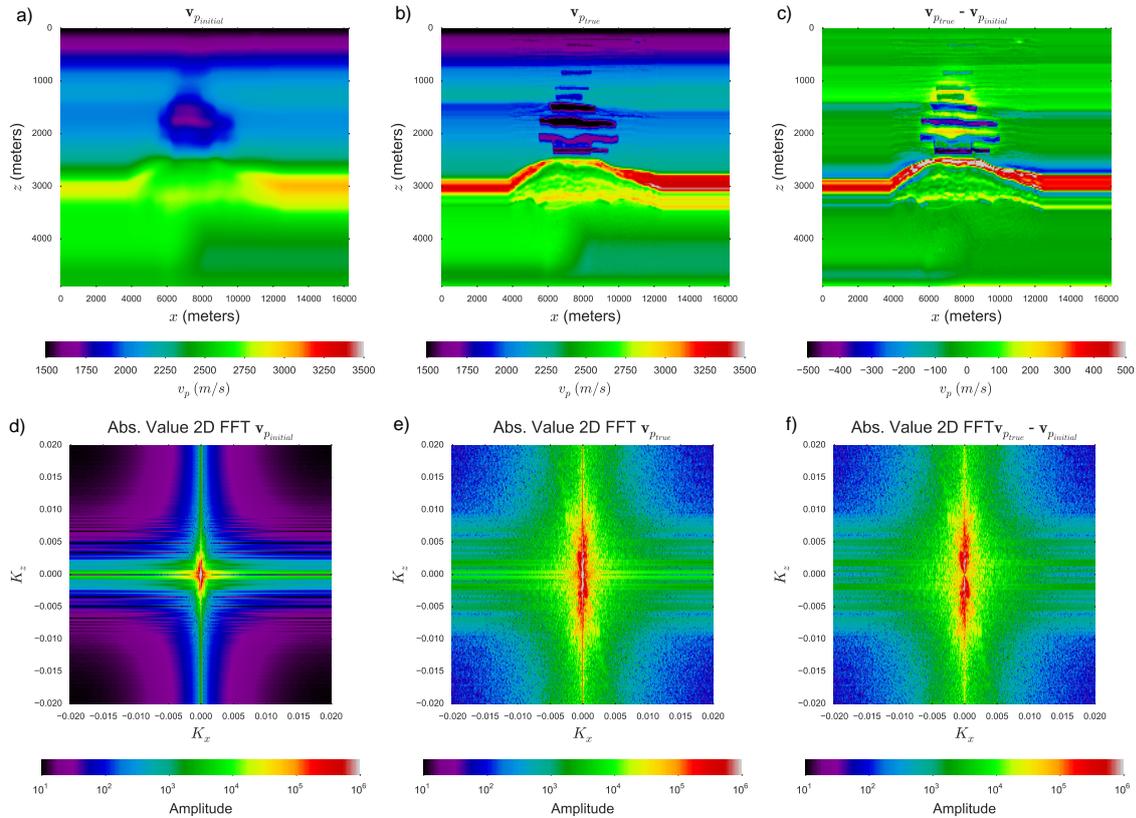
Figure 2.3: a) The initial velocity model ($\mathbf{v}_{p_{initial}}$) used for the Valhall synthetic FWI example. b) The true velocity model ($\mathbf{v}_{p_{true}}$) that we hope to recover from FWI c) The difference between the initial and true velocity model ($\mathbf{v}_{p_{true}} - \mathbf{v}_{p_{initial}}$). The bottom row shows the wavenumber spectrum for d):$\mathbf{v}_{p_{initial}}$ e)$\mathbf{v}_{p_{true}}$ f)$\mathbf{v}_{p_{true}} - \mathbf{v}_{p_{initial}}$.

these desirable locations are expensive due to their proximity to more desirable locations.

The second observation details how much more different and difficult scientific analysis would be in the absence of such clustering phenomena. It states:

"Hell might be a world without spatial dependence since it would be impossible to live there in any practical and meaningful way" - (Longley et al., 2001)

A simple thought experiment that is highlighted by individuals from the Geography/GIS discipline is to imagine a world where the earth's topography had limited, to no spatial correlation. In such a world, tiny mountains and valleys could exist. It is absurd to imagine such a scenario and this is because as Tobler states this would defy the first law of geography.
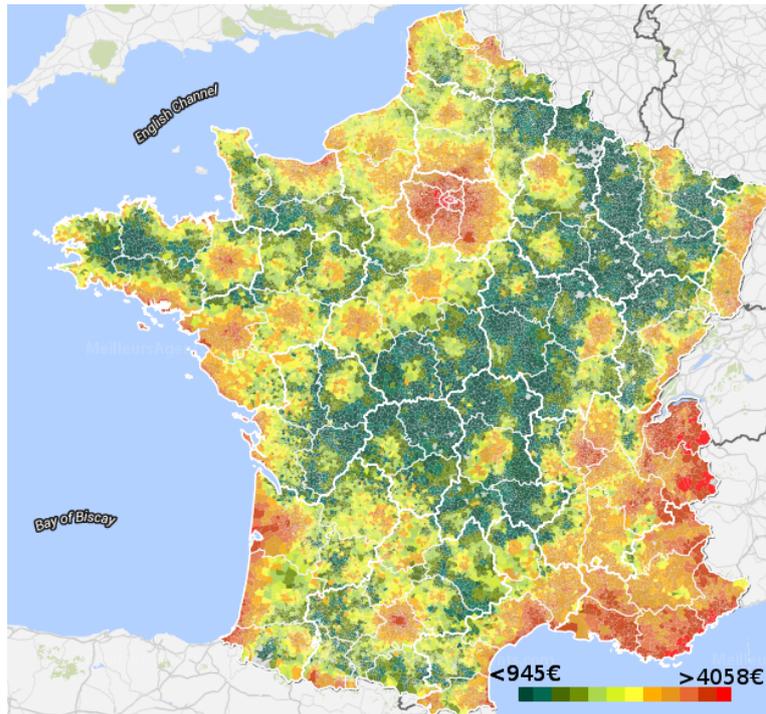
Figure 2.4: A map of the price of French property per m^2. The spatial distribution of property prices is on of many phenomena that obey Tobler's first law of geography. (*Data from meilleursagents.com*)

There are measures that exist and are commonly used in the field of statistical geography, to quantify how much spatial features are correlated. One of the commonly used methods (Longley et al., 2001) is Moran's I (Moran, 1950)

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(u_i - \bar{u})(u_j - \bar{u})}{\sum_i (u_i - \bar{u})^2} \tag{2.1}$$

Moran's I represents a way of quantifying clustering and is defined as a measure of spatial autocorrelation. The values of I are bounded between two extremes -1 and 1.

If we consider a spatial vector, $\mathbf{u}$ with n points, we can quantify how similar values are at one point u_i vs an offset point u_j. This differencing is done relative to the average value $\bar{u}$ of the vector. The weighting function $w_{i,j}$ can be of many forms. In our case we will look to remove the contribution of locations when u_j is not the direct neighbour of u_i (Figure 2.5). Phenomena can have either a positive spatial autocorrelation, a negative spatial autocorrelation or no autocorrelation whatsoever. Examples of these
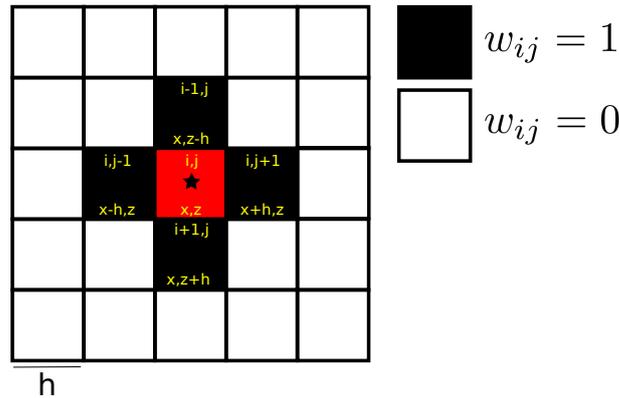
Figure 2.5: The weighting function $w_{i,j}$ used in Figure 2.6. When calculating the Moran's I function, we look at how similar the attributes surrounding the red cell are. It is important to note that, in our case, we only consider the 4 neighbouring cells as all other cells have a weighting value of zero.
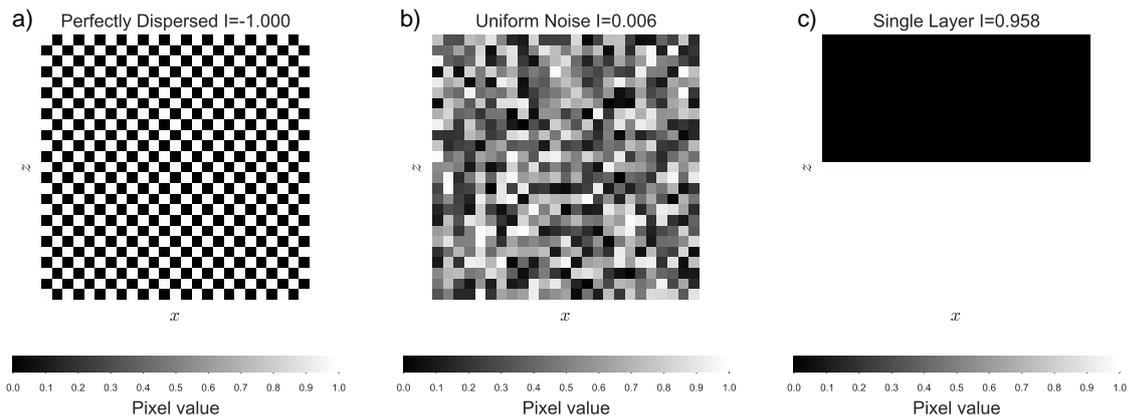

Figure 2.6: The end members for Moran's I classification a) A negative value for Moran's I suggests no value clustering and a perfectly dispersed binary pattern. b) Noise will have no spatial pattern and as a result has a Moran's I value approaching zero. c) An example with two layers is highly clustered and has a positive Moran's I value.

end members are shown in Figure 2.6. Random noise has no spatial pattern and and such has is spatially independent. Moran's I in the case of noise will be approximately zero. A negative Moran's I value occurs when the opposite of clustering occurs and we have a perfectly dispersed pattern (like on a chess board). This dispersed end member can also be seen as periodic. A positive value of I suggest we have clustering and a strong spatial dependence of parameter values. This is shown in the two layers case.

To statistically understand the spatial dependence related to our FWI application
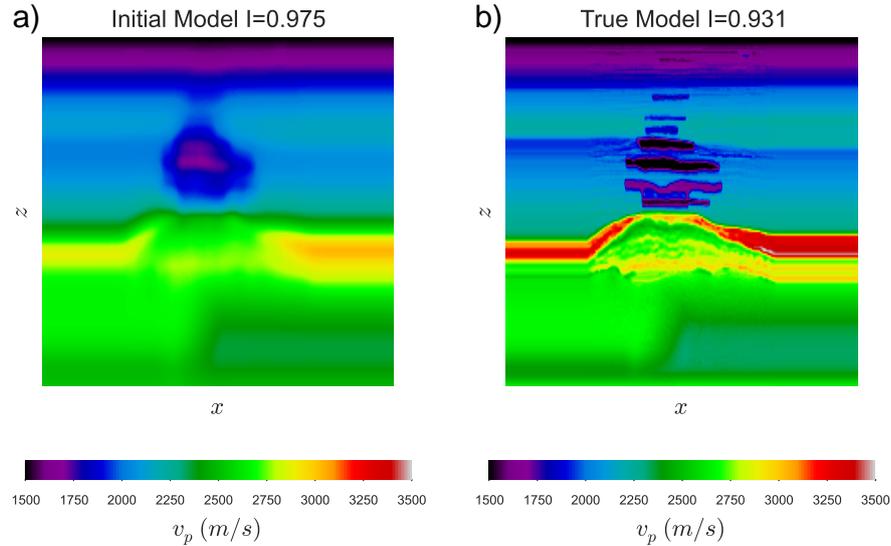
Figure 2.7: Moran's I classification applied to the Valhall model a) $\mathbf{v}_{p_{initial}}$ b) $\mathbf{v}_{p_{true}}$. Note that both have a strong degree of clustering as can be identified from visual inspection.

we perform the same Moran's I calculation on our Valhall model (Figure 2.7). It can be seen that we have a positive autocorrelation in both our $\mathbf{v}_{p_{initial}}$ and $\mathbf{v}_{p_{true}}$. This of course can be seen visually by eye but it is important to note that usually in geology and by extension in our velocity models there is often a strong clustering of parameter values. While it is typically not geological to have highly dispersed or random patterns in our velocity model.

The reasons why these observations are typically true can be highlighted in the study of stratigraphy and one of the key laws of geology, Walther's law (Walther, 1894; Middleton, 1973).

2.1.2   Facies and Walther's law.

Sedimentary stratigraphy is one of the key methods in geology and is a tool of significant importance for understanding spatial distribution of reservoirs in the oil and gas industry. The study of sedimentary stratigraphy examines the phenomena of layering (strata) in rocks and how this layering varies laterally (stratification). A facies in stratigraphy refers to a body of rock with similar characteristics (examples are a sandstone, shale or a chalk formation). Walther's law states that:

In a conformable vertical succession, only those facies that can be observed laterally adjacent to one another can be superimposed vertically.

This observation puts an inherent limitation on the rate rock types (facies) can vary spatially and in depth. The law forces one to consider the depositional origin of each particular rock unit. When considering a facies that is encountered in a vertical section you need to consider the depositional environment under which it was deposited. An example of this sort of thinking can be made when looking at the Valhall example. Much of the overburden at Valhall is made up of fine claystones and siltstones (Leonard and Munns, 1987). This rock type involves grains as small as 4μm. The deposition of these grains occurs when they are transported by rivers away from where they were eroded. The grains can only settle out of suspension in a large stationary body of such as a lake or an ocean (Selley, 2000). As lakes and oceans are typically quite large, Walther's law dictates that around a claystone unit there is likely to be more claystone that will have similar properties and velocity. As geological time (measured in the millions of years) passes the sea level changes and the depositional environments will change perhaps to more complex and variable environments, however knowledge of a particular rock type encountered at a depth, can give clues to how the rocks and by extension velocity field locally around it should behave. A caveat one needs to consider with regards to Walther's Law is that Walther's law only covers the deposition of the strata. After the strata have been deposited there can be other geological effects that occurs such as folding (as is seen at the reservoir level in Valhall), erosion, faulting and rotation (seen in the Marmousi and NWA examples), or intrusion by salt or igneous rocks. These post depositional events can complicate the velocity field. Oil companies exploring for oil and gas within sedimentary basins are well placed to try attempt to understand the level of variation. Much of the oil exploration business in offshore areas involves performing extrapolation of well-based stratigraphic interpretations over large distances using seismic imaging and seismic interpretation methods (Brown, 2004).

2.1.3 Interface and absolute attributes and their spectral characteristics

Our Moran's I analysis of the Valhall model and experience highlighted that subsurface velocity models often exhibit clustering. Walther's law gives clues to the origin of this clustering in sedimentary geology. FWI as opposed to geography/geology can be seen as

more applied, quantitative data fitting tool. As such, instead of just asking if clustering exists we may wish to know quantitatively how much more important are points nearer to a given point than the surrounding points. Well data can help to statistically address this question.

In an oil and gas context seismic data are often acquired to attempt reservoir characterization between wells (Mavko et al., 2009; Avseth et al., 2005) and as a means to decide the best place to drill new wells (Brown, 2004). When the decision to drill wells is made, they often give one an opportunity to acquire very important and precise information. One of the key datasets frequently acquired is wire-line log measurements. The v_p, v_s and ρ logs are of particular interest as they are often represent the parameters we are trying to reconstruct during FWI (in this thesis we will only focus on the reconstruction of v_p). I will refer to these parameters as absolute attributes due to the fact they are actual physical values we look to obtain. We can use these acquired absolute attributes to perform the forward modelling of the 1D reflectivity at the well location. The 1D modelled reflectivity can be seen as an interface attribute. This is due to the fact that although the physical mechanism that describes the partitioning of energy (reflection/transmission) is well understood (Zoeppritz, 1919), the reflectivity value only gives information about the presence of an impedance contrast at given interface, not directly the values of absolute value of impedance. If there is no contrast there is no reflectivity. This is quite different to absolute attributes. Put more simply there is always a velocity or impedance value whether there is a contrast in impedance or not.

The modelled reflectivity at the well is often compared to migrated images to assess the quality of the seismic processing and seismic imaging. The process of comparing the two is known as the seismic-to-well-tie (White and Simm, 2003). Unlike the surface seismic method that is used to position the wells, these measurements at the well are made quite precisely and also quite densely (typically every half a foot) and provide a useful dataset to understand the patterns that occur in both the absolute measurements (v_p, v_s and ρ) and the processed reflectivity.

When we examine the spectral content of the modelled reflectivity (Kazemeini et al., 2008) or alternatively the impedance and velocity spectrum (Lancaster and Whitcombe, 2000; Stefani and De, 2001) we see that the well data suggest that both quantities have interesting and consistent spectral patterns. These patterns in the amplitude spectrum are typically described using a colour based nomenclature. The concept is based on the
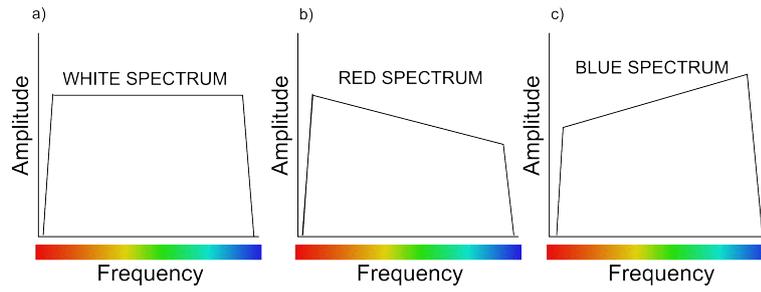
Figure 2.8: The typically used colour based description of the shape of amplitude spectrum. a) A "white spectrum" has approximately equal contribution of all frequency. b) A red spectrum has higher contribution of the lower frequencies as opposed to the higher frequencies. c) A blue spectrum has a higher contribution of higher frequencies. The slope of the curve is typically related to a power-law.

visible spectrum of electromagnetic radiation and described in Figure 2.8. The amplitude of the reflectivity spectrum for instance is typically assumed to have a trend of being on average flat and as such many seismic processing techniques assume a "white reflectivity spectrum" (Yilmaz, 1987). This white term comes from the idea that all frequency components have an equal amplitude. This is much like how white light is composed (equal contribution from all colours). If we analyse the frequency spectrum of the 1D well reflectivity we see, however it has a weak trend of increasing amplitude from low to high frequencies/wave numbers (Figure 2.9). This trend is often used to describe the amplitude spectrum of reflectivity being "blue". If we look at the absolute attributes (i.e. v_p, v_s and ρ) we see that we also have a much stronger trend but in the opposite direction. This "red trend" is typically stronger than the "blue trend" we see in reflectivity logs and provides us with an important corollary between absolute attributes and Tobler's first law. Another way of expressing this in the context of absolute attributes is that for a given point in space all wave numbers contribute but the low wave numbers contribute the most.

We can also see this behaviour when we look at the Valhall synthetic example. The FFT of 1D traces taken in the horizontal and vertical direction show that both the $\mathbf{v}_{p_{true}}$ and $\mathbf{v}_{p_{initial}}$ have an associated red spectrum (Figure 2.10). It makes sense that when we perform constraints on our full waveform inversion we should try to preserve this hierarchy of the importance of lower wave numbers vs the higher wave numbers. This forms the basis for the operators I will look to build.
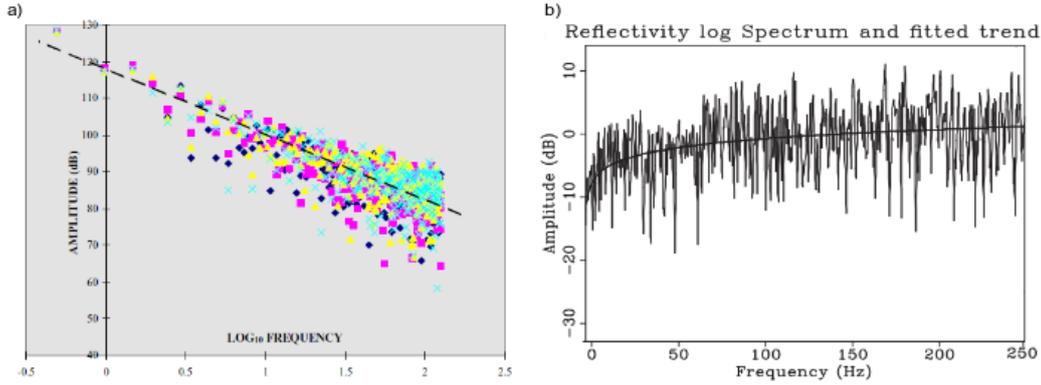
Figure 2.9: a) A red impedance spectrum *Figure taken from (Lancaster and Whitcombe, 2000)* and b) a blue reflectivity spectrum *Figure taken from (Kazemeini et al., 2008).*
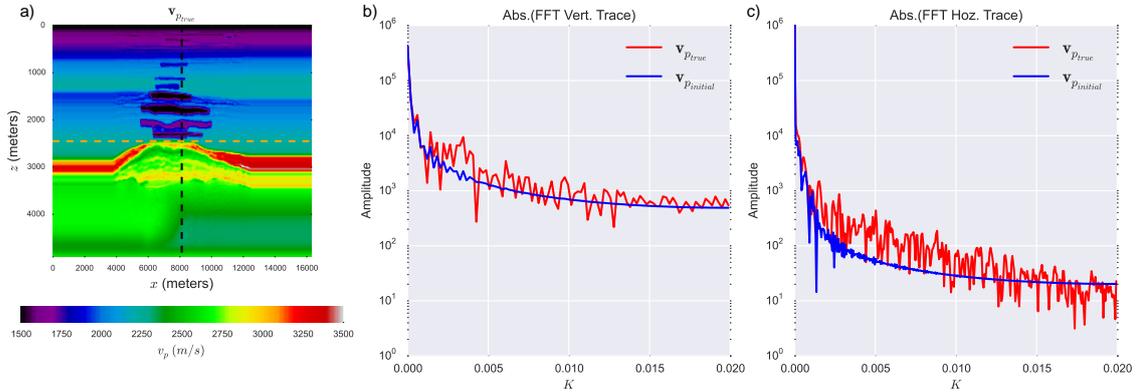

Figure 2.10: a) The Valhall synthetic model with a vertical b) and horizontal c) trace extracted. The plot s in b) and c) show the wavenumber spectrum of the true (red) and initial model (blue).

2.2 The correlation and covariance functions

The previous section details arguments on how sedimentary geology often includes a number of common spatial patterns. These patterns are of importance for FWI investigations as our inversion should be constrained to these commonly observed patterns. Firstly, in general, we should expect parameter values to be clustered as opposed to exhibiting a non-local/periodic behaviour. The basis for this clustering comes from one of the key principles in sedimentary geology, Walther's Law, which limits the variability of of rock types at a given point in space. (Lancaster and Whitcombe, 2000; Stefani and De, 2001) add to the concept of clustering and suggest that the wavenumber spectrum

55

of velocity and impedance logs obey a power law

$$A \propto K^{-\beta}, \tag{2.2}$$

. where the amplitude, A is related to the wavenumber, K with a slope β.

We should use these observations in the context of statistics and more specifically geostatistics to apply meaningful constraints to FWI. When we are looking at the regular spatial grid we often use in FWI applications, the functions that describe how changes at one point should correlate to another can be quite useful.

The correlation function between two points is perhaps one of the simplest geostatistical concepts to understand. Such a function will quantify the coupling between how changes at one point influences changes at an offset point. If we consider a vector, $\mathbf{u}$ which, in this case, we will take to be an extracted vertical trace from the Valhall true model ($\mathbf{v}_{p_{true}}$), then the correlogram is built by comparing the elements of the vector $\mathbf{u}$ to one shifted by a space lag, L ($\mathbf{u}_L$). Figure 2.11 shows that if one constructs a scatter plot of the values of $\mathbf{u}$ which we refer to as the head (the vector without shift) vs the tail $\mathbf{u}_L$ (the vector with shift) the linear correlation coefficient describes how points correlate over this length L. When $L = 0$ we have a correlation coefficient equal to 1 but, as L increases, there is less correlation between the head and tail of the vector. Figure 2.11 shows this scatter plot for 5 different correlation lengths ($L = 25$ m, $L = 125$ m, $L = 250$ m, $L = 500$ m and $L = 1250$ m).

The correlation function is constructed by performing this analysis for a continuum of correlation lengths. When we look at the correlation function from our Valhall extracted trace (Figure 2.12) we see the same trend of decreasing correlation coefficient with increasing correlation length L.

The correlation function, $Corr$ is closely related to the covariance function Cov

$$Corr(L) = \frac{cov(\mathbf{u}, \mathbf{u}_L)}{\sigma_{\mathbf{u}}\sigma_{\mathbf{u}_L}} \tag{2.3}$$

It can be seen that that the correlation function is the covariance function divided by the product of the standard deviations of the vectors ($\sigma_{\mathbf{u}}\sigma_{\mathbf{u}_L}$).
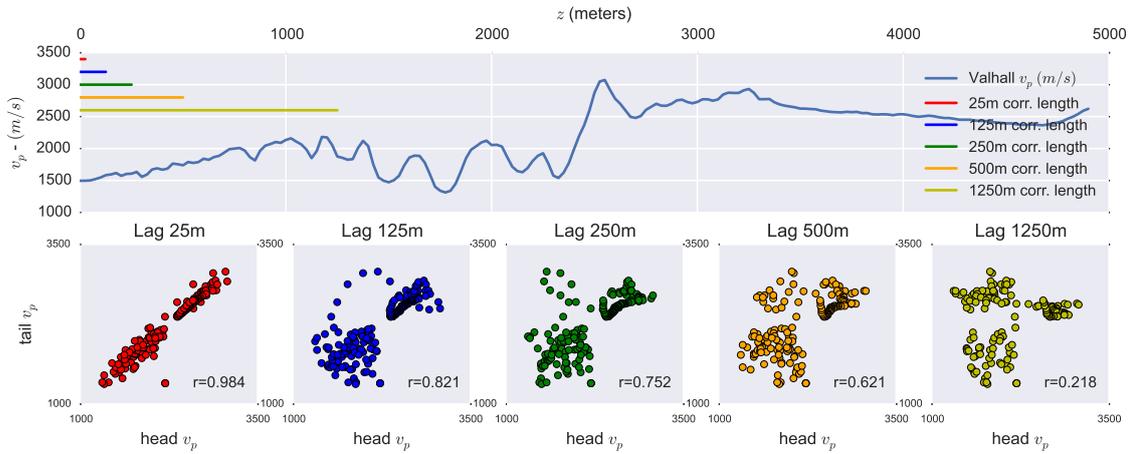
Figure 2.11: Calculation of correlation coefficients for head of the correlation vector compared to the tail. 5 lengths of correlation vector are shown (25 m, 125 m, 250 m, 500 m and 1250 m). The correlation vectors can be seen as a sliding window where the point on the LHS (head) of the vector is compared to the RHS (tail). As the length of this vector increases, the linear correlation coefficient r between the head and the tail decreases.
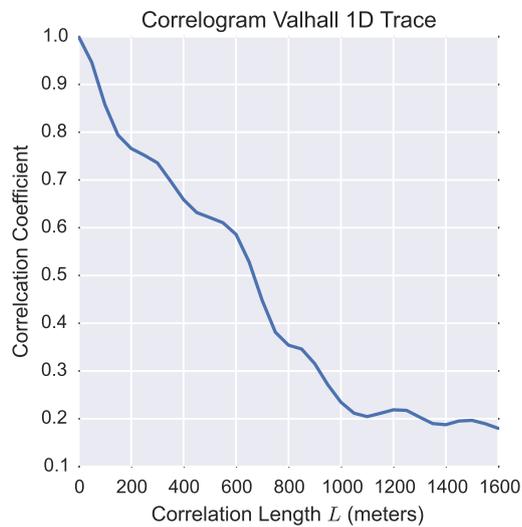

Figure 2.12: The correlation function calculated for increasing correlation length. Note when the correlation length L is zero, the correlation between the head and tail is equal to 1. As the correlation length increases this correlation coefficient decreases.

2.2.1   1D analytical Laplacian covariance and correlation functions

Calculation of the correlation function and the covariance function from the real data as shown previously is quite a straightforward proposition. In FWI we are hoping to reconstruct $\mathbf{v}_{p_{true}}$. It would be preferable to perform this reconstruction with a constraint based on the correlation function. If we continue considering our 1D vertically oriented example, we are interested in understanding for a given point z, how it should be influenced by a point offset from z by a space lag. We refer to this point as the space-lagged point, z'.

There are numerous different potential covariance functions that exist but, in our case, we will be concentrating on the Laplacian covariance function. This covariance is also known as the double exponential covariance

$$Cov_{1D}(z; z') = \sigma^2 e^{-\frac{|z-z'|}{L}}, \tag{2.4}$$

where σ is the standard deviation. The correlation function can then be constructed by dividing by the variance σ^2

$$Corr_{1D}(z; z') = e^{-\frac{|z-z'|}{L}}. \tag{2.5}$$

Normalization

Both the covariance function and the correlation function are of interest. However, for much of the discussion in this chapter, I will focus on the correlation functions. As mentioned previously the covariance function can be obtained by multiplying the correlation function by σ^2. Our desired purpose for this correlation function is to use it to modify the wavenumber spectrum of an input vector. If we have an input vector (we will consider our familiar Valhall 1D vertical trace) and we want an application of our 1D correlation operator $Corr_{1D}(z)$ to our sampled vector $u(z)$, we need to do so by 1D convolution expressed as

$$Corr_{1D}(z) \star u(z) = \int_{-\infty}^{\infty} Cor_{1D}(z, z')u(z')dz'. \tag{2.6}$$

If we apply this process we will find that although we may decrease the influence of the higher wavenumbers of our vector, there will also be scaling for the values of our vector.
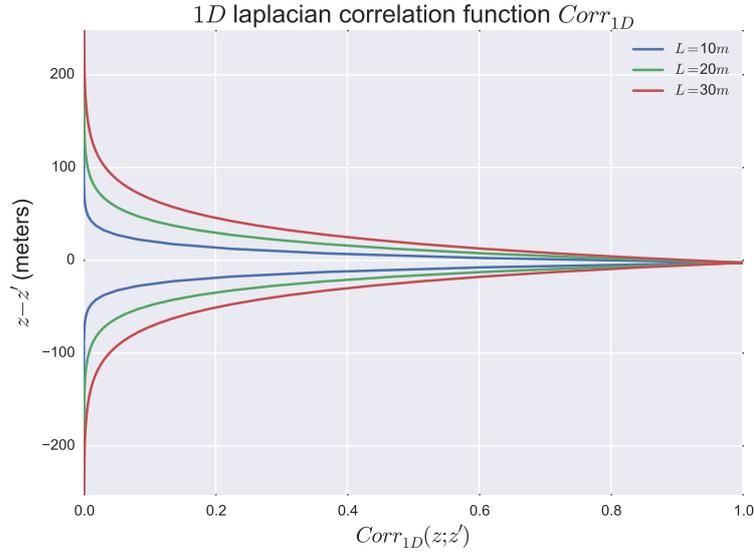
Figure 2.13: Correlation based on the Laplacian distribution described in equation 2.5. The correlation function can also be seen as a special case of the covariance function equation 2.4 when $\sigma = 1$

The values in our vector have a specific significance as they represent absolute velocity values in m/s. To avoid this anomalous scaling we need to normalize our correlation function to ensure it is unit-less, for all values of L. We are able to find the appropriate normalization by finding the scalar a that makes

$$\int\limits_{-\infty}^{\infty} a e^{-\frac{|z-z'|}{L}} dz' = 1 \tag{2.7}$$

true for all values of L.

If we perform this integration analytically, we find that we require a normalization factor of $\frac{1}{2L}$. Our normalized Laplacian correlation function $\overline{Corr_{1D}}(z; z')$ is therefore

$$\overline{Corr_{1D}}(z; z') = \frac{1}{2L} e^{-\frac{|z-z'|}{L}}. \tag{2.8}$$

The behaviour of this normalized operator (Figure 2.14) can be contrasted against the earlier defined operator that is without normalization (Figure 2.13). We see the normalized correlation function $\overline{Corr_{1D}}(z, z')$ will behave as a normalized filter. I define a normalized filter as one where the amplitude spectrum of the filter varies between 1
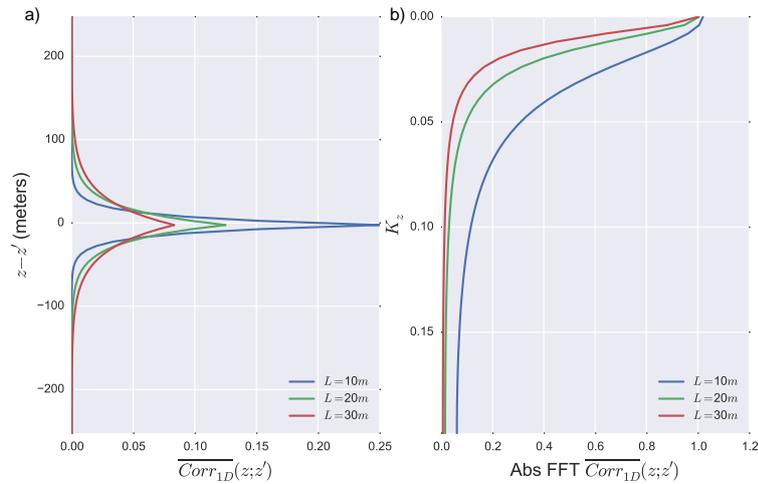
Figure 2.14: The normalized 1D Laplacian correlation function. a) Note at zero space lag $z = z'$ the $\overline{Corr_{1D}}(z, z') = \frac{1}{2L}$. With this correct normalization we can see that our normalized correlation function behaves like a well normalized low-pass filter b).

(full-pass) and 0 (full-reject). Such a filter will not apply a scalar shift to the absolute value of our parameter values but instead alter the wavenumber spectrum attenuating undesired wavenumber components. This normalized operator will also have the property of preserving the norm of the input vector. An additional point to note is that, when we are discussing the application of the operator via spatial convolution, our 1D data vector $\mathbf{u}$ is sampled by a discretization spacing h. We therefore have a discretized version of the equation 2.7 where our normalization factor, a is equal to $h/2L$.

2.2.2 2D analytical Laplacian correlation function

The covariance and correlation functions for the Laplacian distribution can be extended to 2D in a relatively straightforward fashion. In such an extension, we are considering how the point (x, z) is influenced by an offset point (x', z'). Extending equation 2.5 to 2D yields

$$Corr_{2D}(x, z; x', z') = e^{-\frac{\sqrt{(x-x')^2 + (z-z')^2}}{L}}. \tag{2.9}$$

We can see how this isotropic correlation function behaves in the figure 2.15. When there is no space-lag ($x = x'$ and $z = z'$) the function equals unity, however, as the euclidean distance
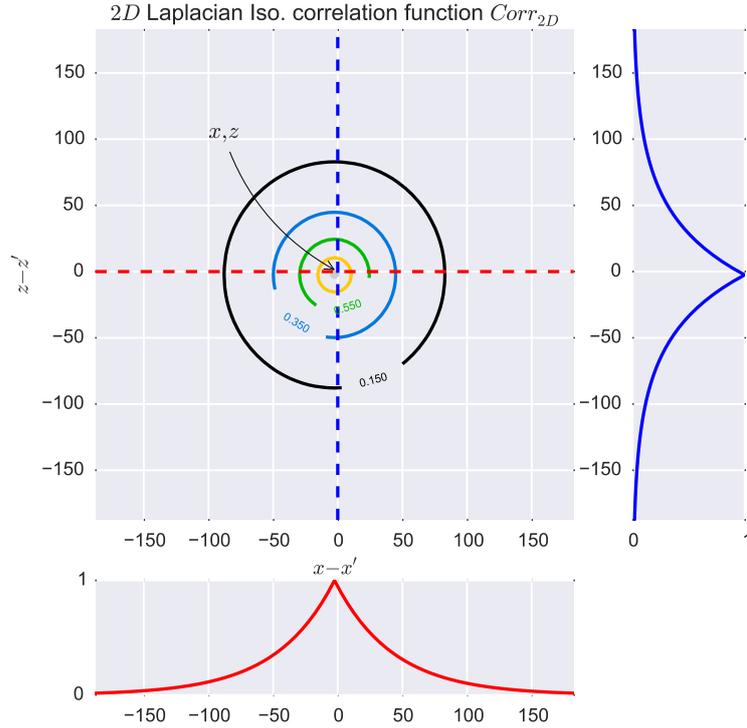
60

Figure 2.15: The 2D correlation function control how points that at some space-lag x', z' influence the points at x, z. When there is zero space lag ($x = x'$ and $z = z'$) the correlation coefficient is equal to 1. As we increase the space-lag the correlation decreases.

$$d = \sqrt{(x - x')^2 + (z - z')^2} \tag{2.10}$$

increases, the correlation weight decreases. This can be seen in the circular contours. One could normalize this function as we did in the 1D case. However, it may be that we would like to have correlation vary as a function of direction as is suggested in the Valhall example. The Valhall examples supports the earlier observations made with reference to clustering and Walther's law. In such a case, correlation lengths in the horizontal direction, should be longer than those in the vertical. When we have non-equal, orthogonal correlation lengths the operator can be said to be anisotropic. The anisotropic operator is expressed as

$$Corr_{2D_{aniso}}(x, z; x', z') = e^{-\frac{1}{L_x}\sqrt{(x-x')^2 + \alpha^2(z-z')^2}} \text{ Where } \alpha = \frac{L_x}{L_z}. \tag{2.11}$$
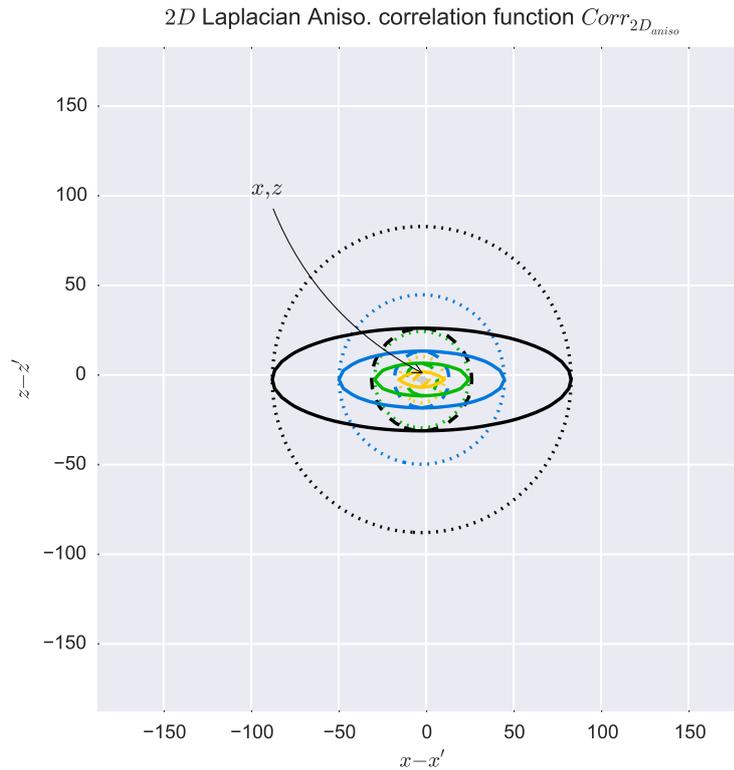
Figure 2.16: The 2D correlation function controls how nearby points (x', z') influence (x, z). The anisotropic correlation function (solid contours) allows different correlation lengths in orthogonal directions. In this case horizontal correlation length (L_x) is 3 times larger than the vertical correlation length. The dotted circle shows the isotropic correlation function for function for $L = 45$ m while the dashed shows the correlation function for $L = 15$ m. Note that the solid lines contours intersect the dotted contours in the x-direction and the dashed contours in the y-direction.

This anisotropic operator has correlation lengths, L_x and L_z orientated in the Cartesian x and z direction respectively. Figure 2.16 shows contours of the operator for a case where $Lx = 3Lz$. The anisotropic operator forms an ellipse. For the operator to be of use as a filter we need to normalize it as was done in the 1D case.

Normalization

In the 2D case as in the 1D case we are interested in the application of the 2D correlation operator to a data vector **u**

$$Corr_{2D_{aniso}}(x, z) \star u(x, z) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} Corr_{2D_{aniso}}(x, z; x', z')u(x, z)dx'dz'. \qquad (2.12)$$

For the correct application of this function we need a normalization scalar b. This scalar needs to satisfy

$$\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} be^{-\frac{1}{L_x}\sqrt{(x-x')^2+\alpha^2(z-z')^2}} dx'dz' = 1 \qquad (2.13)$$

for all values of L_x and L_z. Using analytical integration we find that the normalization factor b is equal to $\alpha/2\pi L_x^2$. Our normalized operator is

$$\overline{Corr_{2D_{aniso}}}(x, z; x', z') = \frac{\alpha}{2\pi L_x^2}e^{-\frac{1}{L_x}\sqrt{(x-x')^2+\alpha^2(z-z')^2}}. \qquad (2.14)$$

If we are applying this to a 2D vector then we need to include the spatial discretization of our vector in the normalization. If the vector is equally sampled at an increment of h in both direction then the coefficient $b = \alpha/\left[2\pi\left(\frac{L_x}{h}\right)^2\right]$. See Appendix A for this derivation.

Rotation

A rotated 2D Laplacian correlation function allows one to have orthogonal correlation lengths that are no longer aligned with the Cartesian grid. These correlation lengths are instead rotated by some angle θ (Figure 2.17). Such a manipulation is obtained by applying a rotation matrix $\mathbf{R}$,

$$\begin{bmatrix} \hat{x}' \\ \hat{z}' \end{bmatrix} = \mathbf{R} \begin{bmatrix} x' \\ z' \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} x' \\ z' \end{bmatrix}. \qquad (2.15)$$

The filtering characteristics of this anisotropic operator are summarized in Figure 2.18.
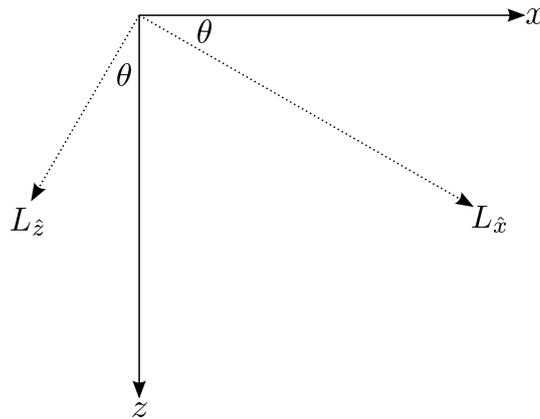
Figure 2.17: Rotated correlation length vectors. The angle, θ rotates the Cartesian correlation lengths L_x and L_z to $L_{\hat{x}}$ and $L_{\hat{z}}$
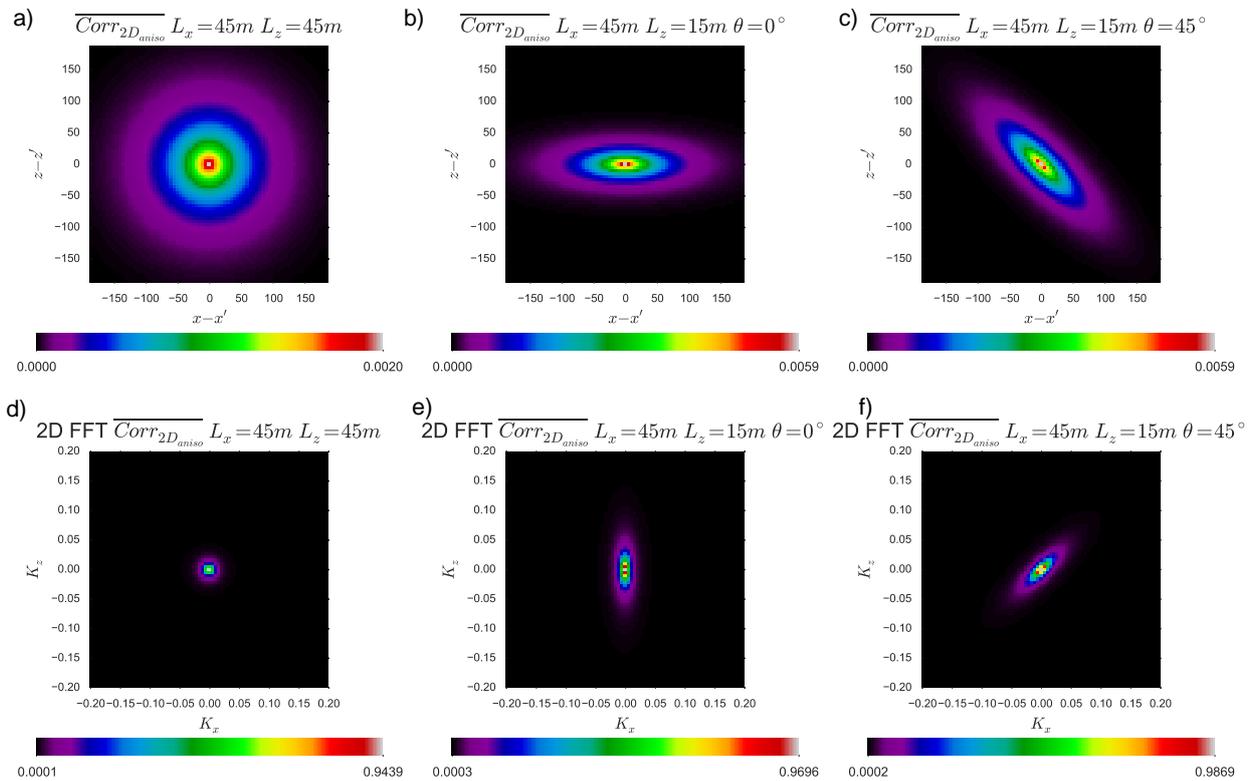
Figure 2.18: The impulse response of the 2D normalized anisotropic Laplacian correlation function (a,b,c) and the associated 2D FFT (d,e,f). When $L_x = L_z$ we have an isotropic filter (a,d), however we may want to have an anisotropic filter where $L_x \neq L_z$ (b,e) and if there is dip we may wish this to not be aligned with the geological dip (c,f)

2.2.3 3D analytical Laplacian correlation function

The extension of the correlation function to 3D is quite straightforward. We start from the expression of the 2D space-lag (eqn 2.9) an add an additional dimension to define the 3D euclidean distance.

$$d = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}. \tag{2.16}$$

If we would like the operator to be isotropic, then we can use the correlation operator

$$Corr_{3D}(x, y, z; x', y', z') = e^{-\frac{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}}{L}}. \tag{2.17}$$

Like the 2D case, however, it is desirable to be able to use different correlation lengths in orthogonal directions. The anisotropic, 3D correlation function is defined by the following expression

$$Corr_{3D_{aniso}}(x, y, z; x', y', z') = e^{-\frac{1}{L_x}\sqrt{(x-x')^2+\alpha^2(y-y')^2+\beta^2(z-z')^2}}. \tag{2.18}$$

Where the coefficients α and β are L_x/L_y and L_x/L_z respectively.

Normalization

In the 3D case as in the 1D and 2D cases we are interested in the application of the 3D correlation operator to a data vector **u** through the expression

$$Corr_{3D_{aniso}}(x, y, z) \star u(x, y, z) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} Corr_{3D_{aniso}}(x, y, z; x', y', z')u(x, y, z)dx'dy'dz'. \tag{2.19}$$

For the normalized implementation of this function we need a factor, c. This scalar needs to make the expression

$$\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} ce^{-\frac{1}{L_x}\sqrt{(x-x')^2+\alpha^2(y-y')^2+\beta^2(z-z')^2}}dx'dy'dz' = 1, \tag{2.20}$$

for all values of L_x, L_y and L_z. As was in the cases 1D and 2D, we calculate the normalization scalar by performing analytical integration of the correlation function.
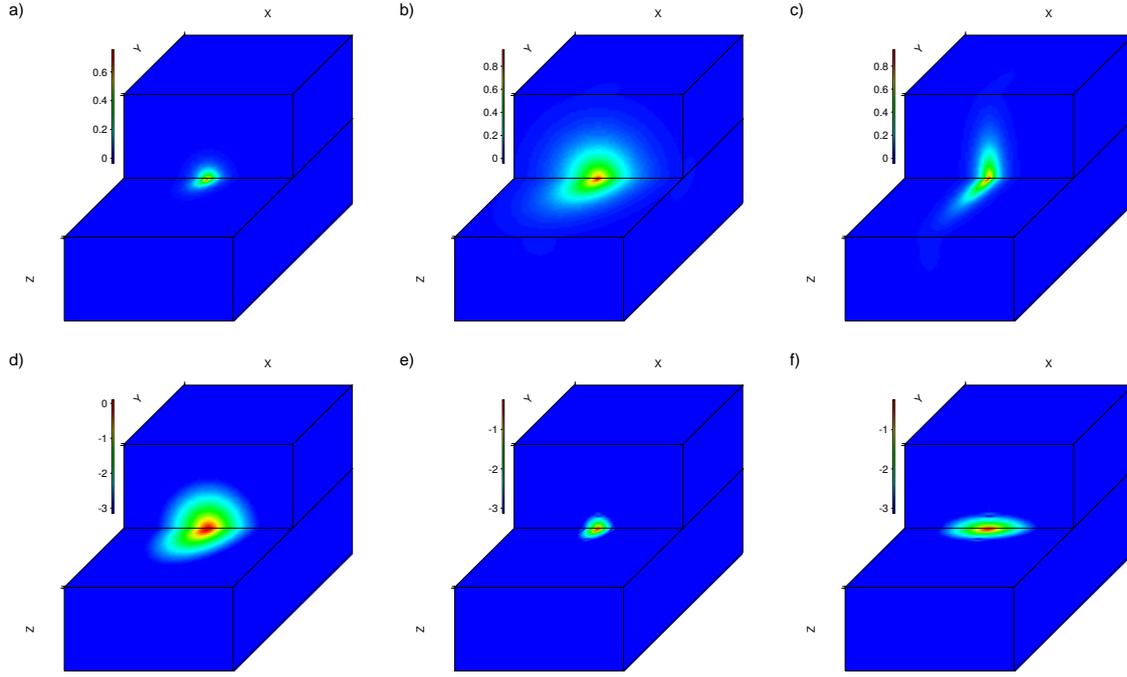
Figure 2.19: The correlation function for a 3D volume $75 \times 75 \times 75$. The top row (a,b,c) shows the 3D anisotropic correlation function $Corr_{3D_{aniso}}$ a) $L_x = L_y = L_z = 15$ m b) $L_x = L_y = L_z = 45$ m c) $L_x = 15$ m, $L_y = L_z = 45$ m. The bottom row (d,e,f) shows the 3D wavenumber spectrum of $\overline{Corr_{3D_{aniso}}}$ for the same L values as above. The display of $Corr_{3D_{aniso}}$ for the top row and $\overline{Corr_{3D_{aniso}}}$ for the bottom row is to make it easier the colourbar limits easier to choose. The bottom plot display the logarithm of the amplitude of the FFT).

When we do this, we find the coefficient c is equal to $8\pi/\alpha\beta L_x^3$. Therefore, our normalized 3D anisotropic correlation function becomes

$$\overline{Corr_{3D_{aniso}}}(x, y, z; x', y', z') = \frac{8\pi}{\alpha\beta L_x^3} e^{-\frac{1}{L_x}\sqrt{(x-x')^2 + \alpha^2(y-y')^2 + \beta^2(z-z')^2}}. \qquad (2.21)$$

The application of this 3D normalized correlation function can be seen in the figure 2.19.

If we are applying this to a vector in a spatial convolution work-flow, we also need to account for the sampling interval of the vector **u**. If your vector is equally in all 3 dimensions at an increment h, then the coefficient c is equal to $8\pi / \left[\alpha\beta\left(\frac{L_x}{h}\right)^3\right]$.

Rotation

To apply the rotation in the 3D case we use the 3 angles described by Euler (Goldstein,
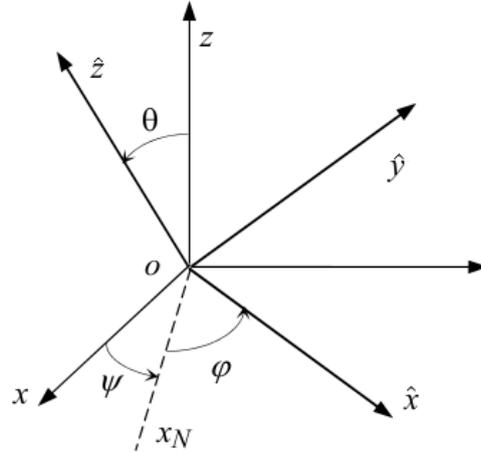
Figure 2.20: The Euler angles are the three angles used to describe the rotation of a rigid body in 3D Euclidean space. The rotations consist of 3 individual transformations, performed in cascade. The first is a rotation around z by an angle φ. This is followed by a rotation around x by an angle θ and finally a second rotation around z by the angle ψ.

1980) to describe the orientation of a rigid body (Figure 2.20).

As such we have a 3×3 rotation matrix

$$
\begin{bmatrix} \hat{x}' \\ \hat{y}' \\ \hat{z}' \end{bmatrix} = \mathbf{R} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}, \tag{2.22}
$$

where the constants $a \to i$ are:

$$
a = cos\psi cos\varphi - sin\psi sin\varphi cos\theta,
$$
$$
b = -cos\psi sin\varphi - sin\psi cos\varphi cos\theta,
$$
$$
c = sin\psi sin\theta,
$$
$$
d = sin\psi cos\varphi + cos\psi sin\varphi cos\theta,
$$
$$
e = -sin\psi sin\varphi + cos\psi cos\varphi cos\theta,
$$
$$
f = -cos\psi sin\theta,
$$
$$
g = sin\varphi sin\theta,
$$
$$
h = cos\varphi sin\theta,
$$
$$
i = cos\theta, \tag{2.23}
$$

The rotation matrix defined in expression 2.22 and 2.23 is composed of 3 rotations performed en-cascade. The first is the rotation around z by an angle φ, followed by a rotation around x by an angle θ and finally a second rotation around z by the angle ψ. With this rotation matrix defined one can calculate the rotated Euclidean distance, $\hat{d}$

$$\hat{d} = \sqrt{(x - \hat{x}')^2 + (y - \hat{y}')^2 + (z - \hat{z}')^2}. \qquad (2.24)$$

2.2.4 Limitations of using the analytical Laplacian correlation functions

We have presented the analytical correlation functions for the Laplacian distribution. The normalized operators can be potentially powerful constraints to the wavenumber spectrum. The application of such constraints becomes computationally intensive for applications in higher dimensions, using long correlation lengths and for the size of inversion models typically required in industrial applications of FWI (well in excess of 10^6 model parameters being inverted). Additionally, we find, that although we may have strategies for applying the application of the correlation matrix to a vector ($\mathbf{Corr}(\mathbf{u})$) the application of the inverse matrix ($\mathbf{Corr}^{-1}(\mathbf{u})$) remains a complex task that is attempted using a number of iterative approaches (Hsieh et al., 2011; Frieman et al., 2007; Guillot et al., 2012). Our focus will be restricted to the efficient application of $\mathbf{Corr}(\mathbf{u})$. This will be explored prior to moving to detail our sparse inverse Laplacian correlation framework.

Generalized spatial convolution of the correlation functions

We have discussed our normalized correlation functions $\overline{Corr_{1D}}$, $\overline{Corr_{2D}}$ and $\overline{Corr_{3D}}$ (we will refer to them for the moment as the generalized normalized Laplacian covariance operator $\overline{Corr}$). Our goal is to apply these operators to spatial vectors where the model parameters were hope to constrain are on a finite, discretized grid. For the moment we will refer to the data we wish to constrain as the vector $\mathbf{u}$. We will also make the distinction between the normalized correlation operator/matrix ($\overline{\mathbf{Corr}}$) shown in bold as opposed to correlation function $\overline{Corr}$. $\overline{\mathbf{Corr}}$ has a size equivalent to the mono-parameter Hessian and as such is quite cumbersome from both a memory and CPU perspective (Figure 2.1 and 2.2). The memory complexity issue is typically mitigated by virtue of the fact that operator $\overline{\mathbf{Corr}}$ is never explicitly formed and we use a spatial convolution approach. This spatial convolution approach can be generalized for any number of dimensions by the brute tensorial strategy shown in the following expression
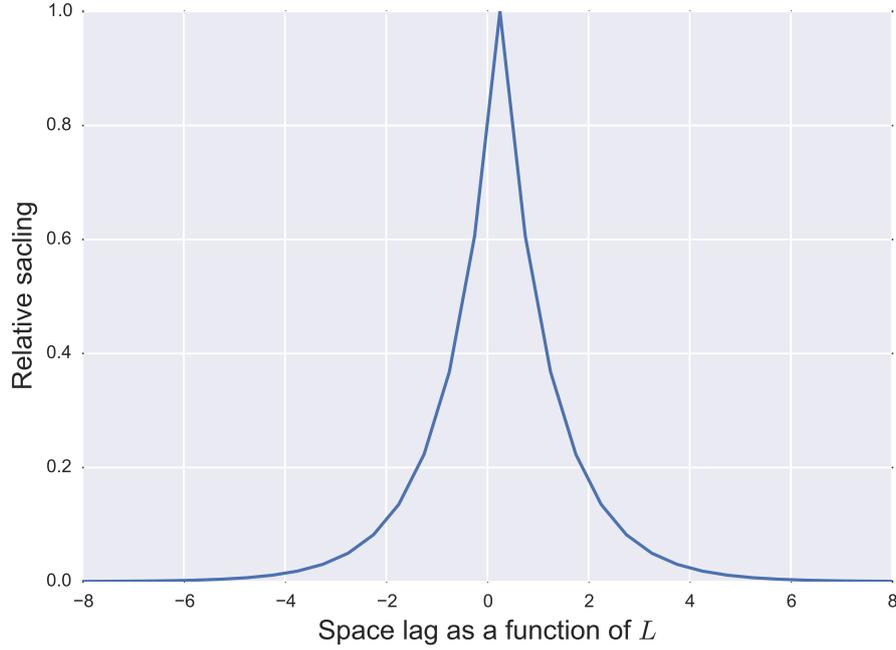
Figure 2.21: The rate of decay of the Laplacian correlation function. The space lag axis, represents the contribution of a point at x' to x for $x' = x + N \times L$ At $4L$ the scaling value is ≈ 0.02

$$\mathbf{u}_{new}(\mathbf{x}) = \int_{\Omega_\infty} d\tau' \overline{\mathbf{Corr}}(\mathbf{x}')\mathbf{u}(\mathbf{x}') \approx \int_{\Omega_{NL}} d\tau' \overline{\mathbf{Corr}}(\mathbf{x}')\mathbf{u}(\mathbf{x}'). \qquad (2.25)$$

Where the domain, Ω, represents the extents of the space we are referencing in $\mathbf{u}$. This may be a 1D length, a 2D area or a 3D volume. A simple way of decreasing the computation time is to not perform the spatial convolution for the entire domain Ω_∞. $\overline{Corr}$ decays rapidly (Figure 2.21) and as such we can perform the summation over a subset of the entire domain Ω_{NL} , where NL represents the number size of the space-lag we will sum over. Values of $4L$ or $8L$ can provide a good level of accuracy as the correlation weight for most correlation functions are quite low at this space-lag.

The spatial convolution approach is only effective at minimizing the memory requirement. Unfortunately, the computational burden is still significant. In 1D expression, 2.25 represents the 1D integration of $-NL$ to $-NL$. However in 2D, it is an area integral ($[2NL]^2$) and in 3D it becomes a volume integral ($[2NL]^3$) that needs to be performed for every element in the vector $\mathbf{u}$. To highlight how the cost of this operations scales we
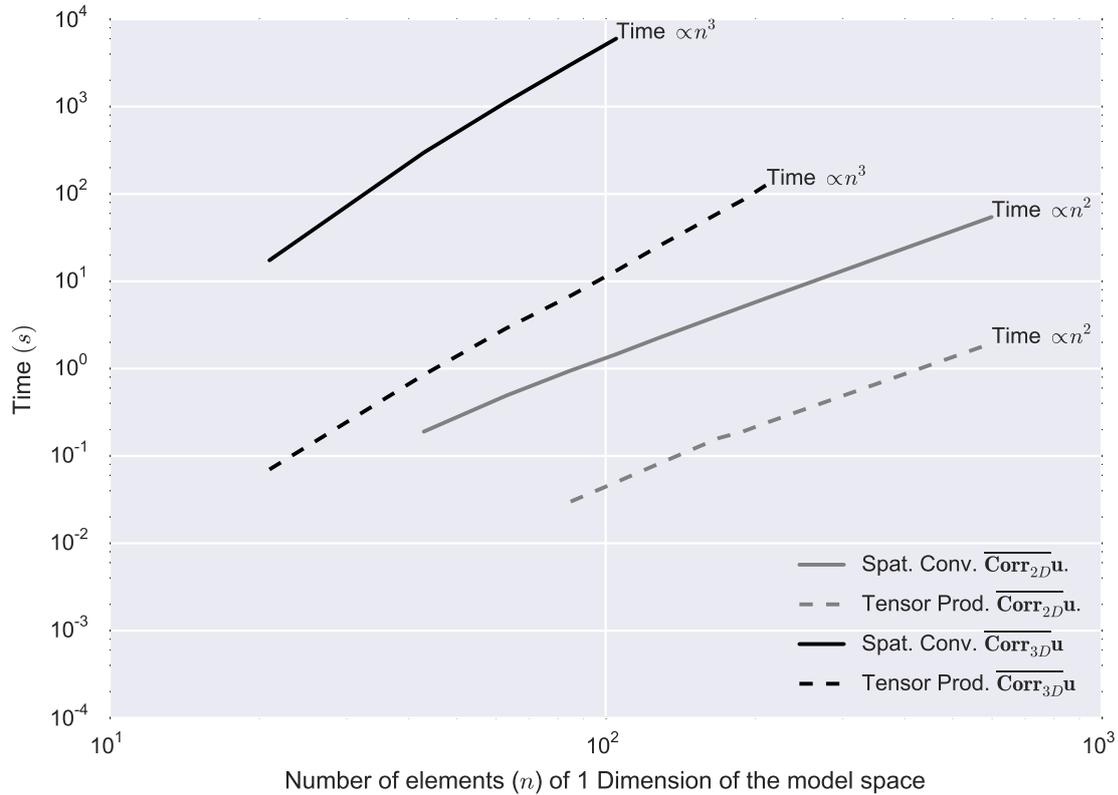
Figure 2.22: The time complexity for calculating $\int_{\Omega_L} d\tau' \overline{\mathbf{Corr}}(\mathbf{x}')\mathbf{u}(\mathbf{x}')$ in the 2D and 3D cases. The solid lines represents the cost for the spatial convolution approach. The dashed lines shows the equivalent cost for a tensorized operator.

.

consider the spatial convolution approach applied to larger and larger domains Ω for the 1D (nelems=n), 2D (nelems=$n \times n$) and 3D (nelems=$n \times n \times n$) cases (Figure 2.22). A second option for the application of a multi-dimensional operator is the tensorial product. The complexity of the tensorial product is also shown in (Figure 2.22).

The tensorial product takes advantage of the fact that for certain multidimensional functions, the spatial convolution process can be equivalently performed in two separate computationally less intensive steps. If we consider our familiar data vector $\mathbf{u}$ we see that if the operator can be tensorized, we can perform the spatial convolution via a tensorized approach expressed as

$$Corr_{2D}(x, z) \star u(x, z) \approx \int\limits_{-NL}^{NL} \int\limits_{-NL}^{NL} Corr_{2D}(x, z; x', z')u(x, z)dx'dz' =$$

$$\int\limits_{-NL}^{NL} Corr_{2D}(x, z; z')u(x, z)\left(\int\limits_{-NL}^{NL} Corr_{2D}(x, z; x')u(x, z)dx' \right) dz'. \qquad (2.26)$$

If we can perform the spatial convolution approach for every element in **u**, we only need to sum over $4NL$ points vs $(2NL)^2$ points in approaches without tensorization. In 3D this saving is even more significant as we need to sum over $8NL$ points as opposed to $(2NL)^3$. An example function that can be tensorized is the Gaussian function (Figure 2.23). The key point to observe is that

$$Corr_{2D_{GAUSSIAN-X}}(x, z; x') \star Corr_{2D_{GAUSSIAN-Z}}(x, z; z') = Corr_{2D_{GAUSSIAN}}(x, z; x', z').$$
$$(2.27)$$

Although it is not shown here, the 3D Gaussian operator can also be tensorized.

Unfortunately the Laplacian correlation function cannot be formed using a tensor approach (Figure 2.24). We see that

$$Corr_{2D_X}(x, z; x') \star Corr_{2D_Z}(x, z; z') \neq Corr_{2D}(x, z; x', z'), \qquad (2.28)$$

and that the function $Corr_{2D_X}(x, z; x') \star Corr_{2D_Z}(x, z; z')$ also does not necessarily represent an operator we should be using to constrain FWI. Looking carefully at the operator we find that

$$Corr_{2D_X}(x, z; x') \star Corr_{2D_Z}(x, z; z') \begin{cases} = Corr_{2D}(x, z; x', z') \text{ for } x - x' = 0 \\ = Corr_{2D}(x, z; x', z') \text{ for } z - z' = 0 \\ \neq Corr_{2D}(x, z; x', z') \text{ for } x - x' \neq 0 \text{ and } z - z' \neq 0. \end{cases}$$
$$(2.29)$$

When the operator is not aligned with the x ($z - z' = 0$) or z axis ($x - x' = 0$) the value of the correlation function is no longer directly controlled by the euclidean distance. In such a case, the wavenumber properties of the operator become strongly angle
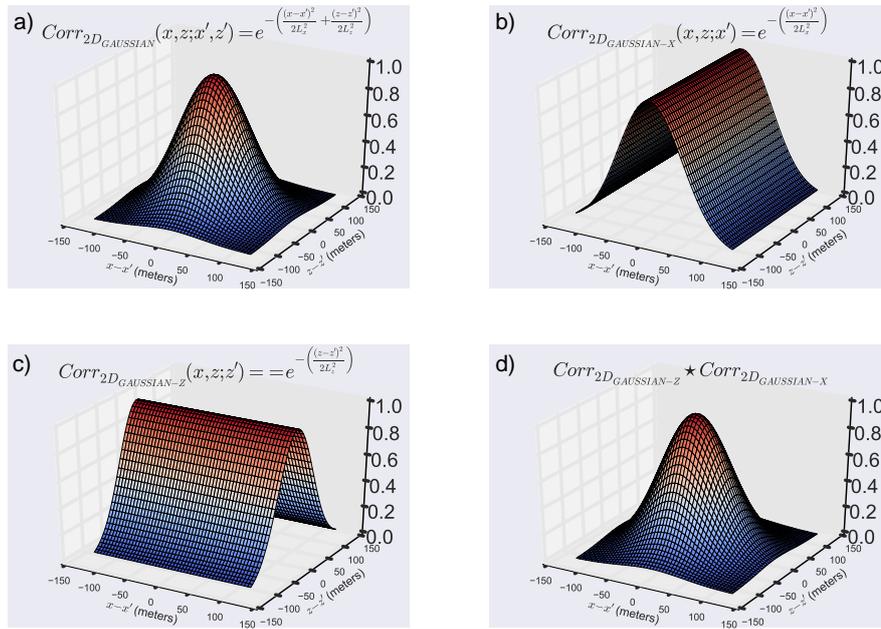
Figure 2.23: The Gaussian function a) is a multi-dimension function that can be applied in a tensorial fashion (Equation 2.26). The correlation function for the X only b) and Z only c) can be used to yield a function d) that is equivalent to the more expensive to apply operator a).

dependent. This was also the case when $L_x \neq L_z$ for our anisotropic Laplacian operators, however the angle dependence formed a smooth ellipse in the wavenumber domain. Figure 2.25 shows the associated wavenumber spectrum of the tensorized Gaussian and Laplacian operators. We see the Laplacian operator has a much milder filtering effect when aligned with the Cartesian grid directions (when one of Cartesian space-lag $x - x'$ or $z - z'$ are equal to zero) than for the other wavenumbers. The transition from this mild to harsher filter is not geologically justified and is likely to introduce artifacts in our inversion results.

Limitations of the Gaussian operator

Although the Gaussian operator can be tensorized, making the application of $Corr_{2D_{GAUSSIAN}}$ and $Corr_{3D_{GAUSSIAN}}$ computationally efficient, there are some limitations. The first and most obvious limitation is, that, as alluded to at the beginning of the chapter, we are often also interested in the inverse correlation $\mathbf{Corr}^{-1}$ and covariance matrices $\mathbf{Cov}^{-1}$. These can theoretically be calculated for any distribution by building the $\mathbf{Corr}$ or $\mathbf{Cov}$ and inverting it numerically. Performing such a task in practice has significant CPU and
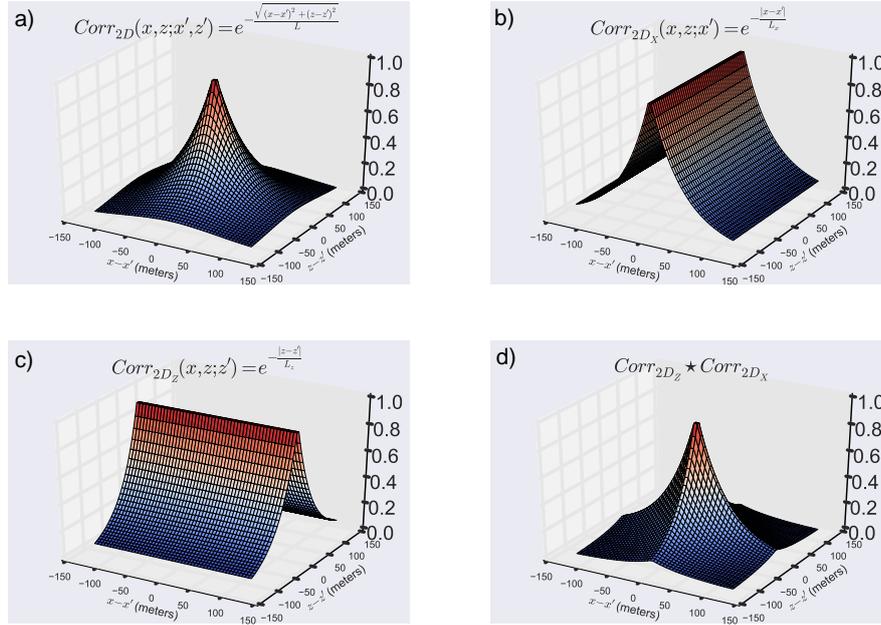
Figure 2.24: The Laplacian function a) is a multi-dimension function that cannot be applied in a tensorial fashion (Equation 2.26). The correlation function for the X only b) and Z only c) cannot be used to yield a function d) that is equivalent to the more expensive to apply operator a).

memory requirements that prohibit their applications. A second potential limitation of the Gaussian distribution is the relative rapid decay it exhibits.

The spatial correlation of features in geology is the reason why constraining the wavenumber spectrum in FWI is of interest. What is important is that the operator we use allows us to constrain the values x, z so they are correctly influenced by distant points at x', z'. If we consider the wavenumber spectrum of the vertical trace we extracted from Valhall (Figure 2.12), we can compare the wavenumber spectrums of our correlation functions. For our operator to be suitable we would like it to have a decay that is at least not wildly different from the decay we expect from the wavenumber spectrum of the Earth (i.e. the extracted Valhall trace). We compare two correlation lengths (50 m and 400 m) and the associated wavenumber spectrum for the Laplacian and Gaussian correlation functions(Figure 2.26). The Valhall model is used as a reference for comparison. To make comparison simple, we scale the raw Valhall trace (red) so the average value is consistent with the average value of the Laplacian operator (trace in purple). This is just for display purposes to allow one to compare the slope of the filter relative to the Valhall trace.
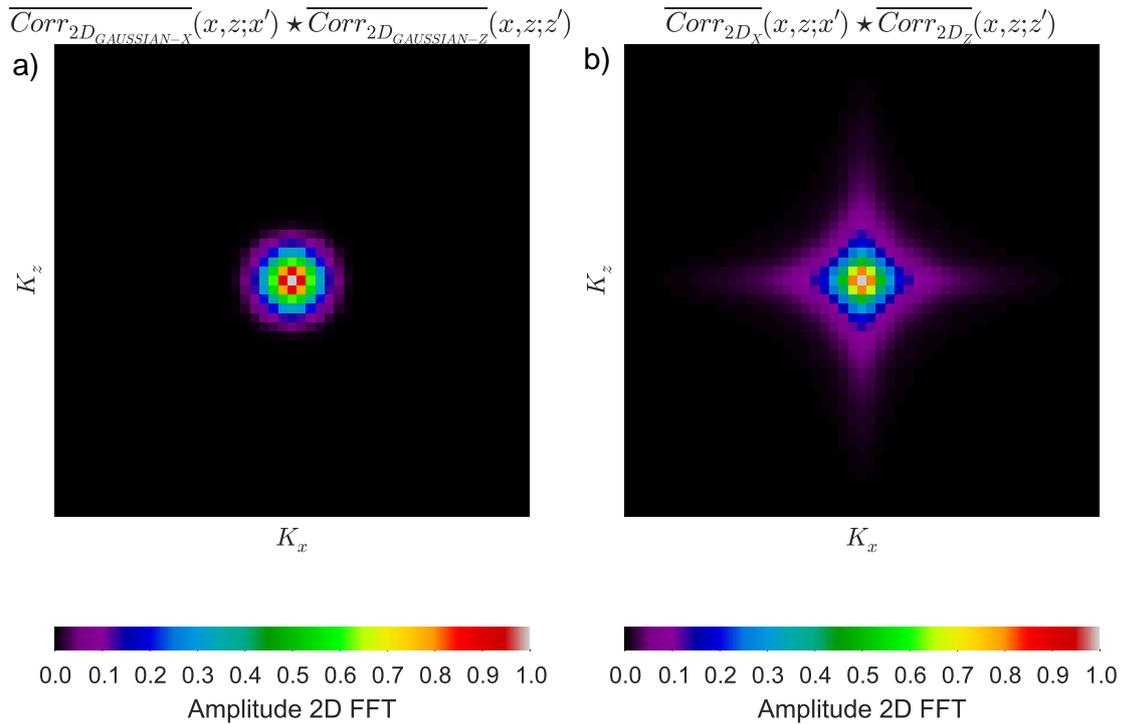
Figure 2.25: The wavenumber spectrum for the tensor application of Gaussian operator a) and tensor application of the Laplacian operator b). The tensorized Gaussian operator has a familiar circular wavenumber spectrum (elliptical if $L_x \neq L_z$). The tensor application of the Laplacian has a strong effect coming from wavenumbers aligned with the Cartesian directions. This operator will cause undesired angle dependent effects when it is applied as a filter.

We see that application of the Laplacian correlation function behaves much like a strong low-pass filter. For $L = 50$ m we see that both the Laplacian and the Gaussian operator have similar amplitude in the wavenumber spectrum from $K = 0 \rightarrow 0.005$. After 0.005 (a wavelength of 200 m) the Gaussian function quickly tapers towards very low amplitude values (note the log_{10} scale). The Laplacian operator on the other hand appears to decrease as a function of wavenumber at a rate roughly consistent with the normalized Valhall trace. For the larger correlation length example ($L = 400$ m), the Gaussian operator decays very rapidly behaving akin to a strong low-pass wavenumber filter. The Laplacian function for this longer correlation length decays more rapidly for low wavenumbers, than $L = 50$ m, however it later flattens off and has a very similar character to spectrum of the normalized true Valhall vertical trace. This operator is in-

teresting as the correlation operator appears to be roughly compatible with wavenumber characteristics that we see in our geological example. The inverse of such an operator would also be quite attractive, as a model space penalty term replacing the frequently used Tikhonov term. The unfortunate issues are the application of $Corr$ is expensive from a CPU perspective and the application of $Corr^{-1}$ is both computationally and memory intensive. In the following part of this chapter, we will introduce our solutions to both of these problems by extending the 1D Laplacian inverse covariance function (Tarantola, 2005). This provides us an operator where we can apply $Corr$ at a computational cost competitive to the tensorized Gaussian and where $Corr^{-1}$ is even cheaper to apply.

2.3 Sparse inverse Laplacian covariance and correlation function

A 1D inverse Laplacian covariance operator is proposed in the appendix of the book Inverse problem theory (Tarantola, 2005, pp. 308-311). Tarantola first describes the well understood Laplacian covariance function

$$Cov_{1D}(x;x') = \sigma^2 e^{-\frac{|x-x'|}{L}}. \tag{2.30}$$

He then introduces the inverse Laplacian covariance function

$$Cov_{1D}^{-1}(x;x') = \frac{1}{2\sigma}\left(\frac{1}{L}\delta(x-x') - L\delta^2(x-x')\right), \tag{2.31}$$

where $\delta(x-x')$ is a delta function (1 at zero space lag and 0 everywhere else) and $\delta^2(x-x')$ is the second derivative of the same delta function. Such an operator appears attractive so we looked to validate his finding and attempt to extend the operator to higher dimensions. In 1D the terms $\delta(x-x')$ and $\delta^2(x-x')$ mean that the discretized operator will be sparse, have low memory requirements and we may be able to solve the linear system

$$\underbrace{\mathbf{A}}_{\mathbf{Cov}_{1D}^{-1}} \underbrace{\mathbf{x}}_{\mathbf{Cov}_{1D}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}} \tag{2.32}$$
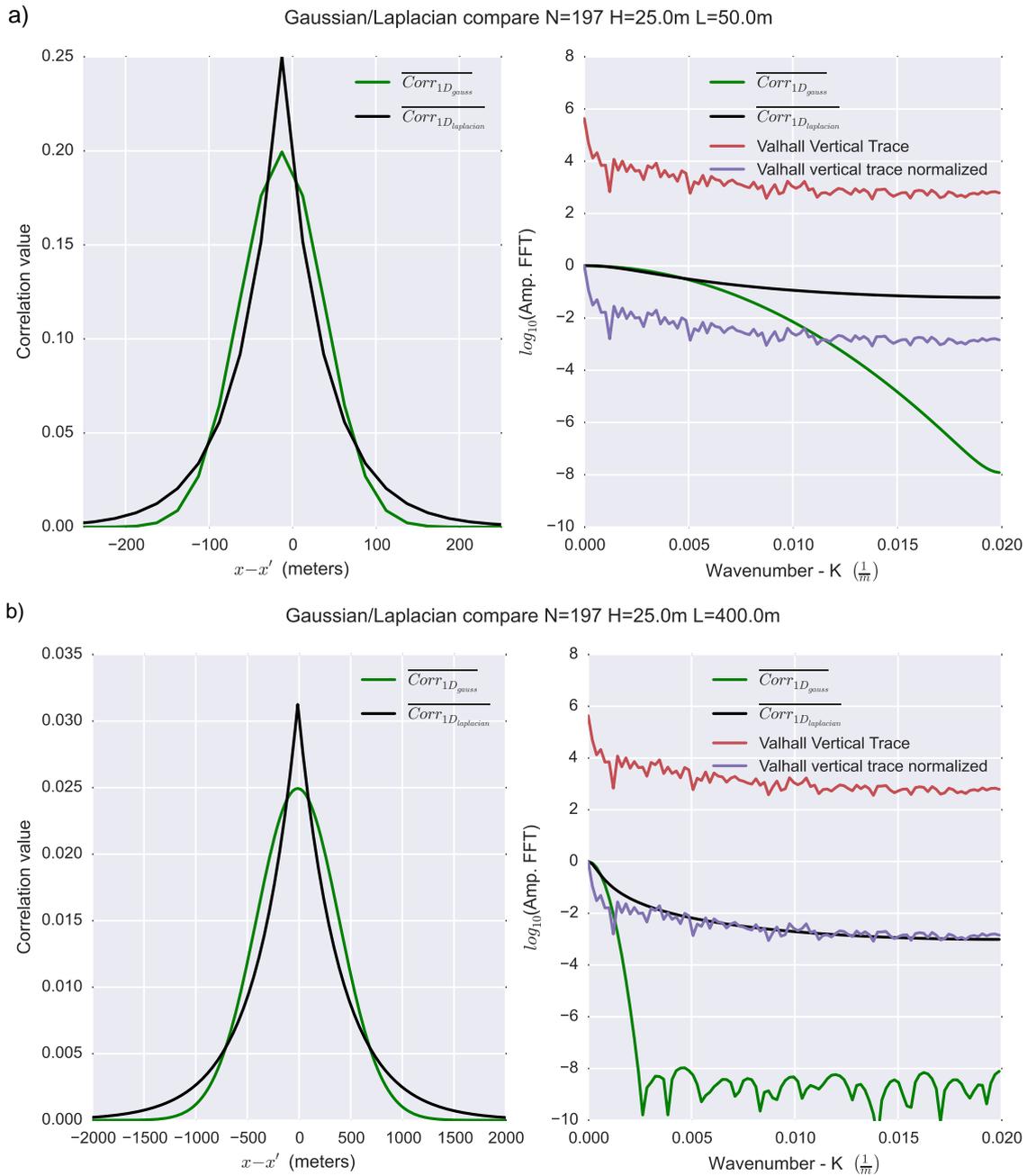
Figure 2.26: The 1D normalized Laplacian correlation operator $\overline{Corr_{1D}}$ (black) and the 1D normalized Gaussian correlation operator $\overline{Corr_{1D_{GAUSSIAN}}}$ (green). A short, 50 m correlation length a) is contrasted with a longer 400 m correlation length b). Note the Gaussian decays quite quickly. The Laplacian function decays more slowly as it is a "long-tailed" distribution. This means that although it decreases, it doesn't decrease at a rate much greater than is suggested by the geological model.

efficiently with an iterative solver such as a sparse conjugate gradient solver. This would allow one to calculate $\mathbf{Cov}_{1D}\mathbf{u}$ without having to use the spatial convolution approach. This section of the chapter will discuss the validation of the 1D operator shown by Tarantola while detailing our efforts to extend the operator to higher dimension and to include dip.

To be consistent with our previous focus on correlation as opposed to covariance we will define the inverse correlation function as one which has no dependence on σ through

$$Corr_{1D}^{-1}(x; x') = \frac{1}{L}\delta(x - x') - L\delta^2(x - x'). \tag{2.33}$$

As a result, we focus on building the sparse matrix $\mathbf{Corr}_{1D}^{-1}$ and solving the linear system

$$\underbrace{\mathbf{A}}_{\mathbf{Corr}_{1D}^{-1}} \underbrace{\mathbf{x}}_{\mathbf{Corr}_{1D}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}} \tag{2.34}$$

to obtain the vector $\mathbf{x} = \mathbf{Corr}_{1D}\mathbf{u}$.

2.3.1 The 1D inverse Laplacian covariance and correlation function

To build the 1D inverse correlation matrix $\mathbf{Corr}_{1D}^{-1}$ we need to discretize $\delta(x - x')$ and $\delta^2(x - x')$. We do this based on a vector $\mathbf{u}$ sampled at evenly spaced increments, h. The first term is quite straight forward and satisfies the following

$$\delta(x - x') = \begin{cases} \frac{1}{h} \text{ for } x = x' \\ 0 \text{ when } x \neq x' \end{cases}, \tag{2.35}$$

while, for the second term we use a second-order centred finite difference stencil

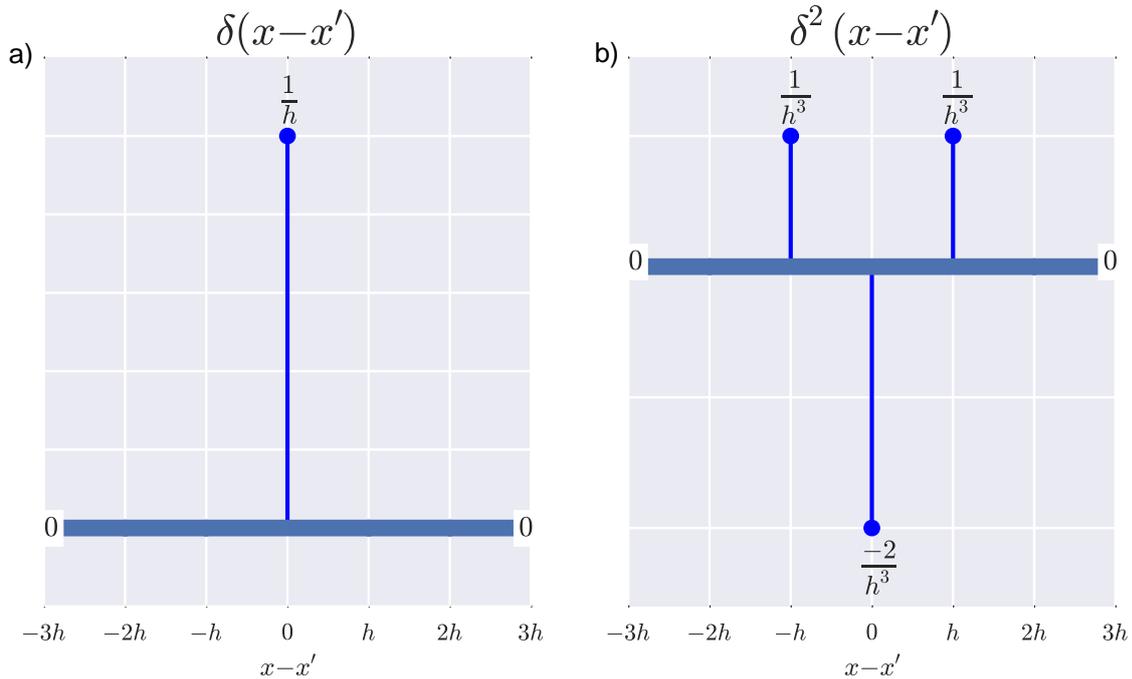$$f''(x) \approx \frac{f(x + h) - 2f(x) + f(x - h)}{h^2}. \tag{2.36}$$

Figure 2.27: The finite discretization of a) $\delta(x-x')$ and b) $\delta^2(x-x')$. This discretization means the operator $\mathbf{Corr}^{-1}$ will be sparse and only require 3 non-zero elements per grid point.

The discretization of $\delta^2(x-x')$ yields

$$\delta^2(x-x') = \delta''(x-x') \begin{cases} \frac{-2}{h^3} \text{ for } x = x' \\ \frac{1}{h^3} \text{ when } x = x' \pm h \\ 0 \text{ when } |x - x'| > h \end{cases} . \qquad (2.37)$$

The discretization values are graphically shown in Figure 2.27. This result suggests that in order to describe the inverse covariance for our vector $\mathbf{u}$ we only need to consider a given cell $(x = x')$ and it's two neighbour points $(x = x' \pm h)$. This is a significant reduction compared to the covariance and correlation functions where we need to consider a large part of or the entire domain Ω. We can describe our inverse correlation matrix $\mathbf{Corr}^{-1}_{1D}$ as shown in Figure 2.28.

Testing the validity of $\mathbf{Corr}^{-1}_{1D}$

We can validate our discretization via some numerical experiments. One of the easiest
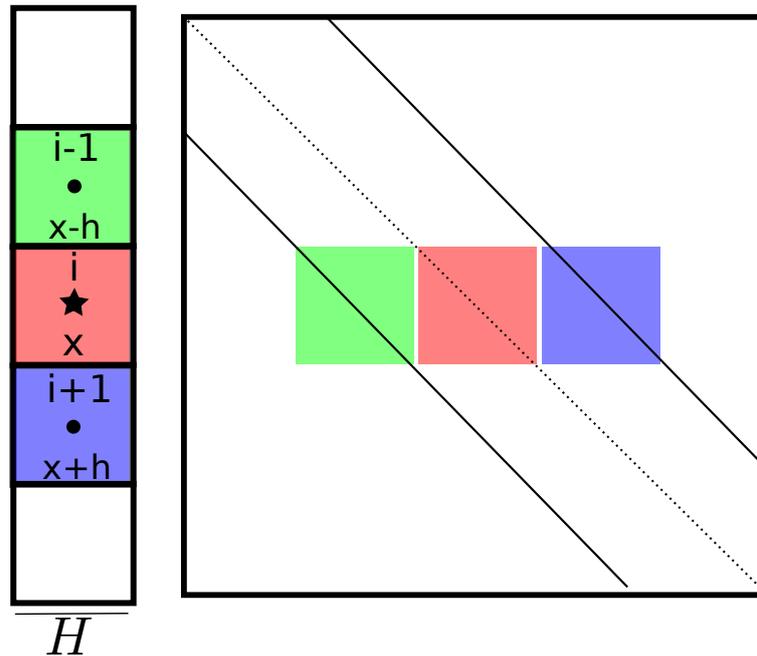
Figure 2.28: The inverse correlation stencil for a 1D vector of length 5. For each element in vector we only need to consider a given point (represented by the diagonal where $x = x'$ represented in pink) and its neighbouring points $x = x' \pm h$.

tests is to take an input vector $\mathbf{u}$ and compare application of $\mathbf{Corr}_{1D}\mathbf{u}$ using the spatial convolution approach vs solving the linear system.

$$\underbrace{\mathbf{A}}_{\mathbf{Corr}_{1D}^{-1}} \underbrace{\mathbf{x}}_{\mathbf{Corr}_{1D}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}}$$

We are able to solve this linear system using a conjugate gradient algorithm, that only requires the non-zero elements of $\mathbf{Corr}_{1D}^{-1}$. At this moment, we will not discuss the rate of convergence of the linear system. This will be discussed for 1D, 2D and 3D cases together at the end of the chapter. The first test we consider is for a vector $\mathbf{u}$ that is a Dirac spike ($\mathbf{u} = \boldsymbol{\delta}$). The vector $\mathbf{u}$ we examine is of length $N = 75$ with a spatial sampling of $h = 5.0$ m This vector is zero, except for the spike which occurs at the 38th element. Figure 2.29 shows the results of this "spike test" using an $L = 25$ m. We see that solving the linear system gives us a result ($\mathbf{x}$) that has a maximum higher amplitude than $\mathbf{Corr}_{1D}\mathbf{u}$. If we normalize ($\mathbf{x}$) by it's maximum value ($\overline{\mathbf{x}}$) we see that this appears identical to ($\mathbf{Corr}_{1D}\mathbf{u}$). It appears there may be some normalization that needs to be performed.
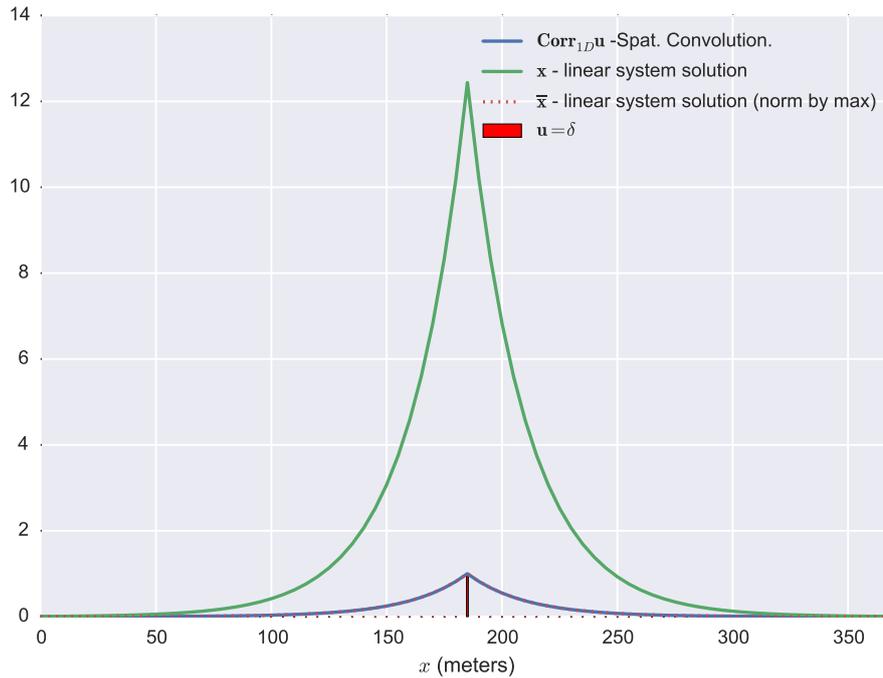
Figure 2.29: Comparison of the result from solving the sparse linear system ($\mathbf{x}$) vs the analytical result ($\mathbf{Corr}_{1D}\mathbf{u}$) calculated by spatial convolution. Note that $\mathbf{x}$ has a much higher maximum value than what is suggested from $\mathbf{Corr}_{1D}\mathbf{u}$. If we compare $\mathbf{x}$ when it is normalized by it's maximum value ($\overline{\mathbf{x}}$) it appears to be equivalent to $\mathbf{Corr}_{1D}\mathbf{u}$.

Normalization

The need to perform some type of normalization should not be alarming as we are required to apply something similar when we considered the analytical correlation functions. This normalization issue was solved for the analytical covariance functions by performing analytical integration. This is not practical in this case as we aren't exactly sure what the associated analytical equation we are simulating when we solve the linear system with $\mathbf{Corr}^{-1}$. There is also the potential that there will be some normalization effects associated with the finite difference discretization. We can look to find the normalization factor required for $Corr_{1D}^{-1}$. This is to ensure that when we solve the linear system $\|\mathbf{x}\|_1 = 1$ for all input values of L and h. The hope is the normalization is something simple, where we only need to modify expression 2.34 by adding a constant a

$$\overline{Corr_{1D}^{-1}(x; x')} = a\left(\frac{1}{L}\delta(x - x') - L\delta^2(x - x')\right). \tag{2.38}$$

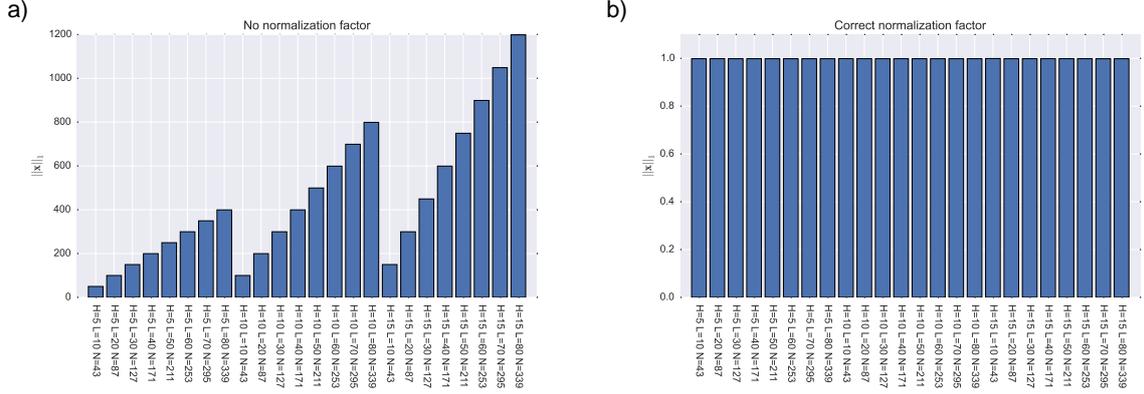I was able to find the answer to this problem numerically by taking the spike test I

Figure 2.30: Comparison of the ℓ_1 norm from what is obtained from solving the linear system involving $\mathbf{Corr}_{1D}^{-1}$ a) as opposed to the one involving $\overline{\mathbf{Corr}_{1D}^{-1}}$ b). The application of the normalization factor $a = hL$ to $\mathbf{Cor}_{1D}^{-1}$ ensures that $\mathbf{x}$ is equivalent to $\overline{\mathbf{Corr}_{1D}}\mathbf{u}$.

have showed for different values of L and h. If we look at the ℓ_1 norm of $\mathbf{x}$ ($||\mathbf{x}||_1$) we know that we would like this result to be unitary no matter what values of L and h we use. I found the appropriate scalar, a via trail and error. The necessary normalization is $a = Lh$ giving us

$$\overline{Corr_{1D}^{-1}}(x; x') = Lh\left(\frac{1}{L}\delta(x - x') - L\delta^2(x - x')\right). \tag{2.39}$$

We can see how $||\mathbf{x}||_1$ is equal to 1 in Figure 2.30. This shows that $\overline{\mathbf{Corr}_{1D}^{-1}}$ is well normalized and that when we can calculate $\overline{\mathbf{Corr}_{1D}}\mathbf{u}$ by solving the sparse linear system

$$\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{1D}^{-1}}} \underbrace{\mathbf{x}}_{\overline{\mathbf{Corr}_{1D}}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}}. \tag{2.40}$$

If we return to our spike example with our normalized inverse covariance operator we can look at what the significance of $\overline{\mathbf{Corr}_{1D}}\mathbf{u}$ and $\overline{\mathbf{Corr}_{1D}^{-1}}\mathbf{u}$ are in terms of the wavenumber spectrum (Figure 2.31). We understand quite well from our work on the analytical correlation functions that the action of $\overline{\mathbf{Corr}_{1D}}$ on $\mathbf{u}$ is akin to a low-pass wavenumber filter. The application of $\overline{\mathbf{Corr}_{1D}^{-1}}$ can be seen as the inverse of a low-pass filter called sometimes a roughening filter. It is important to note that if we apply these operators in cascade to a vector we will remove their effect (for example $\overline{\mathbf{Corr}_{1D}^{-1}}\left(\overline{\mathbf{Corr}_{1D}}\mathbf{u}\right) = \mathbf{u}$).

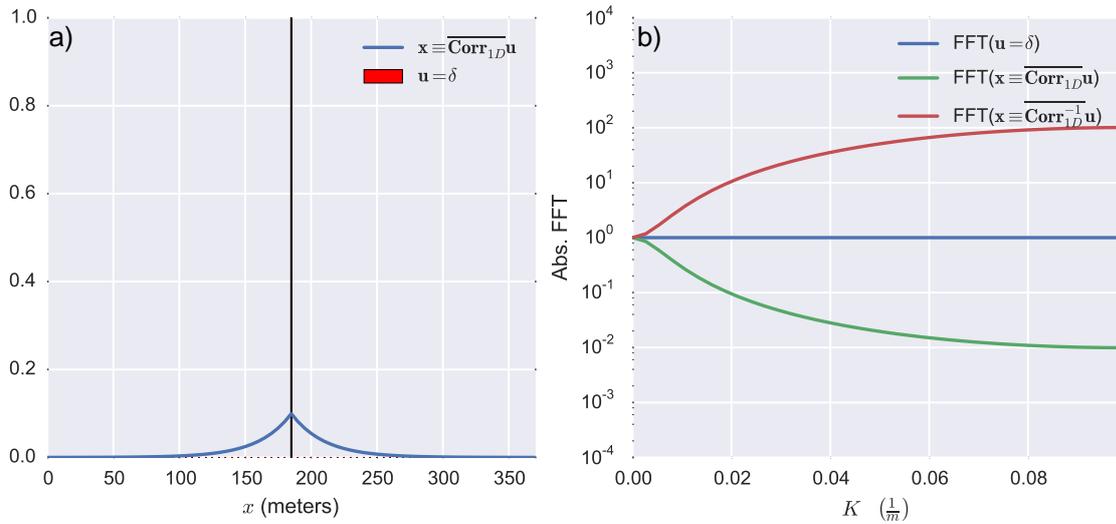**White noise test of $\overline{\mathbf{Corr}_{1D}^{-1}}$**

Figure 2.31: $\overline{\mathbf{Corr}_{1D}}\mathbf{u}$ can be seen as a low-pass filter on the white unit-amplitude spectrum of $\mathbf{u}$. We can also calculate $\overline{\mathbf{Corr}_{1D}^{-1}}\mathbf{u}$. This operator can be see to be the inverse low-pass filter. These filtering and inverse filtering operations are reversible, i.e. $\overline{\mathbf{Corr}_{1D}^{-1}}\left(\overline{\mathbf{Corr}_{1D}}\mathbf{u}\right) = \mathbf{u}$ as well as $\overline{\mathbf{Corr}_{1D}}\left(\overline{\mathbf{Corr}_{1D}^{-1}}\mathbf{u}\right) = \mathbf{u}$

Although we performed a spike test of our inverse covariance matrix we would like to be sure that our normalization and operator works for more complex vectors. A second test that we performed was to apply it to a white noise vector $\mathbf{u} = \boldsymbol{\epsilon}_w$. The vector had a length of $N = 125$ with a sample rate $h = 5$ m. We applied our correlation operators using a correlation length, $L = 10$ m. To test the validity of solving the linear system, we compared it to the analytical solution we have from using the normalized analytical correlation function using the spatial convolution approach (Figure 2.32). The results of both the sparse and spatial convolution methods yield similar results where the high wavenumber content of the noise has been decreased.

2.3.2 The 2D additive inverse Laplacian (AIL) correlation function

We were able to validate the 1D inverse Laplacian function (Tarantola, 2005), but there was no indication whether this operator could be extended to higher dimensions. Our desire was to use the sparse nature of the discretized operators as an efficient means of calculating $\overline{\mathbf{Corr}_{2D}}\mathbf{u}$ by avoiding the expensive spatial convolution option. We are also interested in $\overline{\mathbf{Corr}_{2D}^{-1}}\mathbf{u}$, which if we have a sparse analytical inverse formula for $\mathbf{Corr}_{2D}^{-1}\mathbf{u}$
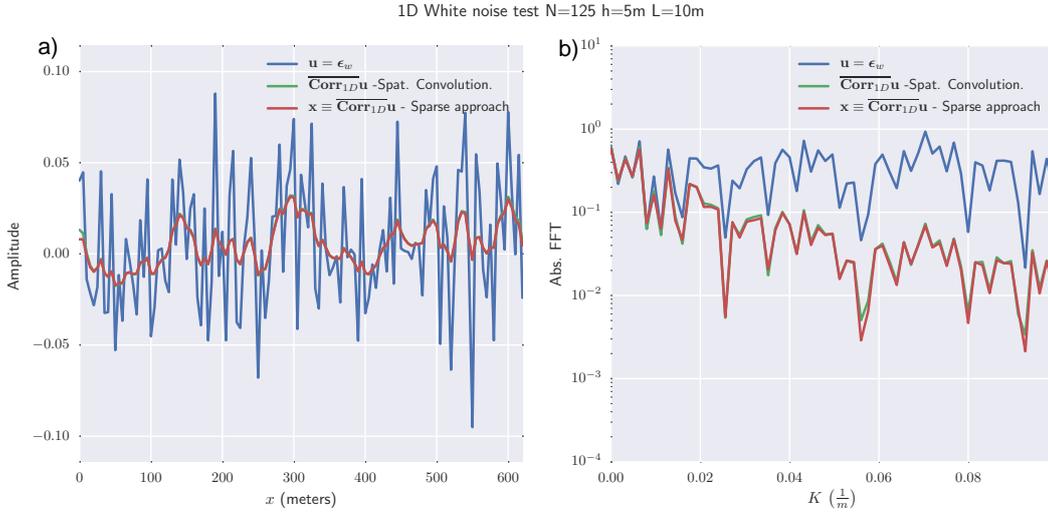
Figure 2.32: Comparison of calculating $\overline{\mathbf{Corr}}_{1D}\mathbf{u}$ using the normalized 1D analytical correlation function (green) vs solving the linear system involving $\overline{\mathbf{Corr}}_{1D}^{-1}$ (red). We see that both results are close to identical and that application of the normalized Laplacian correlation function decreases the high-wavenumber content of the noise (shown in blue). a) Vectors in space domain b) Vectors in the wavenumber domain.

is straight forward to calculate. We found earlier that the Laplacian cannot be tensorized and it was not a surprise when we looked at the $\overline{Corr_{2D_X}^{-1}}(x,z;x') \star Corr_{2D_Z}^{-1}(x,z;z')$ it didn't represent the correct way to go to higher dimensions.

We reasoned that perhaps the addition of two orthogonal inverse covariance functions may provide us with a useful operator. As there was no understood reason why this operator should be related to a 2D Laplacian correlation function we referred to it with the subscript AIL which is to denote that it is the addition of two orthogonal, inverse Laplacian functions 2.41.

$$Corr_{2D_{AIL}}^{-1}(x,z;x',z') = \left(\frac{1}{L_x}\delta(x-x') - L_x\delta^2(x-x')\right) +$$
$$\left(\frac{1}{L_z}\delta(z-z') - L_z\delta^2(z-z')\right). \tag{2.41}$$

To evaluate if this is related to the inverse function we are searching for we need to discretize, $\delta(x-x')$, $\delta(z-z')$, $\delta^2(x-x')$ and $\delta^2(z-z')$. The process to do this is identical to the 1D case, where we used centred finite difference and is summarised in Figure 2.33.
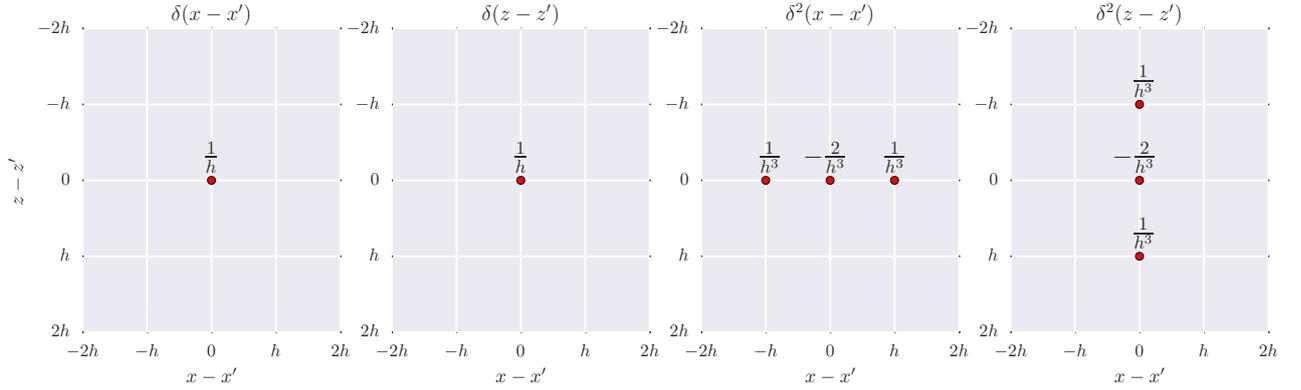
Figure 2.33: The discretization of $\delta(x - x')$,$\delta(z - z')$, $\delta^2(x - x')$ and $\delta^2(z - z')$ using centred finite differences

The matrix that results from this discretization is very sparse and has slightly less than $5n$ non-zero elements (it is slightly less due to the edges present in the 2D model space). To assess the utility of the operator, we need to find what we will obtain from solving the linear system

$$\underbrace{\mathbf{A}}_{\mathbf{Corr}_{2D_{AIL}}^{-1}} \underbrace{\mathbf{x}}_{\mathbf{Corr}_{2D_{AIL}}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}}. \tag{2.42}$$

When we solve this linear system using different values of L_x, L_z and h the value of $||\mathbf{x}||_1$ is not equal to 1. We find numerically that if we modify equation 2.41 to equation 2.43. we have a normalized inverse operator

$$\overline{Corr_{2D_{AIL}}^{-1}}(x, z; x', z') = \frac{hL_x}{2}\left(\frac{1}{L_x}\delta(x - x') - L_x\delta^2(x - x')\right) + $$
$$\frac{hL_z}{2}\left(\frac{1}{L_z}\delta(z - z') - L_z\delta^2(z - z')\right). \tag{2.43}$$

To know whether this operator is of any interest we need to solve the linear system 2.44

$$\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}} \underbrace{\mathbf{x}}_{\mathbf{Corr}_{2D_{AIL}}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}} \tag{2.44}$$
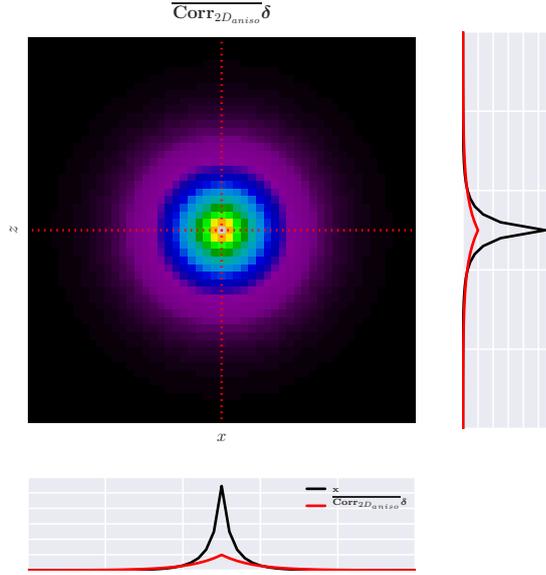
Figure 2.34: Comparison of $\mathbf{x}$ (black) obtained from solving 2.44 vs the application of the normalized 2D Laplacian correlation function $\overline{\mathbf{Corr}_{2D}}$ using spatial convolution (red). The extracted vertical and horizontal traces show the decay of the two results. The additive inverse Laplacian decays ($\mathbf{x}$) much more rapidly at low space-lags than the analytical correlation function.

Our vector $\mathbf{u}$ has 51×51 elements and $h = 5.0$ m. Initially we consider $L_x = L_z = 20$ m. We solve this linear system and we compare the result $\mathbf{x} \equiv \overline{\mathbf{Corr}_{2D_{AIL}}}\mathbf{u}$ to the result from our analytical correlation function $\overline{\mathbf{Corr}_{2D_{aniso}}}\mathbf{u}$ (Figure 2.34). The decay of $\mathbf{x}$ is much faster than $\overline{\mathbf{Corr}_{2D_{aniso}}}\mathbf{u}$. Both operators have norm of 1 by design ($||\mathbf{x}||_1 = 1$ and $||\overline{\mathbf{Corr}_{2D_{aniso}}}\mathbf{u}||_1 = 1$) but our operator does not appear to be related to the Laplacian correlation operator any more. In fact it was discovered that post the initial submission of this manuscript they were instead related to modified Bessel functions of the second kind (explained in more detail at the end of the chapter).

The operator we have here may still be useful as the convergence of the linear system is quite rapid (to be discussed in more detail at the end of the chapter), but we were looking for an operator that had a decay similar to the multi-dimensional analytical Laplacian correlation framework we designed earlier. It appears we find something closer to our goal, if we solve the linear system twice with the output from the first linear system $\mathbf{x}_1$ serving as the input $\mathbf{b}$ for the second linear system solution. This is summarized in Algorithm 1

Application of this to our spike example yields the figure 2.35. We see we have an

---

Algorithm 1 Approximate $\mathbf{Corr}_{2D_{aniso}}\mathbf{u}$ with $\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}$

1: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}} \underbrace{\mathbf{x_{2D_1}}}_{\overline{\mathbf{Corr}_{2D_{AIL}}}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}}$

2: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}} \underbrace{\mathbf{x_{2D_2}}}_{\overline{\mathbf{Corr}_{2D_{AIL}}}\mathbf{x_{2D_1}}} = \underbrace{\mathbf{b}}_{\mathbf{x_{2D_1}}}$

3: {It appears that: $\mathbf{x}_{2D_2}$ is similar to $\overline{\mathbf{Corr}_{2D_{aniso}}}\mathbf{u}$}
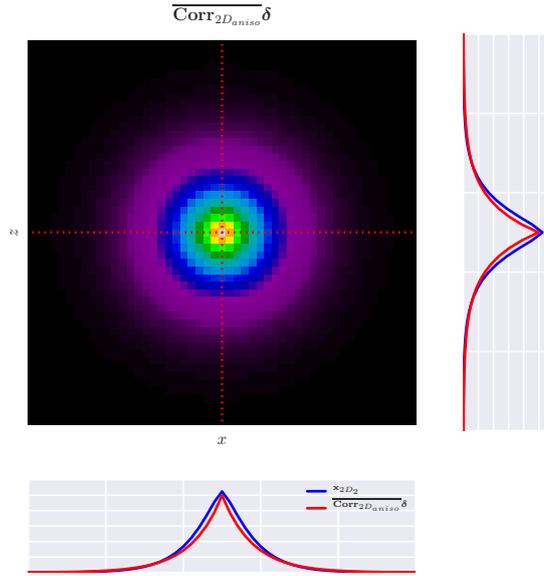

Figure 2.35: Comparison of $\mathbf{x}_{2D_2}$ (blue) obtained from solving 2.44 twice using Algorithm 1 versus the application of the normalized 2D Laplacian correlation function (red). The extracted vertical and horizontal traces show the decay of the two results. Both results show similar decay characteristics and are both well normalized.

operator that is not identical, but quite similar to the normalized analytical correlation function. This result is encouraging and we find that if we test the anisotropic correlation functions ($L_x \neq L_z$ in Figure 2.36) we also get the same level of agreement between $\mathbf{x}_{2D_2}$ and $\overline{\mathbf{Corr}_{2D_{aniso}}}\mathbf{u}$.

Rotation

The inclusion of rotation can be achieved if we consider the same rotation matrix we used to allow rotation with the 2D normalized analytical correlation function (equation 2.15). To include rotation we need to discretize $\delta(x - \hat{x}')$,$\delta(z - \hat{z}')$, $\delta^2(x - \hat{x}')$ and $\delta^2(z - \hat{z}')$.
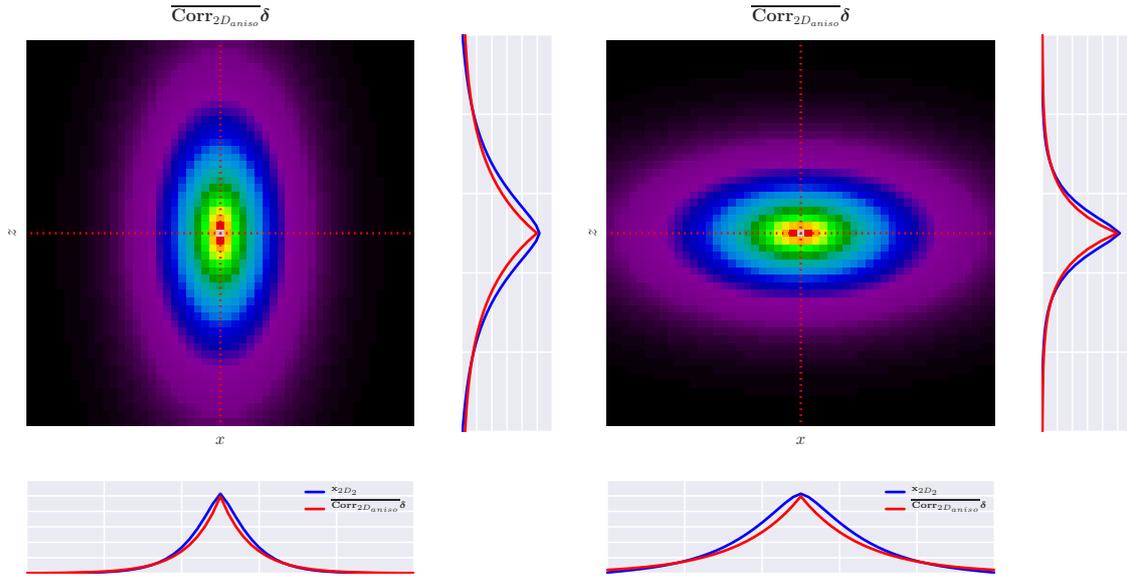
Figure 2.36: Comparison of $\mathbf{x}_{2D_2}$ (blue) obtained from solving 2.44 twice as per Algorithm 1 vs the application of the normalized 2D Laplacian correlation function (red). a) $L_x = 20$ m and $L_z = 40$ m b) $L_x = 40$ m and $L_z = 20$ m. The sparse approach can handle anisotropic operators.

We know that for our finite difference discretization on square grids that

$$\delta(x - \hat{x}') = \delta(x - x') = \delta(z - \hat{z}') = \delta(z - z') \tag{2.45}$$

However, $\delta^2(x - \hat{x}')$ and $\delta^2(z - \hat{z}')$ will now have an angle dependence. To simplify for the following steps we will make the following notation 2.46.

$$\frac{\partial^2}{\partial \hat{x}'^2} = \delta^2(x - \hat{x}') \text{ and } \frac{\partial^2}{\partial x'^2} = \delta^2(x - x')$$

$$\frac{\partial^2}{\partial \hat{z}'^2} = \delta^2(z - \hat{z}') \text{ and } \frac{\partial^2}{\partial z'^2} = \delta^2(z - z') \tag{2.46}$$

So, we want to calculate $\frac{\partial^2}{\partial \hat{x}'^2}$ and $\frac{\partial^2}{\partial \hat{z}'^2}$ on the non-rotated, Cartesian grid that $\frac{\partial^2}{\partial x'^2}$ and $\frac{\partial^2}{\partial z'^2}$ use.

$$\frac{\partial^2}{\partial \hat{x}'^2} = \frac{\partial}{\partial \hat{x}'}\left(\frac{\partial}{\partial \hat{x}'}\right) = \frac{\partial}{\partial \hat{x}'}\underbrace{\left(\frac{\partial}{\partial x'}\frac{\partial x'}{\partial \hat{x}'} + \frac{\partial}{\partial z'}\frac{\partial z'}{\partial \hat{z}'}\right)}_{\text{Product rule}} \tag{2.47}$$

Using our definitions for x' and z'

$$x' = \hat{x}'cos\theta - \hat{z}'sin\theta \tag{2.48}$$

$$z' = \hat{x}'sin\theta + \hat{z}'cos\theta \tag{2.49}$$

we can substitute in $\frac{\partial x'}{\partial \hat{x}'}$ and $\frac{\partial z'}{\partial \hat{z}'}$

$$\frac{\partial^2}{\partial \hat{x}'^2} = \left(\frac{\partial}{\partial x'}cos\theta + \frac{\partial}{\partial z'}sin\theta\right)\left(\frac{\partial}{\partial x'}cos\theta + \frac{\partial}{\partial z'}sin\theta\right) \tag{2.50}$$

$$\frac{\partial^2}{\partial \hat{x}'^2} = \frac{\partial^2}{\partial x'^2}cos^2\theta + \frac{\partial^2}{\partial z'^2}sin^2\theta + 2\frac{\partial}{\partial x'}\frac{\partial}{\partial z'}cos\theta sin\theta \tag{2.51}$$

and we go through the same process for $\frac{\partial^2}{\partial \hat{z}'^2}$

$$\frac{\partial^2}{\partial \hat{z}'^2} = \frac{\partial^2}{\partial x'^2}sin^2\theta + \frac{\partial^2}{\partial z'^2}cos^2\theta - 2\frac{\partial}{\partial x'}\frac{\partial}{\partial z'}cos\theta sin\theta. \tag{2.52}$$

If we write this in the more familiar format we can express $\delta^2(x - \hat{x}')$ and $\delta^2(z - \hat{z}')$ as

$$\delta^2(x - \hat{x}') = \delta^2(x - x')cos^2\theta + \delta^2(z - z')sin^2\theta + 2\delta^1(x - x')\delta^1(z - z')cos\theta sin\theta \tag{2.53}$$

and

$$\delta^2(z - \hat{z}') = \delta^2(x - x')sin^2\theta + \delta^2(z - z')cos^2\theta - 2\delta^1(x - x')\delta^1(z - z')cos\theta sin\theta. \tag{2.54}$$

where $\delta^1(x - x')\delta^1(z - z')$ is equal to the first derivative of the delta function in the x' direction, followed by the first derivative in the z' direction. This can be discretized using centred finite difference (Figure 2.37).

The equations for the 2D rotated normalized additive inverse Laplacian correlation stencil $\overline{\mathbf{Corr}^{-1}_{2D_{AIL-rot}}}$ are shown in Figure 2.38. When θ is equal 0 we see that the green points required for the rotated stencil become zero and our stencil has 5 points. However, when θ is not equal to 0, then we have approximately 9 points per grid cell. This is also shown clearly when we consider the inverse correlation matrices for the rotated and non rotated cases (Figure 2.39).
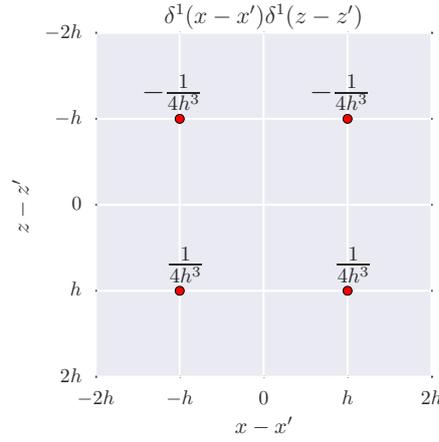
Figure 2.37: The discretization of $\delta^1(x-x')\delta^1(z-z')$ using centred finite difference .
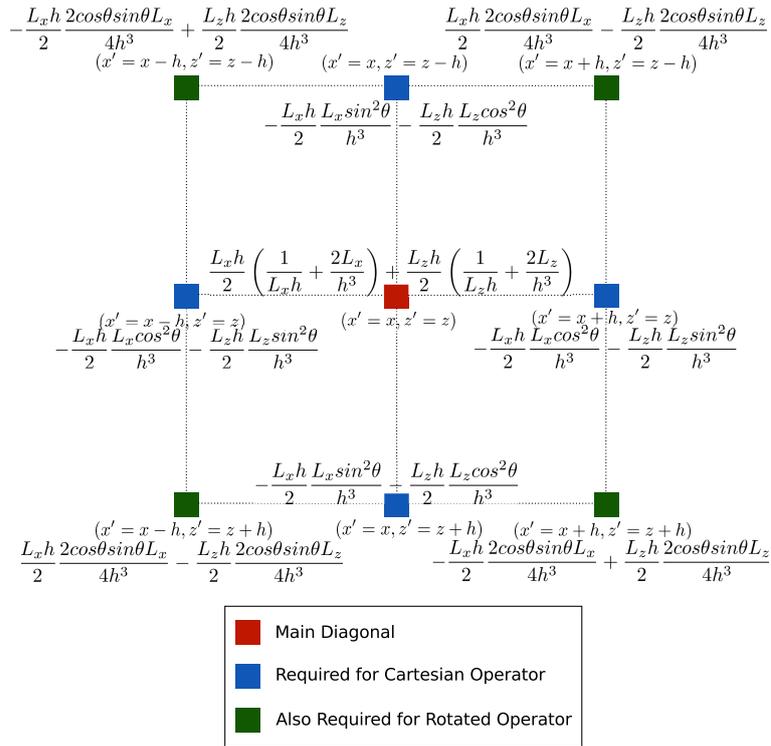
Figure 2.38: The equations describing the points required for the rotated normalized additive inverse Laplacian correlation stencil $\overline{\mathbf{Corr}}^{-1}_{2D_{AIL-rot}}$. Note that the green points are only non-zero when $\theta \neq 0$.

Although we perform a spike test of our 2D inverse additive Laplacian correlation matrix, we would like to be sure that our approach works for more complex vectors and

Figure 2.39: a) $\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}$ b) $\overline{\mathbf{Corr}_{2D_{AIL-rot}}^{-1}}$

that the rotation we proposed is correct. A 51×51 ($h = 5$) white noise vector ($\mathbf{u} = \boldsymbol{\epsilon}_w$) is used as a tes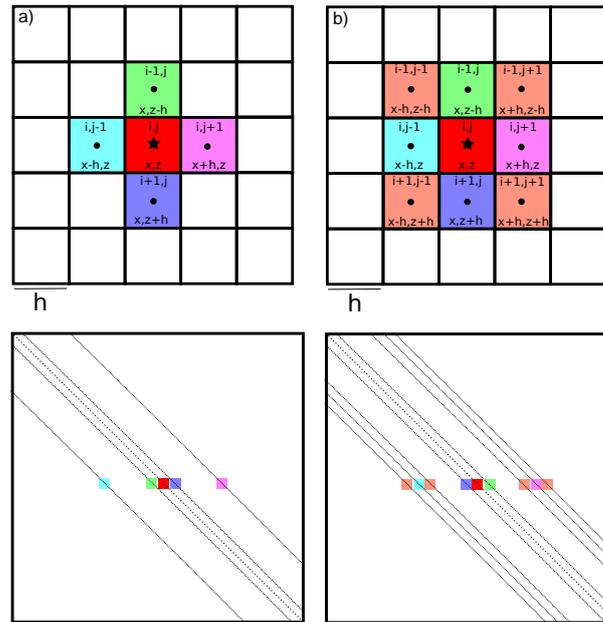t. Correlation lengths $L_z = 10$ m and $L_x = 20$ m are used with $\theta = 45°$. Figure 2.40 shows a comparison between $\mathbf{x}_{2D_2}$ and the application of the normalized correlation function $\overline{Corr_{2D_{aniso}}}$ to $\mathbf{u}$. We see that the results are quire similar for the white noise vector. Even though the two operators have slightly different shapes, they produce visually similar results. A more quantitative comparison of the wavenumber spectrum of these operators will take place after we introduce the 3D extension of our correlation operators.

2.3.3 The 3D additive inverse Laplacian (AIL) correlation function

Given the significant computation burden for calculating $\mathbf{Corr}_{3D}\mathbf{u}$ we were very interested whether $\mathbf{Corr}_{2D_{AIL}}^{-1}$ could be extended to 3D and, if so, how it would behave. The obvious extension was to add a 3rd term that includes the space lag in the y direction $(y - y')$.
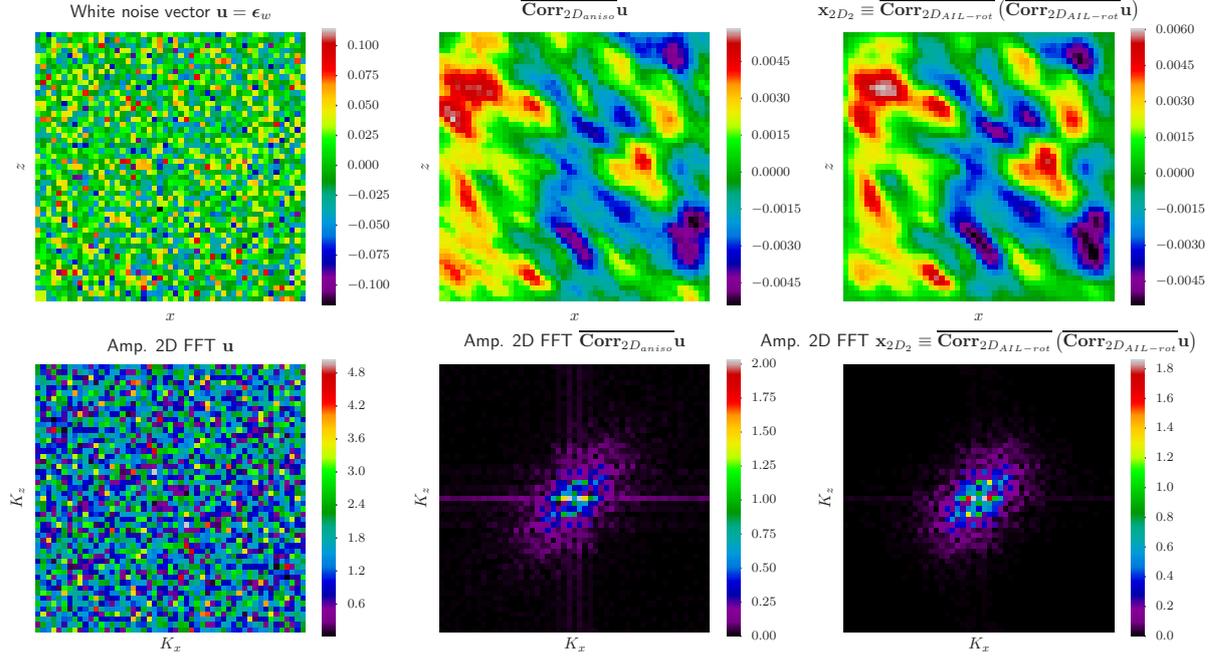
Figure 2.40: Application of the correlation operator to a white noise vector $\left(\mathbf{u} = \boldsymbol{\epsilon}_w\right)$ a). Comparison of $\overline{\mathbf{Corr}_{2D_{aniso}}}\mathbf{u}$ (b,e) and $\mathbf{x}_{2D_2}$ (c,f). Note both results show similar features in the space domain (b,c) and in the 2D wavenumber domain (e,f).

$$Corr^{-1}_{3D_{AIL}}(x, y, z; x', y', z') = \left(\frac{1}{L_x}\delta(x - x') - L_x\delta^2(x - x')\right) +$$
$$+ \left(\frac{1}{L_z}\delta(y - y') - L_z\delta^2(y - y')\right) + \left(\frac{1}{L_z}\delta(z - z') - L_z\delta^2(z - z')\right). \qquad (2.55)$$

This operator requires the discretization of 6 terms and as with the 1D and 2D cases we use a centred finite difference scheme. The discretization of $\delta(x - x') \equiv \delta(y - y') \equiv \delta(z - z')$ which is equal to $1/h$. The three second derivative terms are discretized as shown in Figure 2.41.

With the discretization terms described, we see that the matrix $\mathbf{Corr}^{-1}_{3D_{AIL}}$ requires around 7 non-zero points per element of the vector $\mathbf{u}$ (Figure 2.42).

As in the 1D and 2D case we need to find the values that normalize this operator. We found that if we want to solve the linear system
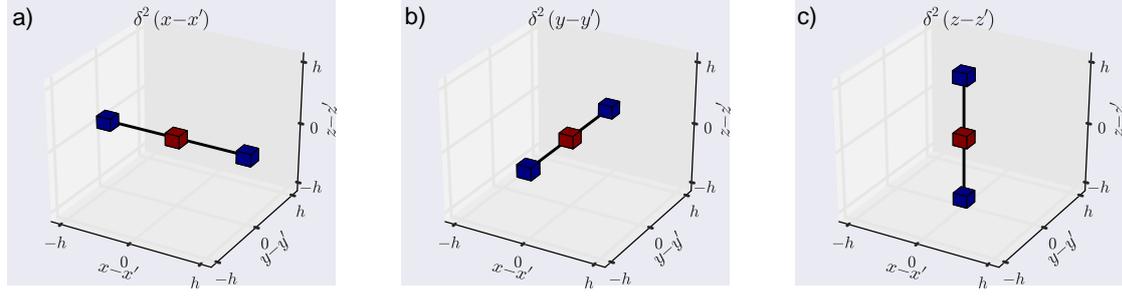
Figure 2.41: The discretization of the δ^2 for the 3 orientations in 3D. The red cube is the diagonal which has a value of $-\frac{2}{h^3}$ while the blue cubes are the off diagonal terms that exist at a space lag of h. These have the value $\frac{1}{h^3}$. The discretization of δ only occurs on the diagonal of the matrix (shown in red) and has a value of $\frac{1}{h}$
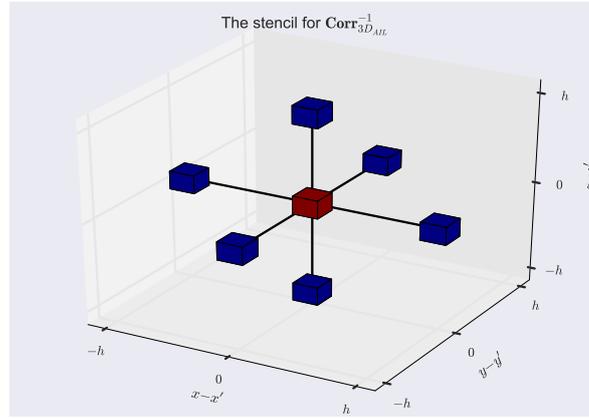
Figure 2.42: The stencil used to build $\mathbf{Corr}^{-1}_{3D_{AIL}}$. The red point represents the diagonal of the $\mathbf{Corr}^{-1}_{3D_{AIL}}$ while the blue points are the cells at a space lag of $\pm h$ away.

$$\underbrace{\mathbf{A}}_{\mathbf{Corr}^{-1}_{3D_{AIL}}} \underbrace{\mathbf{x}}_{\mathbf{Corr}_{3D_{AIL}}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}}. \tag{2.56}$$

We would like that $||\mathbf{x}||_1 = 1$ for which ever value of L or h we use. We find that we can express the 3D normalized inverse additive Laplacian correlation function $\overline{\mathbf{Corr}}_{3D_{AIL}}$ as

$$\overline{Corr^{-1}_{3D_{AIL}}}(x,y,z;x',y',z') = \frac{hL_x}{3}\left(\frac{1}{L_x}\delta(x-x') - L_x\delta^2(x-x')\right) +$$

$$+\frac{hL_y}{3}\left(\frac{1}{L_z}\delta(y-y') - L_z\delta^2(y-y')\right) + \frac{hL_z}{3}\left(\frac{1}{L_z}\delta(z-z') - L_z\delta^2(z-z')\right). \tag{2.57}$$

With the operator discretized we can build the matrix and test it on a 3D vector $\mathbf{u}$. As in the 1D and 2D cases, we will initially consider $\mathbf{u} = \boldsymbol{\delta}$. Our vector $\mathbf{u}$ has dimension $51 \times 51 \times 51$ with a $h = 5$ m. We are able to compare the solution of the linear system, $\mathbf{x}$ with our normalized analytical 3D Laplacian correlation function $\overline{\mathbf{Corr}_{3D_{aniso}}}\mathbf{u}$. When we perform the operation once we get an operator that decays much more rapidly than $\overline{\mathbf{Corr}_{3D_{aniso}}}\mathbf{u}$. This was also the case in the 2D example but if we solve the linear system 3 times with the solution $\mathbf{x}$ of the previous linear system serving as the input $\mathbf{b}$ (Algorithm 2).

Algorithm 2 Approximate $\mathbf{Corr}_{2D_{aniso}}\mathbf{u}$ with $\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}$

1: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{3D_{AIL}}^{-1}}} \quad \underbrace{\mathbf{x_{3D_1}}}_{\overline{\mathbf{Corr}_{3D_{AIL}}}\mathbf{u}} = \underbrace{\mathbf{b}}_{\mathbf{u}}$

2: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{3D_{AIL}}^{-1}}} \quad \underbrace{\mathbf{x_{3D_2}}}_{\overline{\mathbf{Corr}_{3D_{AIL}}}\mathbf{x_{3D_1}}} = \underbrace{\mathbf{b}}_{\mathbf{x_{3D_1}}}$

3: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{3D_{AIL}}^{-1}}} \quad \underbrace{\mathbf{x_{3D_3}}}_{\overline{\mathbf{Corr}_{3D_{AIL}}}\mathbf{x_{3D_2}}} = \underbrace{\mathbf{b}}_{\mathbf{x_{3D_2}}}$

4: {It appears that: $\mathbf{x}_{3D_3}$ is similar to $\overline{\mathbf{Corr}_{3D_{aniso}}}\mathbf{u}$}

If we apply this process we find that we get a result $\mathbf{x}_{3D_3}$ that is similar to $\overline{\mathbf{Corr}_{3D_{aniso}}}\mathbf{u}$. We show a result for $L = 20$ m (Figure 2.43).

To test the operator with a more complex vector we look at a 3D white noise vector $\mathbf{u} = \boldsymbol{\epsilon}_w$ of dimension $51 \times 51 \times 51$. We use $L_x = L_y = L_z = 10$ m for $h = 5$ m. The comparison between the results obtained from $\mathbf{x}_{3D_3}$ and $\overline{\mathbf{Corr}_{3D_{aniso}}}\mathbf{u}$ are very similar (Figure 2.44).

Rotation

To include rotation in the 3D sparse the procedure is similar to the 2D case where we need to discretize $\delta^2(x - \hat{x}')$, $\delta^2(y - \hat{y}')$ and $\delta^2(z - \hat{z}')$ where the $\hat{\ }$ refers to a rotation of the 3D Cartesian coordinate system by the rotation matrix described by expression 2.22. To simplify the following steps we use the following notation for our rotated and un-rotated second derivative of the delta function

$$\frac{\partial^2}{\partial \hat{x}'^2} = \delta^2(x - \hat{x}') \text{ and } \frac{\partial^2}{\partial x'^2} = \delta^2(x - x')$$

$$\frac{\partial^2}{\partial \hat{y}'^2} = \delta^2(y - \hat{y}') \text{ and } \frac{\partial^2}{\partial y'^2} = \delta^2(y - y')$$

$$\frac{\partial^2}{\partial \hat{z}'^2} = \delta^2(z - \hat{z}') \text{ and } \frac{\partial^2}{\partial z'^2} = \delta^2(z - z'). \tag{2.58}$$
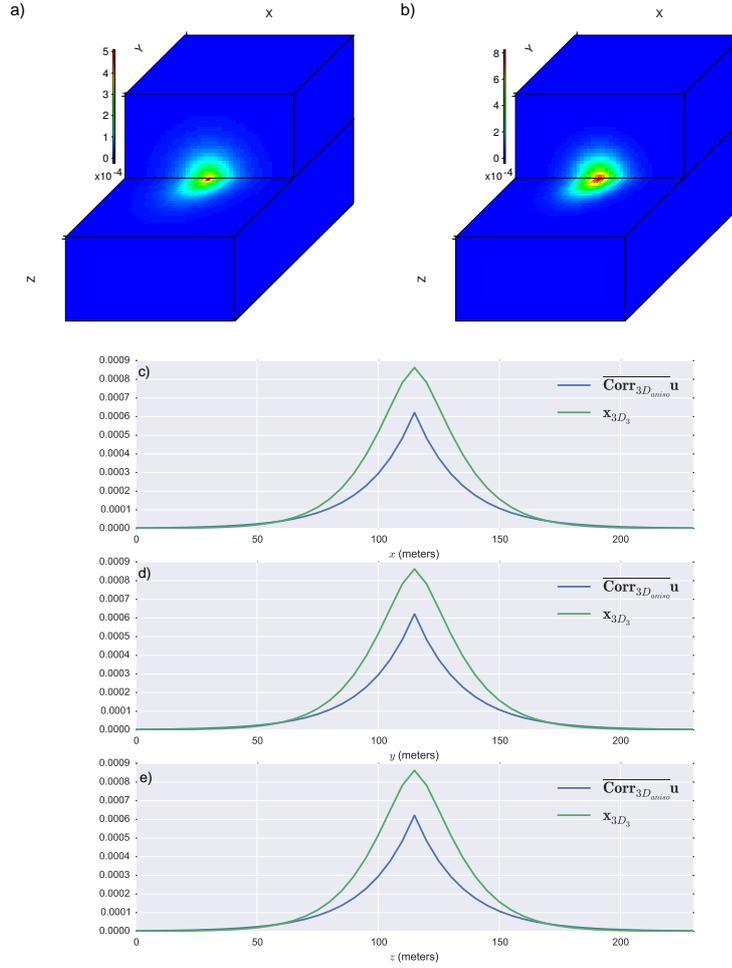
Figure 2.43: 3D vector spike test ($\mathbf{u} = \boldsymbol{\delta}$). a) $\overline{\mathbf{Corr}_{3D_{aniso}}}\mathbf{u}$ and b) $\mathbf{x}_{3D_3}$. Traces extracted from the middle of the model in the x c) y d) and z directions. e) The operators are not exactly the same but another test is used to compare their wavenumber filtering, properties on a more complex vector (Figure 2.44).

So we want to represent $\frac{\partial^2}{\partial \hat{x'}^2}$, $\frac{\partial^2}{\partial \hat{y'}^2}$ and $\frac{\partial^2}{\partial \hat{z'}^2}$ on the non-rotated Cartesian grid that $\frac{\partial^2}{\partial x'^2}$, $\frac{\partial^2}{\partial y'^2}$ and $\frac{\partial^2}{\partial z'^2}$ use.

If we start with

$$\frac{\partial^2}{\partial \hat{x'}^2} = \frac{\partial^2}{\partial \hat{x'}}\left(\frac{\partial^2}{\partial \hat{x'}}\right) = \frac{\partial^2}{\partial \hat{x'}}\underbrace{\left(\frac{\partial}{\partial x'}\frac{\partial x'}{\partial \hat{x'}} + \frac{\partial}{\partial y'}\frac{\partial y'}{\partial \hat{y'}} + \frac{\partial}{\partial z'}\frac{\partial z'}{\partial \hat{z'}}\right)}_{\text{Product rule}} \qquad (2.59)$$
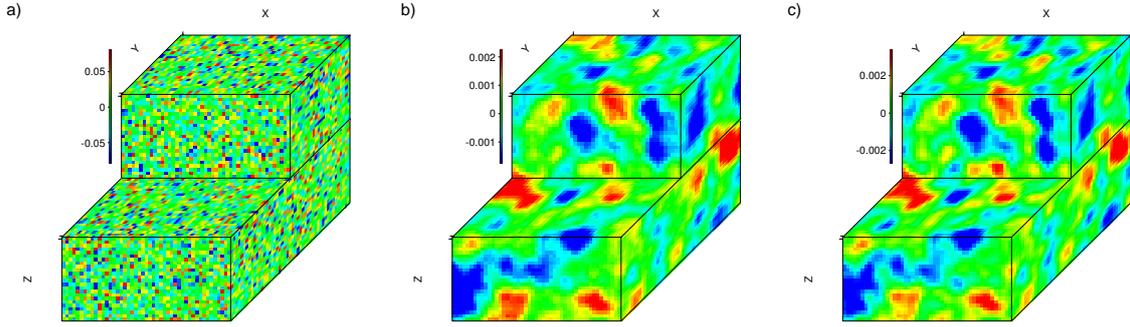
Figure 2.44: A white noise vector $\mathbf{u} = \boldsymbol{\epsilon}_w$ a) is used to compare the application of the brute force approach b)$\overline{\mathbf{Corr}}_{3D_{aniso}}\mathbf{u}$ and the result from solving the sparse linear system b) $\mathbf{x}_{3D_3}$

multiplying the two terms

$$\frac{\partial^2}{\partial \hat{x}'^2} = \frac{\partial^2}{\partial x'^2}\left(\frac{\partial x'}{\partial \hat{x}'}\right)^2 + \frac{\partial}{\partial x'}\frac{\partial}{\partial y'}\frac{\partial x'}{\partial \hat{x}'}\frac{\partial y'}{\partial \hat{x}'} + \frac{\partial}{\partial x'}\frac{\partial}{\partial z'}\frac{\partial x'}{\partial \hat{x}'}\frac{\partial z'}{\partial \hat{x}'} + \frac{\partial}{\partial y'}\frac{\partial}{\partial x'}\frac{\partial y'}{\partial \hat{x}'}\frac{\partial x'}{\partial \hat{x}'}..$$

$$.. + \frac{\partial^2}{\partial y'^2}\left(\frac{\partial y'}{\partial \hat{x}'}\right)^2 + \frac{\partial}{\partial y'}\frac{\partial}{\partial z'}\frac{\partial y'}{\partial \hat{x}'}\frac{\partial z'}{\partial \hat{x}'} + \frac{\partial}{\partial z'}\frac{\partial}{\partial x'}\frac{\partial z'}{\partial \hat{x}'}\frac{\partial x'}{\partial \hat{x}'} + \frac{\partial}{\partial z'}\frac{\partial}{\partial y'}\frac{\partial z'}{\partial \hat{x}'}\frac{\partial y'}{\partial \hat{x}'} + \frac{\partial^2}{\partial z'^2}\left(\frac{\partial z'}{\partial \hat{x}'}\right)^2 \quad (2.60)$$

using our expressions for x', y' and z'

$$x' = a\hat{x}' + b\hat{y}' + c\hat{z}'$$
$$y' = d\hat{x}' + e\hat{y}' + f\hat{z}'$$
$$z' = g\hat{x}' + h\hat{y}' + i\hat{z}' \quad (2.61)$$

where $a \rightarrow i$ are a function of the three rotation angles described in expression 2.23. We can rewrite equation 2.60 as

$$\frac{\partial^2}{\partial \hat{x}'^2} = a^2\frac{\partial^2}{\partial x'^2} + d^2\frac{\partial^2}{\partial y'^2} + g^2\frac{\partial^2}{\partial z'^2} + 2ad\frac{\partial}{\partial x'}\frac{\partial}{\partial y'} + 2ag\frac{\partial}{\partial x'}\frac{\partial}{\partial z'} + 2dg\frac{\partial}{\partial y'}\frac{\partial}{\partial z'}. \quad (2.62)$$

We can perform the same for $\frac{\partial^2}{\partial \hat{y}'^2}$ and $\frac{\partial^2}{\partial \hat{z}'^2}$ and then we have

$$\frac{\partial^2}{\partial \hat{y}'^2} = b^2\frac{\partial^2}{\partial x'^2} + e^2\frac{\partial^2}{\partial y'^2} + h^2\frac{\partial^2}{\partial z'^2} + 2be\frac{\partial}{\partial x'}\frac{\partial}{\partial y'} + 2bh\frac{\partial}{\partial x'}\frac{\partial}{\partial z'} + 2he\frac{\partial}{\partial y'}\frac{\partial}{\partial z'} \quad (2.63)$$

Figure 2.45: The discretization of a) $\delta^1(x-x')\delta^1(x-x')$ b) $\delta^1(x-x')\delta^1(y-y')$ and c) $\delta^1(-x')\delta^1(z-z')$ using centred finite difference. Note the point where the dotted lines cross represents the zero lag point. The green squares have a value of $\frac{1}{4h^3}$ while the yellow squares are $-\frac{1}{4h^3}$

and

$$\frac{\partial^2}{\partial \hat{z}'^2} = c^2 \frac{\partial^2}{\partial x'^2} + f^2 \frac{\partial^2}{\partial y'^2} + i^2 \frac{\partial^2}{\partial z'^2} + 2cf \frac{\partial}{\partial x'} \frac{\partial}{\partial y'} + 2ci \frac{\partial}{\partial x'} \frac{\partial}{\partial z'} + 2fi \frac{\partial}{\partial y'} \frac{\partial}{\partial z'}. \quad (2.64)$$
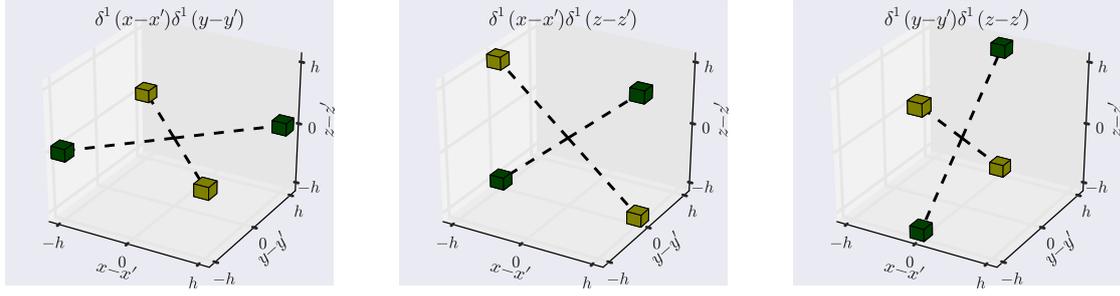
In order to accomodate rotation in our AIL stencil we need to discretize 3 more terms on the Cartesian grid

$$\frac{\partial}{\partial x'} \frac{\partial}{\partial y'} = \delta^1(x-x')\delta^1(y-y'),$$
$$\frac{\partial}{\partial x'} \frac{\partial}{\partial z'} = \delta^1(x-x')\delta^1(z-z'),$$
$$\frac{\partial}{\partial y'} \frac{\partial}{\partial z'} = \delta^1(y-y')\delta^1(z-z'), \quad (2.65)$$

where the superscript [1] denotes the first derivative.

The three extra terms add an additional 12 points to the stencil shown by the green squares and yellow squares in Figures 2.45, These additional points mean that the rotated stencil has 19 points as opposed to 7 in the stencil without rotation. The non-zero points in the stencil can be understood if one considers a rubix cube where the middle of the cube is the point with no space lag $((x', y', z') = (x, y, z))$ and the 8 extreme corners of the rubix are zero (Figure 2.46).

To highlight the behavior of rotated operator we consider two arbitrary vectors of dimension $51 \times 51 \times 51$ (Figure 2.47). The first vector is a unit spike, while the second is a white noise vector. The example shows a case where all 3 correlation lengths (L_x, L_y and L_z) are different and how we can use our rotated stencil to transform these correlation
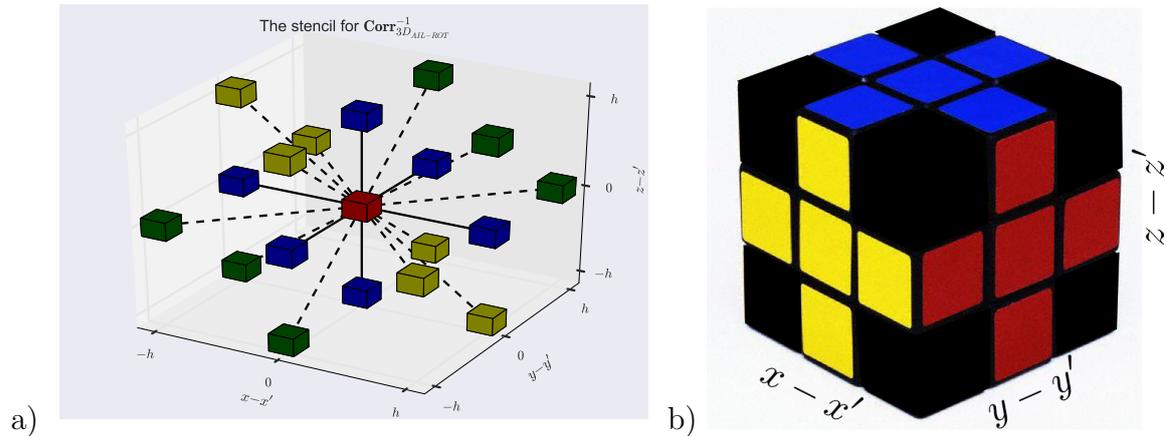
Figure 2.46: a)The stencil for the rotated inverse correlation operator $\mathbf{Corr}^{-1}_{3D_{AIL-ROT}}$ b) The stencil can be visualized if one considers a rubix cube. We have 19 non-zero elements with the middle square being the diagonal. The 8 extreme corners are not involved in the stencil and are blacked out. *NOTE:* The colours on the rubix cube are of no significance and are not related to those in a).

lengths to arbitrary orientations $L_{\hat{x}}$, $L_{\hat{y}}$ and $L_{\hat{z}}$.

2.4   Conclusions

We have shown that we can approximately extend the 1D analytical inverse covariance function (Tarantola, 2005) to higher dimensions. These operators in 2D and 3D are referred to as the additive inverse Laplacian functions. To summarise my result I look at some plots of the wavenumber spectrum of these operators versus the normalized Laplacian analytical correlation functions. I shall discuss the computational issue making the matrix vector product and solving the linear system. Finally I will discuss what the analytical operators correlation operators associted for our 2D and 3D AIL operators. This discovery was found after submission of my PhD manuscript by Phoung-Thu Trinh (Trinh, 2016).

Wavenumber spectrum of AIL operators
The 1D inverse correlation operator proposed by Tarantola is the true inverse of the 1D analytical correlation function. When we extended to higher dimension we found that this is no longer the case. In the 2D case we need to solve 2 linear systems involving $\overline{\mathbf{Corr}^{-1}_{2D_{AIL}}}$. While the 3D case we solve 3 linear systems involving $\overline{\mathbf{Corr}^{-1}_{3D_{AIL}}}$. If we look
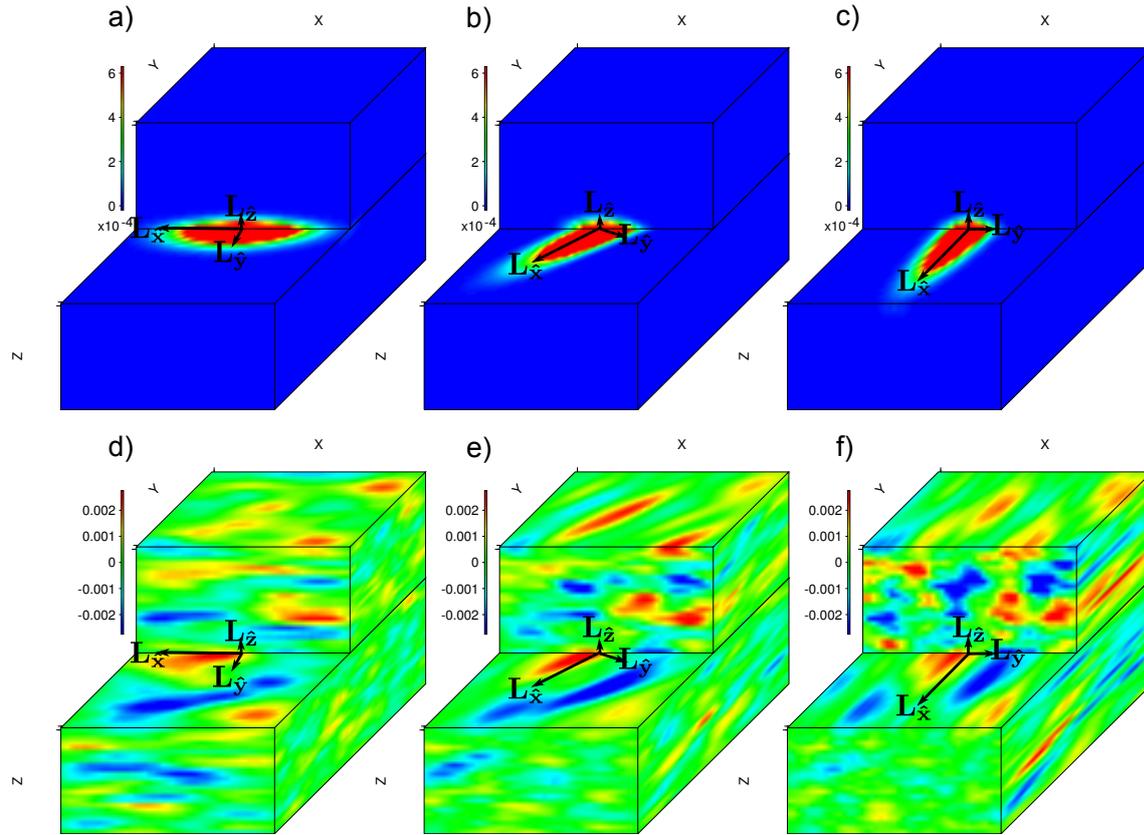
Figure 2.47: The calculation of $\mathbf{x}_{3D_3}$ for two vectors, $\mathbf{u}$. with $L_x = 6h$ $L_y = 2h$ and $L_z = h$. The input vectors for a-c) is a unit spike ($\mathbf{u} = \boldsymbol{\delta}$) and d-f) is a white noise vector ($\mathbf{u} = \boldsymbol{\epsilon}_w$). The correlation lengths are rotated around the z axis by the angle φ to become $L_{\hat{x}}$, $L_{\hat{y}}$ and $L_{\hat{z}}$. a,d) $\varphi = 0°$ b,e) $\varphi = 45°$ c,f) $\varphi = 90°$.

at the wavenumber of a "spike test" for these operators (Figure 2.48) we see that the intermediate results ($\mathbf{x}_{2D_1}$, $\mathbf{x}_{3D_1}$ and $\mathbf{x}_{3D_2}$) have been filtered to a lesser extent. Perhaps these filters would also potentially be of interest. We focus on showing the link of these filters to the Laplacian analytical correlation functions as it provided analytical results to validate the vectors we obtained from solving the sparse linear systems. One can see that both $\mathbf{x}_{2D_2}$ and $\mathbf{x}_{3D_3}$ are not exactly equivalent to our analytical correlation functions with the mismatch being worse in the 3D case.

Computational complexity of calculating AIL results

We will briefly outline a computational cost comparison of filtering use the the computation our AIL framework in 2D ($\mathbf{x}_{2D_2}$) and 3D ($\mathbf{x}_{3D_3}$). To give a fair comparison we compare the cost of solving our sparse linear system multiple times (twice in 2D and three
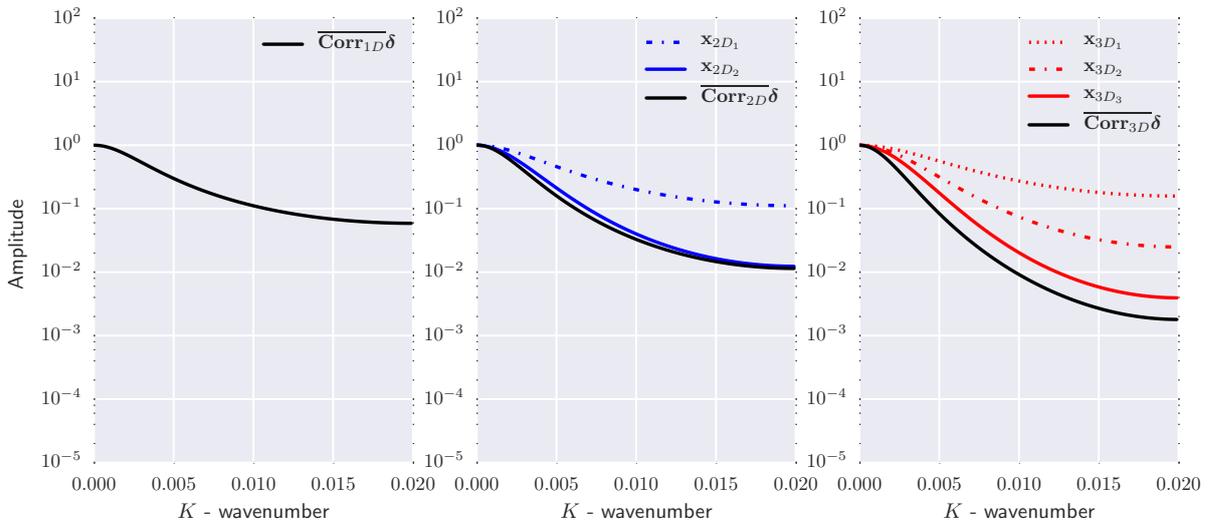
Figure 2.48: The vertical wavenumber of spike tests of operators in 1D (*LEFT*), 2D (*CENTRE*) and 3D (*RIGHT*). All examples have the same L value ($L = 50$ m and $h = 25$ m) in 2D and 3D these represent the vertical wavenumber component calculate from 2D FFT(where $K_x = 0$) and a 3D FFT ($K_x = K_y = 0$).

times in 3D) as opposed to the application of the analytical Laplacian using a spatial convolution approach and the tensorized Gaussian correlation function. It is important to note that although both the alternative approaches have certain positive aspects, they are also both limited by strong negatives. For instance, the spatial convolution approach is able to handle dip, but has a significant computational cost associated. In contrast the tensorized Gaussian approach has low computational cost, but is unable to efficiently handle dips. We argue that solving our AIL linear system allows one to take advantage of both the ability to include dip with a low computational cost. We use a conjugate gradient algorithm to solve the linear system involving our AIL operators, the misfit of our CG algorithm is defined by

$$\frac{||\mathbf{b} - \mathbf{A}\mathbf{x}^k||}{||\mathbf{b}||}, \tag{2.66}$$

where $\mathbf{A}$, $\mathbf{x}$ and $\mathbf{b}$ has been previously described in Algorithm 1 and 2. The value, k refers to the iteration number and convergence is said to be reached when the values in equation 2.66 is less than 0.0001. Such a misfit value is supported by Hestenes and Stiefel (1952). Our computational sensitivity results have all been computed with no multi-threading (single CPU, sequential jobs) on the CIMENT cluster "Froggy".

0   Increasing value of n. Constant value of L. ⟶ 1

Figure 2.49: Computational cost sensitivity attained by increasing the size of the model vector $\mathbf{u}$ while keeping the value of L constant.

0   Increasing value of L. Constant value of n. ⟶ 1

Figure 2.50: Computational cost sensitivity attained by increasing the correlation length, L while keeping the value of n constant.

When considering the computational cost we look at two sensitivities, the sensitivity associated with increasing the number of elements in $\mathbf{u}$ (Figure 2.49) and the sensitivity to increasing the correlation length (Figure 2.50).

The first sensitivity is considered by keeping our correlation lengths constant while we increase the number of the elements in the input vector $\mathbf{u}$ we consider. For simplicity, we consider an identical number of points, n in each dimension. The number of elements in 2D is given by $N_{2D} = n \times n$ and by $N_{3D} = n \times n \times n$ in 3D. For this first sensitivity we use a relatively short correlation length ($L = 10$ m, $h = 5$ m) relative to the spatial extent of the vector $\mathbf{u}$. The computational time comparison is plotted in Figure 2.51 and
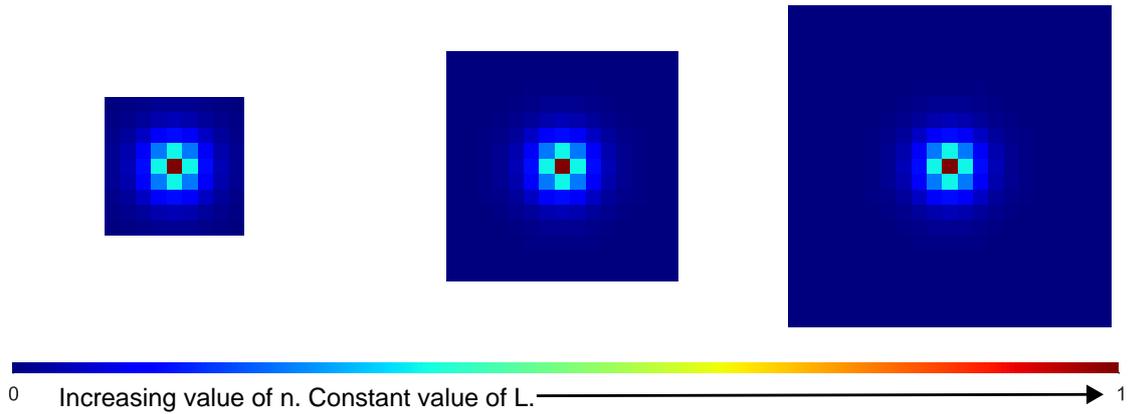
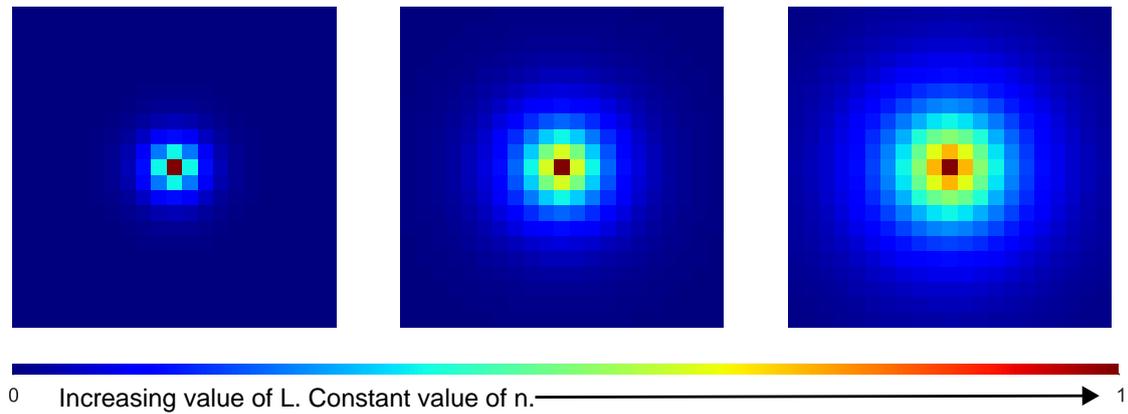Figure 2.51: Computational cost sensitivity described by increasing the size of the model vector $\mathbf{u}$ while keeping the value of L constant. The gray and black arrows highlight the large increase in computation time when comparing AIL and tensorized Gaussian to the spatial convolution approach.

highlighted in Table 2.1. We see that as we increase the value of n our computation time for all the 2D operations increases proportional to n^2. In the 3D case this increase is proportional to n^3. The computational time results for the tensorization and AIL results is quite similar in the 2D and 3D cases, however there is a significant difference in the cost when compared to the spatial convolution approach which has a computation time that is higher by a factor of almost 3 orders of magnitude.

An interesting point to note in the AIL case is that the number of iterations required to reach convergence appears to have no relation to the number of elements in the model vector. The increase in computation time for the AIL operator is associated with an increase in the computation time for each CG iteration as the value of n increases. It is also important to note that different input vectors were tested. One was a

n	2D Brute Time(s)	2D Tensor Time(s)	2D AIL Time(s), N_{it}	3D Brute Time(s)	3D Tensor Time(s)	3D AIL Time(s), N_{it}
21	-	-	-	17.4	0.07	0.17, 59
43	0.19	-	0.02, 39	298	0.85	1.4, 59
63	0.49	-	0.04, 41	1125	2.89	4.41, 59
85	0.95	0.03	0.08, 41	3026	6.83	10.85, 59
105	1.46	0.05	0.11, 41	6000	13.17	20.59, 59
127	2.21	0.08	0.18, 41	11108	24.17	36.25, 59
169	4.04	0.16	0.29, 41	27300	60.63	85.33, 59
189	5.1	0.19	0.37, 41	38702	85.17	119.4, 59
601	54.41	1.95	3.8, 41	-	-	-

Table 2.1: Computational time sensitivity for increasing the size of the input vector **u**. Absent fields highlighted by blue dashes are due to the computation time being too fast to be measured accurately while the red dash highlight memory limitations with computation.

Dirac spike, ($\mathbf{u} = \boldsymbol{\delta}$) while the second was a white noise vector ($\mathbf{u} = \boldsymbol{\epsilon}_w$). It was noticed that both vectors converged in the same number of iterations suggesting that the number of iterations required for convergence of the AIL is independent of n or the input vector **u**.

The second sensitivity (increasing L while keeping n constant) is summarized in Figure 2.52 and Table 2.2. For this case we used $n = 105$, which was a relatively large value while increasing the size of the correlation length. The use of a large vector, **u** relative the correlation length ensures that the computational time estimates are not significantly influenced by edge effects related to the limited computational domain. As we increase the correlation length L we see that we require more iterations to reach our convergence criteria. This increase in the number of iterations appears to be approximately linearly correlated with L. The interpretation for this result may be due to the fact that as we increase the value of L, the AIL operator becomes less diagonally dominant. This linear increase in the number of iterations and by extension the computation time is small in comparison to the spatial convolution approach. The spatial convolution approach computation time increases theoretically proportional to L^2 in 2D and L^3 in 3D. Limitations of the size of the computational domain slightly decrease this in our case but there will be a significant costs associated with spatial convolution approach as the correlation length increases relative to the discretized cell size, h. In the following Chapters we use our AIL operators as filters to constrain FWI where the value of $\frac{L}{h}$ is as high as 30. Application of the spatial convolution approach becomes computationally prohibitive due to
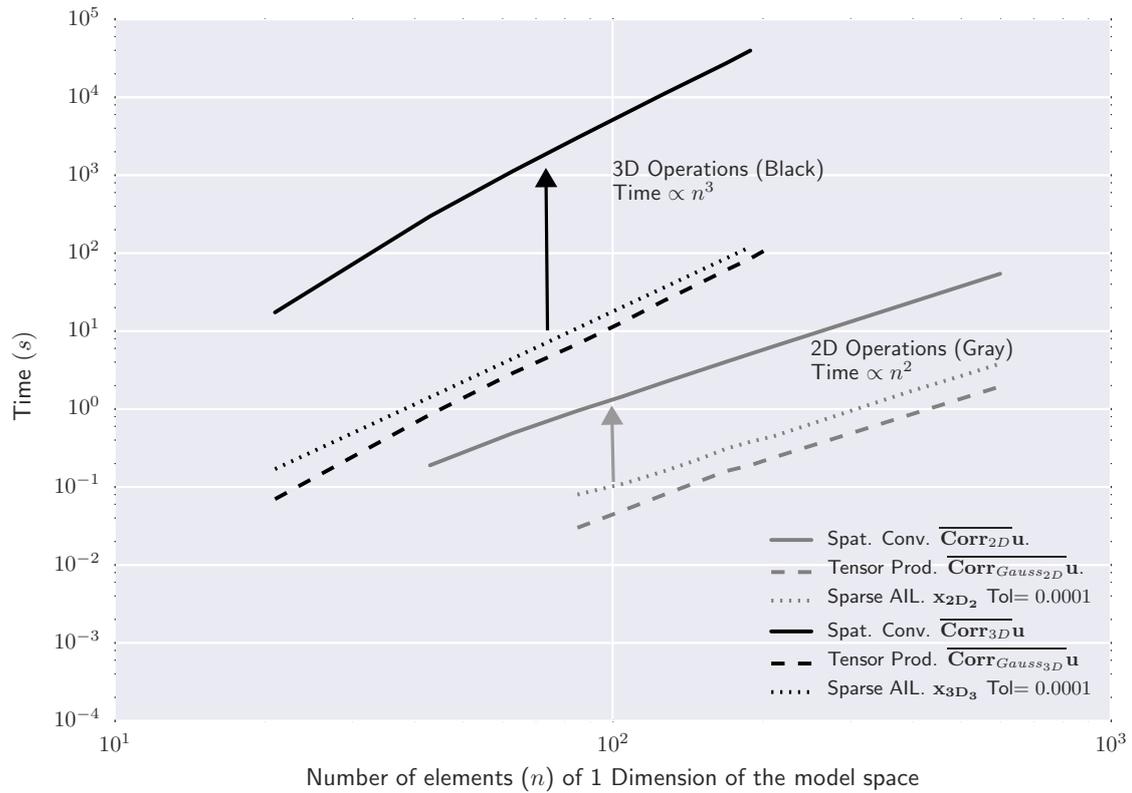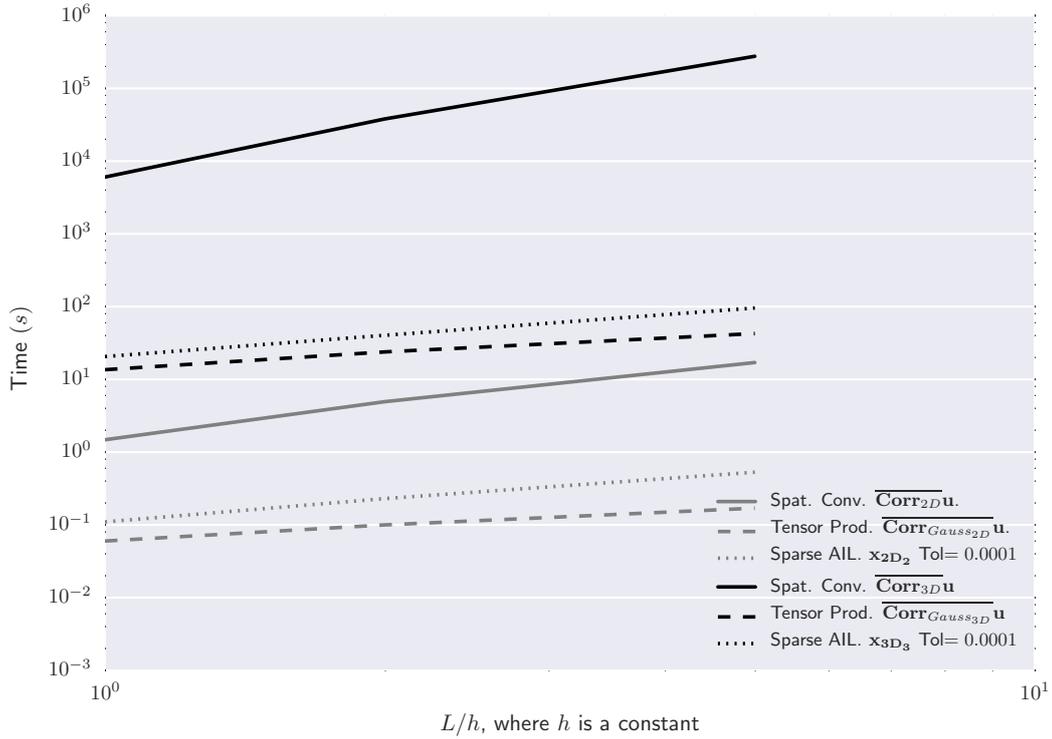
Figure 2.52: Computational cost sensitivity described by increasing the size of the correlation length, L, while keeping the value of n constant.

the excessively high values of n and $\frac{L}{h}$. The tensorized Gaussian approach appears to computationally efficient, but it cannot handle dip.

$\frac{L}{h}$	2D Brute Time(s)	2D Tensor Time(s)	2D AIL Time(s), N_{it}	3D Brute Time(s)	3D Tensor Time(s)	3D AIL Time(s), N_{it}
1	1.48	0.06	0.11, 41	6039	13.59	20.59, 113
2	4.94	0.1	0.23, 78	38030	23.87	40.39, 264
5	17.05	0.17	0.53, 176	276324	42.65	95.8, 449

Table 2.2: Computational time sensitivity for increasing the size of correlation length, L.

Analytical expression of correlation operators:
The 2D correlation function $\overline{Corr_{2D_{AIL}}}(x, z; x', z')$ related to the inverse operator $\overline{Corr_{2D_{AIL}}^{-1}}(x, z; x', z')$ at the time of the submission of the original PhD manuscript this was unknown. Subsequent to submission before the soutenance date, Phuong-Thu Trinh looked to apply these operators in their weak-form (Trinh, 2016). She showed that It can be shown this is the

solution of a partial differential equation which is equivalent to the modified Helmholtz equation with complex wavenumber k

$$\nabla^2 \overline{Corr_{2D_{AIL}}} + k^2 \overline{Corr_{2D_{AIL}}} = 0. \tag{2.67}$$

Due to the radiation condition at infinity (Polyanin and Nazaikinskii, 2002; Abramowitz and Stegun, 1972), the only possible solution of this equation is

$$\overline{Corr_{2D_{AIL}}}(x, z; x', z') = \frac{1}{\pi L_x L_z} K_{\mu=0} \left(\sqrt{\frac{2(x - x')^2}{L_x^2} + \frac{2(z - z')^2}{L_z^2}} \right), \tag{2.68}$$

where $K_\nu()$ is a modified Bessel function of the second kind.

Similar to the 2D case the function she also found that $\overline{Corr_{3D_{AIL}}}$ is characterized by a partial differential equation, which can be transformed to into the modified spherical Bessel equation (Abramowitz and Stegun, 1972). Due to the radiation condition at infinity, the only possible solution of this equation is

$$\overline{Corr_{3D_{AIL}}}(x, y, z; x', y', z') = \frac{1}{4\pi} \frac{3\sqrt{3}}{L_x L_y L_z} \sqrt{\frac{2}{\pi}} \, r^{-1/2} K_{\nu=\frac{1}{2}}(r) \tag{2.69}$$

where
$$r = \sqrt{\frac{3(x - x')^2}{L_x^2} + \frac{3(y - y')^2}{L_y^2} + \frac{3(z - z')^2}{L_z^2}}. \tag{2.70}$$

Chapter summary:

This chapter has introduced the additive inverse Laplacian correlation operator. In the next chapter we will show the utility of this operator in the framework of 2D Frequency domain FWI.

Chapter 3

FWI: Synthetic Example

Contents

Chapter overview:

This chapter will introduce frequency domain full waveform inversion (Pratt and Shipp, 1999; Ravaut et al., 2004; Brenders and Pratt, 2007) and highlight how our sparse Laplacian correlation framework can be used to mitigate against some unwanted features that map into the model parameter reconstruction. First, I will use the Valhall synthetic model introduced in Chapter 1 to introduce the efficient finite difference modelling of monochromatic wavefields using the 2D visco-acoustic wave equation. We will use this forward modelling scheme to show how we can look to efficiently minimize the difference between our measured data $\mathbf{d}_{obs}$ and the modelled data $\mathbf{d}_{mod}(\mathbf{m}^k)$. This modelled data vector is computed in the current model $\mathbf{m}^k$, where k represents the iteration of our

linearized, least squares inversion. Our 2D additive inverse Laplacian operator is used to constrain the wavenumber spectrum of the FWI gradient, $\mathbf{G}^k$ (to be introduced later). We show that the ability to efficiently perform the application of the correlation operator to a vector (the FWI gradient) is a useful and simple to parametrize approach to stabilize FWI. Two common ways that the model parameter (in our case $\mathbf{v}_p$) reconstruction can be impacted are investigated. The first problem we will look at arises from the sparse sampling of shots and frequencies. Decimation of frequencies and shots can decrease the computational cost of the inversion but is also known to allow artifacts to map into the model space (Asnaashari, 2013). The second problem we will investigate is when $\mathbf{d}_{obs}$ is contaminated by white noise. In theses cases we will compare the utility of our AIL operator with Tikhonov regularization.

We will also consider application of FWI to the Marmousi model. Unlike the Valhall model, which to a large extent has horizontally dipping geology, the Marmousi model has tilted and variable dipping geology. We will show how this is not an issue for our sparse additive inverse correlation framework, where it becomes a significant problem for the Gaussian.

3.1 Frequency Domain FWI strategy

3.1.1 Forward Problem

Frequency domain FWI has been acknowledged as a promising approach to build high-resolution velocity models (Pratt and Shipp, 1999; Ravaut et al., 2004; Brenders and Pratt, 2007). One of the key advantages of the frequency domain approach for forward modelling is that it can be quite fast for multiple sources (Pratt and Worthington, 1990; Stekl and Pratt, 1998). This is primarily the case when the number of elements in the modelling grid is not too significant. Our 2D Valhall synthetic is a good candidate for frequency domain forward modelling due to the relatively small number of model parameters we will invert ($394 \times 1304 = 513,776$). To generate the wavefield, we need to solve the 2D acoustic wave equation.

In the frequency domain, the wave equation reduces to a system of linear equations

$$\mathbf{B}(\omega, \mathbf{m}(\mathbf{x}))\mathbf{u}(\omega, \mathbf{x}) = -\mathbf{s}(\omega, \mathbf{x}). \tag{3.1}$$

Where the matrix $\mathbf{B}(\omega, \mathbf{m}(\mathbf{x}))$ is the impedance matrix for our earth model, $\mathbf{m}$ is represented by spatial coordinates $\mathbf{x}$ for the the monochromatic frequency ω. The RHS of the equation represents the sources modelled, $\mathbf{s}$, at positions $\mathbf{x}$ and the frequency ω.

A number of different strategies are available to discretize this problem, however the finite-difference method is typically preferred (Virieux, 1986; Levander, 1988; Graves, 1996; Operto et al., 2007). In more complex environments, such as onshore land environments with complicated surface topography, (free surface) more sophisticated finite-element or finite-volume can be considered to properly account for the boundary conditions through unstructured meshes (Komatitsch and Vilotte, 1998). As our Valhall case is a marine example, with a simple free surface (the water air interface), we have employed the preferred, computationally efficient finite difference method.

The mixed grid stencil (Hustedt et al., 2004) is used in our Valhall case. Perfectly matching layers (Berenger, 1998) are used to attenuate back scattering energy from the model space extremes. The mixed grid stencil combines a Cartesian second-order staggered grid stencil with one orientated at 45° (Jo et al., 1996). The description of the discretization of the staggered grid scheme in the Cartesian and 45° orientations starts by formulating the acoustic wave equation as a first-order hyperbolic system in the time domain using the pressure and particle velocities. This approach is then taken to the frequency domain where the required earth parameters (buoyancy $b(x, z)$ and bulk modulus $\kappa(x, z)$) are discretized on a Cartesian aligned and a 45° rotated orientation. By combining the two discretization it provides a modelling scheme that approximates the accuracy of a 4th order staggered grid scheme while only requiring 9 point as opposed to 13 (Figure 3.1).

To compute the wavefield $\mathbf{u}$ we need to solve the linear system

$$\underbrace{\mathbf{B}(\omega, \mathbf{m}(\mathbf{x}))}_{\mathbf{A}} \underbrace{\mathbf{u}(\omega, \mathbf{x})}_{\mathbf{x}} = \underbrace{\mathbf{s}(\omega, \mathbf{x})}_{\mathbf{b}}. \tag{3.2}$$

The LU direct solver approach is typically viewed as en efficient option for solving the 2D forward problem (Jo et al., 1996; Stekl and Pratt, 1998; Hustedt et al., 2004). The matrix $\mathbf{A}$ is decomposed into upper $\mathbf{U}$ and lower $\mathbf{L}$ triangular matrices such that

$$\mathbf{A}\mathbf{x} = (\mathbf{L}\mathbf{U})\mathbf{x} = \mathbf{L}(\mathbf{U}\mathbf{x}) = \mathbf{b}. \tag{3.3}$$

Once this decomposition is completed, we can efficiently solve our linear system by utilizing the fact that solving a triangular set of equations is quite trivial. We obtain our
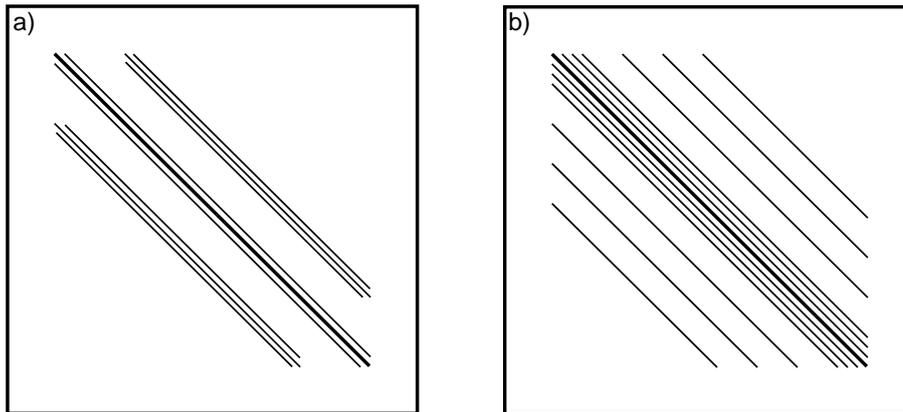
Figure 3.1: The sparsity of the real components of the complex impedance matrix **B** (reproduced from Hustedt et al. (2004)). The impedance matrix is large with the dimension of $(nx \times nz)^2$. The mixed stencil has only $9 \times nx \times nz$ non-zero entries a), while the fourth order staggered grid stencil b) requires $13 \times nx \times nz$ non-zero entries.

wavefield for a given monochromatic frequency **u** by solving the following linear systems

$$\mathbf{Ly} = \mathbf{b} \tag{3.4}$$

$$\mathbf{Ux} = \mathbf{y} \tag{3.5}$$

using forward and back substitutions. This process is quick and once the matrix **A** is factorized, the calculation of multiple shots can be performed rapidly (Marfurt, 1984). Our application of LU decomposition uses the massively parallel MUMPS library (MUMPS-team, 2011) that leverages the multi-frontal methods for LU decomposition (Duff and Reid, 1983). This allows rapid computation via three levels of parallelism (Amestoy et al., 2006).

If we consider our Valhall model, we will simulate 48 equally spaced explosive shots placed at $z = 0$. We use a constant density model in this synthetic case and model with minimal an-elastic attenuation ($Q = 1000$) which is implemented using complex velocities (Toksöz and Johnston, 1981). Three shots of the monochromatic wavefield at $6Hz$ are shown for the true model ($\mathbf{m}_{true}$) and the initial model $\mathbf{m}_0$ (Figure 3.2) .

3.1.2 Inverse Problem

The goal of full waveform inversion is to improve the match between the modelled wavefield $\mathbf{u}_{mod}(\mathbf{m})$ and the observed wavefield $\mathbf{u}_{obs}$. This is achieved using a local, linearized,
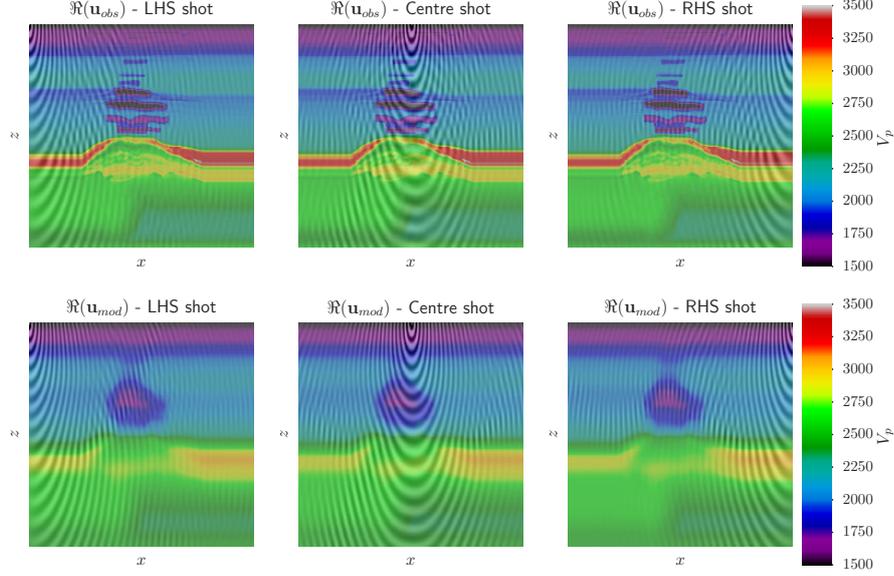
Figure 3.2: The real component of observed $\mathbf{u}_{obs}$ (*TOP*) and modelled (*BOTTOM*) $\mathbf{u}_{mod}(\mathbf{m})$ monochromatic wavefields at $6Hz$. Underlying the wavefields is the velocity model used for their respective calculation.

least squares approach (Tarantola, 2005). The ultimate goal is to iteratively update our current model $\mathbf{m}^k$ (where k represents the kth iteration) so that when we reach convergence after l iterations $\mathbf{m}^l \approx \mathbf{m}_{true}$ (where $\mathbf{m}_{true}$ is the true earth model). In our description the model parameter $\mathbf{m}$ relates to the compressional velocity vector $\mathbf{v}_p$.

It is typically impossible to measure the entire wavefield and we are instead limited to where we have placed receivers (i.e. surface receivers in the Valhall case). We can represent our data $\mathbf{d}$ as

$$\mathbf{d}_{obs} = \mathbf{R}\mathbf{u}_{obs}(\mathbf{x}, \omega) \tag{3.6}$$

$$\mathbf{d}_{mod}^k = \mathbf{R}\mathbf{u}_{mod}^k(\mathbf{x}, \omega), \tag{3.7}$$

where $\mathbf{R}$ is a operator that extracts the wavefield components at the receiver positions. Figure 3.3 shows the data residuals $\Delta \mathbf{d} = \mathbf{d}_{obs} - \mathbf{d}_{mod}^k$ between our observed data and the wavefield modelled using our initial model, $\mathbf{m}^0$. We notice from these residuals that the dominant differences between the two measured wavefields comes from traces offsets from the source locations. It is the goal of FWI to attempt to decrease this misfit.
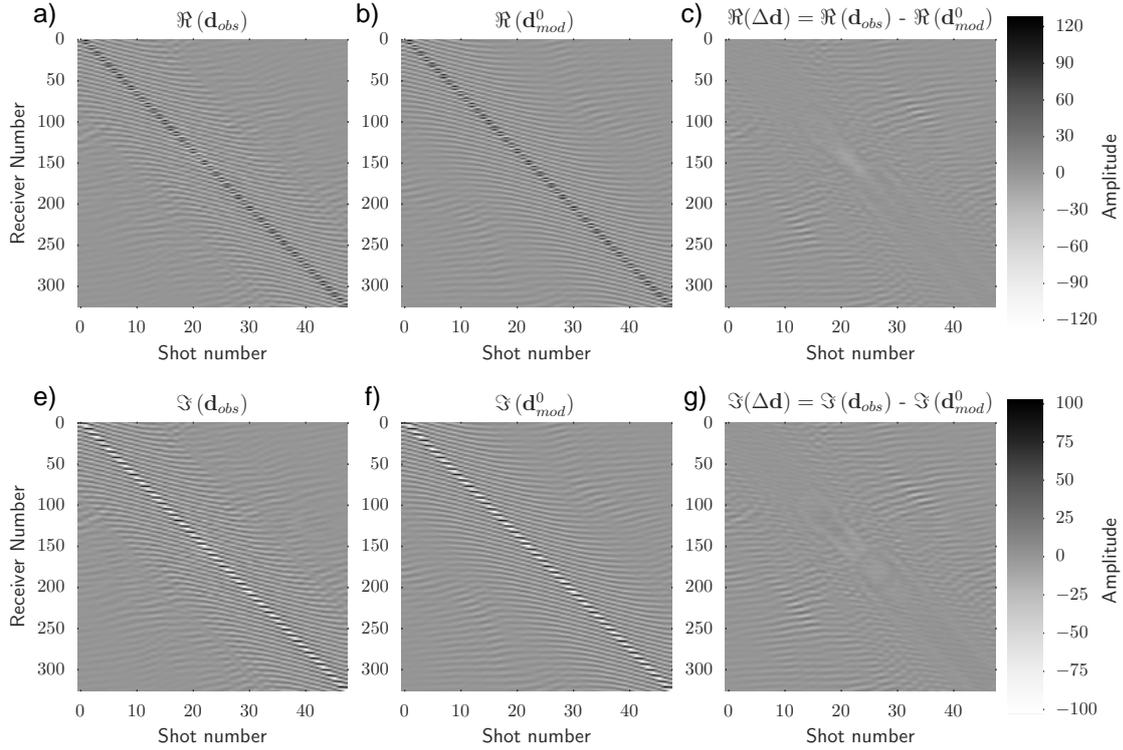
Figure 3.3: The real (a,b,c) and imaginary (d,e,f) components of the observed (a,d) and modelled data (b,e). The residuals (c,f) measured at each of the 326 receivers for the 48 shots, are used to update the initial model, $\mathbf{m}^0$, it is hoped that $\mathbf{m}^0 \to \mathbf{m}_{true}$. The highest amplitude in both the real ($\Re$) and imaginary $\Im$ component for $\mathbf{d}_{obs}$ and $\mathbf{d}_{mod}^k$ occurs close to the source (shown as the diagonal of the vectors). The source to receiver offset increases as function of the row offset from the diagonal. The highest residual values ($\Delta\mathbf{d}$) occur offset from the source location. It is this misfit that FWI will aim to decrease.

In our approach we assess the "goodness of fit" using the length method (Menke, 1984) where we look at the difference between the measured wavefields using the ℓ_2 norm (Tarantola, 1987). Our ℓ_2 objective function $C(\mathbf{m})$ is given by

$$C(\mathbf{m}^k) = \frac{1}{2} \sum_{n\omega} \sum_{ns} \Delta\mathbf{d}^{k\dagger}(\omega) \mathbf{W}_d \Delta\mathbf{d}^k(\omega), \qquad (3.8)$$

where the symbol † represents the complex conjugate. Summation over the number of sources, ns and the number of modelled frequencies $n\omega$ is necessary to calculate the objective function. The operator $\mathbf{W}_d$ is a weighting matrix in the data space. This

weighting matrix is typically diagonal and can be used to provide weighting with respect to the data vector (Tarantola, 2005). An example would be an increasing weighting with offset to try to focus more on the late arrivals.

Our application of full waveform inversion uses a local optimization. In order to attempt to converge to the true earth model, the minimum of the objective function $C(\mathbf{m}^k)$ is searched for in the vicinity of the current model $\mathbf{m}^k$. The linearisation expresses the updated model $\mathbf{m}^{k+1}$ as the sum of the current model plus a perturbation vector $\Delta\mathbf{m}^k$

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \alpha^k \Delta\mathbf{m}^k, \tag{3.9}$$

where α^k represents the step length that is taken along the perturbation vector. We find that $\Delta\mathbf{m}^k$ can be expressed as

$$\Delta\mathbf{m} = - \left[\underbrace{\frac{\partial^2 C(\mathbf{m}^k)}{\partial\mathbf{m}^2}}_{\mathbf{H}^k=\text{Hessian}} \right]^{-1} \underbrace{\frac{\partial C(\mathbf{m}^k)}{\partial\mathbf{m}}}_{\mathbf{G}^k=\text{Gradient}}. \tag{3.10}$$

Fortunately we have strategies to efficiently calculate the gradient $\mathbf{G}^k$ via the adjoint-state method (Plessix, 2006). In this description, the gradient can be seen as the zero-lag correlation between the forward and adjoint wavefields. If we consider a single grid point, i of the gradient for a single frequency and source we see that

$$G_i^k = \Re\left[\mathbf{u}^T \left[\frac{\partial\mathbf{B}}{\partial m_i} \right]^T \mathbf{B}^{-1^T}(\mathcal{P}\Delta\mathbf{d}^k)^* \right], \tag{3.11}$$

where $\frac{\partial\mathbf{B}}{\partial m_i}$ is the sparse matrix representing the radiation pattern of the scattering of the model parameter m_i. The term $\mathcal{P}$ is an operator that places the residual vector at the measured receiver locations and then fills the rest of the computational domain with zeros. The terms $\Re$ and $*$ represent the real component of the complex vector and the conjugate respectively. Due to the spatial reciprocity of Greens functions the vector $\mathbf{B}^{-1^T}$ can be replaced by $\mathbf{B}^{-1}$ if the wave equation is self-adjoint.

$$G_i^k = \Re\left[\mathbf{u}^T \left[\frac{\partial\mathbf{B}}{\partial m_i} \right]^T \underbrace{\mathbf{B}^{-1}(\mathcal{P}\Delta\mathbf{d}^k)^*}_{\mathbf{r}_b=\text{Adjoint field}} \right] \tag{3.12}$$

This back propagated wavefield $\mathbf{r}_b$ comes from the back propagation of all the residuals
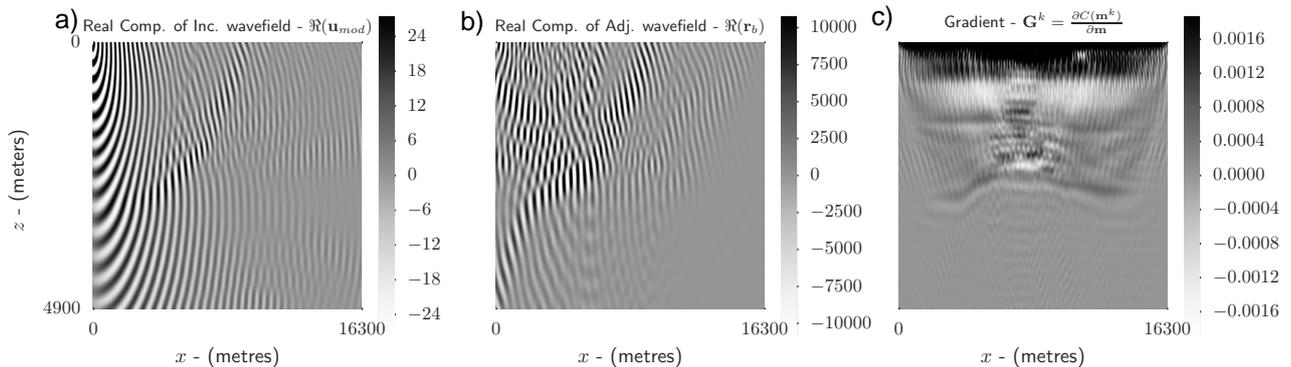
Figure 3.4: a): The incident wavefield $\mathbf{u}^k_{mod}$ for a source at the LHS extreme of the model. The restriction operator $\mathbf{R}$ is applied to allow calculation of the residual vector $\Delta\mathbf{d}^k$ before these residuals are re-injected at the receiver locations (using $\mathcal{P}$) to calculate the back-projected adjoint field $\mathbf{r}_b$ b). The gradient is formed from the combination of these two wavefields 3.12. c). The gradient in this example is computed from summation over all of the 48 shots for the monochromatic frequency $\omega = 6Hz$

associated with one seismic source and is commonly referred to as the adjoint field. In order to calculate the gradient, we therefore need to compute the solution of two forward modelling runs per shot. The first to compute the incident wavefield $\mathbf{u}$ followed by a second to back propagate the residuals between $\mathbf{d}_{obs}$ and $\mathbf{d}^k_{mod}$ to calculate the adjoint field. The underlying basis of this process is similar to the imaging condition of reverse time migration where the recorded wavefield is back projected. In the case of FWI we back project the residuals. In order to calculate the gradient related to the cost function shown in 3.8 we are required to sum over sources (ns) and frequencies ($n\omega$). Figure 3.4 provides an illustration of the incident and adjoint fields well as the gradient for the frequency $\omega = 6Hz$.

Although the gradient can be computed efficiently using the adjoint state method the inverse Hessian $[\mathbf{H}^k]^{-1}$ from our perturbation vector 3.10 is typically never explicitly built, due to the large required computation time. We are able to allow some influence of the inverse hessian using the pseudo-second order L-BFGS optimization (Byrd et al., 1995; Nocedal and Wright, 2006). The non-diagonal hessian $[\mathbf{H}^k]$ can by approximated by using information of the gradient $\mathbf{G}^k$ at previous iterations. Typically the L-BFGS stores a small number of previous gradients (i.e 5-20) and the initial guess required, $\mathbf{H}^0$ can come from the inverse of the diagonal hessian (Brossier et al., 2009a). The step length, α^k is calculated by performing a line-search that satisfies Wolfe's condition (Wolfe, 1969). The use of L-BFGS has been shown to provide improved results when compared to

the first order approaches such as preconditioned steepest descent or conjugate gradient (Brossier et al., 2009b).

3.2 Valhall synthetic application

3.2.1 Optimization results from purely data fitting

We consider the Valhall example and show how FWI performs when we let it minimize the objective function 3.8 with no regularization. We will consider two different cases where the reconstruction of $\mathbf{m}^{true}$ is influenced by the ill-posed nature of the inversion. The first case we will consider involves an analysis of the effects of coarse frequency sampling. The second case will focus on the the influence of noise in our observed data $\mathbf{d}_{obs}$.

Influence of frequency sampling
Although the wavefield modelling examples shown so far consider a single monochromatic frequency ω, FWI typically considers multiple frequencies spanning a desired frequency band. To decrease the non-linearity of the problem and avoid cycle skipping it is common place to start the inversion only focusing on the lower frequencies before introducing higher frequencies later in the inversion (Bunks et al., 1995). We use this approach in our case inverting over 3 successive frequency bands with the output from the previous band serving as the $\mathbf{m}_0$ for the subsequent band. We invert from $4Hz \rightarrow 6Hz$, then $4Hz \rightarrow 8Hz$ and finally $4Hz \rightarrow 10Hz$. As frequency domain forward modelling needs to perform the LU decomposition for each frequency band, the selection of the frequency increment $\Delta\omega$ within a band can significantly change the computation time. A coarse frequency increment has been shown to decrease the quality of the inversion result (Asnaashari, 2013). We will analyse the effect of the frequency increment in this case study, before showing a potentially rapid way of mitigating these artifacts. Full waveform inversion is performed with 3 different values of $\Delta\omega$ ($1Hz$, $0.5Hz$ and $0.25Hz$). Adjusting $\Delta\omega$ changes the total number of frequencies as summarized in Table 3.1

We see from the full waveform inversion results our reconstruction of the Valhall model is optimal for the lowest value of $\Delta\omega$ (Figure 3.5). In the other cases high wavenumber vertical noise has been introduced into the model space. The origin of

	Band 1 4Hz-6Hz	Band 2 4Hz-8Hz	Band 3 4Hz-10Hz
Nfreq $\Delta\omega = 1Hz$	3	5	7
Nfreq $\Delta\omega = 0.5Hz$	5	9	12
Nfreq $\Delta\omega = 0.25Hz$	9	17	23

Table 3.1: Sampling frequency and number of frequencies selected for inversion at each configuration.

this noise can be understood when one looks at the FWI gradient computed for the first iteration of the first band (Figure 3.6). We see the vertically orientated noise is stronger relative to the geological content, the coarser the value of $\Delta\omega$. The wavenumber spectrums shows strong lateral spatial aliasing for all cases, however for $\Delta\omega = 0.25Hz$, the more geologically relevant, low wavenumbers have the most consistent strong amplitude. The strong lateral aliasing is a function of the restriction of only having sources and receivers at the surface. Adding denser sources, receivers and frequencies starts to mitigate this aliasing effect, but at the cost of increased computational cost. In real data cases, this is also limited by the utilized acquisition configuration.

To give some context to the increased computational cost, these jobs were run on 3 nodes with 16 CPUs per node. We summarize these results in Table 3.2 and Figure 3.7. The best result ($\Delta\omega = 0.25Hz$) required the least number of iterations, it however had the highest computational runtime (1440 minutes). This means that the cost for performing a single iteration of the inversion was much higher due to the increased computational cost of the forward modelling and gradient computation.

	$\Delta\omega = 1Hz$	$\Delta\omega = 0.5Hz$	$\Delta\omega = 0.25Hz$
Number of iterations $\Delta\omega = 1Hz$	290	272	199
Total Run Time (min)	612	900	1440

Table 3.2: Number of iterations and time required for $\Delta\omega$ tests

Influence of Noise on the inversion

We also consider a case where the observed data ($\mathbf{d}_{obs}$) is contaminated with white noise. This noise will map into our residual vector (Figure 3.8). We can see in this case the clear residuals we saw in the no noise case (Figure 3.3) are difficult to identify due to the noise. When FWI will look to fit the difference between $\mathbf{d}_{mod}(\mathbf{m})$ and $\mathbf{d}_{obs}$, the added
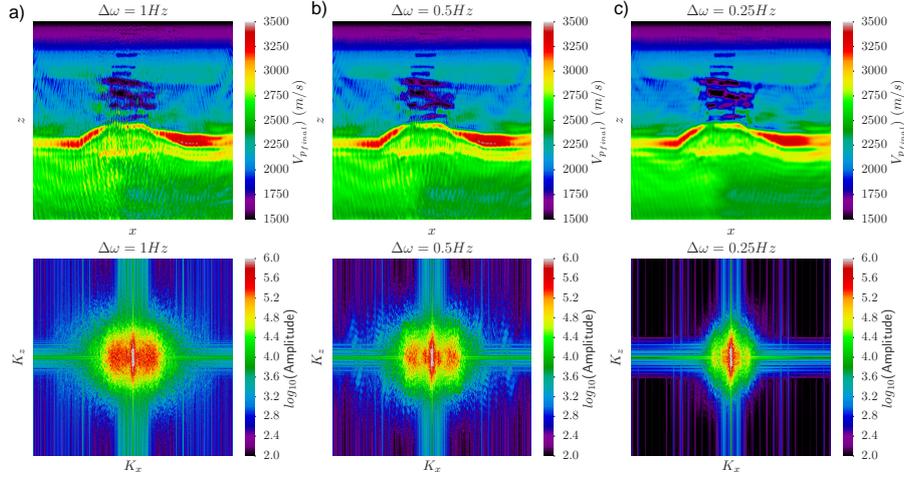
Figure 3.5: The impact of the frequency increment $\Delta\omega$ on the final FWI result. a) $\Delta\omega = 1Hz$, b) $\Delta\omega = 0.5Hz$ and c) $\Delta\omega = 0.25Hz$. The top row shows the final value of **m** while the bottom row shows the wavenumber spectrum of these results. It can be seen that the contamination of high wavenumber noise increases as the value of $\Delta\omega$ increases.
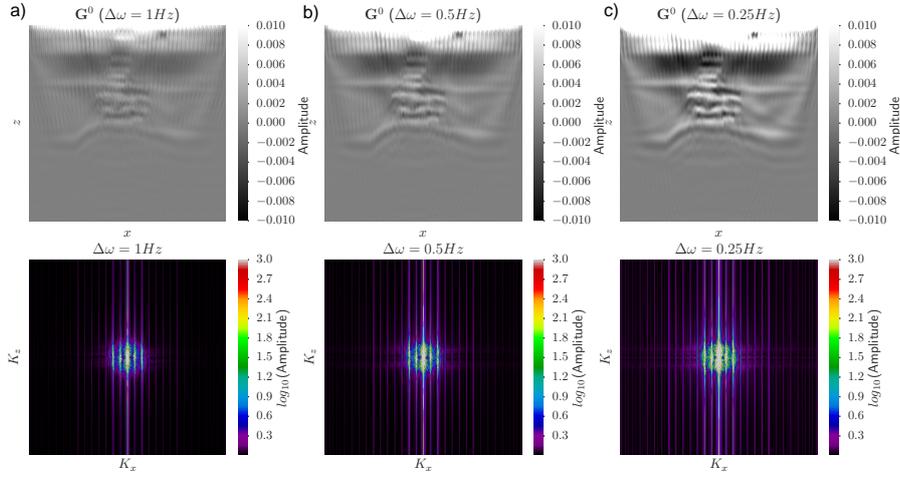
Figure 3.6: The impact of the frequency increment $\Delta\omega$ on the FWI gradient **G**. We show the first gradient $\mathbf{G}^0$ computed in the first band for a): $\Delta\omega = 1Hz$, b): $\Delta\omega = 0.5Hz$ and c): $\Delta\omega = 0.25Hz$. The top row shows the gradients where it can be seen that the geological features we want to add to our initial model become more dominant as $\Delta\omega$ decreases. The wavenumber spectrum shows that we have strongly aliased sampling of the wavenumber spectrum in the horizontal direction.

noise will increase the ill-posed nature of the inversion.

Figure 3.7: The objective function for the 3 bands involved in the test of the effect of frequency sampling on the FWI model. Note: The behaviour where the curve spikes back up to the value of 1.0 occurs when the previous frequency band has converged and the next frequency band is selected.

We run our FWI with exactly the same configuration as the previous no noise FWI results using a frequency step $\Delta\omega$ equal to $1Hz$. The results of this inversion are high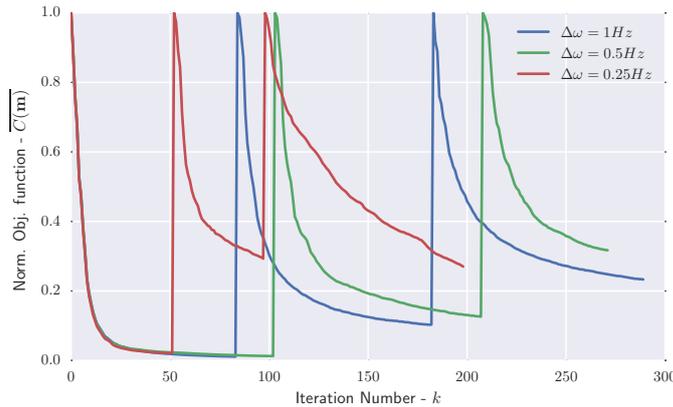ly contaminated by the white noise (Figure 3.9). One thing that is encouraging is that beneath the noise we can see the reconstruction of the individual gas sands has been successful. If we look at the wavenumber spectrum of the inversion result we see that the high wavenumbers have much higher amplitudes than our true model.

To understand the final inversion result it is useful to look at our gradient at the start of the inversion. If we consider the same 6Hz frequency we used to show the incident and adjoint fields in the no noise case (Figure 3.4) we can see the impact of the noise on our back propagated adjoint field at the first iteration $\mathbf{r}_b^0$ (Figure 3.10). Although the adjoint field is strongly contaminated by noise when we sum over sources and calculate the gradient we can still see a strong imprint of the geology we are looking to recover. This geological imprint is accompanied with undesired random noise. It is important to note, at this first gradient, the imprint of the geological signal is much stronger than in the final inversion result. The random noise we see in the gradient maps into a high amplitude circle of wavenumbers than the true model suggests. Ideally we would like to use the geological information that is in this gradient while limiting how much of the random noise can map into the $\mathbf{m}^k$.
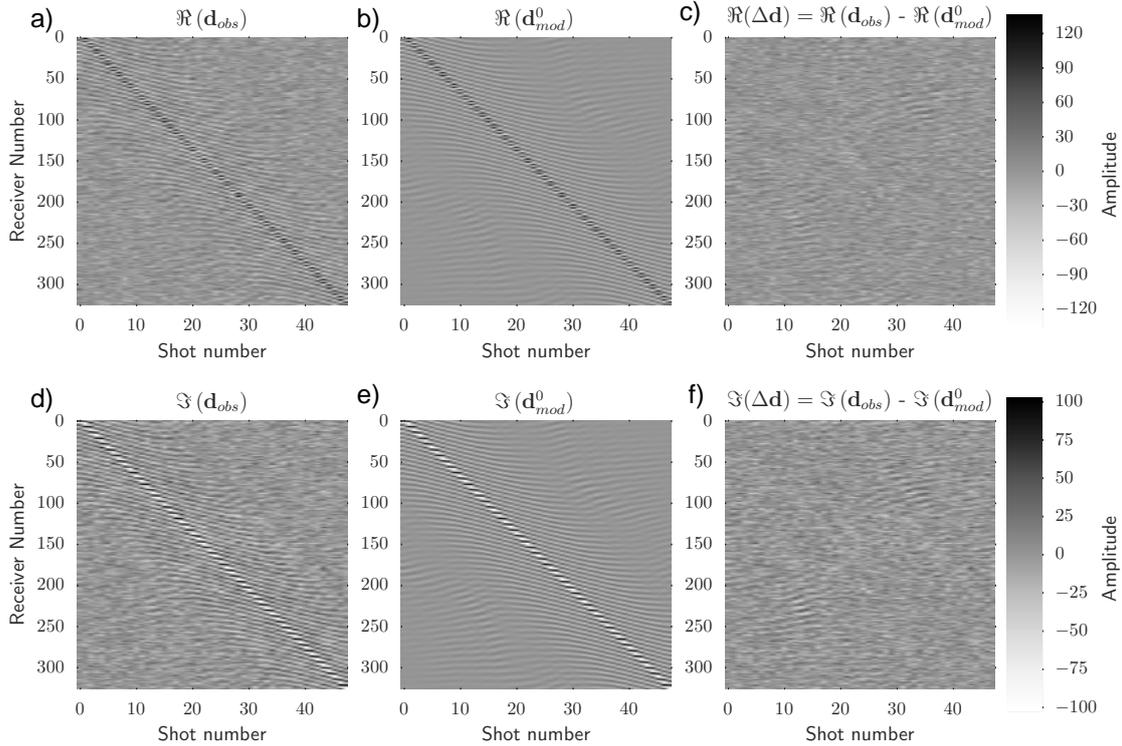
116

Figure 3.8: The effect of white noise contaminating $\mathbf{d}_{obs}$ on our residual vector $\Delta\mathbf{d}^k$. We show the real (a,b,c) and imaginary (d,e,f) components for the 6Hz monochromatic frequency. A comparison of the residual without noise is shown in Figure 3.3

3.2.2 Use of Preconditioning and Tikhonov regularization

The influence of noise and frequency sampling on the non-regularized inversion results have shown some of the issues that can map into our parameter reconstruction. Ideally, we would like to use some form of prior information to stabilize the inversion. The principle of the natural clustering of phenomena in geology was introduced in Chapter 2. We trial two strategies to attempt to promote this clustering in our inversion. The first strategy involves the addition of a penalty term that penalizes roughness in the model via a second derivative stencil operating on the model space. This is an application of the frequently used Tikhonov regularization (Tikhonov and Arsenin, 1977) and is a form of model-space regularization. The second option we will show is the application of an anisotropic smoothing operator to the FWI gradient. This can be seen to similar to the data-space regularization.

Tikhonov Regularization
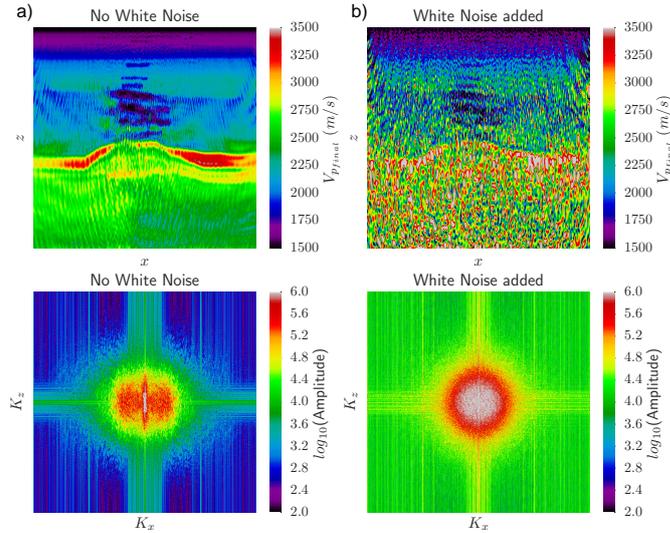
Figure 3.9: The inversion results for the case with no white noise a) and with white noise added b). The upper row shows the inversion results in the space domain $(x - z)$ while the lower row shows the wavenumber domain $K_x - K_z$
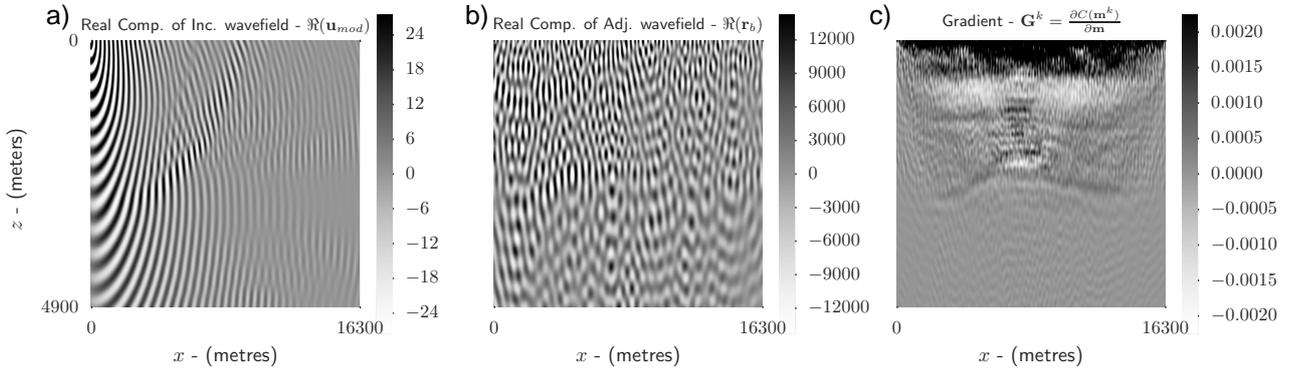
Figure 3.10: The real component of the incident a) and back propagated adjoint b) wave-fields for the case with white noise for a shot on the LHS of the model. We see that the adjoint wavefield contains much more noise than we noticed in Figure 3.4. When we sum over all the shots c) we can see the imprint of the geology we want to reconstruct but also significant amounts of noise.

The addition of a penalty term to the objective function is an often applied approach. We add an additional term to the objective function so our data fitting objective $C(\mathbf{m}^k)_{\text{data}}$ 3.8 is augmented with a model-space penalty term $C(\mathbf{m}^k)_{\text{model penalty}}$

$$C(\mathbf{m})_{reg} = C(\mathbf{m}^k)_{\text{data}} + \frac{\lambda}{2} C(\mathbf{m}^k)_{\text{model penalty}}. \tag{3.13}$$

118

This model-space has penalty term has the general form

$$C(\mathbf{m}^k)_{\text{model penalty}} = \mathbf{m}^T \mathbf{D} \mathbf{m} \qquad (3.14)$$

and will look to emphasize whatever the operator $\mathbf{D}$ is designed to. The most common uses are for it to emphasize smoothness (Tikhonov and Arsenin, 1977; Press et al., 1986) or the total variation (sharp contrasts) (Vogel and Oman, 1996; Vogel, 2002; Guitton et al., 2012; Askan and Bielak, 2008).

In our case we will consider the smoothness operator, however we will modify it allow one to constrain the smoothness in one direction to a greater extend than another. As we have showed that clustering tends to be more prevalent when looking at orientations aligned with the geological dip, this modification is justified. Our regularized objective function then becomes

$$C(\mathbf{m})_{\text{regularized}} = C(\mathbf{m}^k)_{\text{data}} + \frac{\lambda}{2} \left(\underbrace{\mathbf{m}^T a \mathbf{B}_x \mathbf{m} + \mathbf{m}^T b \mathbf{B}_z \mathbf{m}}_{C(\mathbf{m})_{\text{model penalty}}} \right), \qquad (3.15)$$

where the hyper parameter λ controls the weight between the data fitting goal in the the first term and the model-space penalty in the second term. $\mathbf{B}_x$ and $\mathbf{B}_z$ represent the second-order spatial derivative operator matrices with respect to the x and z directions respectively. $\mathbf{B}_x$ and $\mathbf{B}_z$ are applied using finite difference methods, while a and b are scalars that, when non-equal, allow one to penalize roughness in one orientation to a greater extent than another. The addition of this penalty term also modifies the gradient we use for FWI. The gradient I have discussed up to this point will be referred to as $\mathbf{G}^k_{\text{data}}$ and our regularized gradient becomes

$$\mathbf{G}^k_{\text{regularized}} = \mathbf{G}^k_{\text{data}} + \lambda \left(a \mathbf{B}_x \mathbf{m} + b \mathbf{B}_z \mathbf{m} \right) \qquad (3.16)$$

We consider our noised case using the Tikhonov regularization and ran FWI for a range of values for a, b and λ. Figure 3.11 shows the results of the optimization. We see that the inversion is highly sensitive to these parameters and the physical significance of each one is difficult to grasp. a and b can be summarized as $\frac{a}{b}$ which is the ratio of the horizontal second derivative of $\mathbf{m}^k$ vs the vertical second derivative. We know that we would like to have more clustering in the horizontal direction as opposed to the vertical (as the Valhall model has approximately flat geology). The choice of the value
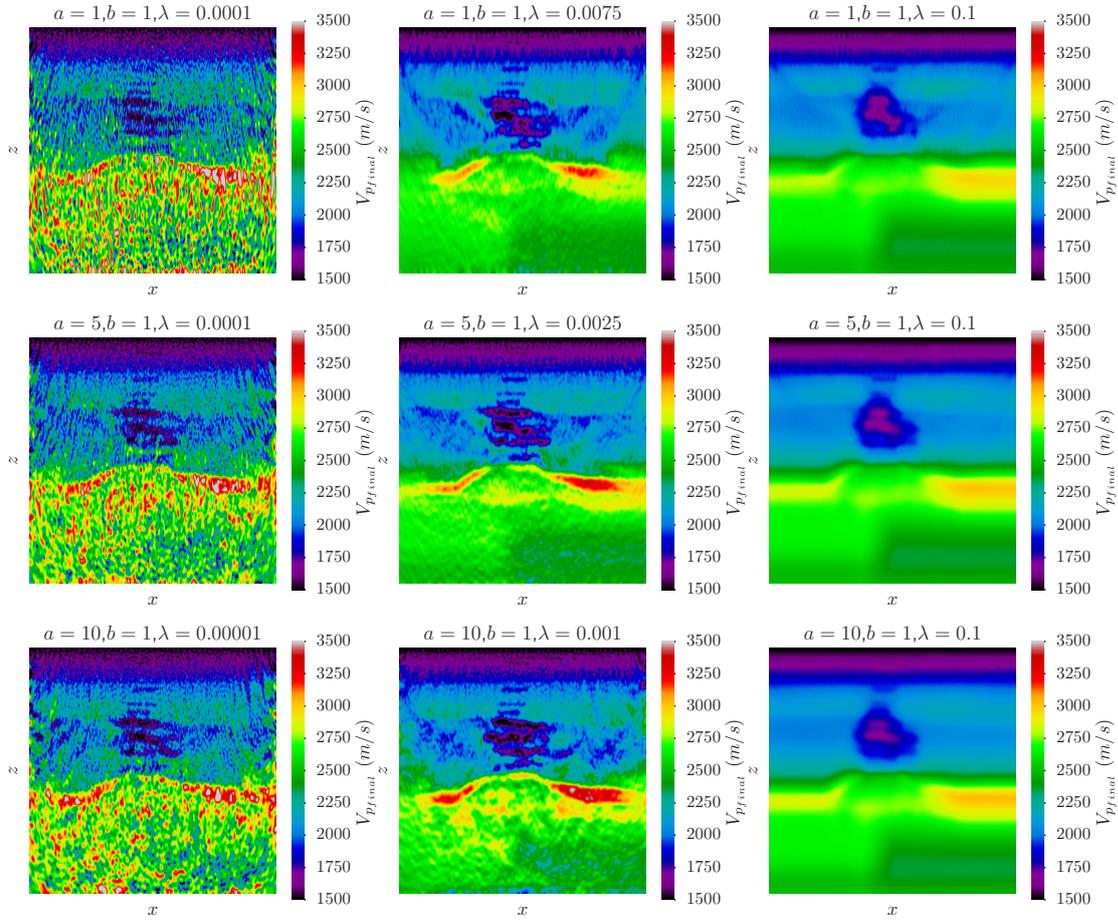
Figure 3.11: Tikhonov results for the noise $\mathbf{d}_{obs}$ case. A sweep of λ, a and b are used to show the wide variety of the sensitivity of the inversion to these parameters.

of this ratio is difficult to ascertain. λ is also a difficult parameter to select. The method of plotting the L-curve using plotting $C(\mathbf{m})_{\text{regularized}}$ vs $C(\mathbf{m})_{\text{tikhonov}}$ at the end of the inversion for a range of values of λ can help in selecting the correct value of λ (Hansen, 1998), but requires one to run FWI multiple times until completion. This is a computationally intensive strategy. We ran the FWI many times to find the optimal value of λ for the given value of $\frac{a}{b}$. We see that if the value of λ is too high then the penalty term prevents the addition of any wavenumber content and the final inversion result looks much like the initial model. If too low a value of λ is applied, the inversion will be contaminated by noise. Our optimal value shows that applying a penalty term that increases the smoothness in the horizontal direction vs the vertical yields the best results.

If we look at the frequency spectrums of the inversion results we see that as we in-
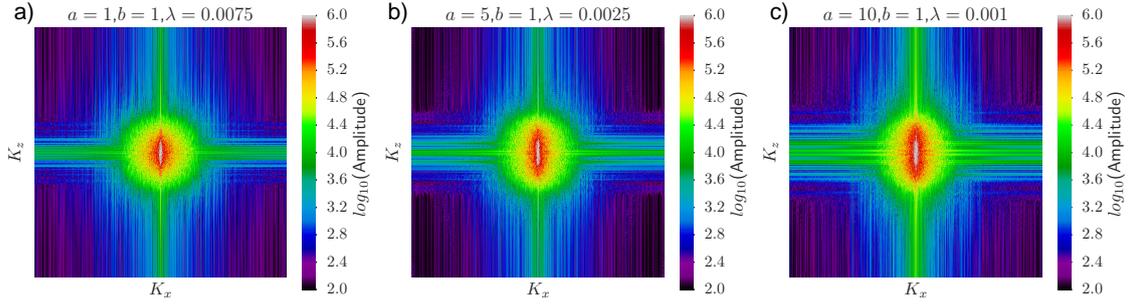
Figure 3.12: Frequency spectrum of Tikhonov results with optimal value of λ. a) $\frac{a}{b} = 1$ b) $\frac{a}{b} = 5$ c) $\frac{a}{b} = 10$. As we increase the penalty on high wavenumber variations in the horizontal direction we add in more stable vertical wavenumbers

crease the value of $\frac{a}{b}$ for the optimal value of λ, the wavenumber spectrums are quite similar (Figure 3.12). We have been able to mitigate against the "high-amplitude circle" that we see in the unconstrained optimization case (Figure 3.9). As we increase the value of the ratio $\frac{a}{b}$ we allow the addition of low to intermediate vertical wavenumbers to the inversion. This can bee seen as an vertically orientated expansion of the red area of amplitude in Figure 3.12.

Gradient Preconditioning by correlation filter

We present an alternative to model-space preconditioning approach where we modify the data FWI gradient $\mathbf{G}^k_{\text{data}}$ by pre-multiplying with a normalized correlation operator $\overline{\mathbf{Corr}}$.

$$\mathbf{G}^k_{\text{preconditioned}} = \overline{\mathbf{Corr}} \left(\mathbf{G}^k_{\text{data}} \right) \tag{3.17}$$

We refer to this gradient as the preconditioned gradient and the key principle behind it is that if we know that there is undesired content that maps into our gradient, why not simply filter them out. The idea of this approach is not new (Ravaut et al., 2004; Operto et al., 2006; Smithyman et al., 2015), but we apply our sparse additive inverse Laplacian operator as a means to calculate $\overline{\mathbf{Corr}}$. Our preconditioned operator is the vector obtained from solving the linear system invoking the 2D sparse additive inverse Laplacian correlation function as detailed in Algorithm 3. We refer to this as the AIL preconditioned gradient ($\mathbf{G}^k_{\text{preconditioned}} = \mathbf{x_{2D_2}}$)

Our correlation lengths L_x and L_z define an isotropic, circular filter when $L_x = L_z$ or an anisotropic elliptical filter when $L_x \neq L_z$. As the geology is relatively flat in the

---

Algorithm 3 Approximate $\overline{\mathbf{Corr}_{2D_{aniso}}}(\mathbf{G}_{\text{data}}^k)$ with $\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}$

1: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}}$ $\underbrace{\mathbf{x_{2D_1}}}_{\overline{\mathbf{Corr}_{2D_{AIL}}}(\mathbf{G}_{\text{data}}^k)}$ $= \underbrace{\mathbf{b}}_{\mathbf{G}_{\text{data}}^k}$

2: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}_{2D_{AIL}}^{-1}}}$ $\underbrace{\mathbf{x_{2D_2}}}_{\overline{\mathbf{Corr}_{2D_{AIL}}}\mathbf{x_{2D_1}}}$ $= \underbrace{\mathbf{b}}_{\mathbf{x_{2D_1}}}$

Valhall example we make the assumption that much of the information we want to reconstruct is the intermediate to high vertical wavenumber content of the gas sands and of the cap-rock that is absent from the initial model. The Valhall model is relatively flat so we can assume this information can be approximately vertically orientated. We select the vertical correlation length to be equal to $h = 12.5$ while we test 3 different horizontal correlation lengths ($L_x = 100$ m, $L_x = 200$ m and $L_x = 300$ m). Initially we will consider the FWI example that is without the white noise added to $\mathbf{d}_{obs}$ ($\Delta\omega = 1Hz$), before considering the noised example.

Before performing FWI, we can look at the gradient at the first iteration prior to applying preconditioning ($\mathbf{G}_{data}^0$) and after application of preconditioning ($\mathbf{G}_{preconditioned}^0$). The results for our additive inverse Laplacian (Figure 3.13) are shown. We see the extent to which the filter removes higher wavenumber information. The removal of this information doesn't appear to radically change the appearance of the gradient vector in the x-z space domain and just attenuates the non-geological vertical lines. When one looks at the wavenumber spectrum, we see there is significant amplitude that we remove when we constrain the correlation length L_x. These wavenumbers are the vertical wraparound seen in the wavenumber spectrum of the unconstrained result. The undesired wavenumbers highlights the significant amount of largely redundant aliased information present in the gradient.

If we compare the final inversion results we see the role of the correlation length L_x (Figure 3.14). In all cases the high wavenumber vertically orientated noise is at least partially attenuated in the final inversion result compared to the result with no correlation filter. The effectiveness of this attenuation improves as we increase the value of L_x. This means that when we compare the wavenumber spectrum of our inversion result with the wavenumber spectrum of the initial model the differences in amplitude occur only in the wavenumber regions we have allowed in the vertical wavenumbers and the

Figure 3.13: The sparse Laplacian preconditioned gradient $\mathbf{G}^0_{\text{preconditioned}}$ from the first iteration of FWI for the gradient preconditioning using a) $L_x = 100$ m, b) $L_x = 200$ m and c) $L_x = 300$ m compared to the d) unconstrained gradient $\mathbf{G}^0_{\text{data}}$. The top row is the gradient in the space domain while the bottom row is in the wavenumber domain.
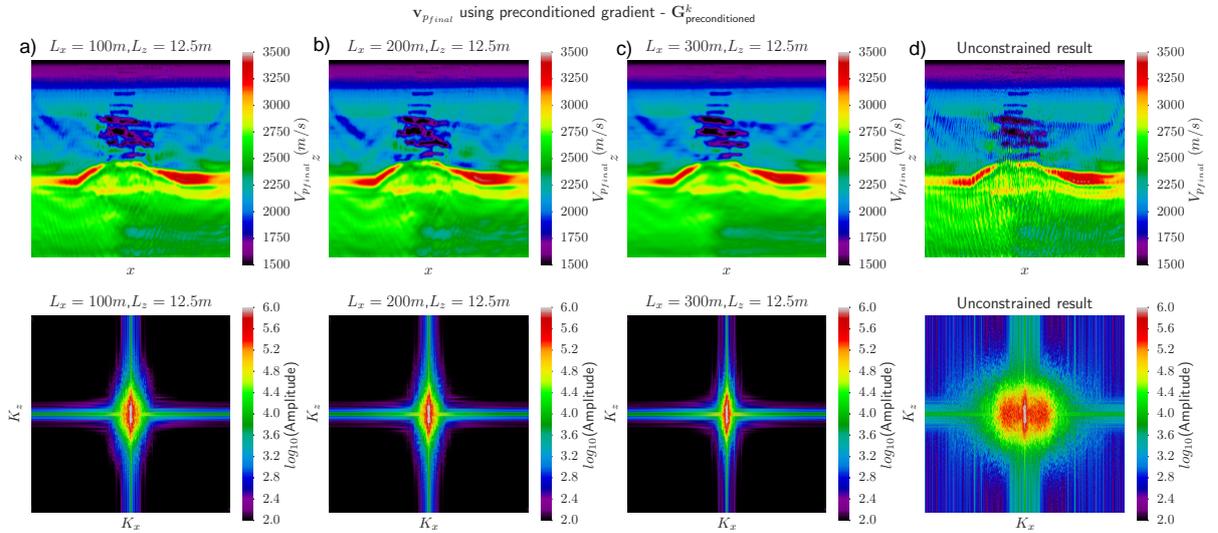
Figure 3.14: The final inversion result for the additive inverse Laplacian preconditioned gradient ($\mathbf{G}^0_{\text{preconditioned}}$) using a) $L_x = 100$ m, b) $L_x = 200$m and c) $L_x = 300$ m compared to the d) unconstrained gradient $\mathbf{G}^0_{\text{data}}$. The top row is the final inversion results in the $x - z$ domain while the bottom row shows the the wavenumber domain.

low horizontal wavenumbers.

123
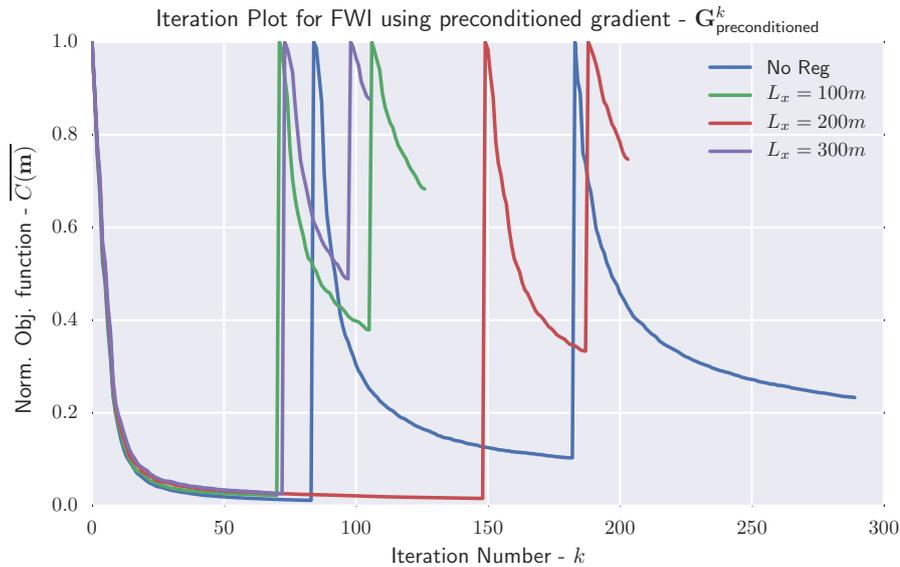
Figure 3.15: The normalized cost function for FWI using the Gaussian preconditioned gradient (*LEFT*) and the additive inverse Laplacian preconditioned gradient (*RIGHT*)

The evolution of the cost function (Figure 3.15) shows that the more optimal results (typically those with a higher value of L_x) require less iterations to reach convergence than the unconstrained inversion. In all cases the biggest decrease of data misfit comes in the first band of the inversion. It appears that, when the preconditioner is applied, the decrease in the subsequent bands is much smaller than the one for the unconstrained inversion. It can be noticed that the $L_x = 300$ m result required approximately one third the iterations to converge as opposed to the unconstrained result.

When we look at the inversion results when we have white noise contaminating the observed data, $\mathbf{d}_{obs}$, the same characteristics can be noticed (Figure 3.16). The preconditioned gradient is required to remove these features. The longer the correlation length the better we are able to attenuate undesired features. It is important to note that in the case of white noise that there is still noise that is left and this has the same wavenumber as the low-high vertically orientated wavenumber that the filter doesn't aggressively attack. This means that the signal to noise of our gradient in these orientations is poor and our inversion result will be still some what contained. The filter however, is able to minimize all the other largely non-geological wavenumbers.

The final inversion results (Figure 3.17) show that we are able to mitigate against
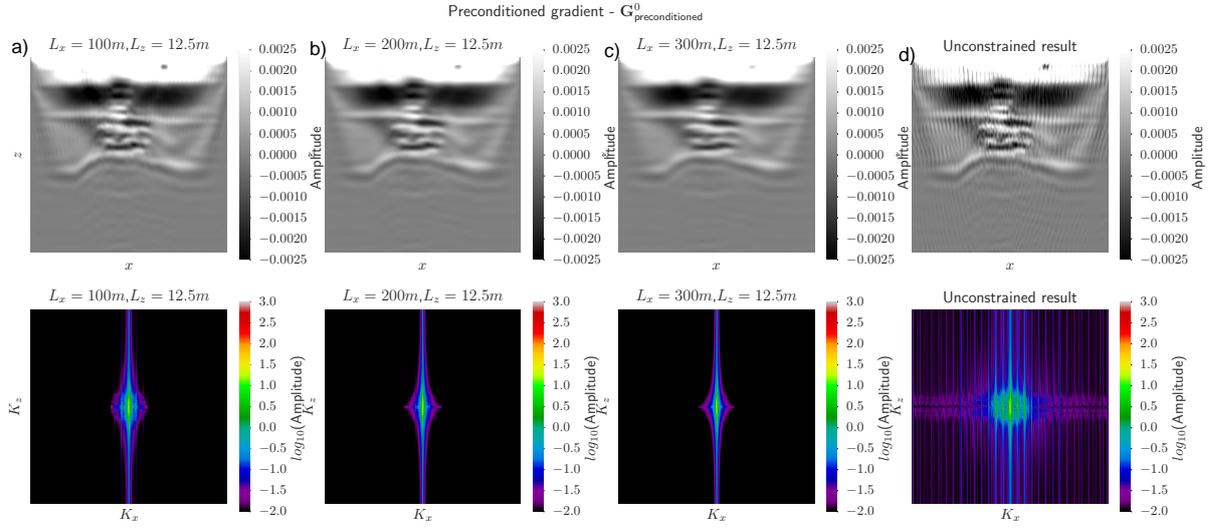
Figure 3.16: The additive inverse Gaussian preconditioned gradient $\mathbf{G}^0_{\text{preconditioned}}$ from the first iteration of FWI in the case where $\mathbf{d}_{obs}$ contains white noise. The gradient preconditioning results for a) $L_x = 100$ m b) $L_x = 200$ m and c) $L_x = 300$ m are shown compared to the d) unconstrained gradient $\mathbf{G}^0_{\text{data}}$. The top row is the gradient in the space domain while the bottom row is in the wavenumber domain.
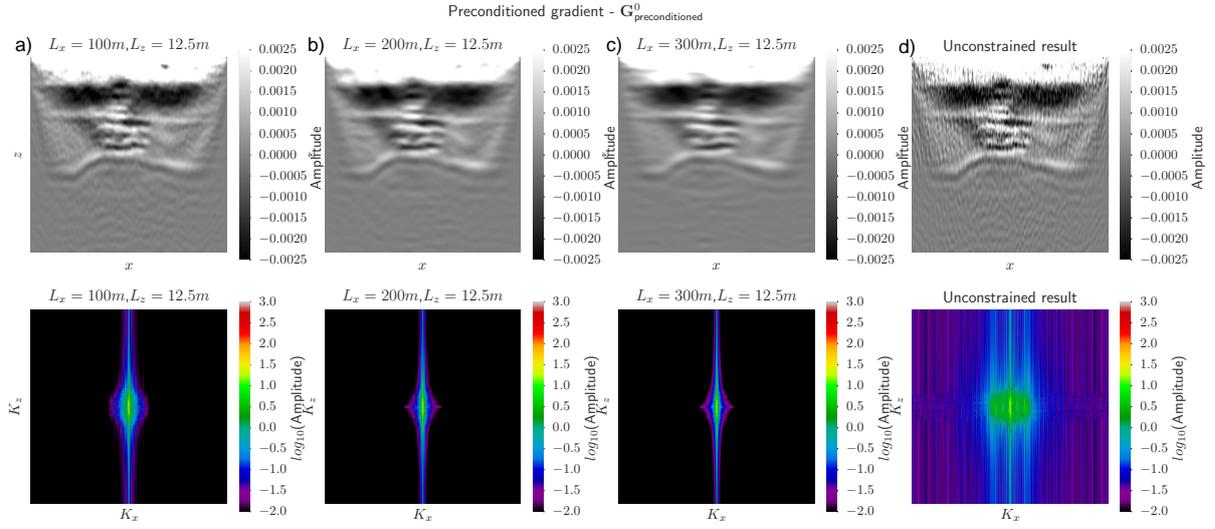
the noise mapping into the inversion via a simple gradient preconditioner. The AIL preconditioned gradient shows a good mitigation against the noise. These results can be contrasted to the Tikhonov results that were shown previously. It is arguable that the wavenumber constraint provides a better result with fewer and more intuitive parameters to choose/calibrate. A value of $L_x = 100$ m decreases the amount of noise that has mapped into the inversion as compared to the unconstrained case and as we increase this value to $L_x = 200$ m and $L_x = 300$ m our inversion result appears improved.

3.2.3 Valhall results summary

To summarize the application of constrained inversion we show a composite (Figure 3.18) of the best results for Tikhonov regularization ($a = 10$, $b = 1$, $\lambda = 0.001$) vs additive inverse Laplacian ($L_x = 300$ m) gradient preconditioning. The initial model and true velocity model and their spectrums are also displayed so a simple comparison can be made. A first pass visual inspection of the inversion result suggest they share similar features. The high-wavenumber content of the gas sands and the high velocity anti-clinal cap-rock appear to quite well recovered considering the strength of the noise that was
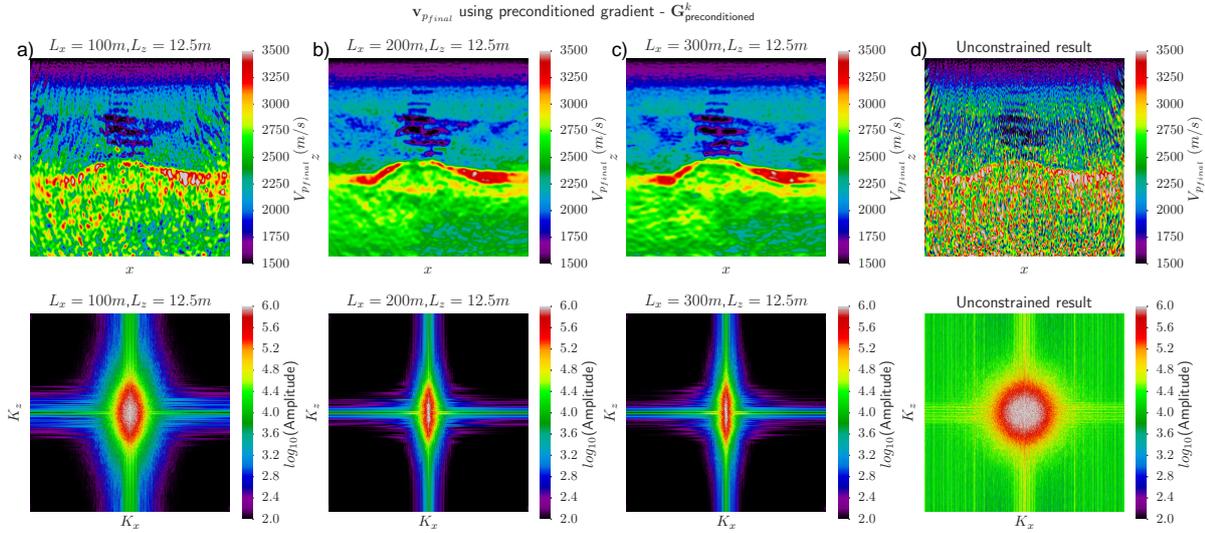
Figure 3.17: The final inversion result for the case with $\mathbf{d}_{obs}$ contaminated with white noise. The inversion is constrained by the additive inverse Laplacian preconditioned gradient ($\mathbf{G}^0_{\text{preconditioned}}$) using a) $L_x = 100$ m, b) $L_x = 200$ m and c) $L_x = 300$ m compared to the d) unconstrained gradient $\mathbf{G}^0_{\text{data}}$. The top row is the final inversion results in the $x - z$ domain while the bottom row shows the the wavenumber domain.

contaminating the inversion. Considering the extracted vertical trace it appears all results are relatively similar, but it could be argued that the additive inverse Laplacian provides the best compromise of fitting the high velocity cap rock and the overlying gas saturated sands.

One of the key advantages of the application of gradient preconditioning is that it represents a form of data-space preconditioning. By applying the anisotropic filter to the FWI gradient we are putting an a priori idea on the spatial continuity of the model update. As we typically start from a smooth initial model this process decreases the ability of particular wavenumbers mapping into our updated model. The inclusion of orthogonal correlation lengths allows one to quickly visually assess whether we have constrained the FWI gradient effectively. The correlation operator acts like a low-pass wavenumber filter and by visually inspecting the FWI gradient at a given iteration it is possible to see whether we have removed undesired information from the descent direction. In contrast the additive penalty term that is employed in Tikhonov regularization is not immediately intuitive. One needs to consider what spatial derivative ratio we should penalize in addition to needing to correctly choose the hyper-parameter λ. I propose that the ease and
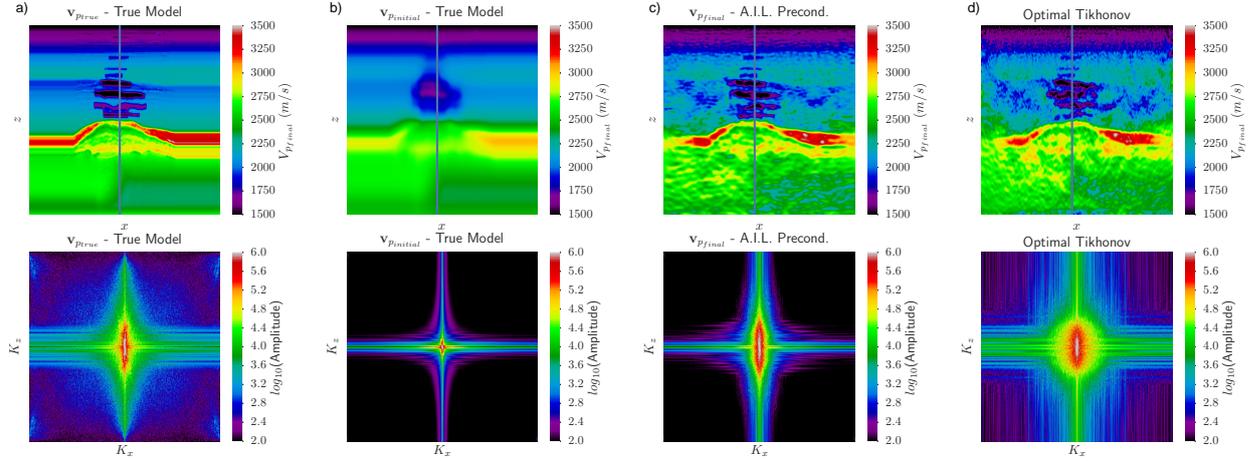
Figure 3.18: Composite of final inversion results and their spectrum. Note the blue vertical line represents the vertical extracted trace displayed in Figure 3.19 a)True Model, b)Starting model c)$\mathbf{v}_{p_{final}}$ -Additive inverse Laplacian and d)$\mathbf{v}_{p_{final}}$ -Optimal Tikhonov

flexibility with which we can parametrize the gradient preconditioning approach make it a more attractive proposition, especially in the presence of strong noise. It is also more intuitive than the Tikhonov regularization to relax the parametrization at later iterations if desired.

When applying the gradient preconditioning approach the additive inverse Laplacian approach is a computationally efficient. A second benefit of the additive inverse Laplacian, is it's ability to utilize non equal correlation lengths that are not orthogonal with the x and z directions. This is not easily feasible with the other correlation functions. We did not explore the value of dip in the Valhall case study, but we will show it by considering the Marmousi synthetic model.

Figure 3.19: Extracted vertical trace from inversion results for the additive inverse Laplacian preconditioned (*TOP*) , and the optimal Tikhonov (*BOTTOM*). The blue line represents the initial model while the black and red are the true and inversion result respectively.
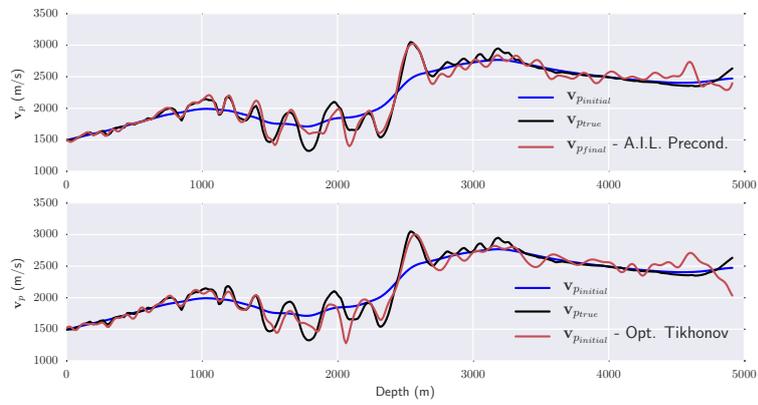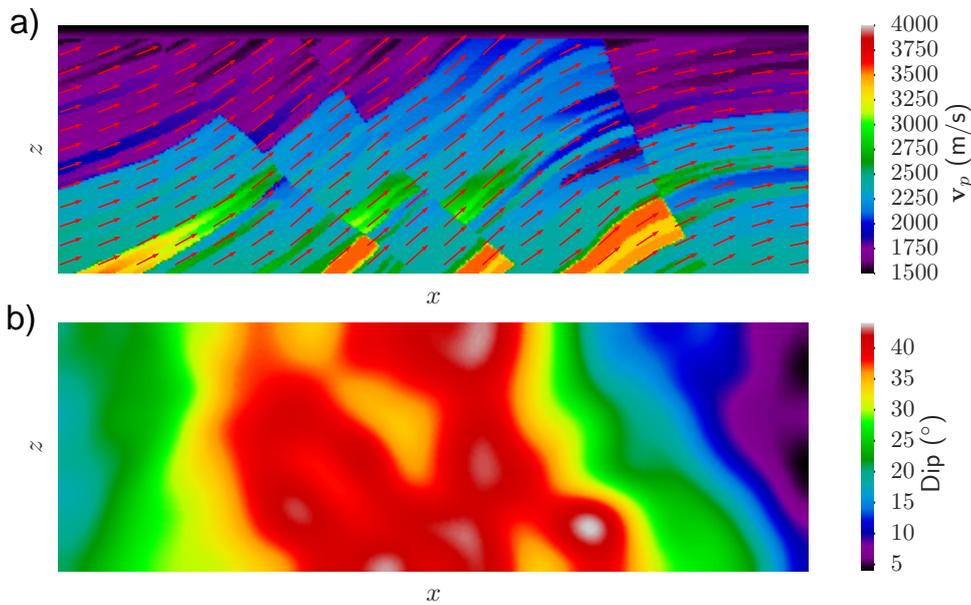
Figure 3.20: The Marmousi synthetic model. a) The true Marmousi velocity model. The red arrows show a vector field representation of the geological dip of the stratigraphy in the model. The scalar value of the dip is shown in the dip vector b). Note the dip values range from 5° to 45°.

3.3 Marmousi - Constrained Inversion

We show an application of applying the additive inverse Laplacian gradient preconditioning approach to a subset of the Marmousi synthetic model. The model is made of 130×390 cells with a grid spacing of $h = 10$ m and is described in Chapter 1. Unlike the previously shown, Valhall model. it is not composed of simple, unfaulted stratigraphy. The Marmousi model is composed of 4 rotated fault blocks, formed from extensional rifting (Martin et al., 2006). These rotated fault blocks have dips that varies between 5° and 45° (Figure 3.20). The faults have a large offset and an approximate en echelon pattern where the same stratigraphic horizon is shifted down as we move from the "hanging-wall" of the fault (left) to the "foot-wall" (right). This dip could be built a priori using migration as is done by Guitton et al. (2012). An alternative strategy would be to try and build it on the fly perhaps from analysis of the FWI gradient. In this case we take it as a known quantity.

Our FWI investigation looks to move from an accurate low wavenumber initial velocity model towards the true velocity model (Figure 3.21). The true model contains many

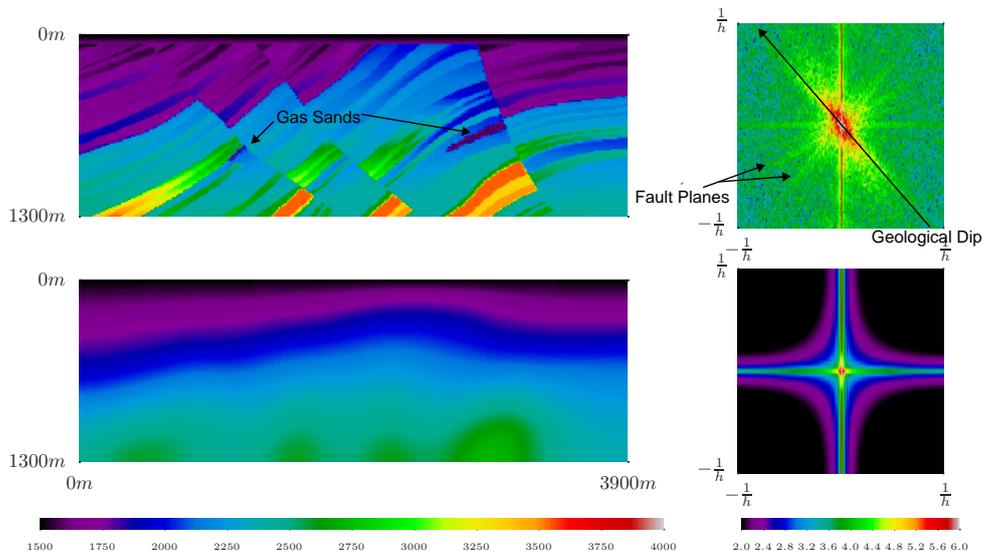Figure 3.21: *TOP:* The true velocity model *BOTTOM:* The initial velocity model. The wavenumber spectrums for each velocity model are shown on the right. Notice there is significant intermediate/high wavenumber information missing from the initial model. When we look at the true model we see that unlike the Valhall synthetic model the wavenumber amplitude are not aligned in the vertical wavenumber orientation. The fault planes can be seen to be approximately orthogonal to this geological dip.

features that we hope to reconstruct: the most important of these is the two highlighted gas charged sands. The initial velocity model $\mathbf{m}_0$ was computed via a 500 m isotropic smoothing of the true model. A constant density model is used for the FWI. A surface acquisition is employed with receivers placed every 10 m and shots performed every 50 m. The inversion of 12 frequency bands is completed simultaneous between $4Hz \rightarrow 20Hz$ ($\Delta\omega = 1Hz$). In order to increase the ill-possedness of the inversion, white noise is added to $\mathbf{d}_{obs}$ (much like in the Valhall case study).

The same additive inverse Laplacian based gradient preconditioning is applied that was employed in the Valhall example. We however also show how including the orientation (dip) of the stratigraphic horizons can improve the final inversion result. In this case we have a horizontal correlation length of 75 m ($L_x = 75$ m) and a vertical correlation length of 10 m ($L_z = 10$ m). The vector containing the dip for each cell of the model allows us to orientated this long smoothing direction parallel with the expected geology. The preconditioned gradient for our Marmousi synthetic at the first iteration of the inversion ($\mathbf{G}^0_{preconditioned}$) highlights the effectiveness of our gradient preconditioning
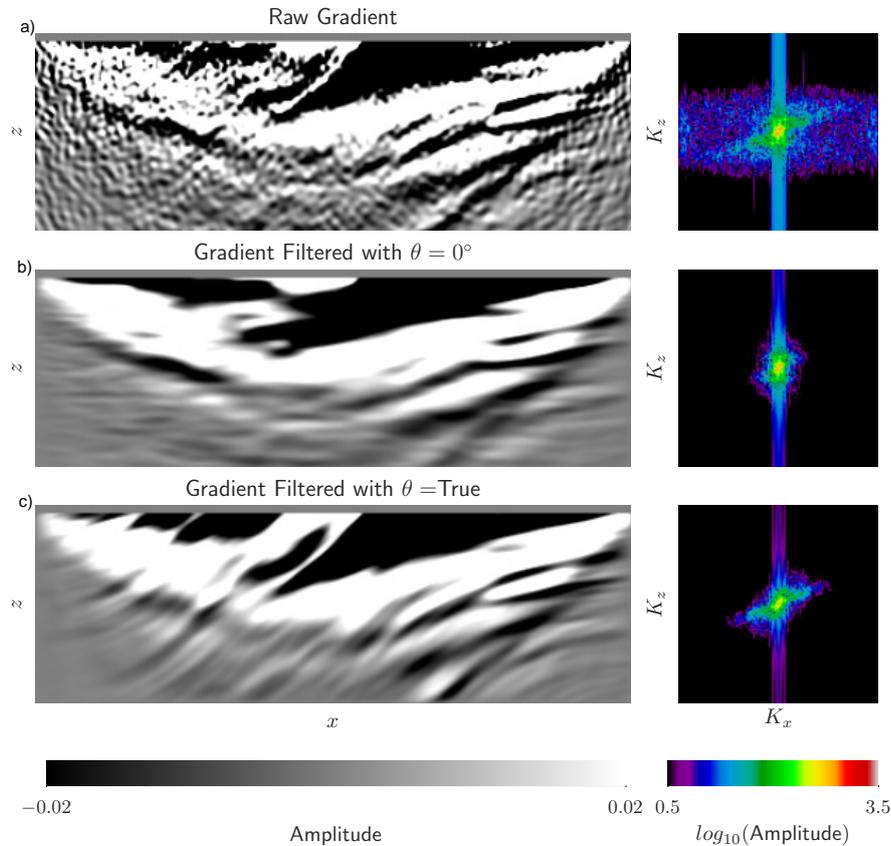
Figure 3.22: The FWI gradient from the first iteration of the inversion. a) The raw gradient with no preconditioning operator is contaminated by noise. If we perform preconditioning using $L_x = 75$m and $L_z = 10$ m (b) we see that we have decreased noise but suppressed much of the sharpness of the dipping reflectors. We are able to remove the noise and keep the reflectors by using the preconditioning with our variable dip vector (d). The wavenumber spectrums are also shown for each of the gradients is shown on the right.

strategy (Figure 3.22). When preconditioning is not applied, our gradient is contaminated by the imprint of the noise from $\mathbf{d}_{obs}$. Applying the preconditioner without dip decreases the noise imprint, but also suppresses much of the dipping information we want to reconstruct in our true model. The variable dip preconditioner shows the best compromise between reducing the noise in the gradient and maintaining the dipping information.

We perform FWI using our 3 different gradient preconditioning options (no preconditioning, preconditioning with out dip and precondition with dip) (Figure 3.23). When we apply no preconditioning, the continuity of events is decreased due the mapping of

a) No Preconditioning

b) Preconditioning $\theta = 0°$

c) Preconditioning $\theta =$True dip

v_p (m/s)
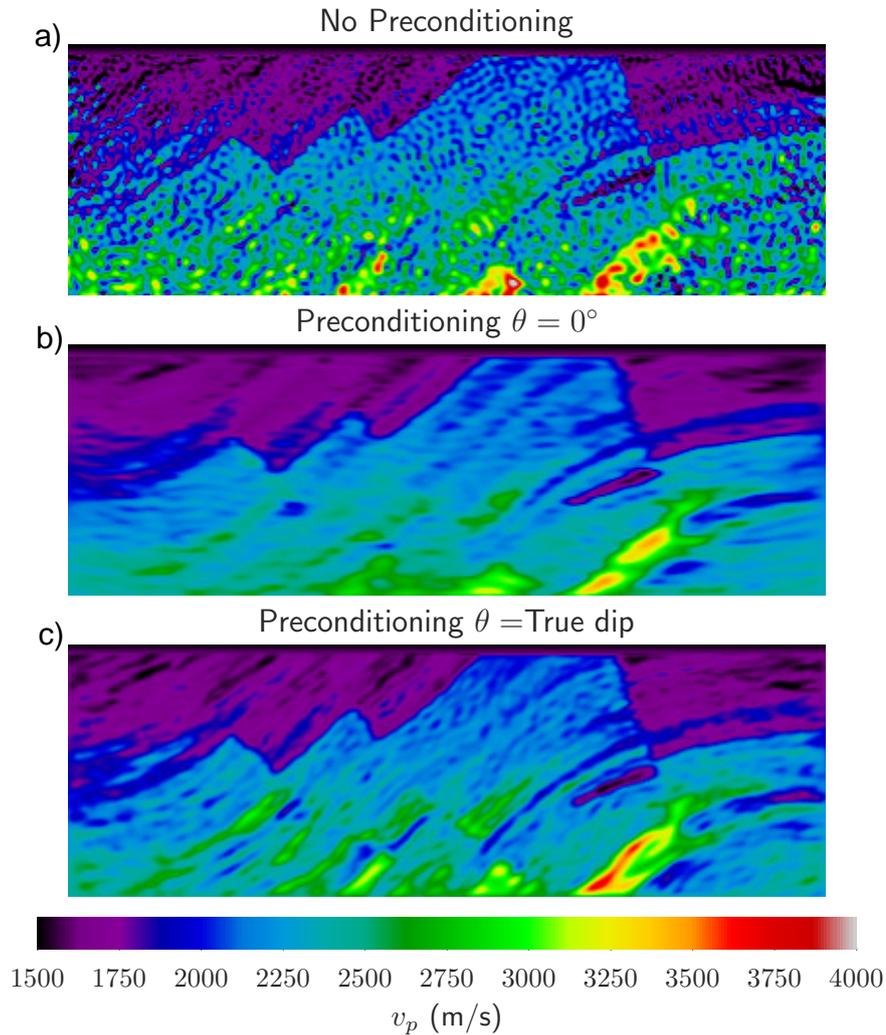
Figure 3.23: The inversion results obtained from the 3 preconditioning strategies: a) No preconditioning, b) Preconditioning zero dip, c) Preconditioning with variable dip. The best result occurs when we precondition with the correct dip field.

noise information from our gradient into our model reconstruction. If we smooth without dip we get a clearer image but decrease the resolution of the stratigraphy orthogonal to dip. By using the dip variant preconditioner we are able to suppress the noise and still reconstruct the fine vertical details in our velocity model. In all the cases the same number of FWI iterations has been performed (30) to make the comparisons fair.

3.4   Conclusion

In this chapter we have introduced frequency domain FWI with two synthetic examples. The importance of the AIL operator as a gradient preconditioner has been highlighted in order to mitigate against non-geological artifacts related to the ill-posed nature of FWI. The AIL operator is relatively simple to parameterize when compared to Tikhonov regularization and behaves as an effective wave-number preconditioner in the examples shown. The inclusion of the true dip in the Marmousi example also highlights the versatility of the operator to handle structurally orientated, spatially variant filtering.

Chapter 4

FWI: Real data example

Contents

Chapter overview: This chapter will complement the work done on frequency domain FWI by introducing the time domain approach. The NWA-006 Broadseis dataset (introduced in Chapter 1) will be investigated with the focus being on two key points. Firstly, the role of the initial model will be discussed, where we compare inversion results from a crude initial model as opposed to one built from a reflection tomography workflow. Effort is made to show where the inversion results for the two cases are similar and where there are vast differences. Two main regions will be identified. These regions appear to be related to the penetration of diving waves, with the inversion result

being significantly more sensitive to the starting model below the limit of diving wave penetration. The second point we address is the role of the AIL correlation framework: it plays an important role in removing undesired high-wavenumber artifacts from our model parameter reconstruction.

4.1 Time domain FWI strategy

The full waveform inversion results in this chapter have been computed using the SEIS-COPE tool TOYxDAC_TIME. This FWI algorithm allows one to perform VTI wave propagation and inversion under an acoustic approximation. In this study, we have performed VTI wave propagation, but have only looked to invert for the velocity and density parameters. A description of the VTI forward modelling engine (Forward Problem) and how to calculate the first derivative of the misfit function (Inverse Problem) are detailed before we look more specifically at the NWA-006 Broadseis line.

4.1.1 Forward Problem

The generalized time domain wave equation is described as

$$\mathbf{M}(\mathbf{x})\frac{\partial^2 \mathbf{u}(\mathbf{x},t)}{\partial t^2} = \mathbf{A}(\mathbf{x})\mathbf{u}(\mathbf{x},t) + s(\mathbf{x},t), \tag{4.1}$$

where $\mathbf{M}$ and $\mathbf{A}$ are the mass and stiffness matrices respectively, while $\mathbf{s}$ represents the source function and $\mathbf{u}$ is the wavefield to be modelled. In our discussion, we are working with offshore marine streamer data and will use the acoustic approximation. The field $\mathbf{u}$, will refer to the pressure wavefield and $\mathbf{s}$ is an acoustic impulsive air-gun source. The system of second-order equations is conveniently recast as a first-order hyperbolic velocity-stress system (Levander, 1988; Virieux, 1986). These P-SV equations can be simplified to handle the 2D acoustic VTI wave equation. Our system of equations becomes

$$\frac{\partial u_x(x,z,t)}{\partial t} = \kappa(x,z)\left[(1+2\epsilon(x,z))\left(\frac{\partial v_x(x,z,t)}{\partial x}\right) + \sqrt{1+2\delta(x,z)}\frac{\partial v_z(x,z,t)}{\partial z} + s(x,z,t)\right]$$

$$\frac{\partial u_z(x,z,t)}{\partial t} = \kappa(x,z)\left[\sqrt{1+2\delta(x,z)}\left(\frac{v_x(x,z,t)}{\partial x}\right) + \frac{\partial v_z(x,z,t)}{\partial z}\right]$$

$$
\begin{aligned}
\frac{\partial v_x(x,z,t)}{\partial t} &= b(x,z)\frac{\partial u_x(x,z,t)}{\partial x} \\
\frac{\partial v_z(x,z,t)}{\partial t} &= b(x,z)\frac{\partial u_z(x,z,t)}{\partial z} \, ,
\end{aligned}
\tag{4.2}
$$

where the Bulk modulus and buoyancy are represented by $\kappa(x,z)$ and $b(x,z)$) while the parameters that described VTI anisotropy are $\delta(x,z)$ and $\epsilon(x,z)$ (Thomsen, 1986). The terms $v_x(x,z,t)$ and $v_z(x,z,t)$ are the horizontal and vertical particle velocities, while the pressure components are $u_x(x,z,t)$ and $u_z(x,z,t)$. These pressure components are non-physical in themselves, but are used to separate the horizontal and vertical derivatives and are also used to take into account the CPML (Komatitsch and Martin, 2007). The true pressure field is calculated by

$$
u(x,z,t) = u_x(x,z,t) + u_z(x,z,t).
\tag{4.3}
$$

In time-domain forward modelling we need to discretize both the temporal and spatial derivatives. This is in contrast to the frequency domain, where only the spatial derivative must be discretized. The spatial discretization is performed by the staggered-grid approach using either 4th or 8th order FD stencils (Levander, 1988; Fornberg, 1988). The spatial discretization stencil is shown in Figure 4.1. The temporal discretization is performed explicitly using a second-order leapfrog method. An interpolation strategy is used in order to correctly locate the sources and receivers on the finite difference grid (Hicks, 2002), thanks to the linearity of the wave equation.

4.1.2   Inverse Problem

With the means of calculating the pressure wavefield in the time domain we can now focus on trying to set up the inverse problem. Although we started by discussing frequency domain FWI in Chapter 3, the initial applications of FWI were performed in the time domain (Tarantola et al., 1984; Tarantola, 1987). To introduce the gradient we start by considering a single offset trace for a single source. In such a case, we can represent the difference between the measured $\mathbf{d}_{obs}$ and modelled wavefield $\mathbf{d}_{mod}(\mathbf{m})$ as

$$
\Delta d_{s,r}(t) = \underbrace{d_{s,r}(t)}_{\mathbf{d}_{obs}} - \underbrace{R_r u_s(x,z,t)}_{\mathbf{d}_{mod}},
\tag{4.4}
$$

where our data vectors are for the source, s. The restriction operator is denoted by R_r and only extracts the wavefield at the receiver r. Our ℓ_2 objective function can therefore

$$\begin{array}{l} \bullet \; v_x, b \\ \blacksquare \; v_z, b \\ \blacktriangledown \; u, \kappa, \delta, \epsilon \end{array}$$
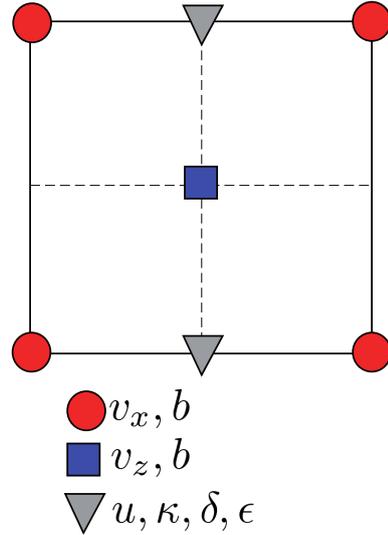
Figure 4.1: Staggered grid for Virieux-Lavander stencil adapted to the Acoustic VTI case

be defined as

$$C(m) = \frac{1}{2} \sum_{N_r} \sum_{N_s} \int_0^T ||d_{s,r}(t) - R_r u_s(x, z, t)||^2, \tag{4.5}$$

where the calculation of the objective function requires a summation over N_s shots, T time samples and N_r receiver locations. In this illustrative single source receiver case $N_s = N_r = 1$, however in practice, many sources and receivers are required to adequately illuminate the subsurface.

Gradient computation

We can hope to use the local optimization approaches detailed in Chapter 1 and applied using frequency domain FWI in Chapter 3. We intend to use the same quasi-Newton approach but require an efficient calculation of the first derivative of the misfit function, which is the FWI gradient. From expression 4.5, we can define our gradient as

$$\frac{\partial C(m)}{\partial m} = \int_0^T \sum_{N_r} \sum_{N_s} \left(\underbrace{\frac{\partial \left[R_r u_s(x, z, t) \right]}{\partial m}}_{\mathbf{J}} \right)^T (d_{s,r}(t) - R_r u_s(x, z, t)) dt \tag{4.6}$$

where the Jacobian or Fréchet derivative matrix is denoted by the symbol $\mathbf{J}$. The calcu-

lation of $\mathbf{J}$ depends on the number of parameters in our model space m and is usually considered to be too computationally prohibitive to compute numerically. In order to avoid this high computational cost, the adjoint state approach is employed where all contributions for receivers associated for one source are combined simultaneously (Chavent, 1974). This approach is attractive and has been applied in many problems in Geophysics (Lailly, 1983; Chavent and Jacewitz, 1995; Tromp et al., 2005; Plessix, 2006; Chavent, 2009).

The development of the adjoint state method will not be completed here. Please refer to (Plessix, 2006; Brossier et al., 2013). The resultant gradient for a single source is

$$\frac{\partial C(m)}{\partial m} = G(x,z) = \sum_{N_s} \int\limits_0^T \left[u^T(x,z,t) \left(\frac{\partial A}{\partial m} \right)^T \lambda(x,z,T-t) \right] dt, \qquad (4.7)$$

where the operator $\partial A/\partial m$ represents the radiation matrix and $\lambda(x,z,T-t)$ is the adjoint field.

4.2 Broadseis Data Analysis

4.2.1 Initial model construction

The fact that FWI is performed as a local optimization approach means the the construction of an accurate initial model is of key importance. Without an accurate initial model, our modelled seismograph will be cycle-skipped at our lowest recorded frequency and FWI will be unable to remedy this kinematic mismatch. In our synthetic examples presented in Chapter 3, this was not an issue as our initial models were constructed by simply smoothing the true models. In real data cases, the construction of the initial model is crucial (Virieux and Operto, 2009). A number of approaches exist, that can be used to build this low-wavenumber, kinematically accurate initial velocity model. First Arrival Travel-time tomography (FATT) (Nolet, 1987; Hole, 1992; Zelt and Barton, 1998) is often proposed as a potentially attractive solution and performs non-linear inversion of the first-arrival travel-times to produce smooth models of the subsurface. A limitation of the FATT approach occurs when there are low-velocities zones which limit the penetration of diving waves. Laplace-domain/Laplace-Fourier-domain FWI (Shin and Cha, 2008; Shin and Ha, 2008; Shin and Cha, 2009) is a second method that also has shown some good results. We however perform the construction of our initial model using the

reflection tomography approach (see Woodward et al., 2008, for a good overview). We decide to utilize the reflection tomography in our case, due to two reasons. The first reason is that reflection tomography is often used as the velocity modelling tool of choice in real data situations with limited offsets as is the case in our streamer configuration (Houbiers et al., 2012; Sirgue et al., 2011; Manuel et al., 2014). When the depth to the target reservoir is large (at NWA-006 > 3 km), the short offsets acquired in streamer configurations prevent diving waves from penetrating to the target level. A second reason for the use of reflection tomography, was due to the support afforded to the SEISCOPE consortia by CGG. With strong skills in reflection tomography, they have offered to help with diverse refinements of a crude initial velocity model using anisotropic reflection tomography.

4.2.1.1 Crude initial model construction

Our initial velocity model that was to serve as the input to reflection tomography was constructed using the sparse stacking velocities $\mathbf{v}_{stk}$ fields provided by CGG. These velocities were interpreted using the interactive velocity analysis workflow (Yilmaz, 2001, pp. 311-319). Prior to undertaking depth imaging, a pre-stack time migration (PreSTM) processing workflow was performed by CGG. This time imaging workflow can offer a first pass image and can be suitable when there are not significant lateral velocity variations. The processing workflow involves sorting data from the common shot domain that is used in FWI to the common mid-point domain (Figure 4.2). The key processes of the workflow are to remove noise from coherent (multiples, refracted energy) and incoherent sources (swell and ambient noise) as well as applying a number of correction (statics, noisy trace editing) prior to imaging using of the primary reflected wavefield using a Kirchhoff time migration algorithm (Schneider, 1978; Berryhill, 1979; Berkhout, 1980). This time imaging approach can provide accurate results when the velocity structure is not highly complicated. However, when we have rapid, lateral velocity variations, the time imaging assumption breaks down. Strong lateral velocity variations occur in the NWA case due to the rapidly varying water bottom topography. This give rise to significant ray bending at the sea-floor boundary. Our CMP processing assumptions become invalidated when we have this strong ray bending, and as a result, data in the CMP domain has strong non-hyperbolic move-out that cannot be corrected with time imaging algorithms. In such cases, depth imaging and velocity model building workflows are required to refine the velocity model. The pre-stack time migrated section can still however provide an interesting and useful image. In the case of the NWA-006 example

we can see much of the reflectivity we expect to recover with a depth imaging workflow. However, the positioning and focusing of the reflectors is likely to be compromised. This is especially true under the shelf-break. A succinct example of the kind of imaging effects we can expect to suffer is presented from another area of the North Western Shelf of Australia in the Seismic Data Analysis textbook (see Yilmaz, 2001, pp. 1597-1604).

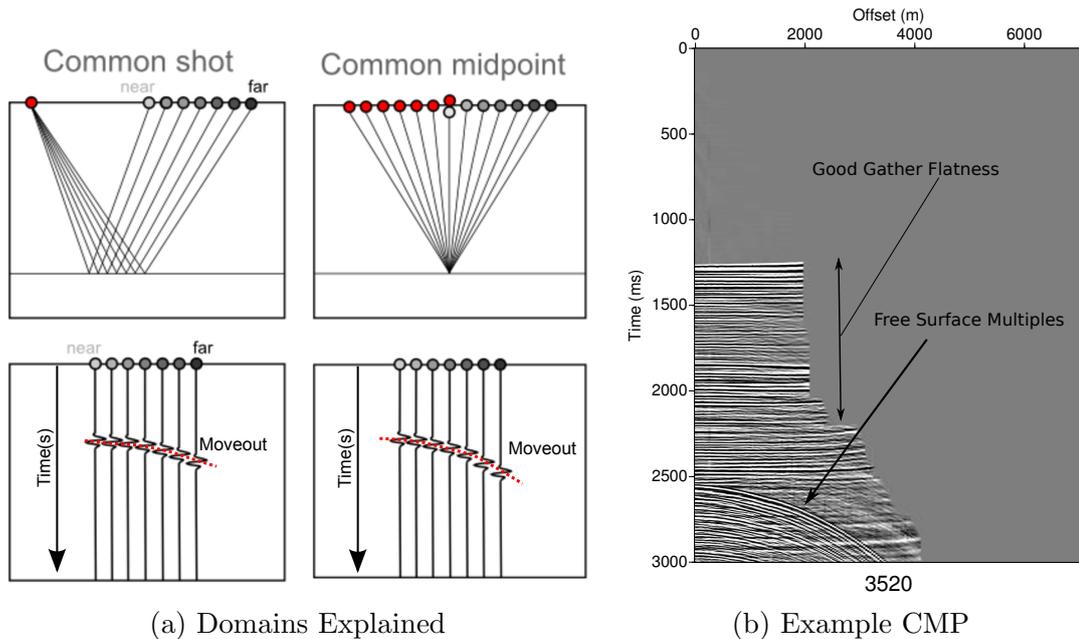(a) Domains Explained   (b) Example CMP

Figure 4.2: a) The common shot (CS) and common mid point (CMP) domains. In the common shot domain, seismic information is measured at offset receivers for a given shot (red), the mid point between source and receivers occur at a different positions in the earth. The CMP domain involves using sources and receivers pairs that share a common mid point. Notice that the wiggle plot also shows the associated travel-time. As source to receiver offset increases, this change in travel time is referred to as move-out. b) An example CMP from a deep-water section of the NWA-006 line that has been corrected for normal move-out. Super-critical reflections and transmitted waves have been removed. Notice many events are flat. A large amount of non-flat energy comes into the CMP at 2500 ms. This is the first water bottom multiple.

Quality Control (QC) of subsurface stacking velocities:

The stacking velocities for the NWA-006 PreSTM were interpreted at CDP locations on a 2 km grid. Typically, care needs to be taken in order to ensure that the velocity trend corresponding with primary energy is interpreted. To ensure this is done, an iterative workflow of multiple attenuation via 2D SRME and Radon de-multiple is common place in order to ensure that the velocity trend interpreted corresponds to the true primary
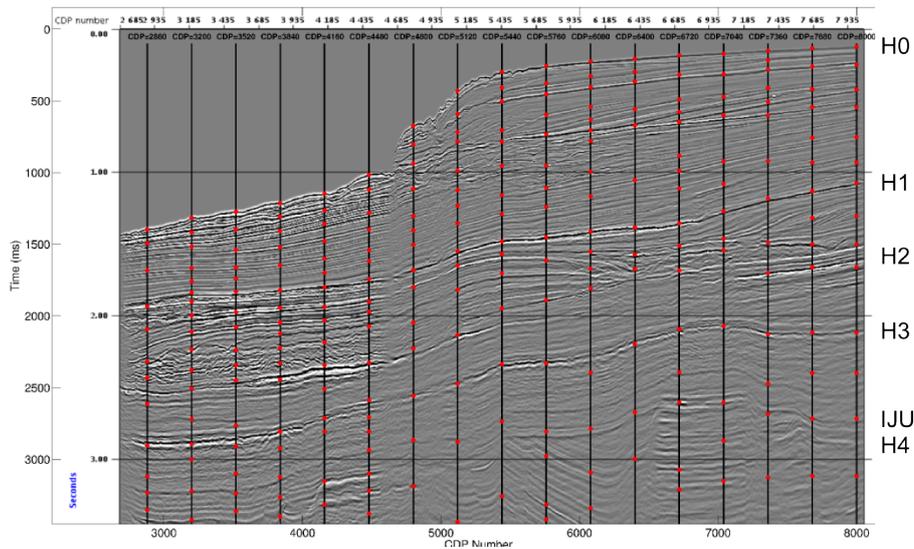
Figure 4.3: QC of RMS Stacking velocity picks. The PreSTM migrated time image is overlain with velocity pick locations (shown in red). It can be seen that the picks are made consistently along what appear to be primary reflectors. The approximate horizon locations described in Chapter 1 are highlighted on the right.

trend and that multiple and diffracted multiple energy is not selected. This form of workflow was applied in this PreSTM example (personal communication with CGG). Although I had no intention of re-picking the velocities, I wanted to perform a quick QC to validate that they were consistent with the PreSTM migrated section and the raw field data. In the first QC, I overlay the picks on a migrated image obtained from the Conventional Kirchhoff PreSTM(Figure 4.3). We can see that the velocity picks in this case are picked consistently on what appear to be primary, reflection events. A second QC was made by looking at the individual velocity analysis locations. The velocity picks were plotted on the computed CMP velocity analysis semblance. These semblances were computed using the Seismic Unix routine, suvelan_nccs (Normalized cross correlation sum based semblance analysis). A montage of CMP locations (Figure 4.4) highlights that the stacking velocities coincide with expected semblance picks, while also highlighting the difficulties of correctly picking the correct velocity trend in the shallow section on the shelf. This shallow trend was likely picked with care after application of an iterative multiple attenuation workflow. With these QCs complete, I was comfortable with the subsurface velocity picks.

XBT derived Water column velocities:

142
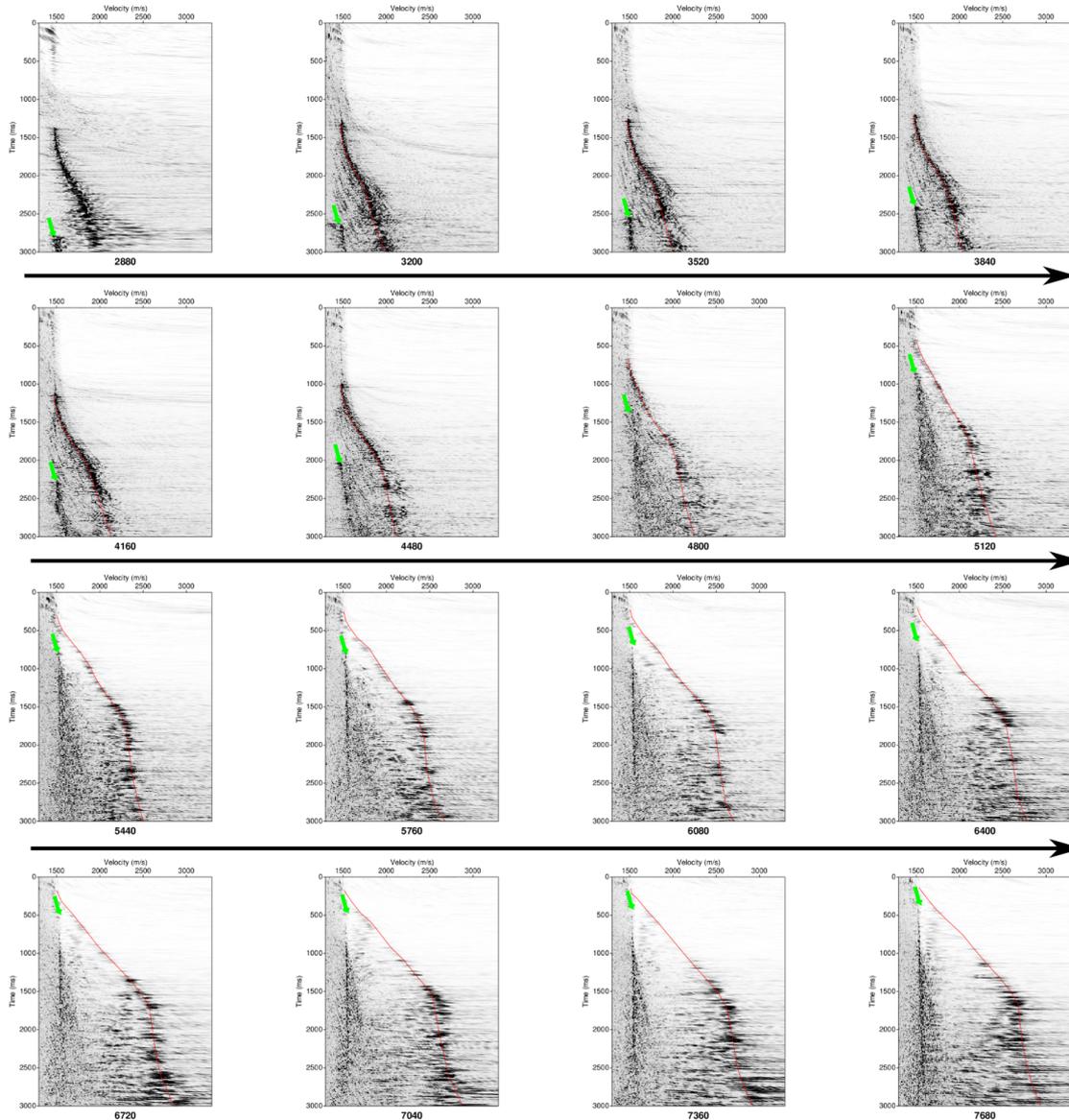
Figure 4.4: Semblance analysis calculated using suvelannccs. The red lines show the interpreted stacking velocities. The CMP numbers are indicated on the bottom and as we move to shallower water (indicated by the black arrow) we can see it gets harder to pick the primary trend from semblance. The green arrow indicates the first free-surface water bottom multiple.

Although one can make a constant velocity assumption for the water column. this can be incorrect, especially in regions where the water depth has rapid variations. A typically employed strategy to ensure an accurate water column velocity model is to utilize data acquired from an expendable bathythermograph (XBT) to build a relationship of velocity

143

Figure 4.5: XBT derived TWT velocity function

versus water depth. We were able to obtain some nearby XBT measurements (offset from the seismic line by less than 2000 km), acquired by the Australian equivalent of CNRS, the Commonwealth Scientific and Industrial Research Organization (CSIRO). The XBT measurements are combined with the seawater salinity to allow calculation of an empirical water temperature to velocity transform (Medwin, 1975) (Figure 4.5). We there then able to compute the interval velocity as a function of TWT
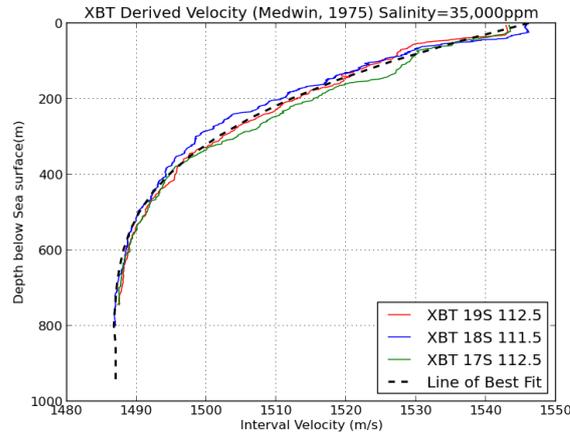
$$V_{INT}(t) = -48.492t^3 + 155.27t^2 - 165.54t + 1546, \tag{4.8}$$

where t refers to the TWT.

Merging velocity sources and gridding

To build the final interval velocity model, we need to merge the velocities function from the water column with the subsurface stacking velocities. The workflow to do so is described in Figure 4.6. The first step involves loading the time velocity pairs from the interactive velocity analysis. A water-bottom horizon interpreted in TWT is used to digitize the water-bottom as an interval velocity value using expression 4.8. Above the water-bottom the XBT function is used to represent the water column velocity as a function of time. Below the water-bottom the stacking velocity picks are converted to interval velocity using Dix Equation (Dix, 1955)
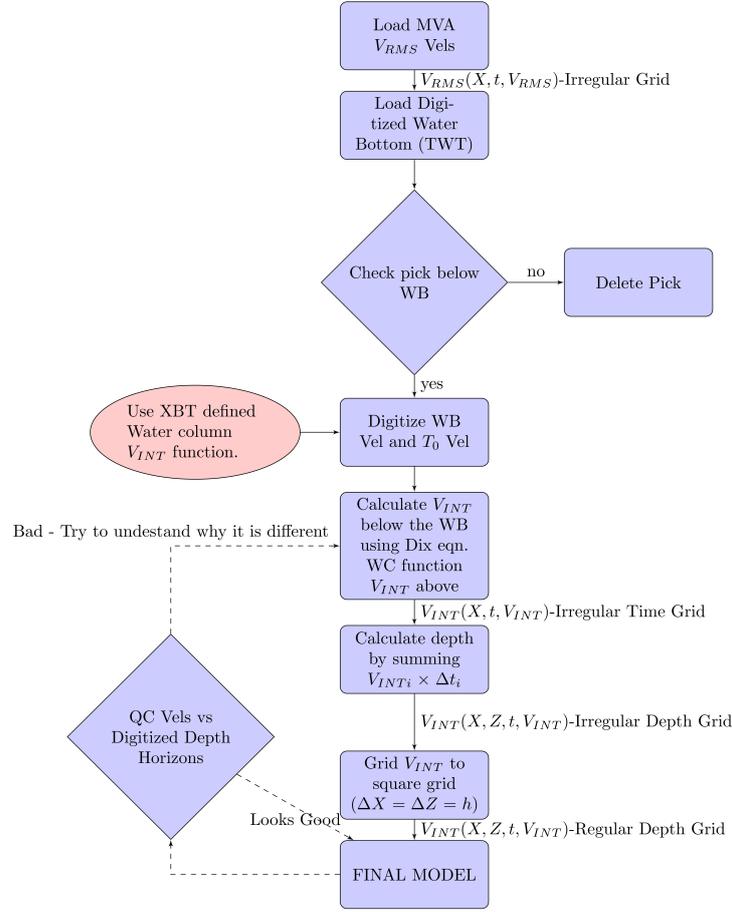
Figure 4.6: Workflow to build crude initial $\mathbf{v}_{INT}$ model from XBT function, water bottom horizon and interpreted stacking (RMS) velocities.

$$v_{INT_n} = \left[\frac{v_{stk_n}^2 t_n - v_{stk_{n-1}}^2 t_{n-1}}{t_n - t_{n-1}} \right]^{\frac{1}{2}}, \tag{4.9}$$

where the number of picks (n) at each CMP location is represented by a time-velocity pair (t is the interpreted, two-way-time (seconds) and v_{stk} is the stacking velocity). With all of the picks available the sparse picks are converted to depth and then gridded on a 25 m rectangular grid with a cubic interpolation algorithm. After griding some mild isotropic smoothing is used to decrease griding artifacts. The resultant initial model is shown in Figure 4.7.

Figure 4.7: The crude initial $\mathbf{v}_{INT}$ model field computed by the workflow described in Figure 4.6. The black point represent the sparse irregular picks that are use to construct the velocity model through cubic griding. Black vertical lines highlight the position of two vertical wells.

4.2.1.2 Anisotropic Reflection Tomography (CGG)

To attempt to improve the low wavenumber velocity reconstruction, 2D reflection tomography was performed by CGG. Significant research was conducted into finding what data was available from the Aus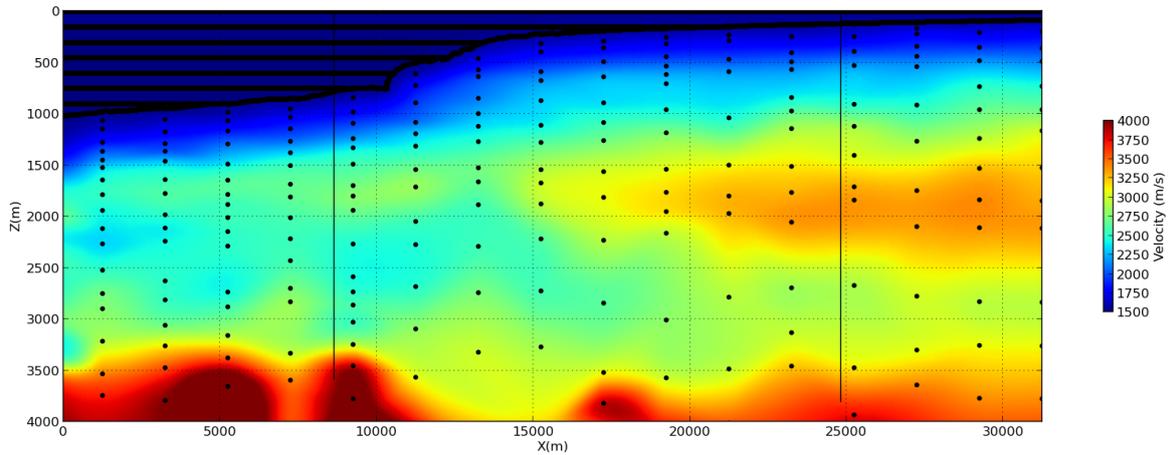tralian "open-file" petroleum databases. These databases compiled by the Australian Department of Mines and Petroleum (DMP) and Geo-science Australia (GA) permitted key additional prior information to both QC the reflection tomography results and also providing meaningful constraints. Well data in addition to an open-file 3D seismic interpretation report (Walton, 2008) was particularly useful. The report contained stratigraphic interpretations and horizon interpretation from a 3D Anisotropic pre-stack depth migration project and was used to perform the geological description in Chapter 1. One of the lines displayed in the report was a seismic section that was coincident with the western potion of the NWA-006 line. This provided good structural control (Figure 4.8). Although the Broadseis line we have is only 2D, if our seismic imaging is performed with an accurate kinematic model, we should hope to have a similar structure at depth at the gas reservoir target as shown in Figure 4.8.

The reflection tomography was performed in two stages and the resultant anisotropic velocity models are shown in Figure 4.9. Firstly, the delta and epsilon models were constructed after which a vertical velocity-only, global anisotropic update was performed.

There were a number of depth imaging processing reports accessible on the Australian government website that suggested the need for inclusion of VTI anisotropy to accurately describe the kinematics of the subsurface. The near surface in this region is made of a mix of soft carbonates and silts: these rocks do not have significant layering present and as a result are approximately isotropic. In the deeper parts of the line there are marine and deltaic shales. These shale layers contain thin laminations that lead to vertical transverse isotropy (VTI). A seismic marker was interpreted that coincided with the approximate top of these marine pelagic shales (H2-Toolonga Calcilucite). Below the Toologna Calcilucite $\delta = 0.08$ and $\epsilon = 0.16$, while above both were set to zero. After the δ and ϵ models were built they were used as constraints in an anisotropic, vertical velocity-only global reflection tomography update. No attempt at updating the anisotropic parameters was made. Some of the key differences that can be seen in the reflection tomography model occur between 2 km and 3 km where there is apparently a continuous velocity inversion in the reflection tomography initial model. This velocity inversion is also supported by well velocities.

We assessed the kinematic improvement of the reflection tomography model, by looking at the improvement of the velocity trends at the two wells and by looking at common angle gathers. The well comparison is shown in Figure 4.10. Some of the key improvements in the initial model are highlighted by blue arrows. In the overburden to the East of the line at the WTR-4 well location we see that the match between the well and seismic velocities in the more carbonate rich lithologies (H0-H2) has been improved quite significantly. This is shown by a systematic increase in the velocities. The velocities in this region initially were picked using NMO velocity analysis. In shallow water, the strong free surface and inter-bed multiple trend will tend to drive picking to a lower velocities trend than the true primary trend. The reflection tomography appears to be more robust to correctly identifying the primary trend. The second key improvement occurs in the more lithic section above the reservoir (H2-H4). In this region a strong velocity inversion occurs to the East as the velocity contrast between the carbonate rich lithologies and the silt/shale lithologies is significant. The NMO velocity analysis has difficulty in picking up this velocity inversion while we see in Figure 4.9 that the velocity inversion occurring at approximately 2500 m depth is consistent from the left to the right of the line. It appears that both the reflection tomography and the crude initial model significantly underestimate the velocity below the IJU in the Triassic section.

To assess the kinematic improvement we have obtained from this anisotropic reflection tomography model we performed an anisotropic "pseudo-RTM". Using the same forward

modelling engine that we will later use for FWI (TOYxDAC_TIME), we can calculate the FWI gradient for the I_p parameter. The adjoint source that is back propagated is not the residual wavefield as shown in expression 4.7. We instead back-propagate the true shot record. This pseudo-RTM workflow was performed for every 5th shot record (approximately every 100 m). We can perform this process for a number of source to receiver offset limited bins (each bin is for an offset range of 150 m). With these image gathers, we convert from the offset domain to the angle domain using Hampson Russell software using the frequently used offset to angle angle transform approximations (Todd and Backus, 1985). Although the offset to angle conversion is not precise as it would require ray-tracing, it still allows an efficient means for limiting the data to the pre-critical reflections from which we can calculate angle stacks (Figure 4.11) and common angle gathers (Figure 4.12). When looking at these angle gathers and stacks, it is important to note that in order to keep the events at the water bottom, the offset header is modified artificially. The minimum offset bin that comes from our image gather workflow is 375 m. This is due to the fact that the closest receiver on the streamer is over 200 m from the source. As we apply a 50 degree outer angle mute to our angle gathers to limit it to pre-critical reflection, if we do not modify the offset header, the shallow water the water-bottom event would not be visible as the data measured at the first receiver is offset by more than 50°. By subtracting offset 375 m from the offset header we get well muted CAGs and stacked image but will see non-flat post-critical and diving wave information in our gathers close to the water-bottom. The apparent lack of gather alignment in this area is an artifact of this.

The angle stack which is limited to 30° shows a target reservoir structure, which is much more consistent with the structural image shown in the open-file interpretation report. In this case, both of the gas reservoirs structure exhibit monotonic dip, from right to left of the image after application of reflection tomography. Prior to reflection tomography, this was not the case. The common angle gathers (Figure 4.12) show a decrease in the residual move-out present as highlighted by the coloured arrows.

a)

b)
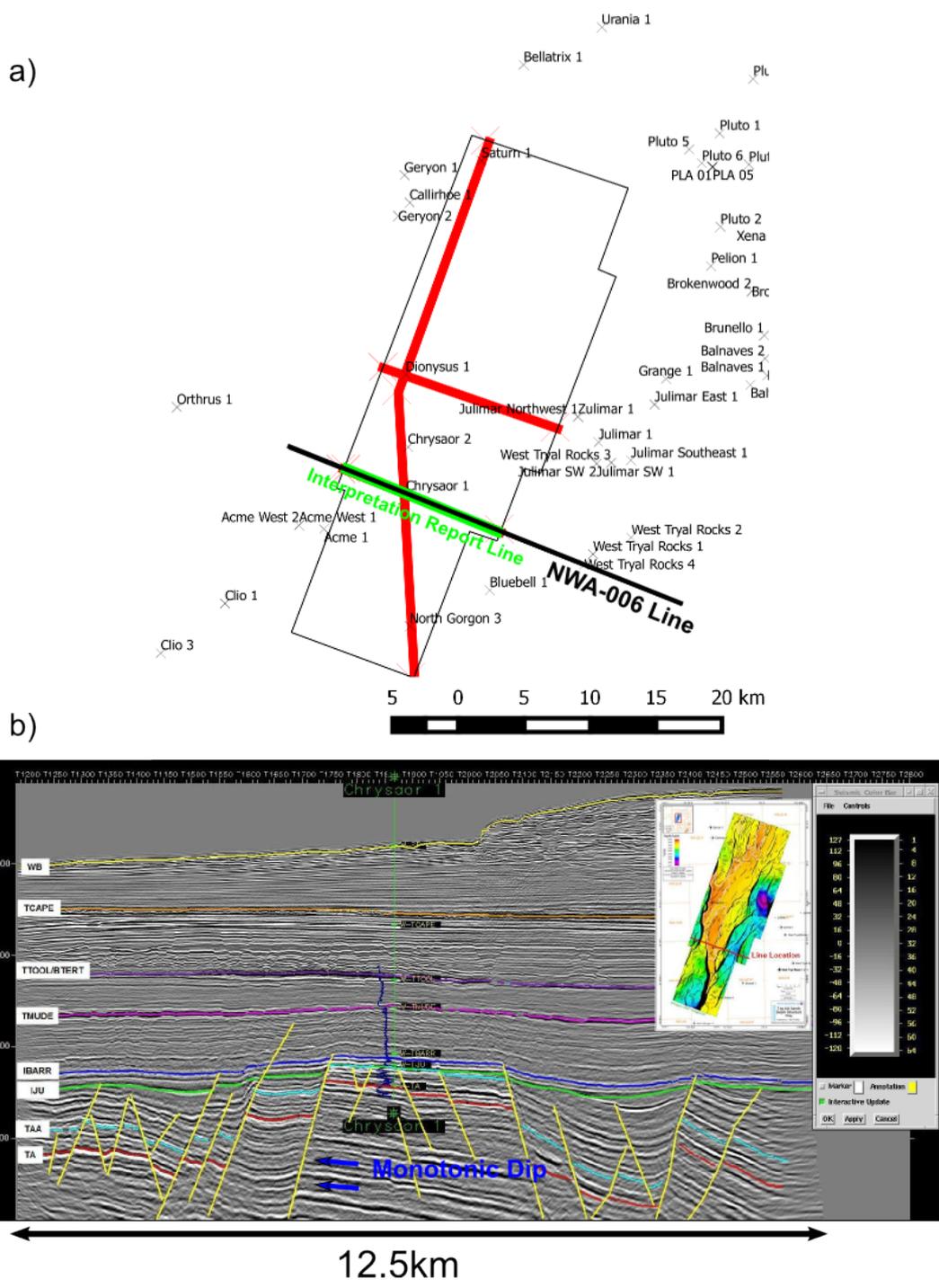
Figure 4.8: a) The Chrysaor-East Anisotropic PreSDM project is highlighted by the polygon. Three 2D lines are displayed in the report (Red and green). b) The green line was coincident with a portion of the NWA-006 line and showed the expected structure in the shallow part and at depth. Note the relatively homogeneous dip in the Chrysaor-1 structure.
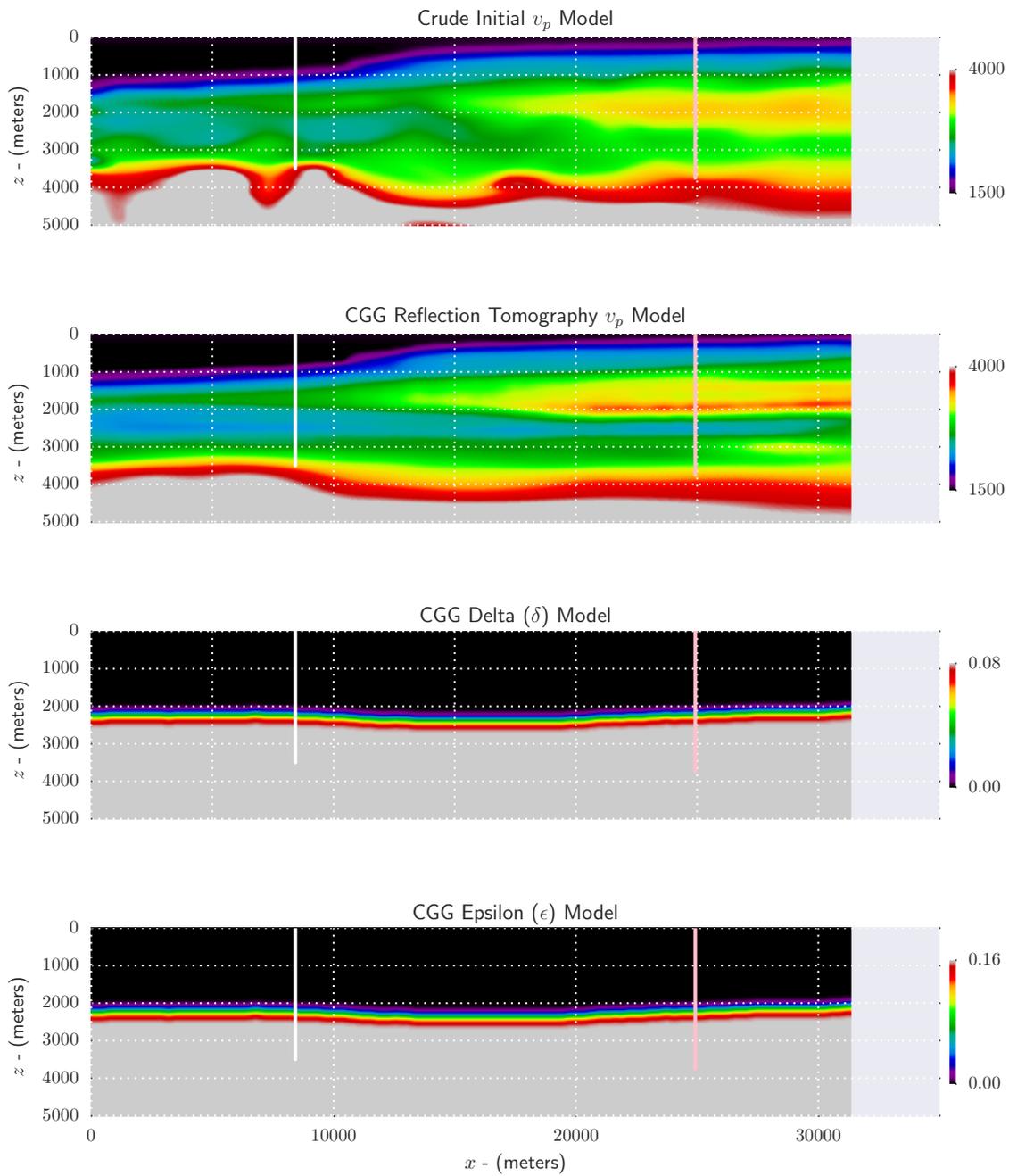
Figure 4.9: a) Crude initial velocity model built from stacking velocities, b) The reflection tomography FWI initial model built from a global velocity only tomographic update with the anisotropic parameters c) δ and d) ϵ set prior to the update. The location of the Chrysaor-1 (white) and WTR-4 (pink) wells are also annotated.
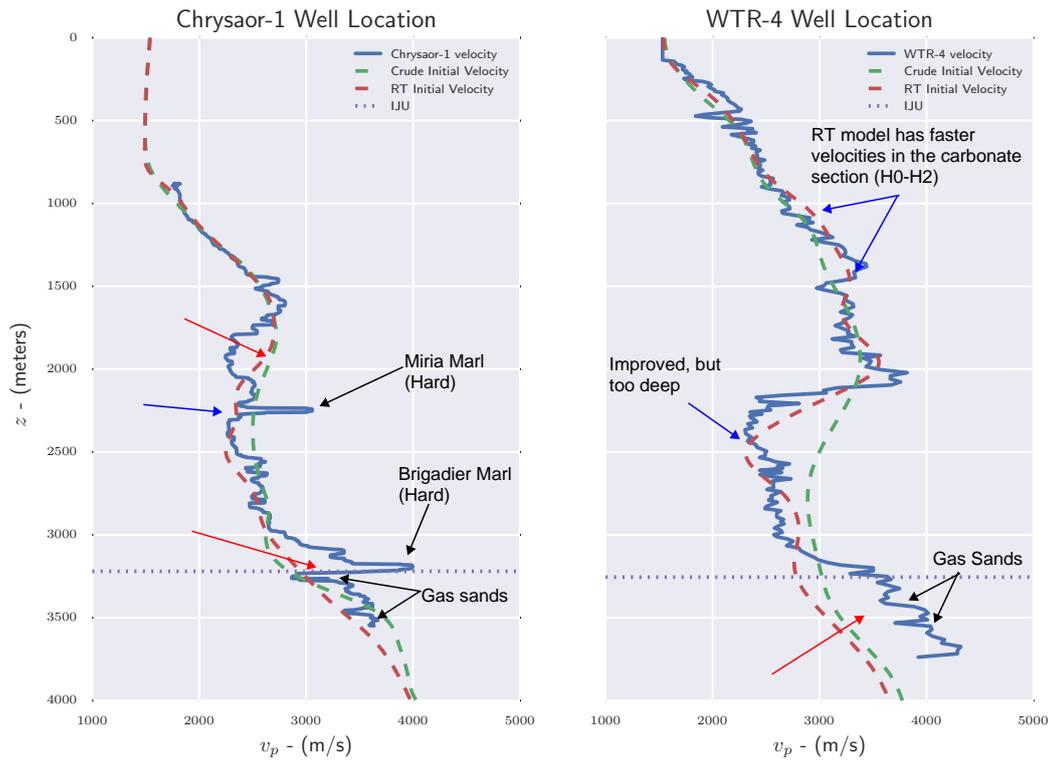
Figure 4.10: Comparison of the crude initial model (green) and reflection tomography (green) to the VSP/Sonic well velocity (blue) at Chrysaor-1 and WTR-4. Blue arrows highlight clear improvements in the velocity model while the red arrows highlight residual mismatch. The black arrows highlight the gas sand and key stratigraphy.
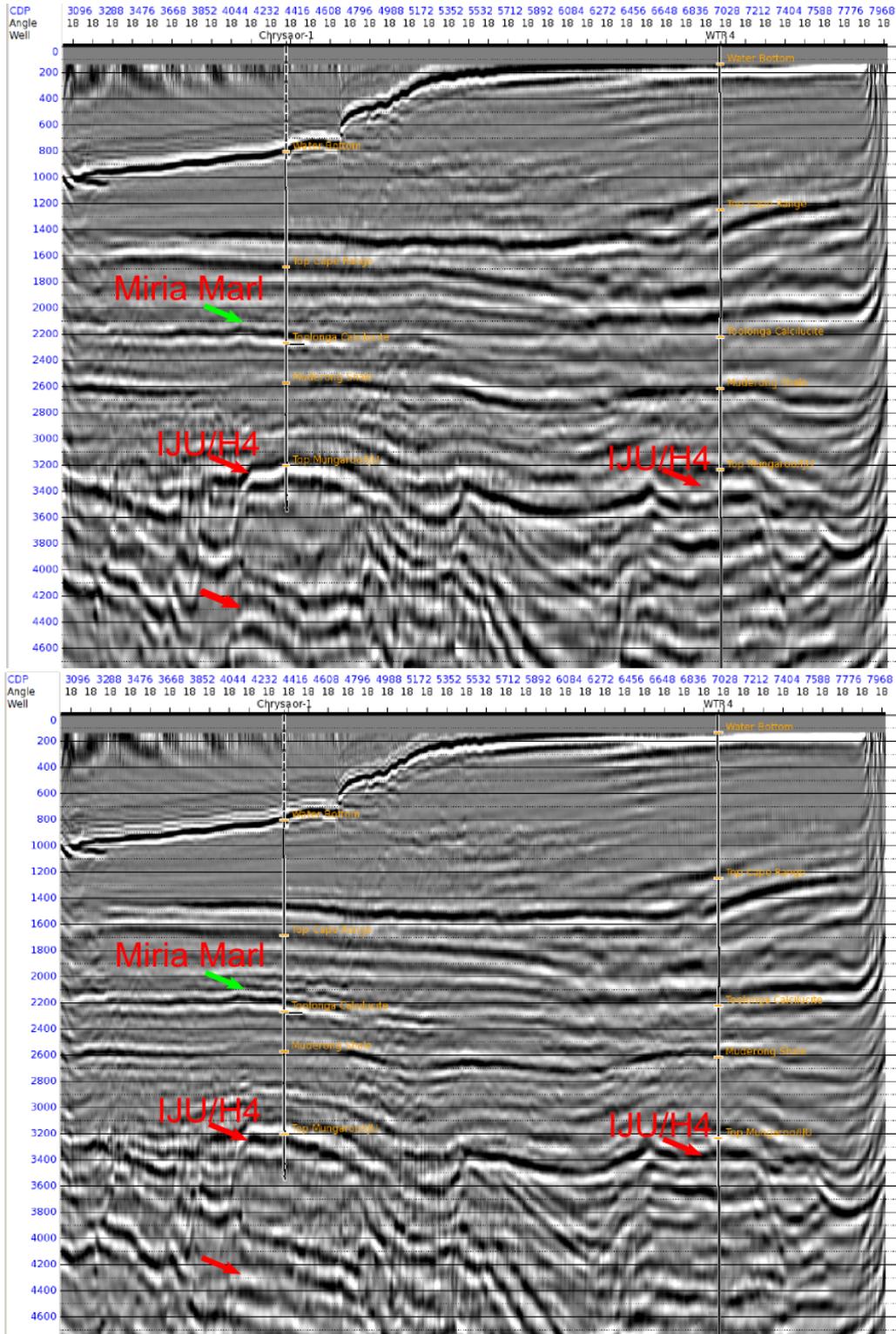
Figure 4.11: 0-30° stack computed using (*TOP*) crude initial model and (*BOTTOM*) the anisotropic reflection tomography model. The red arrows indicate where the reservoir structure is much more consistent with the WAPIMS interpretation report. A green arrow highlights the higher amplitude Miria Marl.

Figure 4.12: Common angle gathers computed using the (*TOP*) crude initial model and (*BOTTOM*) the anisotropic reflection tomography model. The red arrows highlight improvements in the shallow section under the shelf break. The green arrow highlights improvements to gather alignment at the Chrysaor-1 structure.
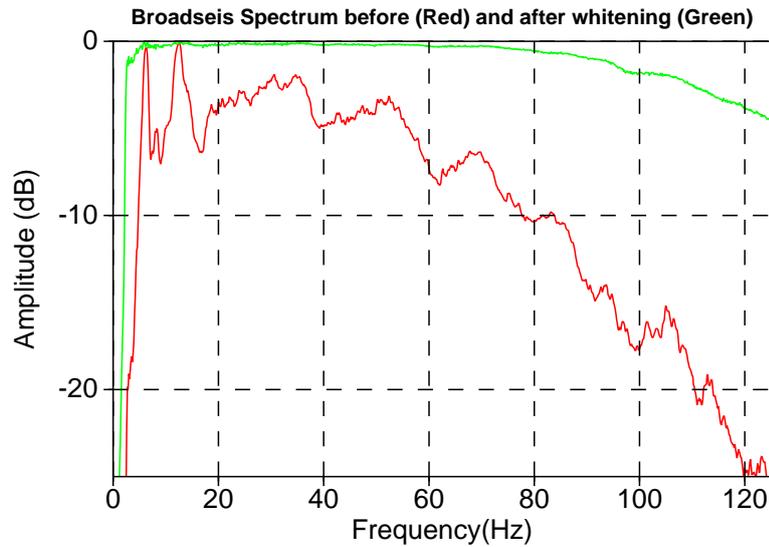
Figure 4.13: The Broadseis spectrum before (red) and after (green) minimum phase spectral whitening.

4.2.2 Full waveform inversion

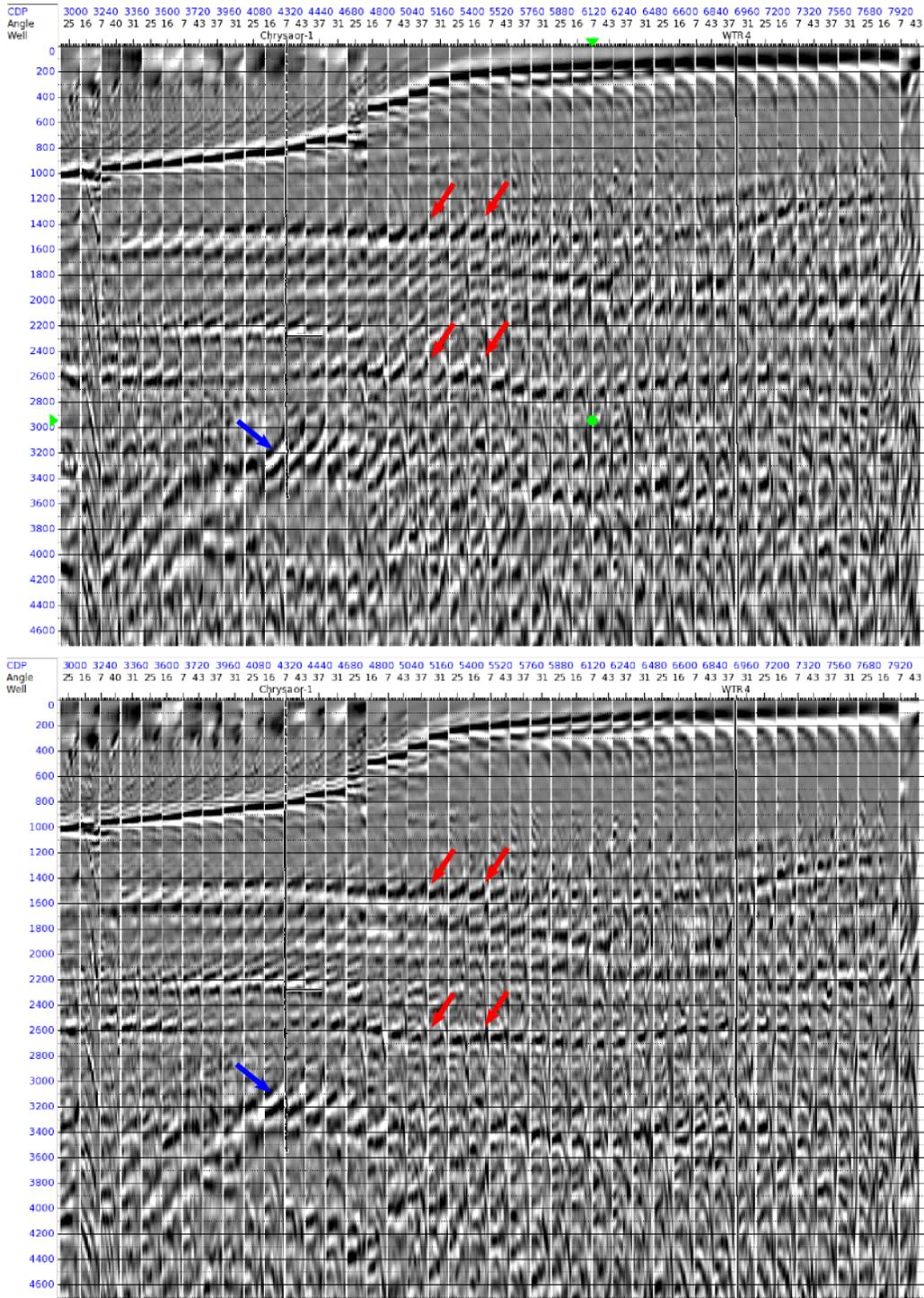The focus of the real data FWI investigation was to look at the role FWI could perform in improving the imaging of the NWA-006 line. There was special interest whether the additive inverse Laplacian preconditioning that was successfully applied on synthetic data in Chapter 3 would be of use in a real data situations. Another potentially interesting possibility was to investigate the impact a Broadseis streamer acquisition would have with respect to the Conventional streamer. Although such an investigation was of interest it was not included in the body of the thesis. In this chapter we will focus on the Broadseis dataset due to the extended low frequencies available for inversion.

4.2.2.1 FWI preprocessing

Only a minimal preprocessing workflow was applied, prior to performing FWI. The first step was to attempt to locate "noisy-traces" that were characterized by anomalously high amplitudes. These noisy traces were removed from the input data set as to not contaminate the FWI result. It was then important to "flatten" the frequency spectrum, due to our desire to apply the same multi-scale FWI workflow (Bunks et al., 1995) used on the synthetic datasets. The application of the multi-scale workflows requires one to apply band-pass filters to the true data. The amplitude spectrum of the real data shows some quite rapid amplitude variations. The most significant of which is two notches

which occur below $20Hz$, right within the FWI frequency band we would like to use. To flatten the spectrum, minimum phase spectral whitening is applied (Figure 4.13). The application of this whitening workflow was identical to what was applied in other FWI real data applications such as Ravaut et al. (2004).

This resultant frequency spectrum is much flatter especially within the $0 \rightarrow 20Hz$ region we will be using our FWI application. The final correction applied is an approximate 3D to 2D correction in the form of a gain described as a power law of time, T^2.

Before performing the inversion, it is important to understand some of the data quality issues associated with the dataset. A shot record taken over the Chyrsaor gas field location is useful to highlight the frequency dependent signal to noise content of the dataset(Figure 4.14). Four butterworth bands are taken that show the data quality as we move from the lower frequencies to higher frequencies. The lower frequencies are often used to decrease non-linearity in FWI, (Sirgue, 2006), this is especially the case when the offsets are limited (as is the case in this marine streamer study). There are limitations to how low we can go as swell noise and limitations of the recording equipment prevent us from approaching $0Hz$. In this study, we use Band 2, which has a peak frequency of approximate 2.5Hz as our starting band for FWI. The coloured arrows in Figure 4.14 show some of the key events present in the data. We perform FWI using the entire shot record and will be looking to fit all of these features including the sub-critical reflections. Note that there is also noise that appears to have come from out of the plane (orange arrow). It is expected that this has come from an offset seismic survey occurring at the same time. This in addition to noise from swell noise are features that we should decrease due to the inherent summation that occurs when we calculate the FWI gradient.

4.2.2.2   FWI without AIL preconditioning

Prior to discussing the role the additive inverse Laplacian can provide to stabilize the inversion results, we highlight the FWI workflow that will be applied. The only additional difference between this FWI workflow and what is applied for AIL preconditioned results will be the fact that the FWI gradient will be preconditioned by the AIL correlation operator.

Our initial model was re-gridded to a cell size of $h = 25$ m. Velocity is inverted in a multi-scale fashion where 6 second-order butterworth frequency bands (Table 4.1, Figure
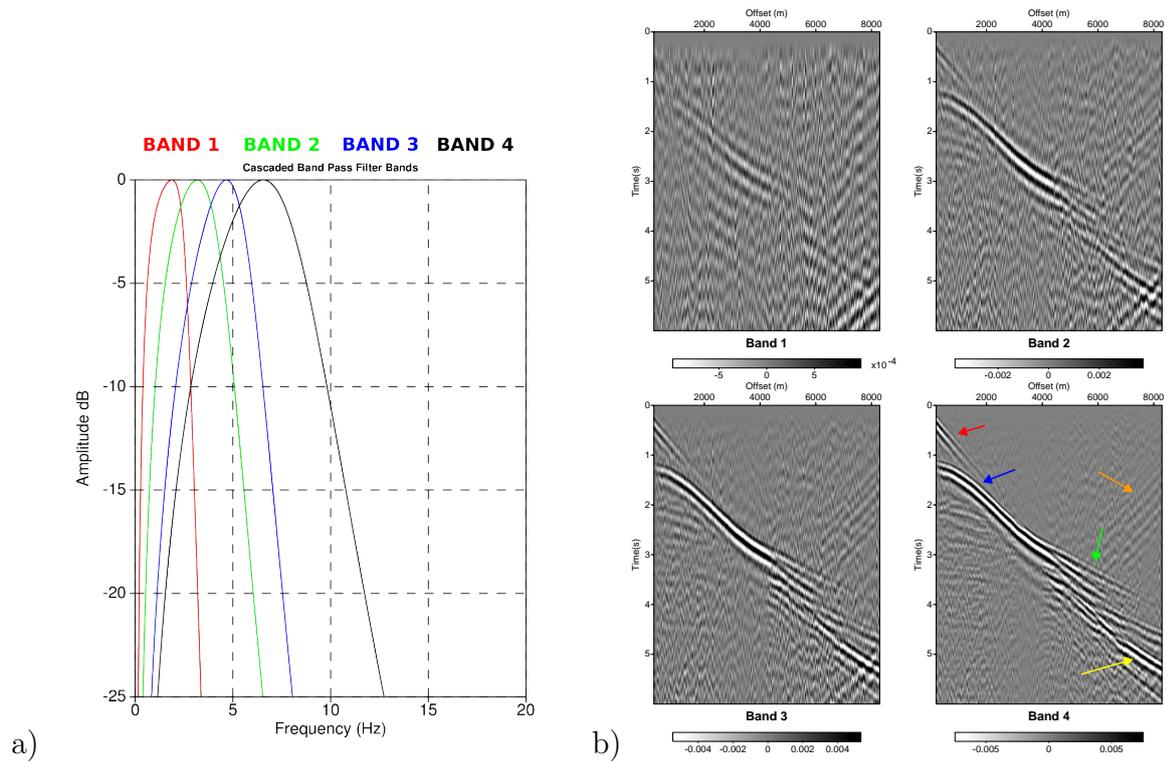
Figure 4.14: a) 4 Band pass filter bands to help asses signal to noise in different frequency ranges. A preprocessed shot record for the different frequency bands. Note the lower frequency bands have a lower signal to noise ratio than the higher bands. The coloured arrow highlight key seismic events. *RED:* The direct wave. *BLUE:* The water bottom reflection. *GREEN:* The first arrival. *YELLOW*: Post critical reflection events. *ORANGE:* Potentially noise from another survey.

4.15) are inverted with the output from the previous band serving as to the current band. At the start of each band, the density model is set to a constant value of $\rho = 1000kg/m^3$. Prior to updating the subsurface parameters (velocity and density) a source estimation is required. The entire shot record is used for the source estimation which is performed in the frequency domain by using the fact that the relationship between the seismic wavefield and the source is linear (Pratt, 1999)

$$s(\omega) = \frac{\mathbf{d}^{\dagger}_{mod}(\omega)\mathbf{d}_{cal}(\omega)}{\mathbf{d}^{\dagger}_{mod}(\omega)\mathbf{d}_{mod}(\omega)}, \tag{4.10}$$

where the source function, s is calculated as a function of frequency ω and the complex conjugate transpose is denoted by the symbol $\dagger$. Once the source individual frequency components are estimated, the source function is transformed back into the time domain
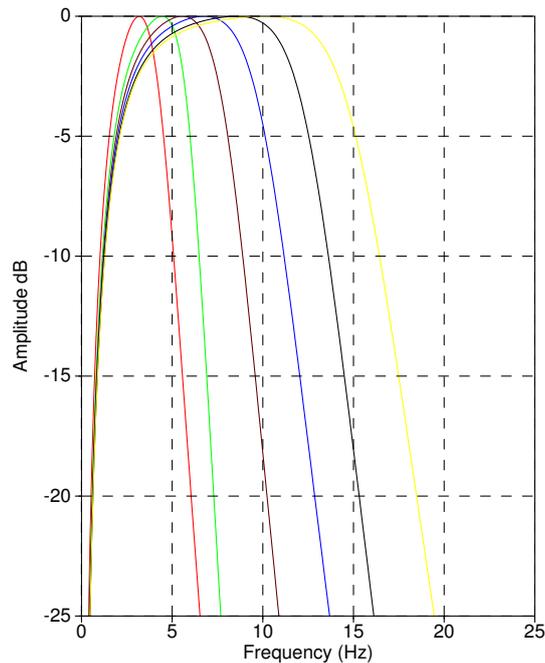
Figure 4.15: The six butterworth second order frequency bands used as part for FWI. The four frequencies describe the band pass filter being applied and correspond to those shown in Figure 4.1.

using an inverse FFT. The beginning and end of this source function are then tapered to stabilize the forward modelling and to remove spurious information at later lag times. It is important to note that the source is only estimated once for each frequency band at the start of the inversion.

BAND	STOP LOW	PASS LOW	PASS HIGH	STOP HIGH
Band 1	$0.1Hz$	$1.5Hz$	$3.0Hz$	$6.5Hz$
Band 2	$0.1Hz$	$1.5Hz$	$4.5Hz$	$8.0Hz$
Band 3	$0.1Hz$	$1.5Hz$	$6.0Hz$	$11.0Hz$
Band 4	$0.1Hz$	$1.5Hz$	$7.5Hz$	$14.0Hz$
Band 5	$0.1Hz$	$1.5Hz$	$9.5Hz$	$17.0Hz$
Band 6	$0.1Hz$	$1.5Hz$	$11.5Hz$	$20.5Hz$

Table 4.1: Broadseis FWI Butterworth Frequency Bands.

The final inversion results obtained from the reflection tomography initial model are shown in Figure 4.16. We can see that both the velocity and density updates appear correlated and are approximately consistent with the geological structure we have seen

on migrated images. It appears the velocity update has a lower wavenumber spectrum than the density update. This is to be expected as the density will be more sensitive to the shorter illumination angles and will have a wavenumber content similar to a migrated section. A number of features has been added after FWI, two of the most visible being the addition of the Miria Marl and the velocity contrast at the IJU boundary. The Miria Marl in the shallow can be seen to correlate well with the velocities at the Chrysaor-1 well. It appears that the velocity inclusion in the FWI model comes in at deeper depth than is suggested by the well, (approximately 100 m). However, this is not unexpected as the velocity in the FWI initial model is overestimated just above the marl event. A strong imprint of the Miria marl is also visible on the density result. The second point of interest is the sharpening of the IJU formation at depth. At the Chrysaor structure, the IJU reflector is marked by perhaps the strongest velocity contrast on the line as the lithology changes from the hard Brigadier marl (absent at WTR-4) into the very soft upper Mungaroo gas sand. There is evidence of both the anomalously fast marl and anomalously slow gas sand and their depths are approximately consistent with the wells. To try to improve the imaging of these feature it may be worthwhile performing FWI at higher frequencies to improve these results. The current velocity changes do improve the alignment of the common angle gathers in this region (Figure 4.17).

A composite of the shot record before and after FWI at the Chrysaor-1 well shows the effectiveness of the data fitting of the FWI engine (Figure 4.18). The short spread reflection events are highlighted on the true data which appear to match well with the modelled shot record. The character of the far offset diving waves is also much more in agreement in the FWI updated model than the initial velocity model. The displayed shot record is extracted at the Chrysaor-1 location and is for the final frequency band (Band 6) included in the inversion. The evolution of the objective function shows that, for each band of the inversion the objective function has decreased (Figure 4.19) with the minimum decrease being around 6% for the second band and the maximum around 15% for the 4th band.

The right-hand side of the line, towards the WTR-4 well has had only minor updates, although these appear to be consistent with the geological strata. A potential reason for the lack of strong updates in this region is two-fold. Firstly, the Miria marl is absent at WTR-4 and, as such, there is lack of a large strong overburden velocity contrast. Well data suggests that the IJU/reservoir velocity contrast at the West Tryal rocks is much
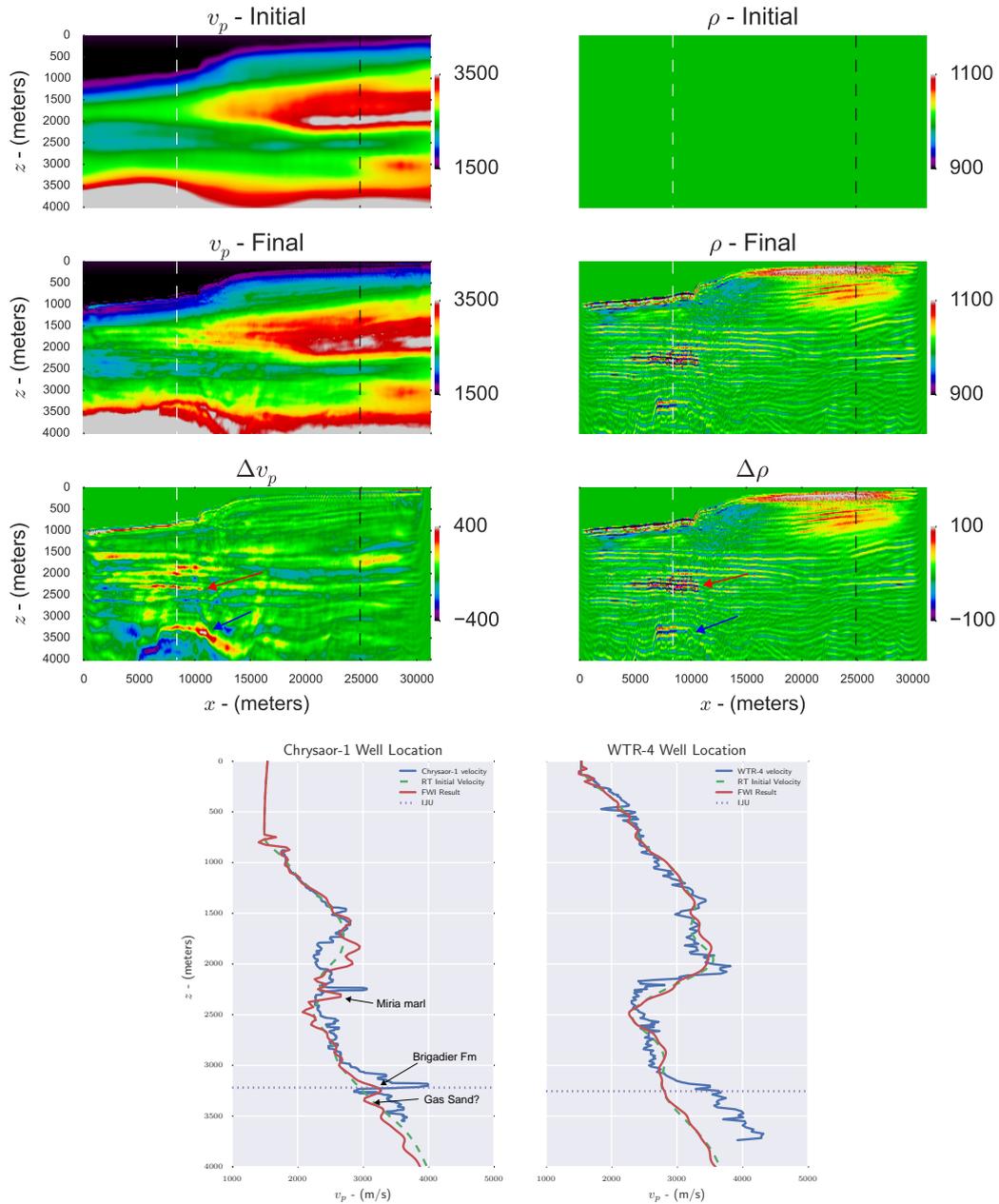
Figure 4.16: FWI using reflection tomography initial model. The arrows highlight the Miria Marl at H2 and the IJU (Brigadier Marl to Triassic gas sand).

weaker than at Chrysaor-1 due to an apparent absence of the Brigadier marl. It is also noticeable that the mismatch between the well velocity and initial velocity model is also much more significant than at the Chrysaor-1 well. This could also significantly affect the results.
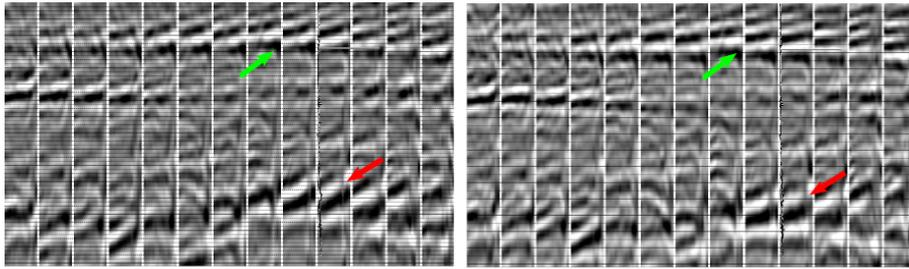
Figure 4.17: The common angle gathers computed using (*LEFT*) Reflection Tomography initial model (*RIGHT*) velocity model after FWI. The green arrow highlights the Miria marl while the red arrow highlights the IJU reflector at the Chrysaor structure.
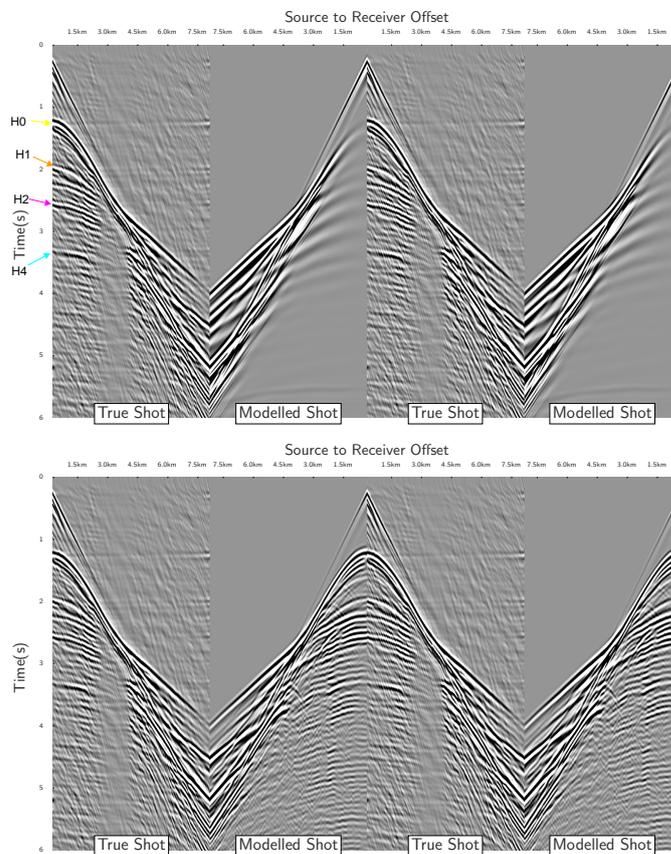
Figure 4.18: Shot taken from the approximate Chrysaor-1 location before (*TOP*) and after (*BOTTOM*) FWI using the reflection tomography initial model. The key stratigraphic horizons are highlighted in the shot record (H3 appears too weak to see on an individual shot record.)

Sensitivity of FWI result to the initial model

To illustrate the importance of the initial model and also the role of diving wave energy,
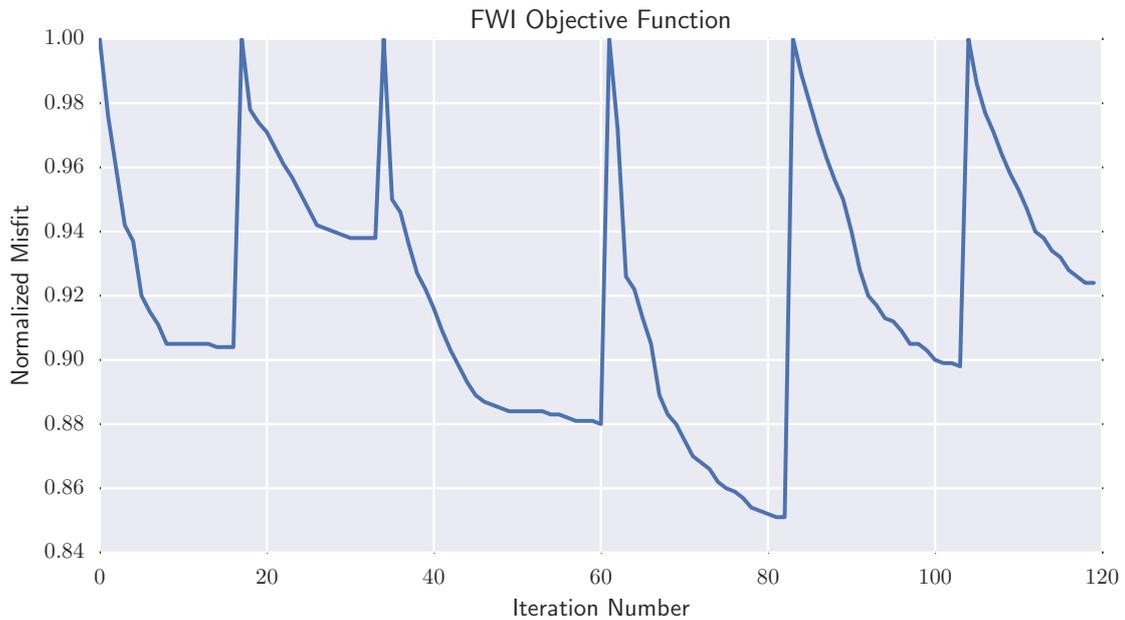
Figure 4.19: FWI objective function for the inversion using the reflection tomography initial model.

we perform FWI using our crude velocity model. This initial model was proposed as the initial model for reflection tomography, but not FWI. The FWI inverted model using this crude initial model is shown in Figure 4.20. The result has some undesirable approximately vertically orientated features (highlighted with black arrows). These features cross many geological strata layers and are FWI artifacts that are an unrelated to the true velocity field. Other issues with the inversion are the fact that it does not provide good focusing of the anomalous velocity features at the Miria Marl and the top of the IJU at Chrysaor-1. When we compare the inversion results to the well data, this difference is easier to identify (Figure 4.21). The IJU velocity contrast comes in at a much deeper depth (200 m) than the real event and in addition to this the Miria marl is no longer identified as a simple thin feature as noticed in the well. Blue arrows do highlight some improvement that can be noticed at WTR-4, in the overburden carbonate rich sediments (H0-H2).

A direct comparison of the two FWI velocity models (Figure 4.22) shows that the FWI results are virtually identical in the shallow section. Use of a back-ray-tracing Ekional equation algorithm provided by Stéphane Operto allowed us to map the expected maximum penetration of diving waves. We see that the shallow zone where inversion results
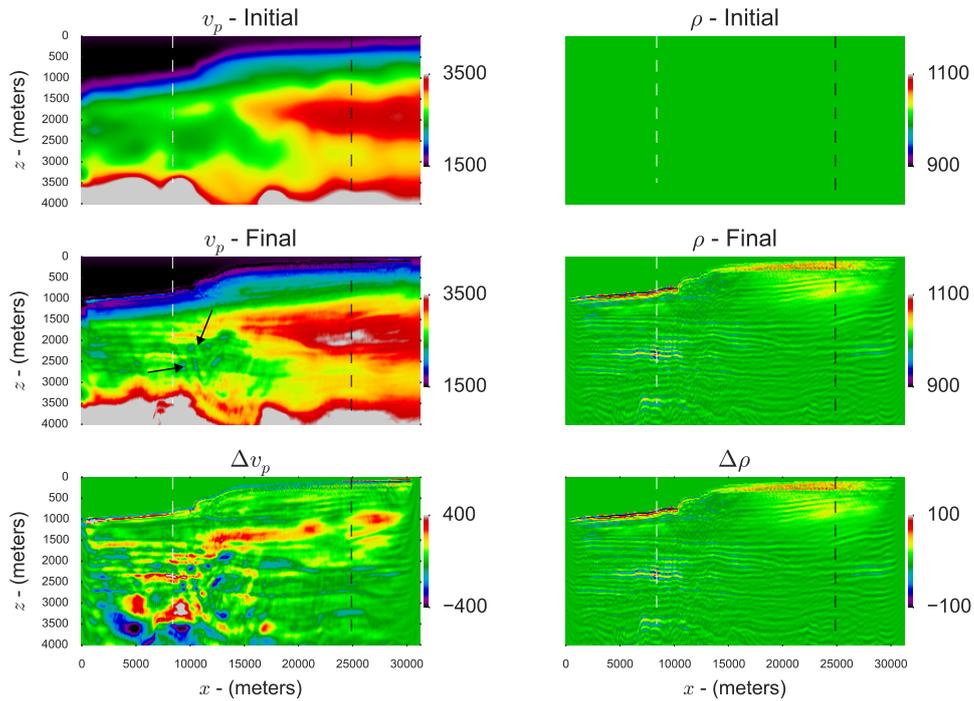
Figure 4.20: FWI using the crude initial model.

are close to identical corresponds to the zone illuminated by diving waves. This zone is well highlighted in the deeper section and also at the edge of the model where our shots do not have significant enough offsets to contain diving waves. The combination of diving waves (low wavenumbers) and short spread reflections (high-wavenumbers) make this region of the line quite stable for the inversion. As we move below the diving wave zone, there is no longer stabilizing the low-wavenumber contribution and the inversion becomes significantly more non-linear and dependent on the initial model.

I calculate common image gathers using the FWI updated crude model (Figure 4.23). The improvement of gather alignment is quite significant, this improvement is greater than what was obtained from solely reflection tomography on the initial model (Figure 4.12). It appears that significant improvements to gather alignment can be gained by correcting the shallow velocity structure. The success of using FWI to correct the shallow velocity structure in this region has been previously discussed by Manuel et al. (2014) (results in Figure 4.24). In this example a similar area was put through a 3D FWI workflow to successfully update the shallow velocity structure. The shallow velocity structure in this 3D example, is the same carbonate rich stratigraphy (H0-H2) that we
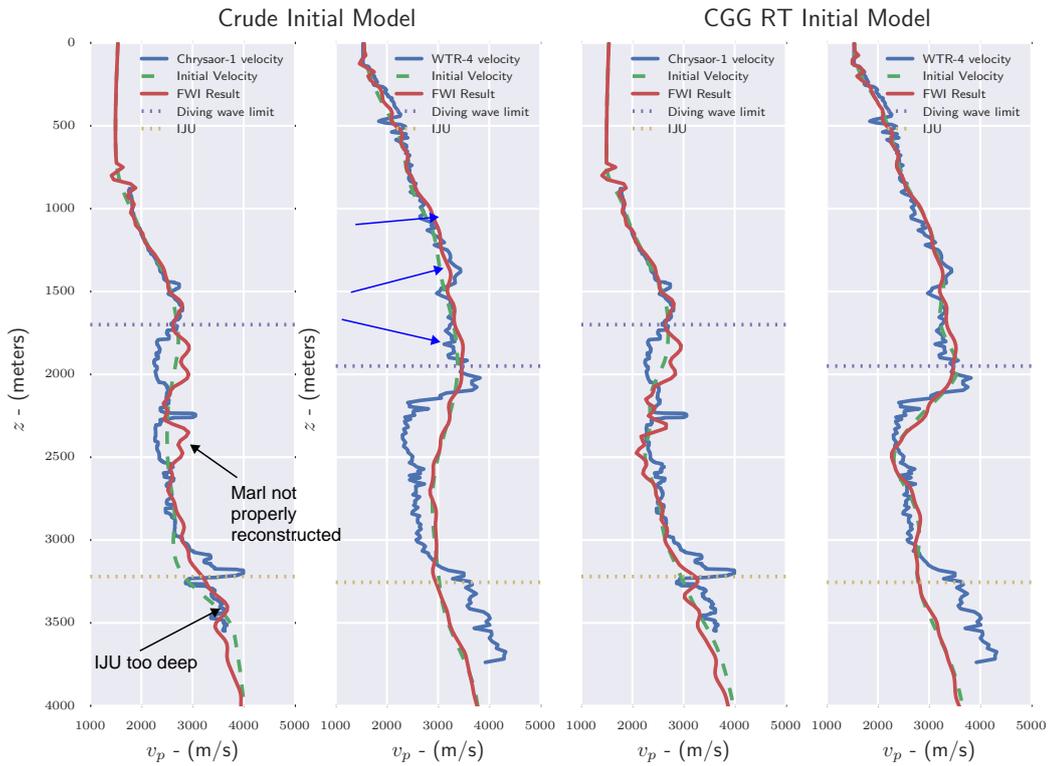
Figure 4.21: The FWI results from the crude initial model (*LEFT*) and the reflection tomography initial model (*RIGHT*). Blue arrows highlight apparent improvement of the velocity.

see at the Broadseis line. Although the updated velocity model is significantly improved in the shallow section, the author states:

"Due to the relatively short cable length and limited frequency content, FWI is only converging successfully in the areas where turning ray energy is available. Where it is not available reflection tomography is required in the workflow to assist in successful convergence."

These findings are consistent with what we have experienced with this Broadseis line. In our case there is also additional complexity due to the fact our 2D FWI cannot completely correct for 3D effects. It is likely that to improve the performance of the inversion longer offset acquisition and/or better initial model building results are required.
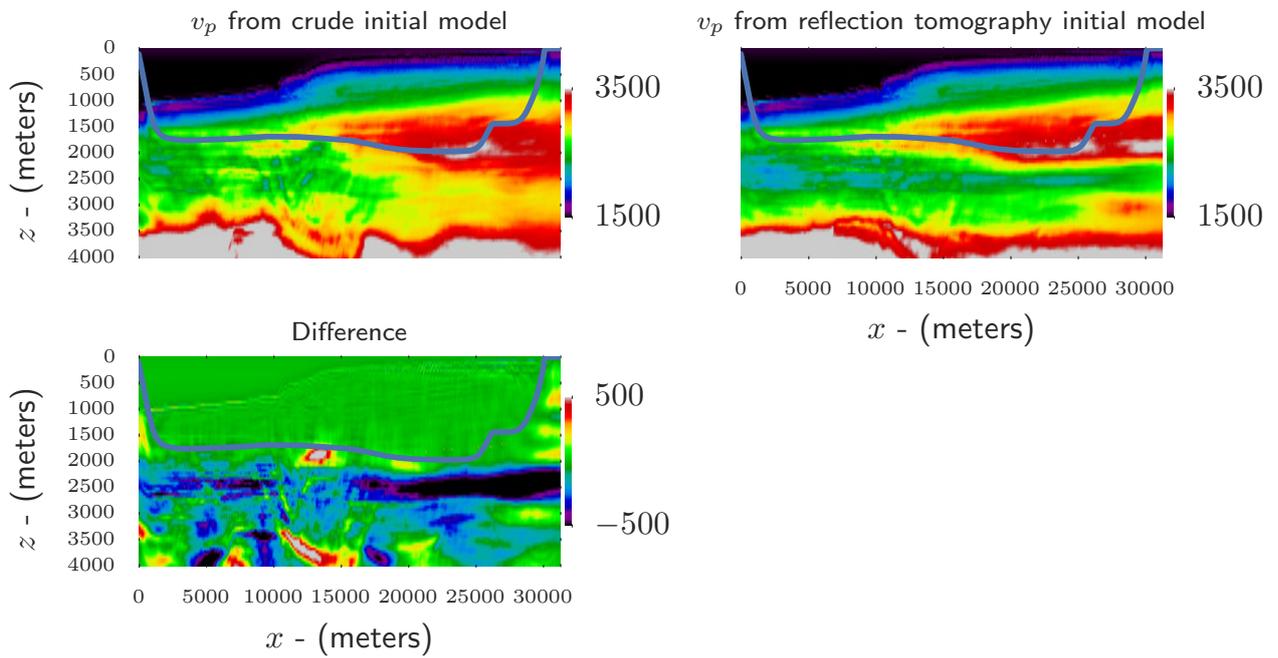
Figure 4.22: The final inversion result obtained using the (*TOP*) crude initial model (*CENTRE*) CGG reflection tomography initial model (*BOTTOM*) The difference between the two results, with the blue line highlighting the penetration of diving waves calculated using a back-ray-tracing algorithm.
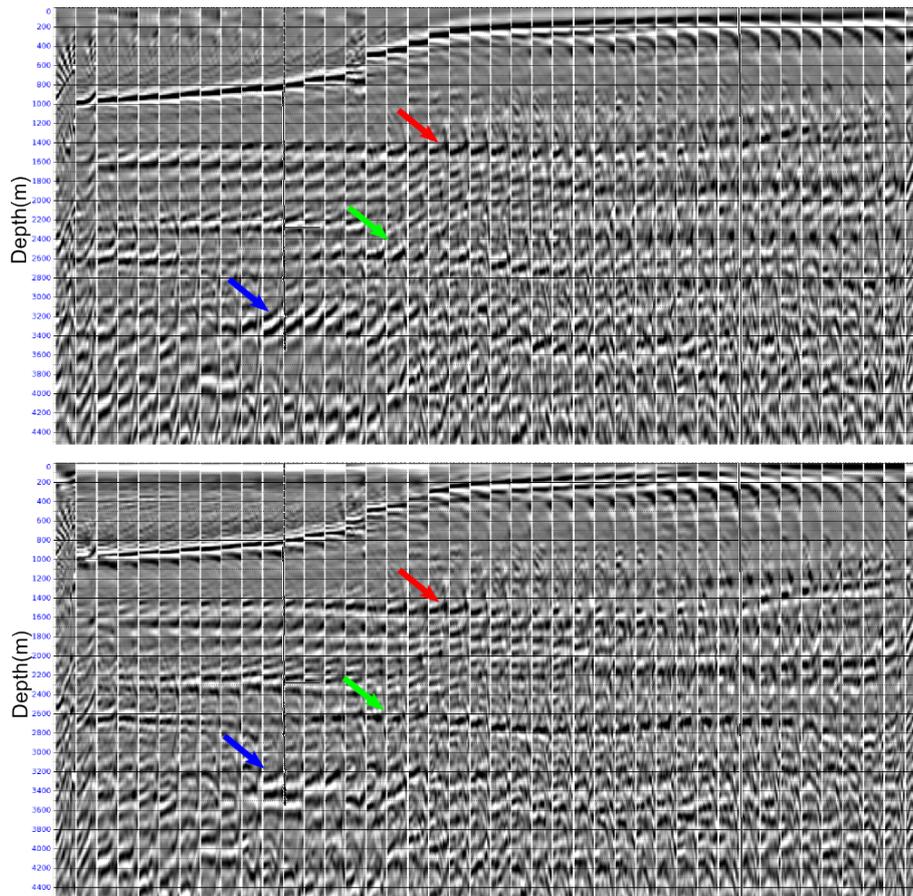
Figure 4.23: Common Angle Gathers computed using the crude initial velocity model and the result after the application of FWI. There is a significant improvement in gather flatness at H1(Red), H3 (Green) and H4(Blue).
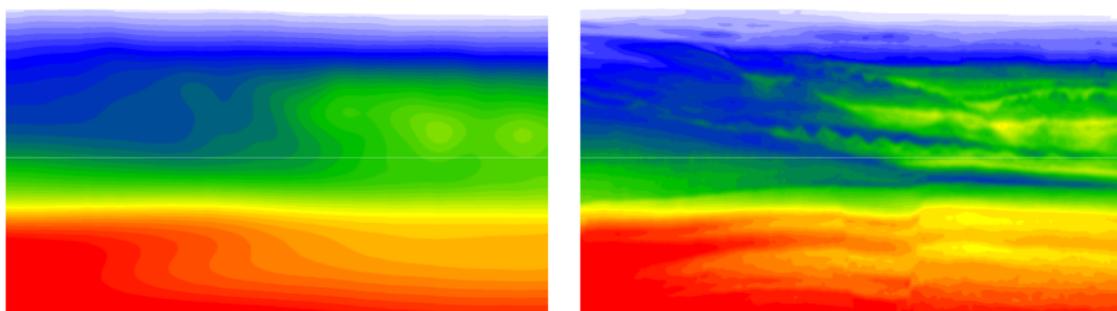
Figure 4.24: 3D FWI example taken from Manuel et al. (2014) in the same region as the NWA-006. The initial model (*LEFT*) and the FWI model (*RIGHT*). Notice that the most significant updates are in the overburden section (H0-H2).

4.2.2.3 FWI AIL preconditioning

The NWA-006 inversion results appeared to be quite stable with only some minor high-wavenumber artifacts in the shallow section, near the water-bottom. The synthetic FWI examples detailed in Chapter 3 suffered from much more significant artifacts. These artifacts were introduced from either the addition of white noise to the observed data vector, $\mathbf{d}_{obs}$ or from aliased artifacts coming from an inappropriate frequency increment $\Delta\omega$. The increase in the frequency increment was investigated as it decreases the cost of frequency domain FWI. To highlight the potential utility of the AIL framework for time-domain FWI, I will use it to maintain the stability of the inversion when we increase the shot spacing involved in the inversion. The ability to be able to decrease the number of shots is of significant interest in time domain FWI. The time domain approach, unlike the frequency domain, has no efficient strategy for dealing with multiple sources. As we decrease the number of sources involved in the inversion, we decrease the number of CPU hours required for the calculation of the gradient and to perform the line search. In the NWA-006 example I have used 1 CPU per shot. As we perform shot decimation, the CPU hours required for gradient building and the line search decreases linearly with the shot increment.

The Figure 4.25 highlights some of the artifacts that are introduced to the model-space as we perform FWI. Our "stable" FWI results shown previously was performed using every 5th shot record. We investigate the results for shot increments of every 10th and every 20th shot. As we increase this shot increment, spatial aliasing effects occur at each computed gradient. The first computed gradient from the first bunks band of the inversion is shown for reference in Figure 4.26. The aliased artifacts in the gradient map into the model-space at each iteration. This information results in a undesired non-geological, vertically orientated aliased footprint. When we look at the evolution of the objective function for these different shot decimation cases, we see that as we increase the shot increment, it takes longer to for the inversion to reach convergence (Figure 4.27). A comparison of using every 5th shot versus every 20th suggests an increase in the required number of iterations of approximately 200%. After performing these additional iterations, it is however apparent that the final inversion result is inferior to the inversion with a denser sampling of shots.

To attempt to mitigate against the undesired, aliased imprint we used our AIL frame-
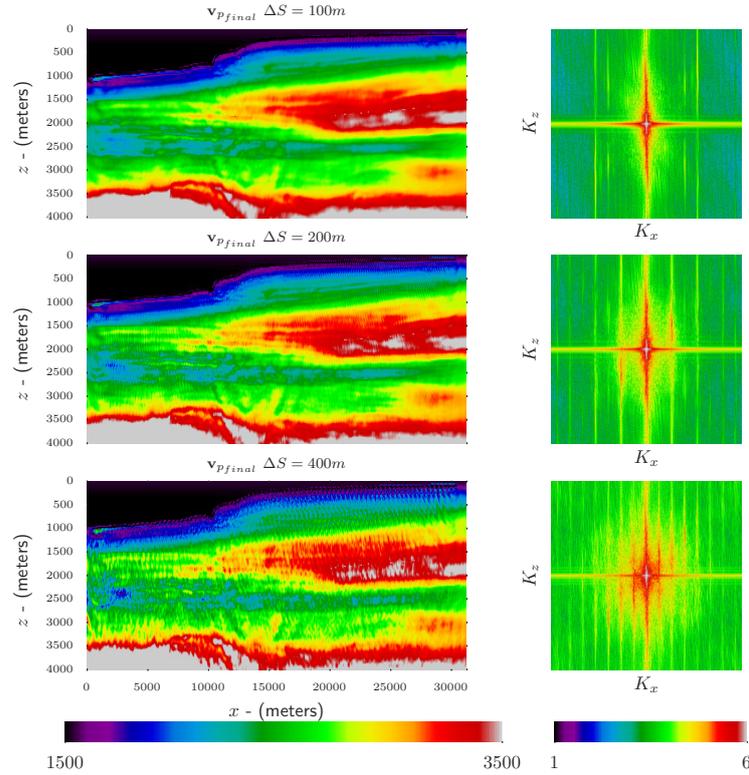
Figure 4.25: The final inversion results when FWI is performed with every 5th (TOP), 10th (CENTRE) or 20th (BOTTOM) shot. The increased shot spacing results in aliased artifacts in the final inversion results.

work to precondition the gradient. Our preconditioned gradient involves solving one linear system involving the $\overline{\mathbf{Corr}^{-1}_{2D_{AIL}}}$ (Algorithm 4).

Algorithm 4 Calculate the preconditioned gradient $\mathbf{G}^k_{prec} = \overline{\mathbf{Corr}_{2D_{AIL}}}(\mathbf{G}^k_{\text{data}})$

1: $\underbrace{\mathbf{A}}_{\overline{\mathbf{Corr}^{-1}_{2D_{AIL}}}} \underbrace{\mathbf{x_{2D_1}}}_{\overline{\mathbf{Corr}_{2D_{AIL}}}(\mathbf{G}^k_{\text{data}})} = \underbrace{\mathbf{b}}_{\mathbf{G}^k_{\text{data}}}$

In other examples in Chapter 2 and 3 we have solved the linear system twice. The decision to solve it only once in this case to preserve some of the high vertical wavenumbers. A vertical correlation length, L_z of 12.5 m is used for all cases. As the aliasing artifacts increase with increasing shot spacing a higher horizontal correlation length is required for the coarser shot spacing. We use horizontal correlation lengths of 100 m, 300 m and 750 m for the every 5th, 10th and 20th shot cases respectively. The maximum dipping reflector in this seismic line was identified in the Triassic section in one of the tilted grabens. This dip was 5° and was deemed to be insignificant and, as such, our correlation lengths are

Figure 4.26: The gradient at the first iteration of FWI when it is performed with every 5th (TOP), 10th (CENTRE) or 20th (BOTTOM) shot. The increased shot spacing results in aliased artifacts in the gradient.
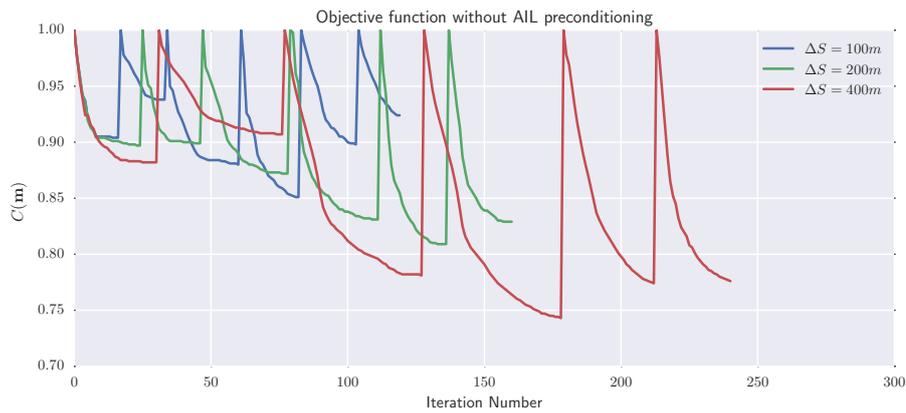
Figure 4.27: The evolution of the objective function for different levels of shot decimation

left to be aligned with the Cartesian directions (the dip appears greater in the velocity models due to vertical exaggeration). The preconditioned gradients shown in Figure 4.28

have decreased aliased artifacts. It is important to note that the wavenumber spectrum for gradient show similar vertical wavenumber (K_z) content while we have filtered some more horizontal wavenumber K_x information in the stronger aliased cases. The final inversion results (Figure 4.29) show very similar vertical wavenumber content to each other. The example with the most densely acquired shots has the most significant horizontal wavenumber attenuation. However as we have performed filtering parallel with the expected geological dip (approximately zero), it is arguable that there has not been significant attenuation of geological horizontal wavenumber information. It is important that the low velocity anomaly at the Chrysaor structure is still identifiable. The objective function (Figure 4.30) shows that a similar number of iterations is required for each of the 3 shot densities. As the cost of gradient computation and line search decreased linearly with the factor of shot decimation, this means that the preconditioned shot decimation results are able to provide good FWI performance, with a fraction of the computational cost.

4.3 Conclusions

In this chapter we have detailed the FWI workflow performed on the 2D NWA-006 Broadseis line taken across the North Western Continental shelf margin. The application of FWI improved the kinematics of the velocity model and improved alignment could be identified on common angle gathers computed using RTM. Addition there was a slightly improved velocity match at the wells. The dataset also provided an illustration of the importance of the initial model for FWI and the need for diving waves to correct the overburden velocity model. As the Broadseis data was of strong signal to noise we made the inversion problem more ill-posed by perform shot decimation. The decimation of shots can decrease the computation cost of each iteration of FWI. Using the AIL framework we were able to precondition the gradient by smoothing along the geological structure to remove high-wavenumber artifacts while still maintaining the key geological structure and velocity information. Although a zero-dip assumption was made for this case-study, due to the relatively flat geology, our AIL operators can handle spatially variant dip as shown in the Chapter 3.
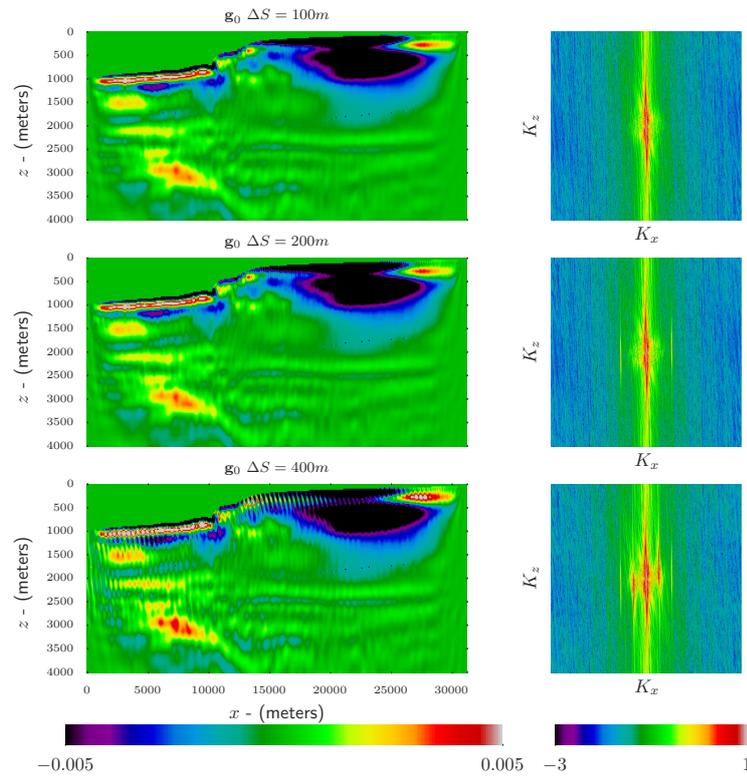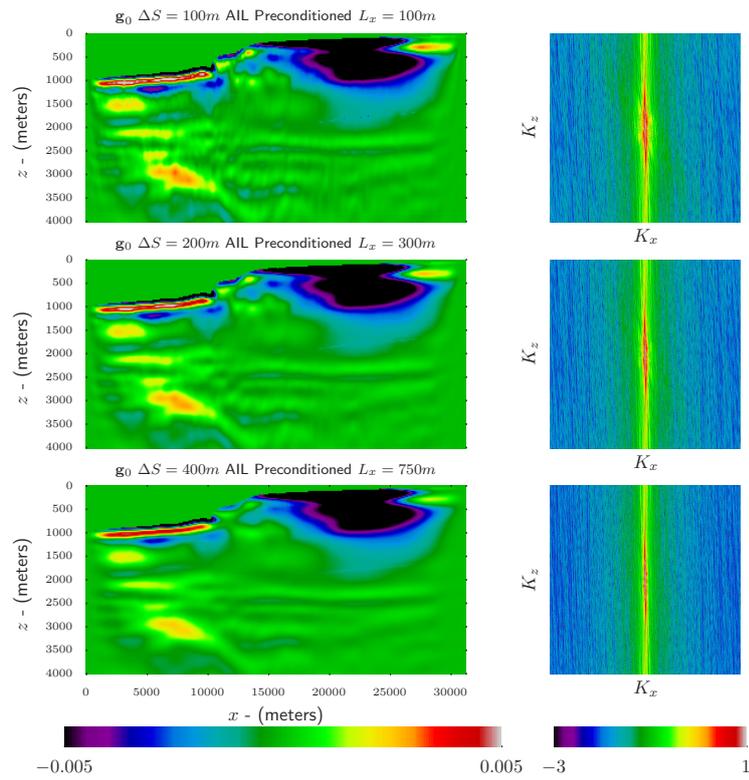
Figure 4.28: The gradient at the first iteration of FWI when it is performed with every 5th (TOP), 10th (CENTRE) or 20th (BOTTOM) shot. The aliasing artifacts related to increased shot spacing are mitigated using horizontal correlation lengths of 100 m, 300 m and 750 m respectively.

Figure 4.29: The final inversion results from FWI when it is performed with every 5th (TOP), 10th (CENTRE) or 20th (BOTTOM) shot. The aliasing artifacts related to increased shot spacing are mitigated using horizontal correlation lengths of 100 m, 300 m and 750 m respectively.
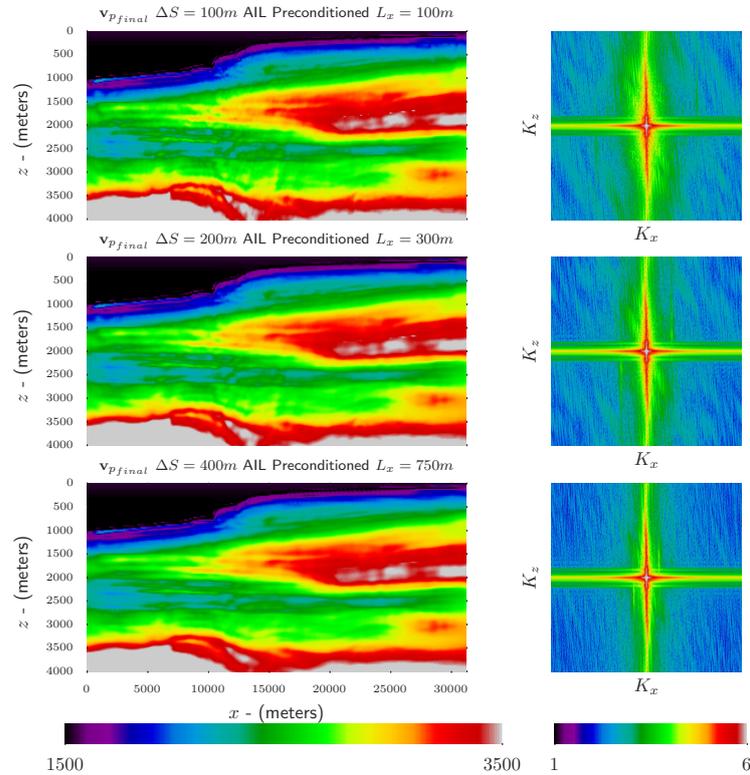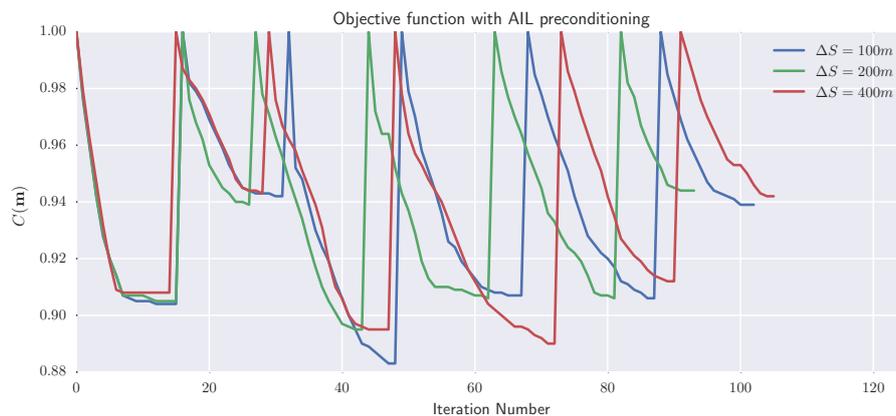

Figure 4.30: The evolution of the objective function for different levels of shot decimation when AIL preconditioning is applied

Conclusions and perspectives

In this dissertation, I have introduced a new, flexible and computationally efficient wavenumber filtering operator based on the extension of the 1D inverse Laplacian correlation function (Tarantola, 2005). We design a framework for the application of the inverse operator for an efficient calculation of the product of correlation operator times a vector. Such a computation is often quite intensive when there are many model parameters and when it is performed in higher dimensions. A generalization of the extension of the operator to higher dimensions comes from the addition of orthogonal, 1D inverse Laplacian correlation operators (i.e. 2 for 2D and 3 for 3D) for which the appropriate normalization was found to ensure that the norm of the vector before and after application of the correlation operator is preserved. The normalized inverse operator in 2D and 3D is referred to as the additive inverse Laplacian (AIL).

Typically in full-waveform inversion, the inversion starts from an accurate, low-wavenumber, initial model inorder to avoid cycle skipping phenomena. The addition of intermediate to high wavenumber is achieved via the FWI engine as intermediate to high wavenumbers are added orthogonal to the orientation of geological dip. The ill-posed nature of the FWI problem, unfortunately means that other wavenumbers can also map into the reconstructed model parameters. These features will not have a geological origin. Two of the key examples of such features are ambient noise and spatial aliasing artifacts involved in the FWI imaging condition. To attempt to mitigate the influence that these artifacts have on the final model, we solve the linear system involving our AIL operator and the FWI gradient at each iteration. Using orthogonal correlation lengths, that are longer along the direction of geological dip versus orthogonal to it, we can mitigate against the ill-posed inversion artifacts, at each inversion iteration. This can be achieved, while still preserving the desired intermediate/high wavenumber geological signal.

The solution of the linear system involved in our AIL operator has been compared to the analytical, anisotropic 2D and 3D Laplacian operators. It is shown that the solution of 2 linear systems well approximates the anisotropic 2D Laplacian operator while 3 linear system solutions is required in the 3D case. The important ability to locally align correlation lengths on a rotated coordinate system is facilitated in the discretization step of our AIL operator. In 2D the AIL operator that includes dip required 9 points as opposed to 5 for the cartesian stencil, while in 3D the stencil required increases to 19 points from 7. The implementation of this dip is of key importance as it allows us to filter along the geological dip, to minimize the attenuation of desirable features, while maximizing the attenuation of undesirable artifacts coming from the ill-posed nature of the inversion. The solution of our linear system is rapid with the application of the AIL filter taking a similar amount of time to apply as the multi-dimensional tensorized filtering approach. One of the key advantages of the AIL filter is that it can handle dip, where such tensorized approaches cannot.

The potential role the AIL stencil has for FWI is investigated in the 3 FWI case studies (2 synthetic and 1 real data). The first case is performed using the Valhall synthetic example. In this case the application of an AIL based low-pass filter is compared to additive Tikhonov regularization. The goal is to mitigate the influence of white noise in the observed data on the final FWI result. It is shown that, although all methods can help to mitigate against the ill-posed nature, the AIL operator is the simplest to be parameterized. It was also shown that the AIL operator could mitigate against spatial aliasing artifacts that arise from the coarse frequency increment used in the inversion.

The Marmousi example showed how spatially variable dip could be used to align the filtering along the geological strata. In the Marmousi case, the dip varied from $0 \rightarrow 45°$ and the alignment of the filtering along geological dip allowed the best compromise between attenuation of ill-posed artifacts in the gradient and preservation of the thin stratigraphic information. The efficiency of the AIL stencil allowed this local dip-filtering otherwise it would not have been computationally feasible without significantly increasing the cost of the FWI workflow.

The final example considered was a real data case study taken from the North Western Continental shelf margin (NWA-006 2D line). Access to this dataset was provided by CGG. Unlike the synthetic dataset, a high quality initial model was not guaranteed and

I focused on the construction of an initial model from before performing FWI without AIL preconditioning. The application of FWI on this line improved the match between well data and also the common angle gather alignment computed from "pseudo-RTM". The importance of the diving waves on FWI in this case were highlighted by looking at two initial starting models. Both starting models converged to the same solution in the shallow diving wave section. However, in the deeper section, the inversion result appeared significantly more non-linear and became subject to cycle-skipping problems. It appears that resolution of the velocity model in the shallow section did significantly improve the imaging results, both in the shallow and the target gas reservoir sections. It was noted, however, that the best result came from the most accurate initial velocity model. This initial model was built in collaboration with CGG using 2D anisotropic reflection tomography.

The inversion results on the NWA-006 data-set were relatively stable in the initial FWI configuration and there was limited imprint from the ill-posed nature on the inversion result. To increase the presence of ill-posed artifacts, shot decimation was performed and the AIL operator was used to attempt to decrease the spatially aliased footprint that resulted within the gradient at each iteration. It was possible to preserve the most important, vertically orientated intermediate/high wave-number information added into the final inversion even with a shot decimation of 400%. It was also noted that approximately the same number of iterations were required to reach convergence as in the non-decimated shot case. This similar number of iterations meant there was a decrease in computational cost approximately linearly correlated with the amount of shot decimation applied.

I suggest the following future work to complement the findings of this thesis. Firstly, the application of the AIL operator to a 3D FWI case study. The AIL operator has been shown to be computationally efficient in 3D, while successful 2D AIL preconditioned FWI results suggest that there should be utility in applying the preconditioner in 3D FWI. The operator is most likely to be useful to either, limit the number of shots required in 3D Time domain FWI, or, alternatively to decrease the frequency increment required in 3D Frequency domain FWI. The ability to decrease either of these parameters, will have a significant influence on the total computation time required for FWI. If relatively robust inversion results are still obtainable after decimation this could be of significant value to decrease the computational cost of 3D FWI.

A second application that has not been discussed is the role of the inverse Laplacian operator times a vector. As we have been primarily focused with the use of low-pass wavenumber filtering potential of solving the linear system I have not investigated the importance of this non-diagonal inverse correlation operator. Prior based, damped-least squares inversion, such as the work performed by Asnaashari et al. (2013) typically use a diagonal approximation of the inverse covariance matrix. The AIL operator represents a low-cost, non-diagonal example of such an operator and may be of potential interest.

Bibliography

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover publications, Inc., New York.

Alkhalifah, T. (2015). Scattering-angle based filtering of the waveform inversion gradients. *Geophysical Journal International*, 200(1):363–373.

Alkhalifah, T. and Choi, Y. (2012). Taming waveform inversion non-linearity through phase unwrapping of the model and objective functions. *Geophysical Journal International*, 191:1171–1178.

Amestoy, P. R., Guermouche, A., L'Excellent, J. Y., and Pralet, S. (2006). Hybrid scheduling for the parallel solution of linear systems. *Parallel Computing*, 32:136–156.

Amundsen, L. (1991). Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion. *Geophysics*, 56:2027–2035.

Askan, A. and Bielak, J. (2008). Full anelastic waveform tomography including model uncertainty. *Bulletin of the Seismological Society of America*, 98(6):2975–2989.

Asnaashari, A. (2013). *Quantitative 4D seismic imaging in complex media using 2D full waveform inversion.* PhD thesis, ISTerre, Université Joseph Fourier.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J. (2013). Regularized seismic full waveform inversion with prior model information. *Geophysics*, 78(2):R25–R36.

Avseth, P., Mukerji, T., and Mavko, G. (2005). *Quantitative Seismic Interpretation: Applying Rock Physics Tools to Reduce Interpretation Risk.* Cambridge University Press, Cambridge, UK.

Barber, P. M. (1988). The exmouth plateau deep water frontier: a case history. the north west shelf, australia. In Purcell, P. and Purcell, R., editors, *The North West Shelf,*

Australia. Proceedings of Petroleum Exploration Society of Australia Symposium, pages 173–187.

Barkved, O., Heavey, P., Kjelstadli, R., Kleppan, T., and Kristiansen, T. G. (2003). Valhall field - still on plateau after 20 years of production. In *SPE Offshore Europe Conference*.

Barkved, O., Heavey, P., Kommedal, J. H., van Gestel, J.-P., ve, R. S., Pettersen, H., Kent, C., and Albertin, U. (2010). Business impact of full waveform inversion at valhall. *SEG Technical Program Expanded Abstracts*, 29(1):925–929.

Barkved, O., Mueller, M., and Thomsen, L. (1999). Vector interpretation of the Valhall 3D/4C OBS dataset. In *Extended Abstracts, 61th Annual EAGE Conference & Exhibition, Helsinki, Finland*.

Baysal, E., Kosloff, D., and Sherwood, J. (1983). Reverse time migration. *Geophysics*, 48:1514–1524.

Ben Hadj Ali, H., Operto, S., and Virieux, J. (2008). Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data. *Geophysics*, 73(5):VE101–VE117.

Berenger, J. P. (1998). A perfectly mathed layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200.

Berkhout, A. J. (1980). *Seismic migration-Imaging of acoustic energy by wave field extrapolation*. Elsevier Science Publ. Co., Inc.

Berryhill, J. R. (1979). Wave equation datuming. *Geophysics*, 44:1329–1333.

Beydoun, W. B. and Tarantola, A. (1988). First Born and Rytov approximation: Modeling and inversion conditions in a canonical example. *Journal of the Acoustical Society of America*, 83:1045–1055.

Biondi, B. and Symes, W. (2004). Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging. *Geophysics*, 69(5):1283–1298.

Bonnans, J. F., Gilbert, J. C., Lemaréchal, C., and Sagastizábal, C. A. (2006). *Numerical Optimization, Theoretical and Practical Aspects*. Springer series, Universitext.

Born, M. (1926). Quantenmechanik der stossvorgänge. *Zeitschrift für Physik*, 38:803–827.

Bourgeois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P., and Versteeg, R. (1991). Marmousi, model and data. In *The Marmousi Experience*, pages 5–16. Eur. Ass. Expl. Geophys.

Brenders, A. J. and Pratt, R. G. (2007). Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International*, 168:133–151.

Brossier, R., Métivier, L., Operto, S., Ribodetti, A., and Virieux, J. (2013). Vti acoustic equations: a first-order symmetric pde: Technical report 50 - seiscope project. Technical report, ISTerre, Université Joseph Fourier.

Brossier, R., Operto, S., and Virieux, J. (2009a). Robust frequency-domain full-waveform inversion using the l_1 norm. *Geophysical Research Letters*, 36:L20310.

Brossier, R., Operto, S., and Virieux, J. (2009b). Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC105–WCC118.

Brown, A. R. (2004). *Interpretation of Three-dimensional Seismic Data*. American Association of Petroleum Geolologist and the Society of Exploration Geophysicists.

Bube, K. P. and Nemeth, T. (2007). Fast line searches for the robust solution of linear systems in the hybrid l_1/l_2 and huber norms. *Geophysics*, 72(2):A13–A17.

Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.

Byrd, R. H., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208.

Castellanos, C., Métivier, L., Operto, S., Brossier, R., and Virieux, J. (2015). Fast full waveform inversion with source encoding and second-order optimization methods. *Geophysical Journal International*, 200(2):720–744.

Chavent, G. (1974). Identification of parameter distributed systems. In Goodson, R. and Polis, M., editors, *Identification of function parameters in partial differential equations*, pages 31–48. American Society of Mechanical Engineers, New York.

Chavent, G. (2009). *Nonlinear least squares for inverse problems.* Springer Dordrecht Heidelberg London New York.

Chavent, G. and Jacewitz, C. A. (1995). Determination of background velocities by multiple migration fitting. *Geophysics*, 60(2):476–490.

Claerbout, J. (1971). Towards a unified theory of reflector mapping. *Geophysics*, 36:467–481.

Claerbout, J. (1985). *Imaging the Earth's interior.* Blackwell Scientific Publication.

Claerbout, J. and Doherty, S. (1972). Downward continuation of moveout corrected seismograms. *Geophysics*, 37:741–768.

Claerbout, J. F. (1976). *Fundamentals of Geophysical Data Processing.* McGraw-Hill Book Co.

Claerbout, J. F. (1992). *Earth souding analysis.* Blackwell Scientific Publications.

Clapp, R., Biondi, B., and Claerbout, J. (2004). Incorporating geologic information into reflection tomography. *Geophysics*, 69:533–546.

Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic non-linear waveform inversion: application to real data. *Geophysics*, 55:527–538.

Dai, Y. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, 10:177–182.

Debens, H. A., Warner, M., Umpleby, A., and da Silva, N. V. (2015). Global anisotropic 3d fwi. In *SEG Technical Program Expanded Abstracts*.

Dix, C. H. (1955). Seismic velocities from surface measurements. *Geophysics*, 20:68–86.

Docherty, P., Silva, R., Singh, S., Song, Z.-M., and Wood, M. (2003). Migration velocity analysis using a genetic algorithm. *Geophysical Prospecting*, 45:865–878.

Duff, I. S. and Reid, J. K. (1983). The multifrontal solution of indefinite sparse symmetric linear systems. *ACM Transactions on Mathematical Software*, 9:302–325.

Fomel, S. (2002). Applications of plane-wave destruction filters. *Geophysics*, 67(6):1946–1960.

Fomel, S. and Claerbout, J. F. (2003). Mutlidimensional recursive filter preconditioning in geophysical estimation problems. *Geophysics*, 68(2):1–12.

Fornberg, B. (1988). Generation of finite difference formulas on arbitrarily spaced grids. *Mathematics of Computation*, 51:699–706.

Frieman, J., Gould, S., and Koller, D. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9:432–441.

Gao, Z., Gao, J., Pan, Z., and Zhang, X. (2014). Building and initial model for fwi using a global optimization scheme. In *SEG Technical Program Expanded Abstracts*.

Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *Geophysics*, 51(7):1387–1403.

Gazdag, J. (1978). Wave equation migration with the phase-shift method. *Geophysics*, 43:1342–1351.

Glennie, K. (1998). *Petroleum Geology of the North Sea: Basic Concepts and Recent Advances.* Blackwell Science Ltd.

Goldstein, H. (1980). *Classical Mechanics.* Addison-Wesley Publishing Company. Inc.

Graves, R. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society of America*, 86:1091–1106.

Guillot, D., Rajaratnam, B., Rolfs, B., Maleki, A., and Wong, I. (2012). Iterative thresholding algorithm for sparse inverse covariance estimation. *Advances in Neural Information Processing Systems*, 25.

Guitton, A. (2012). Blocky regularization schemes for full waveform inversion. *Geophysical Prospecting*, 60(5):870–884.

Guitton, A., Ayeni, G., and Dìaz, E. (2012). Constrained full-waveform inversion by model reparameterization. *Geophysics*, 77(2):R117–R127.

Guitton, A. and Symes, W. W. (2003). Robust inversion of seismic data using the Huber norm. *Geophysics*, 68(4):1310–1319.

Ha, T., Chung, W., and Shin, C. (2009). Waveform inversion using a back-propagation algorithm and a Huber function norm. *Geophysics*, 74(3):R15–R24.

Hadamard, J. (1902). Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52.

Haining, R. P. (2009). Spatial autocorrelation and the quantiative revolution. *Geographic Analysis*, 41:364–374.

Hale, D. (2007). Local dip filtering with directional laplacians. Technical report, Centre for Wave Phenomena, Colorado School of Mines.

Hale, D. (2013). Dynamic warping of seismic images. *Geophysics*, 78(2):S105–S115.

Hansen, C. (1998). *Rank-defficient and discrete ill-posed problems - Numerical aspects of linear inversion*. Society for Industrial and Applied Mathematics - Mathematical modeling and computation.

Harlan, W. (1995). Regularization by model reparameterization. http:www.billharlan.com/pub/papers/regularization.pdf.

Hestenes, M. R. and Stiefel, E. (1952). Methods of conjugate gradient for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436.

Hicks, G. J. (2002). Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions. *Geophysics*, 67:156–166.

Hocking, R. M. (1988). Regional geology of the northern carnarvon basin. In Purcell, P. and Purcell, R., editors, *The North West Shelf, Australia. Proceedings of Petroleum Exploration Society of Australia Symposium*, pages 97–114.

Hole, J. A. (1992). Nonlinear high-resolution three-dimensional seismic travel time tomography. *Journal of Geophysical Research*, 97:6553–6562.

Holland, J. H. (1975). *Adaption in Natural and Artificial Systems*. University of Michigan Press.

Houbiers, M., Wiarda, E., Mispel, J., Dmitry, N., Vigh, D., Thompson, M., and Hill, D. (2012). 3d full-waveform inversion at mariner - a shallow north sea reservoir. In *Expanded Abstracts*.

Hsieh, C., Sustik, M. A., Dhillon, I. S., and Ravikumar, P. (2011). Sparse inverse covariance matrix estimation using quadratic approximation. *Advances in Neural Information Processing Systems*, 24:1–18.

Hustedt, B., Operto, S., and Virieux, J. (2004). Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling. *Geophysical Journal International*, 157:1269–1296.

Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., and Trezeguet, D. (1989). Wavelengths of Earth structures that can be resolved from seismic reflection data. *Geophysics*, 54(7):906–910.

Jo, C. H., Shin, C., and Suh, J. H. (1996). An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator. *Geophysics*, 61:529–537.

Kazemeini, H., Fomel, S., and Juhlin, C. (2008). Prestack spectral blueing: A tool for increasing seismic resolution. In *SEG Expanded Abstracts*.

Khalil, A., Sun, J., Zhang, Y., and Poole, G. (2013). Rtm noise attenuation and image enhancement using time-shift gathers. In *SEG Expanded Abstracts*.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

Komatitsch, D. and Martin, R. (2007). An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics*, 72(5):SM155–SM167.

Komatitsch, D. and Vilotte, J. P. (1998). The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America*, 88:368–392.

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Bednar, R. and Weglein, editors, *Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia*, pages 206–220.

Lancaster, S. and Whitcombe, D. (2000). Fast-track 'coloured' inversion. In *SEG Expanded Abstracts*.

Leonard, R. and Munns, J. (1987). *Valhall Field in Geology of Norwegian Oil and Gas Fields*. Graham and Trotman.

Levander, A. R. (1988). Fourth-order finite-difference P-SV seismograms. *Geophysics*, 53(11):1425–1436.

Longley, I., Buessenschuett, C., Clydsdale, L., Cubitt, C. J., Davis, R. C., Johnson, M. K., Marshall, M. N., Murray, A. P., Somerville, R., Spry, T. B., and Thompson, N. B. (2002). The north west shelf of australia - a woodside perspective. In Keep, M. and Moss, S., editors, *The Sedimentary Basins of Western Australia: Proceedings Western Australian Basins Symposium*, volume 3, pages 28–88.

Longley, P., Goodchild, M., Maguire, D., and Rhind, D. (2001). *Geographics Infromation Systems and Science.* John Wiley and Sons.

Lou, Y. and Schuster, G. (1991). Wave-equation traveltime inversion. *Geophysics*, 56:645–653.

Luo, S. and Sava, P. (2011). A deconvolution-based objective function for wave-equation inversion. *SEG Technical Program Expanded Abstracts*, 30(1):2788–2792.

Manuel, C. D., Hampson, G., Chagalov, D., Bevc, D., Xu, T., Lim, B. J., and Vigh, D. (2014). Full waveform inversion on the nw shelf of australia. In *76th EAGE Conference and Exhibition*.

Marfurt, K. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49:533–549.

Martin, G. S., Wiley, R., and Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for Marmousi. *The Leading Edge*, 25(2):156–166.

Mavko, G., Mukerji, T., and Dvorkin, J. (2009). *The Rocks Physics Handbooks, Tools for Seismic Analysis in Porous Media; Second Edition.* Cambridge University Press, Cambridge, UK.

Medwin, H. (1975). Speed of sound in seawater: A simple equation for releastic parameter. *Journal of Acoustic Society of America*, 58:1318–1319.

Menke, W. (1984). *Geophysical Data Analysis: Discrete Inverse Theory.* Academic Press, Inc., Orlando, USA.

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S., and Virieux, J. (2014). Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures. *Geophysical Prospecting*, 62:1353–1375.

Métivier, L. and Brossier, R. (2016). The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. *Geophysics*, 81(2):F11–F25.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2015). Measuring the misfit beteen seismograms using an optimal transport distance: Application to full waveform inverson. *Geophysical Journal International (Submitted)*.

Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013). Full Waveform Inversion and the truncated Newton method. *SIAM Journal On Scientific Computing*, 35(2):B401–B437.

Middleton, G. V. (1973). Johannes walther's law of correlation of facies. *Geological Society of America Bulletin*, 84:979–988.

Mora, P. R. (1987). Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228.

Mora, P. R. (1988). Elastic wavefield inversion of reflection and transmission data. *Geophysics*, 53:750–759.

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, pages 17–23.

MUMPS-team (2011). *MUMPS - MUltifrontal Massively Parallel Solver users' guide - version 4.10.0 (May 10, 2011)*. ENSEEIHT-ENS Lyon, http://www.enseeiht.fr/apo/MUMPS/ or http://graal.ens-lyon.fr/MUMPS.

Nash, S. G. (2000). A survey of truncated Newton methods. *Journal of Computational and Applied Mathematics*, 124:45–59.

Nocedal, J. (1980). Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, 35(151):773–782.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.

Nolet, G. (1987). *Seismic tomography with applications in global seismology and exploration geophysics*. D. Reidel publishing Company.

Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., and Virieux, J. (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the

visco-acoustic vertical transverse isotropic approximation. *Geophysical Journal International*, 202(2):1362–1391.

Operto, S., Virieux, J., Amestoy, P., L'Éxcellent, J.-Y., Giraud, L., and Ben Hadj Ali, H. (2007). 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. *Geophysics*, 72(5):SM195–SM211.

Operto, S., Virieux, J., Dessa, J. X., and Pascal, G. (2006). Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough. *Journal of Geophysical Research*, 111(B09306):doi:10.1029/2005JB003835.

Peacock, H. B. (1942). How can geophysicists best serve? *Geophysics*, 7:259–267.

Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.

Plessix, R. E. (2009). Three-dimensional frequency-domain full-waveform inversion with an iterative solver. *Geophysics*, 74:WCC149–WCC157.

Polyanin, A. D. and Nazaikinskii, V. E. (2002). *Handbook of linear partial differential equations for engineers and scientists*. CRC Press.

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I : theory and verification in a physical scale model. *Geophysics*, 64:888–901.

Pratt, R. G. and Shipp, R. M. (1999). Seismic waveform inversion in the frequency domain, part II: Fault delineation in sediments using crosshole data. *Geophysics*, 64:902–914.

Pratt, R. G. and Worthington, M. H. (1990). Inverse theory applied to multi-source crosshole tomography. Part I: acoustic wave-equation method. *Geophysical Prospecting*, 38:287–310.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Veltering, W. T. (1986). *Numerical recipes : the art of scientific computing*. Cambridge University Press.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O., and Kommedal, J. (2011). On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International*, 187:1495–1515.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., and dell'Aversana, P. (2004). Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical Journal International*, 159:1032–1056.

Rosland, B., Tree, E., Barkved, O., and Kristiansen, P. (1999). Acquisition of 3d/4c obs data at valhall. In *61th Annual Conference*. EAGE.

Sambridge, M. and Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3):1–29.

Sambridge, M. S. (1999a). Geophysical inversion with a neighbourhood algorithm - I. Searching a parameter space. *Geophysical Journal International*, 138:479–494.

Sambridge, M. S. (1999b). Geophysical inversion with a neighbourhood algorithm - II. Appraising the ensemble. *Geophysical Journal International*, 138:727–746.

Sava, P. and Fomel, S. (2006). Time-shift imaging condition in seismic migration. *Geophysics*, 71(6):S209–S217.

Scales, J. A. and Smith, M. L. (1994). *Introductory geophysical inverse theory*. Samizdat press.

Scales, J. A. and Snieder, R. (1997). To bayes or not to bayes. *Geophysics*, 62:1045–1046.

Schneider, W. A. (1978). Integral formulation for migration in two and three dimensions. *Geophysics*, 43:49–76.

Selley, R. (2000). *Applied Sedimentology*. Academic Press.

Sen, M. K. and Stoffa, P. L. (1995). *Global Optimization Methods in Geophysical Inversion*. Elsevier Science Publishing Co.

Shah, N. K. (2014). *Seismic Full Waveform Inversion for wrapped and unwrapped phase*. PhD thesis, Imperial College London.

Shin, C. and Cha, Y. H. (2008). Waveform inversion in the Laplace domain. *Geophysical Journal International*, 173(3):922–931.

Shin, C. and Cha, Y. H. (2009). Waveform inversion in the Laplace-Fourier domain. *Geophysical Journal International*, 177:1067–1079.

Shin, C. and Ha, W. (2008). A comparison between the behavior of objective functions for waveform inversion in the frequency and laplace domains. *Geophysics*, 73(5):VE119–VE133.

Sirgue, L. (2006). The importance of low frequency and large offset in waveform inversion. In *68th EAGE Conference & Exhibition, Vienna, EAGE*, page A037.

Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, 28:65–70.

Sirgue, L., Denel, B., and Gao, F. (2011). Integrating 3d full waveform inversion into depth imaging projects. In *SEG Expanded Abstracts*.

Smithyman, B. R., Peters, B., and Herrmann, F. J. (2015). Constrained waveform inversion of colocated vsp and surface seismic data. In *EAGE Exapanded Abstracts*.

Stefani, J. and De, G. (2001). On the power-law behaviour of subsurface heterogeneity. In *SEG Expanded Abstracts*.

Stekl, I. and Pratt, R. G. (1998). Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators. *Geophysics*, 63:1779–1794.

Stolt, R. H. (1978). Migration by fourier transform. *Geophysics*, 43:23–48.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Tarantola, A. (1987). *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, New York.

Tarantola, A. (2005). *Inverse Problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics, Philadelphia.

Tarantola, A. (2006). Popper, bayes and the inverse problem. *Nature Physics*, 2:492–494.

Tarantola, A. et al. (1984). The seismic reflection inverse problem. *Inverse problems of acoustic and elastic waves*, pages 104–181.

Thomsen, L. A. (1986). Weak elastic anisotropy. *Geophysics*, 51:1954–1966.

Thomsen, L. A., Barkved, O., Haggard, B., Kommedal, J. H., and Rosland, B. (1997). Converted wave imaging of valhall reservoir. *Presented at the 59th EAGE conference & Exhibition, Geneva, EAGE.*

Tikhonov, A. and Arsenin, V. (1977). *Solution of ill-posed problems.* Winston, Washington, DC.

Tindale, K., Newell, N., Keall, J., and Smith, N. (1998). Structural evolution and charge history of the exmouth sub-basin, northern carnarvon basin, western australia. In Purcell, P. G. and Purcell, R. R., editors, *The Sedimentary Basins of Western Australia 2. Proceedings of Petroleum Exploration Society of Australia Symposium,* volume 2, pages 447–472.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240.

Todd, C. P. and Backus, M. M. (1985). Offset-dependent reflectivity in a structural context. In *SEG Expanded Abstracts.*

Toksöz, M. N. and Johnston, D. H. (1981). *Geophysics reprint series, No. 2: Seismic wave attenuation.* Society of exploration geophysicists, Tulsa, OK.

Trinh, P. T. (2016). Bessel smoothing filters: Analytical expression and weak forumlation in sem: Technical report 101 - seiscope. Technical report, ISTerre, University Grenoble Alpes.

Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160:195–216.

Verrier, G. and Branco, F. C. (1972). La fosse teritaire et le gisement de quenguela-nord. *Revue de l'Institut Francais du Petrole*, 27:51–72.

Vigh, D. and Starr, E. W. (2008). 3D prestack plane-wave, full waveform inversion. *Geophysics*, 73:VE135–VE144.

Vigh, D., Starr, E. W., and Kapoor, J. (2009). Developing Earth model with full waveform inversion. *The Leading Edge*, 28(4):432–435.

Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite difference method. *Geophysics*, 51:889–901.

Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.

Vogel, C. (2002). *Computational methods for inverse problems.* Society of Industrial and Applied Mathematics, Philadelphia.

Vogel, C. R. and Oman, M. E. (1996). Iterative methods for total variation denoising. *Society for Industrial and Applied Mathematics Journal on Scientific Computing*, 17(1):227–238.

Walther, J. (1894). Einleitung in die geologie als historische wissenschaft. *Lithogenesis der Gegenwart*, 3:535–1055.

Walton, C. (2008). Wa-14r & wa-15r interpretation report: 2008 reprocessed chrysaor east/gorgon 3d. Technical report, Chevron Report submitted to the Australian Government - Department of Mines and Petroleum (DMP).

Wang, H., Singh, S., Audebert, F., and Calandra, H. (2015). Inversion of seismic refraction and reflection data for building long-wavelengths velocity models. *Geophysics*, 80(2):R81–R93.

Warner, M. and Guasch, L. (2014). Adaptative waveform inversion - FWI without cycle skipping - theory. In *76th EAGE Conference and Exhibition 2014*, page We E106 13.

Warner, M., Stekl, I., and Umpleby, A. (2007). Full wavefield seismic tomography - iterative forward modelling in 3D. In *Expanded Abstracts*, page C025. EAGE.

White, R. and Simm, R. (2003). Tutorial: Good practice in well ties. *First Break*, 21:75–83.

Wolfe, P. (1969). Convergence conditions for ascent methods. *SIAM Review*, 11.

Woodward, M. J., Nichols, D., Zdraveva, O., Whitfield, P., and Johns, T. (2008). A decade of tomography. *Geophysics*, 73(5):VE5–VE11.

Yilmaz, Ö. (1987). *Seismic data processing.* Society of Exploration Geophyscists.

Yilmaz, O. (2001). *Seismic data analysis.* Society of Exploration Geophysicists: processing, inversion and interpretation of seismic data.

Zelt, C. and Barton, P. J. (1998). Three-dimensional seismic refraction tomography: a comparison of two methods applied to data from the Faeroe basin. *Journal of Geophysical Research*, 103(B4):7187–7210.

Zhou, W., Brossier, R., Operto, S., and Virieux, J. (2015). Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation. *Geophysical Journal International*, 202(3):1535–1554.

Zoeppritz, K. (1919). Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen, über Erdbebenwellen VIIb. *Nachrichten von der Königlichen Gesellschaft der Wissenscaften zu Göttingen, Mathematisch-Physikalische Klasse*, KI:66–84.

Appendix A

Laplacian correlation function normalization

2D Laplacian correlation function normalization

We need to find the scalar, a that makes

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} a e^{-\frac{1}{L}\sqrt{(x-x')^2+(z-z')^2}}\,dx'dz' = 1, \tag{A.1}$$

true for all values of L. For simplicity we redefine $x = x - x'$ and $z = z - z'$.

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} a e^{-\frac{1}{L_x}\sqrt{x^2+z^2}}\,dxdz = 1, \tag{A.2}$$

and to allow anisotropic operators we define $\alpha = \frac{L_x}{L_z}$. $z' = \alpha z$ and therefore $dz' = \alpha dz$.

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} a e^{-\frac{1}{L_x}\sqrt{x^2+z'^2}}\,dx\frac{dz'}{\alpha} = 1, \tag{A.3}$$

expressing using polar coordiantes

$$\frac{a}{\alpha}\int\limits_{0}^{2\pi} d\theta \int\limits_{0}^{\infty} a e^{-\frac{\rho}{L_x}}\rho d\rho = 1, \tag{A.4}$$

where $\rho = \sqrt{x^2 + z'^2}$.

$$\frac{2\pi a}{\alpha} \int_0^\infty a e^{-\frac{\rho}{L_x}} \rho d\rho = 1. \tag{A.5}$$

We define $u = \rho$, $du = d\rho$, $v = -L_x e^{-\frac{\rho}{L_x}}$ and $dv = e^{-\frac{\rho}{L_x}} d\rho$. Using integration by parts we can show

$$\frac{2\pi a}{\alpha} \left([uv]_0^\infty - \int_0^\infty -L_x e^{-\frac{\rho}{L_x}} d\rho \right) = 1 \tag{A.6}$$

$$\frac{2\pi a}{\alpha} L_x \int_0^\infty e^{-\frac{\rho}{L_x}} d\rho = 1 \tag{A.7}$$

$$\frac{2\pi a L_x}{\alpha} [-L_x e^{\frac{\rho}{L_x}}]_0^\infty = 1 \tag{A.8}$$

$$a = \frac{\alpha}{2\pi L_x^2} \tag{A.9}$$

3D Laplacian correlation function normalization

We need to find the scalar, a that makes

$$\int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty a e^{-\frac{1}{L}\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}} dx'dy'dz' = 1, \tag{A.10}$$

true for all values of L. For simplicity we redefine $x = x - x'$, $y = y - y'$ and $z = z - z'$.

$$\int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty a e^{-\frac{1}{L}\sqrt{x^2+y^2+z^2}} dxdydz = 1, \tag{A.11}$$

To define anisotropic operators we define $\alpha = \frac{L_x}{L_y}$, $\beta = \frac{L_x}{L_z}$ and therefore $y' = \alpha y$, $dy' = \alpha dy$, $z' = \beta z$ and $dz' = \beta dz$.

$$\frac{a}{\alpha\beta} \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-\frac{1}{L_x}\sqrt{x^2+y'^2+z'^2}} dxdy'dz' = 1. \tag{A.12}$$

We use rewrite in spherical coordinates

$$\frac{2\pi a}{\alpha\beta} \int_0^\pi sin\theta d\theta \int_0^\infty r^2 dr e^{-\frac{r}{L_x}} = 1, \tag{A.13}$$

where $r = \sqrt{x^2 + y'^2 + z'^2}$.

Following similar steps as were performed in the 2D case we are able to obtain.

$$\frac{8\pi a L_x^3}{\alpha\beta} = 1 \rightarrow a = \frac{\alpha\beta}{8\pi L_x^3} \tag{A.14}$$

Appendix B

Supplementary Publications

Contents

This appendix contains two supplementary publications, both are conference expanded abstracts. The first was presented at the SEG in 2015 and highlights how horizon information could be useful for defining spatially variant fields of correlation lenghts. A gradient preconditioning operator was used but was based on combining 2 1D tensorized laplacian correlation functions and was unable to handle spatially variant dip. An interesting extension to this work would be to either show an application using the sparse additive inverse laplacian (which would allow inclusion of spatially variant dip) or by replacing the tensorized laplacian approach with a tensorized gaussian approach.

The second abstract details the early efforts of FWI applied to the real NWA-006 Broadseis data example. Careful effort was made to provide an initial FWI result would allow it to be used as part of the scale separation work of Wei Zhou (Zhou et al., 2015). Careful windowing, initial model building and amplitude normalization were applied to focus a velocity only full waveform inversion update only on the diving wave data. The hope was this would allow a quick application of the joint diving and reflected wave inversion workflow presented by Wei, however, unfortunately time pressures made this synergy impossible.

Laplacian based Preconditioning of FWI: Using prior information from seismic reflection data.

Paul Wellington*, Romain Brossier, Stéphane Garambois and Jean Virieux
ISTerre, Univ. Grenoble Alpes, France

SUMMARY

Full waveform inversion (FWI) is an ill-posed data-fitting technique that can benefit significantly from preconditioning and/or regularization. We propose to help constrain the FWI problem by using the readily available, imperfect information derived from seismic migration, namely the approximate position in space of the dominant seismic reflection events. A 2D Laplacian filtering framework is described that allows us to modify the local wavenumber content of the FWI gradient via non-stationary filtering. This workflow is applied to a 2D synthetic dataset to highlight its effectiveness. The application of the filtering adds little to no time to the FWI workflow and should be extended straightforwardly to a 3D application.

INTRODUCTION

Full Waveform Inversion (FWI) is a powerful tool that allows one to converge from an initial low wavenumber model of the subsurface to a more complete and accurate high-wavenumber representation (see Virieux and Operto, 2009, for a review). FWI considers the entire time series to be interpreted based on the two-way wave propagation. The FWI problem, as a non-linear inverse problem, is an ill-posed problem taking into account the fact that the acquisition geometry is generally limited to the near surface. For efficiency when considering least-squares minimization, a local linearized optimization is considered starting from a sufficiently accurate initial model (Tarantola, 1987).

An efficient local optimisation approach performs FWI by minimizing the data cost function, $C_d(\mathbf{m})$ using a Quasi-Newton L-BFGS approach with the data gradient $G_d(\mathbf{m})$ computed efficiently using the adjoint state method (Brossier et al., 2009). Without regularisation, this approach will minimize the difference between the observed $\mathbf{d}_{obs}$ and modeled data $\mathbf{d}_{cal}$; however, due to the nature of wave propogation, the limited frequency content, the presence of noise and/or illumination issues, the image reconstruction may suffer from a large null space.

Effective application of preconditioning and regularization techniques can be powerful in narrowing the null-space contribution. Regularization terms based on velocity model constraints (Asnaashari et al., 2012), regularization weights based on the seismic image (Castellanos, 2014) or model-space preconditioning based on prior knowledge of the local dip field (Guitton et al., 2012) have shown promise in both synthetic and real data examples. The approach we utilize here uses a preconditioning filter on the FWI gradient with the parameters of the filter designed from prior information.

METHODOLOGY

FWI acts as a data-fitting process where the model vector $\mathbf{m}$ is iteratively updated so that the synthetically modelled data $\mathbf{d}_{cal} = \mathbf{d}(\mathbf{m})$ matches the observed field data $\mathbf{d}_{obs}$. The full waveform inversion typically minimizes the misfit function based on the ℓ_2 norm of the differences (1).

$$C(\mathbf{m}) = \frac{1}{2}||\mathbf{d}_{cal} - \mathbf{d}_{obs}||^2. \tag{1}$$

Starting from an initial model $\mathbf{m}_0$, we update a current model $\mathbf{m}_n$ at the iteration n with a pertubation model $\Delta\mathbf{m}_n$ to define the new model $\mathbf{m}_{n+1} = \mathbf{m}_n + \Delta\mathbf{m}_n$. In such a case we need to look at the shape of the misfit function around the current model $\mathbf{m}_n$ (see Virieux and Operto, 2009, for an explanation). The model perturbation is given by

$$\Delta\mathbf{m} = -\underbrace{\left[\frac{\partial^2 C(\mathbf{m}_n)}{\partial\mathbf{m}^2}\right]^{-1}}_{Hessian=\mathbf{H}}\underbrace{\frac{\partial C(\mathbf{m}_n)}{\partial\mathbf{m}}}_{Gradient=\mathbf{G}}. \tag{2}$$

The updated model $\mathbf{m}_{n+1}$ can then be expressed as

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n\mathbf{H}_n^{-1}\mathbf{G}_n. \tag{3}$$

where the step length α_n along the perturbation model vector attempts to speed up the convergence. In our approach, we calculate the gradient using the adjoint state method (Plessix, 2006) and converge to a solution using the quasi-newton LBFGS method (Brossier et al., 2009).

Analytical Laplacian Smoother

A 2D analytical laplacian smoothing filter that smooths a point (x,z) based on contributions away from the point at (x',z') is defined (4) (where L_X and L_Z are correlation lengths in the horizontal and vertical, directions given in meters).

$$S_{2D}(x,z;x'z') = \underbrace{\left[\frac{2}{L_X}\exp^{-\frac{|x-x'|}{L_X}}\right]}_{S_X(x,x')} \times \underbrace{\left[\frac{2}{L_Z}\exp^{-\frac{|z-z'|}{L_Z}}\right]}_{S_Z(z,z')}. \tag{4}$$

If we wish to discretize this filter to provide a smoothing framework for each cell in the model space, we can define $\mathbf{S}_{2D} = \mathbf{S}_X\mathbf{S}_Z$ where $\mathbf{L}_X$ and $\mathbf{L}_Z$ become vectors that allow the correlation length to change for each point in the model space. We propose to perform the FWI workflow using this filter to precondition the gradient $\mathbf{G}' = \mathbf{S}_{2D}\mathbf{G}$. At low values of L, for a given smoothing direction only the very high wavenumber component will be attenuated. As the input values of L increase, the smoother operator will act as a low-pass wavenumber smoother. $\mathbf{L}_X$ and $\mathbf{L}_Z$ can vary smoothly in space and do not have to be equal, allowing a powerful strategy for anisotropic, non-stationary wavenumber filtering.

The operator $\mathbf{S}_{2D}$ could be very large and banded with a similar dimension to the Hessian operator and as such is never explicitly defined. We are interested in the application of $\mathbf{S}_{2D}$ to our

gradient vector $\mathbf{G}$. It is equivalent to perform this operation by splitting $\mathbf{S}_{2D}$ into the horizontal and vertical components of the smoothing operator through the expression

$$\mathbf{S}_{2D}\mathbf{G} = \mathbf{S}_X\left(\mathbf{S}_Z\mathbf{G}\right). \qquad (5)$$

By using this strategy we can then efficiently perform the non-stationary smoothing operation with negligible CPU/memory requirements (Algorithm 1).

Algorithm 1 Efficient application of $\mathbf{S}_{2D}$ to $\mathbf{G}$

1: ▶ **Input:** The FWI Gradient, $\mathbf{G}$ with dimensions $(n \times m)$ correlation length vectors $\mathbf{L}_X$ and $\mathbf{L}_z$ with dimensions $(n \times m)$ Where n is number of cells of the model space in the vertical direction and m is the number of cells in the models space in the horizontal direction.
2: **for** $i = 1$ **to** m **do**
3:    **for** $j = 1$ **to** n **do**
4:       Calculate current cell of temporary array $\tilde{\mathbf{G}}_Z[j,i] = \mathbf{S}_Z\mathbf{G}$ at $[j,i]$ using $L_Z = \mathbf{L}_\mathbf{Z}[j,i]$ for points from $\mathbf{G}[j - 4L_Z, i] \to \mathbf{G}[j + 4L_Z, i]$.
5:    **end for**
6: **end for**
7: Define $\mathbf{G}_Z = a\tilde{\mathbf{G}}_Z$ where $a = \sqrt{\frac{\tilde{\mathbf{G}}_Z^T \tilde{\mathbf{G}}_Z}{\mathbf{G}^T\mathbf{G}}}$
8: **for** $i = 1$ **to** n **do**
9:    **for** $j = 1$ **to** m **do**
10:       Calculate current cell of temporary array $\tilde{\mathbf{G}}_{XZ}[i,j] = \mathbf{S}_X\mathbf{G}_Z$ at $[i,j]$ using $L_X = \mathbf{L}_X[i,j]$ for points from $\mathbf{G}[i, j - 4L_X] \to \mathbf{G}[i, j + 4L_X]$.
11:    **end for**
12: **end for**
13: Define $\mathbf{G}' = \mathbf{G}_{XZ} = b\tilde{\mathbf{G}}_{XZ}$ where $b = \sqrt{\frac{\tilde{\mathbf{G}}_{XZ}^T \tilde{\mathbf{G}}_{XZ}}{\mathbf{G}_Z^T\mathbf{G}_Z}}$

NUMERICAL EXAMPLE: 2D SYNTHETIC VALHALL MODEL

We perform a test of the Laplacian smoother using a synthetic model of the Valhall field (Figure 1). In this model there are low-velocity gas saturated sands in the shallow section that overly an anti-clinal oil reservoir structure that rests just below 2500m depth. The true model is discretized with a 25m grid spacing and the initial model has been derived by applying a 375m gaussian slowness smoother to the true model. After performing this smoothing, the high wavenumber content of the model including the top and base of individual gas sands is no longer evident. Frequency-domain full waveform inversion has been applied to attempt to recover this missing high-wavenumber content.

Our FWI strategy involves using a Bunks frequency sweeping approach (Bunks et al., 1995) where we invert over 3 successive frequency bands ($4Hz \to 6Hz$, then $4Hz \to 8Hz$, then $4Hz \to 10Hz$) with the inversion result from the previous band serving as the input to the next band. The density model is calculated using Gardner's law. A fixed-spread acquisition is simulated with receivers every 50m and shots every 250m. Gaussian white noise is added to the true modelled data in order to increase the ill-posed nature of the inversion.
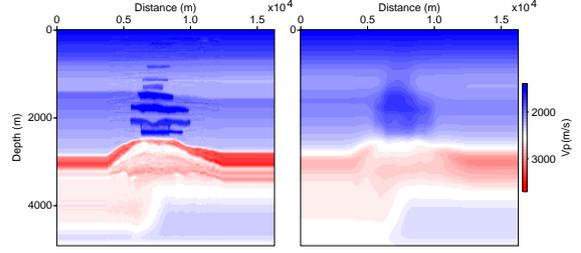

Figure 1: True Vp model(LEFT) and Initial Vp Model (RIGHT).

Prior Correlation Length Vectors

Seismic reflection events were interpreted from an RTM image built using the initial velocity model. The euclidian distance between a given point in the model space and the nearest interpreted horizon pick is used to calculate a transition between a low value of L which is used if the grid cell is at a horizon point (75m) and a higher value used further away (1000m). The transition between these two points is managed using a Gaussian function to allow a slow transition between the two extremes (Figure 2). In the case showed the $\mathbf{L}_Z$ and $\mathbf{L}_X$ models were equivalent except the minimum value of $\mathbf{L}_Z$ was made to be 25m vs 75m for $\mathbf{L}_X$. This was deemed reasonable as much of the layering in this example was horizontal and we wished to preserve the vertical high-wavenumber content of the low velocity sands.

Results

A comparison is made between the FWI inversion results with no preconditioning, preconditioning using a constant value of L ($\mathbf{L}_X = \mathbf{L}_Z = 200m$) and one using the variable L model shown in Figure 2. At the first iteration of the inversion, there are significant differences in the FWI gradient for each case (Figure 3). When no preconditioning is applied, it is possible to see the high wavenumber velocity perturbations that correspond to the gas sands. There is significant noise as we move away from these gas sands. The noise amplitude is constant, as such away from reflectors where the model should be smooth, noise dominates the gradient. The second gradient image has been preconditioned using a constant value of L. In this case, the noise has been attenuated, but so has been the high wavenumber information that corresponds to the gas sands. The optimal gradient comes from when one uses a large value of L away from reflectors and a much smaller one near the reflectors (as shown in the 3rd image).

The final inversion models (Figure 4) show similar characteristics to what is seen in the gradient. When no preconditioning is applied, details of the high wavenumber gas sands are improved but there is also significant noise contaminating the final model. The result obtained using the constant L value is an inversion that is not contaminated by noise, but is without the high-wavenumber detail of the gas sands. The optimal result is obtained using the variable L preconditioner. This is also shown from an extracted vertical trace (Figure 5).
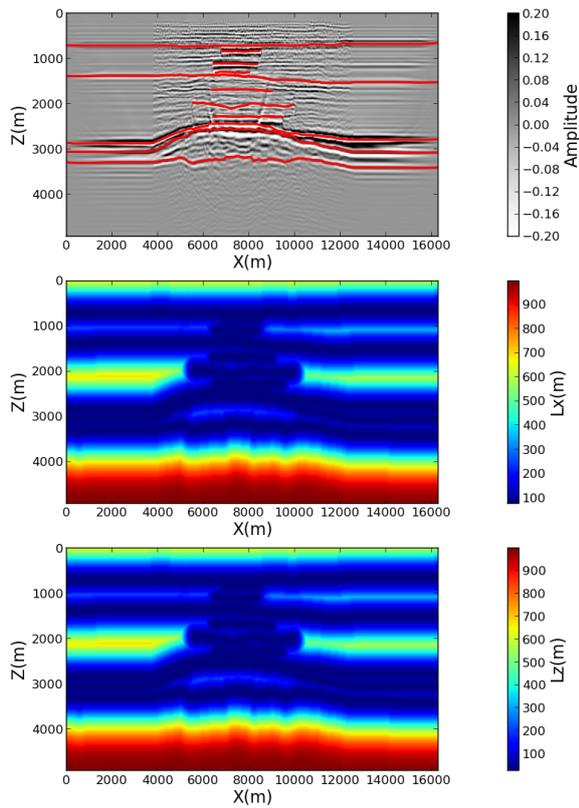
Figure 2: RTM Image with key horizons interpreted in red. (TOP) The estimated $\mathbf{L}_X$ (CENTER) and $\mathbf{L}_Z$ (BOTTOM) models. L were models built from horizon information interpretted on the RTM image.

CONCLUSIONS

We have shown that, by preconditioning the FWI gradient by non-stationary Laplacian filtering, it is possible to decrease the null-space contribution. Our approach adds little to no computation time to the FWI workflow and could be scaled to 3D applications with ease. Future work could involve the automation of the preconditioning workflow, tuning the preconditioner for each individual frequency band and the inclusion of additional discontinuities such as faults.
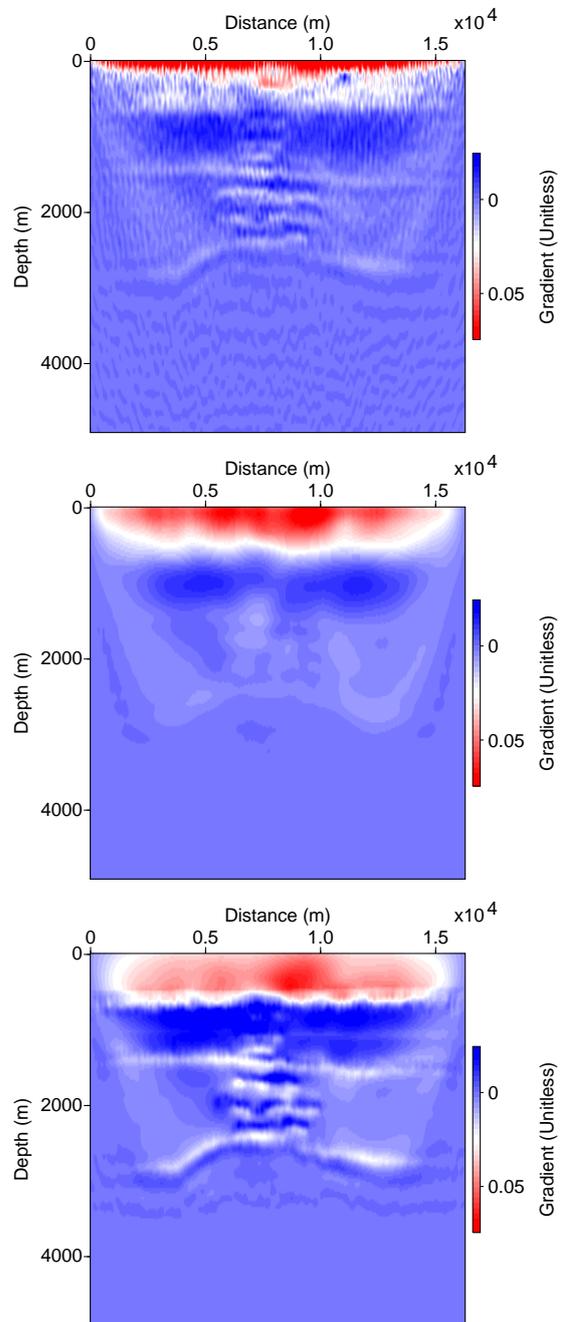
ACKNOWLEDGMENTS

Figure 3: Comparison of FWI Gradients at the first iteration: No Preconditioning(TOP) Constant (200m) L Preconditioning Result (CENTER) Variable L Preconditioning Result (BOTTOM).
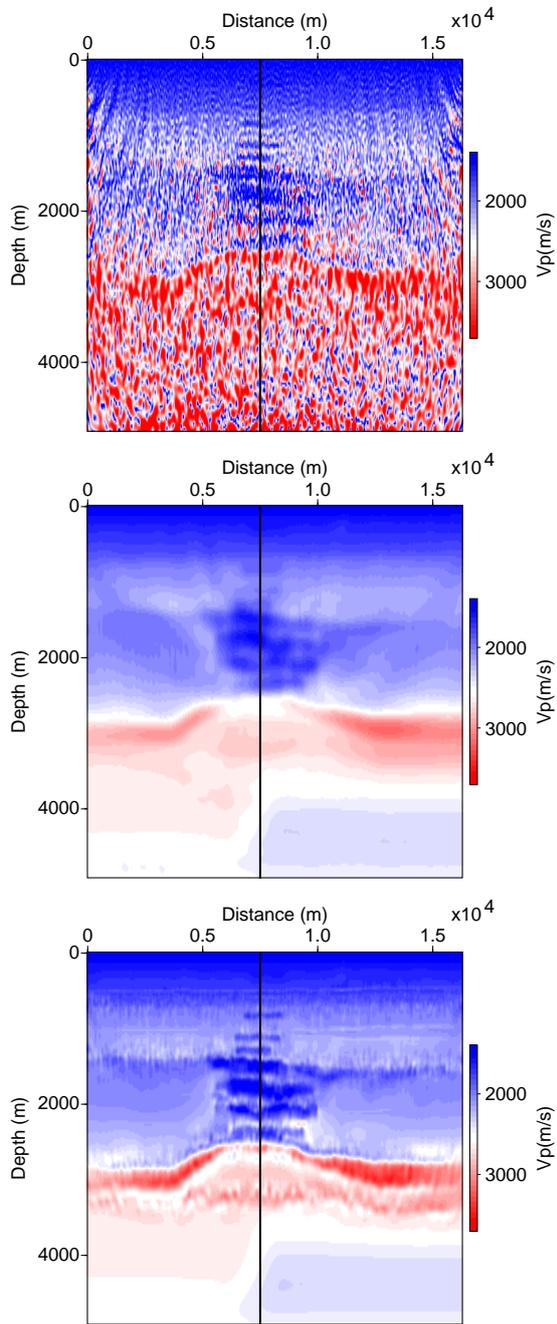
Figure 4: Comparison of Final Vp Inversion Results: No Preconditioning(TOP) Constant (200m) L Preconditioning Result(CENTER) Variable L Preconditioning Result (BOTTOM). The black line represents the extracted trace in Figure 5.
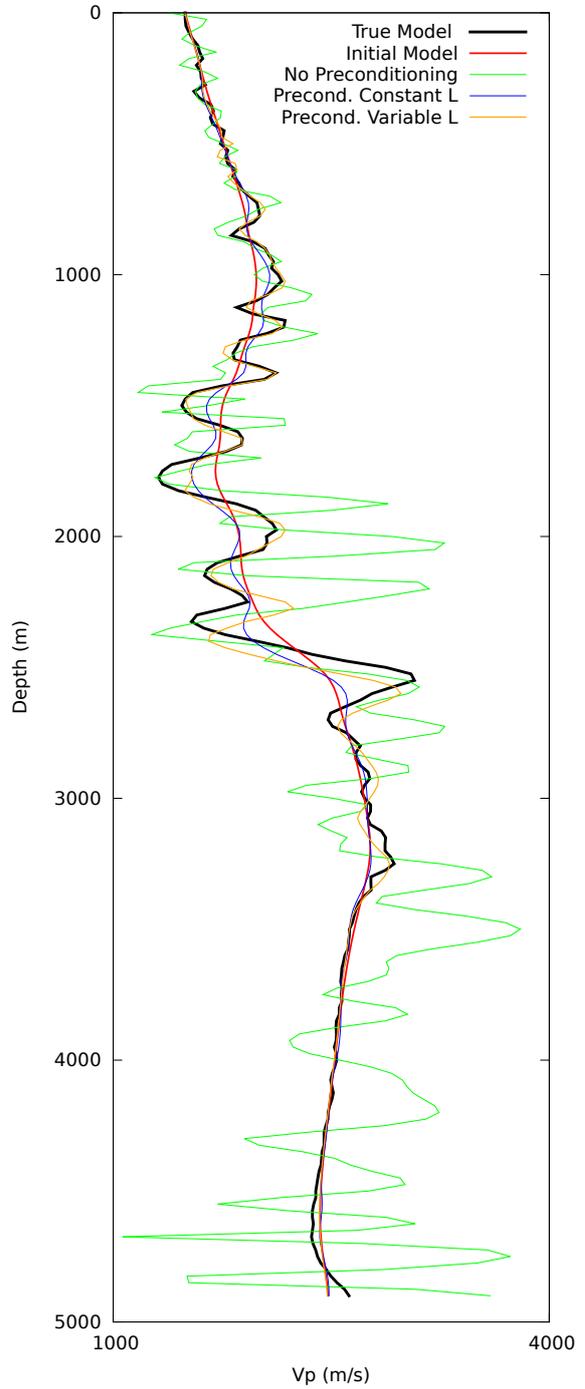

Figure 5: Inversion extracted trace comparison.

2D Acoustic time domain Full Waveform Inversion:
A Broadband application in the Carnarvon Basin, Australia

Paul Wellington, Romain Brossier, Stéphane Garambois, Jean Virieux

Introduction

Full waveform inversion (FWI) is a wavefield data fitting procedure that is increasingly finding acceptance in academia and industry as a means to extract quantitative information from seismograms (see Virieux and Operto, 2009, for a review). FWI considers the entire time series to be interpreted based on two-way wave propagation. The FWI problem, as a non-linear inverse problem, is an ill-posed problem taking into account that the acquisition geometry is generally limited to near the free surface. For efficiency when considering least-squares minimization, a local linearized optimization is considered from an initial model (Tarantola, 1987). In order to avoid cycle-skipping problems which may prevent converging to the global minimum, the initial model should predict travel times of phases in the observed data within half the period (Beydoun and Tarantola, 1988) which could be expressed as a relative time error depending on the number of wavelengths to be propagated.

In order to avoid "cycle-skipping" phenomena prohibiting a successful inversion, a combination of three strategies has been implemented:

1) Use of ray based methods (Migration Velocity Analysis or Reflection Tomography) (Woodward et al., 2008) to improve the accuracy of the low wavenumber content of the initial model. 2) Broadband seismic acquisition to allow the recording of very low frequencies ($\approx 2\,Hz$), reducing the number of wavelengths to be propagated for a given distance between the source and the receiver. 3) Multi-step hierarchical inversion from lower to higher frequencies in order to mitigate the cycle-skipping problem and to avoid local minima (Bunks et al., 1995; Sirgue and Pratt, 2004).

In this work, we present an application of acoustic FWI to a 2D real dataset transecting the North-Western Australian Continental shelf where these three strategies have been applied. In this region, the water depth changes from values lower than 100 m to high values above 1000 m over just few kilometers. This dataset was acquired with a Broadseis (CGG) seismic acquisition strategy. We have constructed an initial model built from kinematically imperfect pre-stack time migration (PreSTM) velocity analysis. We show via Kirchhoff pre-stack depth migration (PreSDM) that, after application of FWI, the velocity model is kinematically more accurate through an analysis of common image gathers (CIG). These improvements in the kinematics occur both in the overburden and at the target level where two previously discovered gas fields are located.

We first consider the construction of the initial model. Then we will apply the FWI workflow, before concluding with inversion quality control based on CIG analysis.

Initial Model construction

Initial models are built for both P-wave velocity V_p and density ρ. We shall invert for a single parameter, (P-wave velocity) but we require a realistic ρ initial model that honors the sharp contrast between the water column and the subsurface. A Gardner law is applied during the inversion procedure. A multi-parameter inversion following the strategy proposed by Zhou et al. (2014) will be considered at a later stage.

V_p Initial Model Building

The V_p initial model is constructed by merging information from two data sources and is displayed in Figure 1. These datasources are: 1) An expendable bathythermographic (XBT) derived velocity function within the water column. 2) Stacking velocities that are the result of a pre-stack time migration (PreSTM) processing workflow.

Public-domain bathythermographic information is available from the Commonwealth Scientific and Industrial Research Organisation (CSIRO). This bathymetric information connected with the knowl-
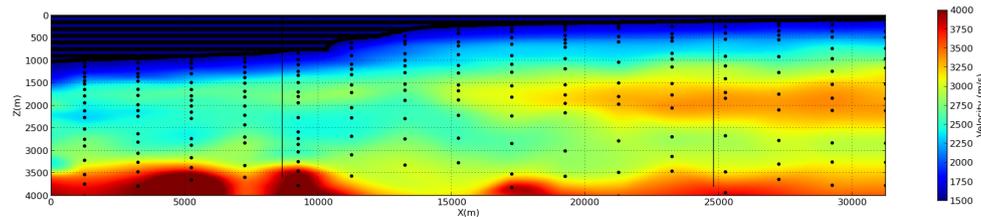
Figure 1 *The initial velocity model. Above the waterbottom interface, the velocity is calculated from XBT information while, below, the seafloor V_{rms} is interpreted every 2 km. The black dots represent position of data points before gridding. The two well paths are marked with vertical black lines.*

edge of the sea water salinity ($\approx 35ppk$) allows an estimation of the velocity in the water column with respect to the depth via an empirical velocity/temperature transform (Medwin, 1975). Stacking velocities obtained from migration velocity analysis performed as part of the PreSTM workflow are converted to interval velocities using the Dix equation. A water-sediment interface is interpretted in time from the PreSTM data after CDP stacking. The velocity that would be predicted by the XBT ($V_{INT}(t) = -48.492t^3 + 155.27t^2 - 165.54t + 1546$ where t is in seconds) velocity column function at the interpreted waterbottom TWT is included to provide a transition between these two zones. Above this interface, the XBT interval velocitiy is used while, below it, the Dix converted MVA RMS velocities are used. All this information is provided on an irregular velocity grid in two-way time. One-dimensional integration of the interval V_{int} pairs converts this irregular two-way-time grid to an irregular depth grid. A regridding step will produce a regular 25 m square gridded velocity model through a cubic polynomial interpolation. A 300 m isotropic gaussian filter in slowness will remove unwanted high frequency variations.

ρ Initial Model Building

Once the initial velocity model is constructed, Kirchhoff PreSDM is performed. The waterbottom depth event is detected and interpreted on this stacked, migrated section. Two wells were intersected by the 2D seismic line (one shallow on the continental shelf above 100 m and another one at a depth greater than 800 m). The depth migration predicted the waterbottom depth of these two wells to within 10m of the true value. Above the waterbottom interface, we consider a constant density of $1000kg/m^3$ while the Gardner law of the expression $\rho = 310V_p^{0.25}$ is used below this interface. The density contrast at this interface is approximately 50% and, as such, it controlled the amplitude of the waterbottom reflections.

Full waveform inversion

Full waveform inversion is applied, focusing on primarily diving wave energy. We consider isotropic wave propagation using a finite difference approach and perform both forward modelling and inversion of the data in the time domain(TDTD FWI). This allows an approach similar to the one proposed by (Bunks et al., 1995), called 'TDTD Bunks FWI'. Initially, an approximate 3D correction is applied to the data (T^2 gain) in addition to F-K filtering and spectral whitening. We consider different second-order butterworth frequency bands starting with a low frequency ($\approx 3\,Hz$) up to higher frequencies ($\approx 15\,Hz$). We proceed through 6 frequency bands by extending the high frequency limit while keeping always the low one fixed. The output of the velocity from the previous frequency band serves as the input to the next band.

Source functions are estimated in the deepwater portion of the line using the direct wave information at near offsets (500-1700m). Once the source information is obtained, synthetic records are computed in the initial model($\mathbf{d_{mod}}_{initial}$). The near-offset phase and amplitude match is quite good suggesting that the position and the impedance contrast of the water bottom is well represented. We notice significant amplitude discrepancies between between $\mathbf{d_{obs}}$ and $\mathbf{d_{mod}}_{initial}$. Therefore, at the start of the FWI for

each frequency band, the ratio of the average amplitude between $\mathbf{d_{obs}}$ and $\mathbf{d_{mod}}_{initial}$ is used to normalize this amplitude discrepancy so that the FWI focused on matching phases of data events rather than amplitude differences. This pragmatic approach is justified as the amplitude difference is influenced by visco-elastic and anisotropic effects that are not yet considered in the forward modeling in our time-domain acoustic finite difference modelling engine. The FWI gradient is computed using the adjoint state method (Plessix, 2006) and the inversion is iterated using a preconditioned steepest descent algorithm. A weak regularization is applied through a gaussian wavenumber filter over the gradient. After each frequency band attained convergence, the inverted velocity model is used as the starting model for the next iteration, while the density model is recomputed using the following expression $\rho = 310 V_p^{0.25}$ (Gardner equation) to ensure that features added to the initial model are present in the updated density model as shown in the following algorithm.

Algorithm 1 Algorithm for FDTD Bunks FWI for $\mathbf{V_{p}}_{(i)}$ from frequency band $i \rightarrow N_{band}$

1: **for** i to N_{band} **do**
2:     Bandpass filter preprocessed Data using current band to calculate $\mathbf{d_{obs}}$.
3:     **if** i=1 (First Band) **then**
4:         Use initial models $\mathbf{V_p} = \mathbf{V_p}_{initial}$ and $\boldsymbol{\rho} = \boldsymbol{\rho}_{initial}$
5:     **end if**
6:     **if** $i \neq 1$ **then**
7:         Set $\mathbf{V_p} = \mathbf{V_p}_{i-1}$
        Set $\boldsymbol{\rho}(AboveWB) = 1000 kg/m^3$ & $\boldsymbol{\rho}(BelowWB) = 310 \mathbf{V_p}^{0.25}$
8:     **end if**
9:     Estimate source from watercolumn direct wavefield
10:     Forward model using $\mathbf{V_p}$ and $\boldsymbol{\rho} \rightarrow \mathbf{d_{mod}}_{initial}$
11:     Calculate average amplitude ratio ($\boldsymbol{\alpha}$) for each trace $\mathbf{d_{obs}}$ vs $\mathbf{d_{mod}}_{initial}$
12:     **while** FWI above convergence criteria **do**
13:         Perform FWI minimizing on current band $C = \|\boldsymbol{\alpha}\mathbf{d_{obs}} - \mathbf{d_{mod}}\|_2$
14:     **end while**
15:     $\mathbf{V_p}_{final}$ from FWI $\rightarrow \mathbf{V_p}_i$
16: **end for**

Results

The full waveform inversion appears to behave in a quite stable manner. The primary changes to the velocity model are a sharpening of the waterbottom and also a sharpening of the top of a high velocity zone that becomes visible below 1km depth (Figure 2). Gather horizontal alignment on CIGs obtained from Kirchhoff PreSDM is significantly improved across almost the entire line (Figure 3). This is true in the overburden as well as at the depth of previously known gas reservoir sands.

Conclusions

A successful 2D real data application of isotropic acoustic TDTD full waveform inversion has been presented. From a crude initial model built from standard time processing, we have been able to converge to a stable velocity model that improves CIG flatness in the overburden and reservoir target level.
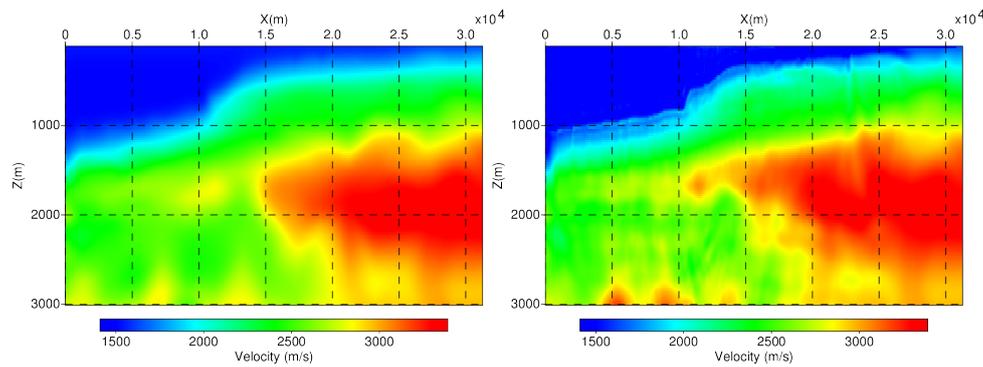
Acknowledgements

Figure 2 *The initial velocity model is shown in the left panel wil the final velocity model is shown on the right panel. Same color scale is used in both figures*
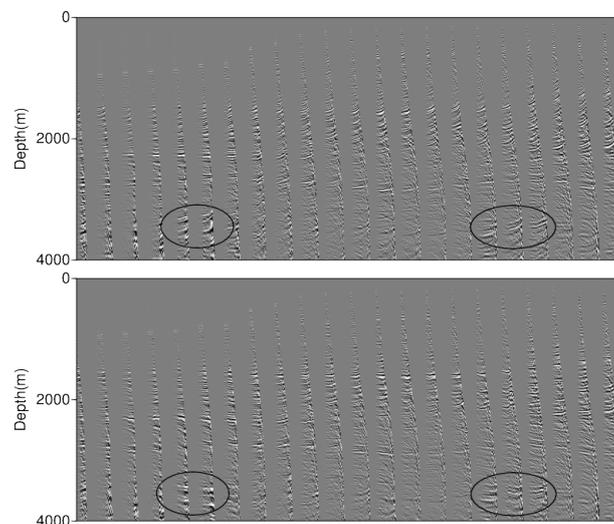

Figure 3 *Common image gathers formed from Kirchhoff PreSDM using the velocity model before (top panel) and after FWI (bottom panel). Note the full waveform inversion improves horizontal alignment of the events in the overburden and at the top reservoir level of the two gas fields (shown in black circles).*

References

Beydoun, W.B. and Tarantola, A. [1988] First Born and Rytov approximation: Modeling and inversion conditions in a canonical example. *Journal of the Acoustical Society of America*, **83**, 1045–1055.

Bunks, C., Salek, F.M., Zaleski, S. and Chavent, G. [1995] Multiscale seismic waveform inversion. *Geophysics*, **60**(5), 1457–1473.

Medwin, H. [1975] Speed of sound in seawater: A simple equation for releastic parameter. *Journal of Acoustic Society of America*, **58**, 1318–1319.

Plessix, R.E. [2006] A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, **167**(2), 495–503.

Sirgue, L. and Pratt, R.G. [2004] Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, **69**(1), 231–248.

Tarantola, A. [1987] *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, New York.

Virieux, J. and Operto, S. [2009] An overview of full waveform inversion in exploration geophysics. *Geophysics*, **74**(6), WCC1–WCC26.

Woodward, M.J., Nichols, D., Zdraveva, O., Whitfield, P. and Johns, T. [2008] A decade of tomography. *Geophysics*, **73(5)**, VE5–VE11.

Zhou, W., Brossier, R., Operto, S. and Virieux, J. [2014] Acoustic multiparameter full-waveform inversion of diving and reflected waves through a hierarchical scheme. *Expanded Abstracts, 84th Annual SEG Meeting (Denver)*.