2D Full Waveform Inversion in time domain for Ground Penetrating Radar Data

Jheyston Omar Serrano Luna

This dessertation is submited for the degree of

Doctor en ingeniería

Advisor:

Ph.D Ana Beatriz Ramirez Silva

Co-advisor:

Dr-Ing. Sergio Alberto Abreo Carrillo

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2021

Dedicatoria

En primer lugar, quiero agradecer a Dios por darme la oportunidad de alcanzar este logro. A mi esposa Katherine, por todo su apoyo incondicional y motivación durante cada etapa en el desarrollo de este proyecto de investigación. A mis padres Jaime y Elizabeth que siempre estuvieron pendientes de cada etapa del proyecto. A todo el grupo de investigación CPS, en particular a mis directores: profesora Ana y profesor Sergio; que siempre estuvieron orientando, discutiendo y guiando este proceso de formación. Gracias a los revisores de esta tesis, todos sus comentarios, sugerencias y recomendaciones fueron valiosas para la tesis. A mis amigos Iván, Anderson y Carlos Niño por las salidas a tomar tinto donde siempre buenas discusiones de investigación salieron. A doña Janet, mi tío Jaime y mi madrina Esperanza. A mis hermanos Firson y Jaime Enrigue, mis amigos Carlos Muriel y Ricardo Herrera. Finalmente, y no menos importante, a mis abuelos Saul, Edelmira, Transito y Horacio.

Acknowledgments

This research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-17-1-0530. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. The authors gratefully acknowledge the support of CPS research group of Industrial University of Santander, Geophysics laboratory of the Physics School and Ecopetrol S.A.

Contents

List of Figures

List of Tables

Glossary

Notation	Description	Notation	Description
CMP	Common midpoint	D_z	Derivative in z-direction
CPML	Convolutional Perfectly Matched Layer	du	Initial condition for $\dot{s}$
DC	the mean value of a signal	$\mathbf{d}_{mod}$	Modeled data
FFT	Fast fourier transform	$\mathbf{d}_{obs}$	Observed data
FWI	Full Waveform Inversion	$E[\cdot]$	Expected value
GPR	Ground Penetrating Radar	$\mathbf{E}_i$	Electrical field in i-direction
GPU	Graphics Processing Units	$\widetilde{\mathbf{E}_y}$	Electric field in y-direction with radiation pattern
GS	Gaussian Smoothing	E_r	Electric field in r position
IFFT	Inverse fourier transform	E_θ	Electric field in θ position
L-BFGS	Limited-memory Broyden-Fletcher-Goldfarb-Shanno	E_ϕ	Electric field in ϕ position
MPI	Message Passing Interface	f_c	Central frequency
MTV	Modified Total Variation	$g(...)$	Gradient
PSNR	Peak signal to noise ration	$\mathbf{g}_l^k$	the best global position in PSO
PSO	Particle Swarm Optimization	$\widetilde{\mathbf{H}_z}$	Magnetic field in z-direction with radiation pattern

Notation	**Description**		
$\mathbf{H}_i$	Magnetic field intensity in i-direction	σ_0	Regularization value of conductivity
$\mathbf{J}_s$	Current density	Δt_{air}	The air duration pulse
k_i	Wavenumber in i-direction	Δ_r	Longitudinal resolution
ℓ	Number of gradients and models in L-BFGS	Δ_l	Lateral resolution
Msi	Magnetic current densities in i-direction	μ_0	Vacuum permeability
$\mathbf{m}^k$	Unknown parameters	λ_{MTV}	Weight regularization in MTV
Offset	Distance between T_x and R_x	Ξ_H	Radiation plane H
$\mathbf{p}^k$	the best local position in PSO	Ξ_E	Radiation plane E
P_{CPML}	Total points in CPML	$\varepsilon_r(hdp)$	Permittivity of the coating
$\mathbf{r}$	Vector position $\in \mathbb{R}^2$	$\varepsilon_r(PCB)$	Permittivity of the PCB
R_{xobs}	Observed data with radiation pattern	$\sigma r(abs)$	Conductivity of the absorbent barrier
R_{xmod}	Moleded data with radiation pattern	$\varepsilon_r(abs)$	Permittivity of the absorbent barrier
R_{corr}	Correlation coefficient	$\mathfrak{r}$	Distance to the target
R_x	Receiver position	ε	Permittivity
r_{opt}	Matrix to uptate $\mathbf{m}^k$ with L-BFGS	σ	Conductivity
s	Discretized fields	β	Scale conductivity parameter
T_x	Transmitter position	ρ	standard deviation
t	Time	μ	Permeability
u	Initial conditions for s	ε_0	Vacuum permittivity
$\mathbf{v}_p$	Phase velocity	Φ	Cost function
W_p	Duration of the radar pulse	$\lambda.$	Adjoint fields
W	Forward operator in matrix form	ω	Angular frequency
w_{pso}	Inertial weight in PSO	η	Noise
ε_r	Relative permittivity	$\mathscr{L}$	The lagrangian
σ_r	Relative conductivity	Φ_A	Alternativa cost function
μ_r	Relative permeability	Φ_{GS}	Gaussian cost function
α_k	Step size	σ_0	Regularization value of conductivity
Δz	Spatial step in z-direction	Φ_{TV}	TV cost function
Δx	Spatial step in x-direction	Φ_{MTV}	MTV cost function
Δt	Time step	λ_{TV}	Weight regularization in TV
		Δh	Spatial step

Resumen

Título: Inversión de onda completa 2D en el dominio del tiempo para datos de radar de penetracion terrestre *

Autor: Jheyston Omar Serrano Luna **

Palabras Clave: Inversión de onda completa, radar de penetración terrestre, unidades de procesamiento grafico.

Descripción: En esta tesis doctoral se presenta una metodología de inversión de onda completa para adquisiciones de offset-corto y un solo-canal. En este tipo de adquisiciones se tienen diferentes desafíos, pero sin lugar a duda la más importante de ellas es su falta de bajos números onda en el proceso de inversión. Las adquisiciones de offset-corto y un solo canal son más sensibles al punto de partida comparadas con adquisiciones multi-offset. Se utiliza este tipo de adquisición ya que permite portabilidad en zonas de difícil acceso así como una rápida recolección de datos permitiendo reducir los tiempos de procesamiento y costos. En esta tesis doctoral se ha propuesto una función de costo alternativa para compensar la diferencia de amplitud entre el dato recolectado y el dato modelado de forma automática. Tres restricciones son utilizadas en la función de costo, las cuales son: Gaussian , variación total y variación total modificada. Tanto los regularizadores como la función de costo alternativa han sido evaluadas en datos sintéticos y recolectados. Los regularizadores permiten converger a modelos más suaves que mantienen las principales estructuras y son más estables en escenarios ruidosos. Usando la regularización Gaussian se logra reducir el ruido incoherente en los datos cuando se aplica en la dirección de los scans. El software libre gprMax junto con la medición de campo y la técnica de optimización global permiten obtener los parámetros internos de la antena así como sus patrones de radiación. Finalmente, el patrón de radiación se tiene en cuenta en la propagación de un frente de onda bidimensional

---

* Tesis doctoral

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y telecomunicaciones. Director: Ana Beatriz Ramirez Silva, Doctorado en Ingeniería Eléctrica. Codirector: Sergio Alberto Abreo Carrillo, Doctor en Ingeniería Electrónica.

y se incluye en el proceso de inversión para datos sintéticos.

Abstract

Title: 2D Full Waveform Inversion in time domain for Ground Penetrating Radar Data *

Author: Jheyston Omar Serrano Luna **

Keywords: Full Waveform Inversion, Ground Penetrating Radar, Graphics processing units.

Description: Full Waveform Inversion methodology for short-offset and single-channel acquisitions is proposed in this thesis doctoral. This kind of acquisition has different challenges, but without a doubt, the most important of them is its lack of low wavenumbers in the inversion process. Single-channel and short-offset acquisitions are more sensitive to the starting point compared to multi-offset acquisitions. This kind of acquisition is used since it allows portability in areas of difficult access, in addition to rapid data collection, reducing processing times and costs. An alternative cost function is proposed to compensate for the difference of amplitude between the collected and modeled data. Three constraints are used in the cost function: Gaussian, Total Variation, and Modified Total Variation. The regularizations and the alternative cost function have been evaluated in synthetic and collected data. The use of regularization helps to converge in smooth models where the main subsurface layers are kept and are stable with noisy data.. The incoherent noise is reduced using Gaussian regularization in the scans-direction. The free software gprMax together with the field measurement and the global optimization technique allows obtaining the internal parameters of the antenna and its radiation patterns. Finally, the radiation pattern is taken into account in the propagation of a two-dimensional wavefront, and it is included in the inversion process for synthetic data.

---

* Doctoral Thesis

** Faculty of Physicomechanical Engineering. School of Electrical, Electronic and Telecomunications Engineering. Advisor: Ana Beatriz Ramirez Silva, Ph.D in Electrical Engineering. Co-advisor: Sergio Alberto Abreo Carrillo, Doctor in electronic engineering.

Introduction

How to obtain a geophysical method that allows characterizing the properties of internal materials of the subsurface from a set of observations in a free surface?, this is the question formulated by experts in geophysics for many years. The oil-deposits have been discovered using traditional methods based on seismic ray-tracing. However, these approximations allow obtaining a smooth version of the subsurface parameters. In 1983, The Full Waveform Inversion (FWI) is developed by Lailly (Lailly and Bednar, 1983), but until 1984 is adapted in the seismic context by Tarantola (Tarantola, 1984). FWI is a local optimization technique that allows estimating physical subsurface parameters as $\mathbf{V}_p$, $\mathbf{V}_s$, ρ, λ, among others, from a set of observations at the surface. FWI has been applied in different contexts like seismic (Bunks et al., 1995), seismology (Blom et al., 2020), medical imaging (Agudo et al., 2018), and electromagnetics (Serrano et al., 2018), (Mozaffari et al., 2020), (Lambot et al., 2004a), (Klotzsche et al., 2019), (Klotzsche et al., 2010), among others.

In geophysics, the Ground Penetrating Radar (GPR) is a standard geophysical method used to identify shallow materials of the subsurface ($<$100m). The GPR method is based on the electromagnetic waves, and in the electromagnetic properties of subsurface materials like permittivity, permeability, and conductivity. The GPR method has several applications such as ice characterization, mines and tunnels detection, pavement condition assesment, soil characterization for agricultural use and oil and gas exploration (Daniels, 2004). The GPR method usually works with frequencies from 10 MHZ to 2.4 GHz (Daniels, 1996).

The use of GPR, together with FWI, has many challenges, such as the starting point, the lack of sufficient illumination in the acquisition that produces high wavenumbers in the estimation of the parameters, the cross-talk between parameters, the high computational cost, and finally, the inverse problem is ill-conditioned and ill-posed. In this doctoral thesis, a methodology is proposed to mitigate the problem of the noise contamination of the acquired data, showing advantages and disadvantages in synthetic and collected data. In addition, the inversion algorithm is implemented in high-performance architectures using GPUs and communication protocols to obtain electromagnetic parameters of the soil using less computational time.

The main contributions of this doctoral thesis are: 1. The estimation of the internal parameters of a shielded antenna through a global optimization technique (i.e. PSO). The importance of incorporating the antenna's radiation patterns during a 2D Full Waveform Inversion. 2. The Full Waveform Inversion methods include alternative cost functions. A new cost function is proposed to address scenarios where the source energy is unknown or where the DC level affects the electromagnetic parameters' estimation. A Gaussian regularization promotes convexity and reduces the incoherent noise present in the data. Two regularizations on the parameters: TV and MTV, where we have reached a better fit in synthetic and collected data than the traditional FWI. The regularization TV and MTV are stable and avoid incorrect values with noisy data. 3. The algorithm optimization methods to reduce the computational cost of the GPR-FWI. A horizontal resolution rule to reduce the execution times by 2x without substantially affecting the inversion results. A shot decomposition scheme, together with a hybrid architecture of CPUs-GPUs, speeds up 8x the execution time. In FWI-3D, a new computational strategy avoids the complete gradient computation, reducing the computational resources required by up to 9.5x.

The document initially introduces the basic concepts of GPR and FWI. In chapter 3, the forward and inverse problems are introduced. This chapter also studies the behavior of CPML, numerical dispersion, numerical stability, discretization using finite differences, the computational cost required by an FWI and ends by presenting the problems associated with the lack of wave numbers. In chapter 4, the alternative cost function seeks that observed and modeled data comparable in amplitude. Next, the Gaussian regularization is included in *scans*-directions to reduce incoherent noise, promote convexity in the cost function; and finally, the use of constraints such as Total Variation (TV) and Modified Total Variation (MTV) are studied. Experiments are carried out in each section to identify advantages in the estimation of synthetic models and tunning up the regularization parameters. Chapter 5, presents results with data collected in Colombia, where the regularizations and the alternative cost functions are tested in the estimation of the permittivity and conductivity parameters. In chapter 6, a methodology to find the internal parameters and radiation patterns in the E and H plane for a 400 MHz shielded antenna that uses global optimization is implemented using PSO and the free software gprMax. In the last chapter, a new computational strategy for FWI-3D is introduced, and the use of applying machine learning together with FWI to improve the estimation of electromagnetic parameters.

Figure 1 shows the general overview with some stages and chapters in this thesis. The three initial steps are: to generate an initial model based on the information in the data, to make an estimate of the source, and to collect the raw data. Next, preprocessing stage is carried out to highlight the information of interest on the raw data; to generate a modeled data based on the finite difference method where the radiation pattern, the source estimation, and the initial models are included. Based on the observed and modeled data, alternatives cost functions and regularizations are proposed on the data or on the parameters to reduce the noise in the parameter estimation. A backpropagation stage is carried out together with the mathematical formulation to define the gradients that allow updating the parameters to be estimated. Finally, a quality control stage is carried out to determine if the parameters obtained converge in models with a geological sense.

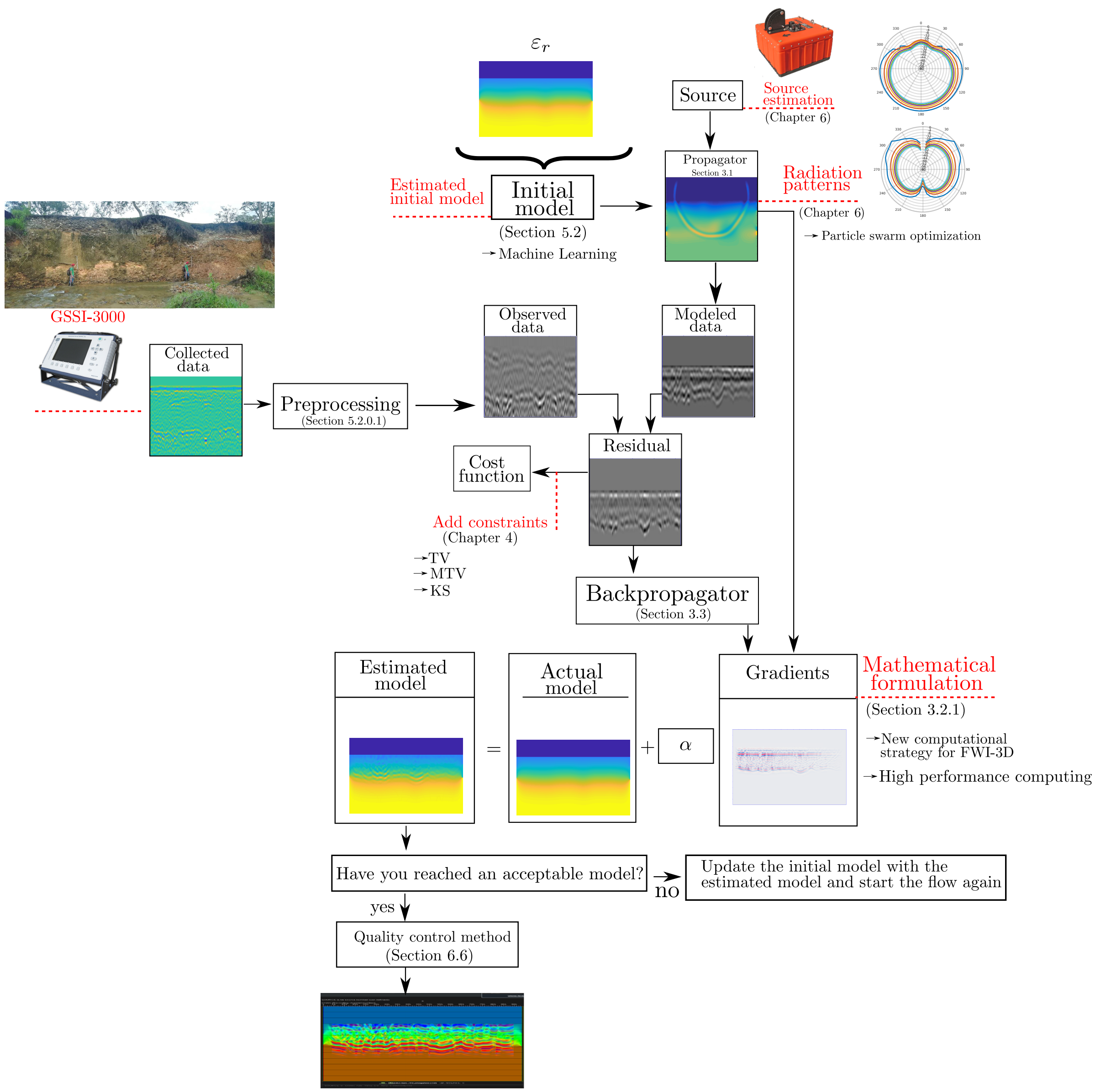

Figure 1. General overview of FWI methodology for collected data in a single-channel and short-offset acquisition.

1. Objectives

Main Objective

To formulate a methodology of inversion for estimating the permittivity, permeability, and conductivity parameters from real data for ground penetrating radar using single-channel and single short-offset acquisition.

Specific Objectives

To develop the mathematical FWI formulation in time domain for GPR applications.

To develop and implement an algorithm for FWI in time domain using a cluster of GPUs.

To evaluate the proposed FWI algorithm using synthetic and real GPR data.

2. Ground penetrating radar

This chapter provides a background GPR method. A compilation of the basic concepts of GPR, the state-of-the-art in the use, the types of waves obtained in GPR acquisitions, their horizontal resolution, and the electrical and magnetic properties are presented.

2.1. Ground penetrating radar

The Ground Penetrating Radar (GPR) is an electromagnetic method with applications in geophysics, civil engineering, and archeology (Daniels, 2004), (Miller et al., 2010). The GPR method is based on Maxwell's equations using a transmission antenna, a receiving antenna, and the unit control. The unit control synchronizes both antennas, where the transmitter antenna generates an electromagnetic pulse according to its source radiation pattern (Persico, 2014). The changes in the magnetic and electrical properties, due to buried objects or the presence of different layers, make the energy to scatter. A receiving antenna on the surface measures the energy according to a receiver radiation pattern. The energy of the transmitted electromagnetic pulse that is measured at the surface is called a scan . In a shielded antenna, the distance between the transmitter antenna T_x and receiver antenna R_x is fixed. The control unit and the antennas move along in the scanning line direction, measuring several scans. Figure 2 shows the principle of operation in the GPR method for a shielded antenna for a single scan.

In the GPR method, it is possible to separate some events when the distance between the transmitter and receiver antenna is compared with the wavelength of the electromagnetic pulse. Figure 3 shows the trajectories between the transmitter and receiver antennas using the rays theory.

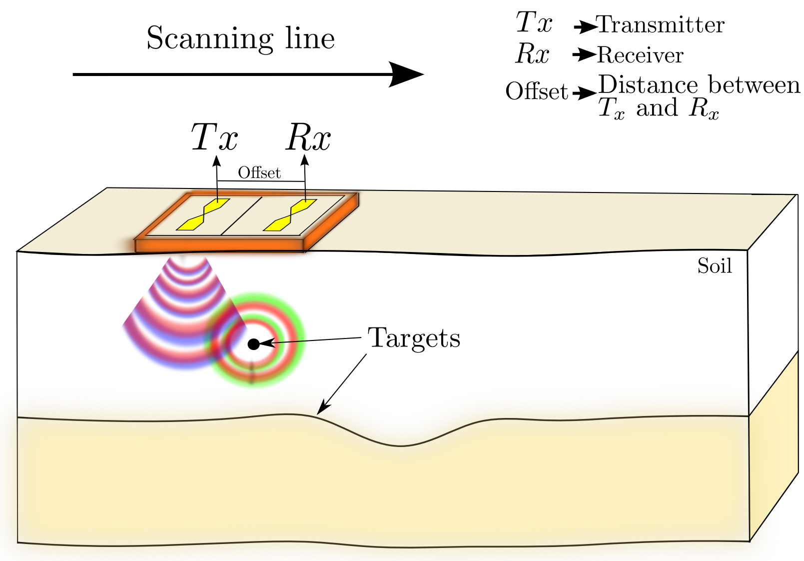

Figure 2. The principle of operation in GPR acquisition, adapted from Persico (2014) Figura 1.1.

The trajectories are: air-wave (AW); ground-wave (GW); reflected wave (RW), and critically refracted wave (CRW). The AW trajectory is a direct wave, and its velocity is the speed of light. The GW trajectory is propagated slower than the AW, and it is a surface wave. The RW is generated by the changes in the electromagnetic parameters of the subsurface (relative permittivity, relative permeability, or conductivity). Part of the energy is transmitted, and other is reflected according to Snell's law, and the CRW is produced by a critical angle in the Snell's law.

A radargram is a set of scans after a GPR acquisition. Figure 4 shows a real preprocess radargram with the main trajectories in GPR data: AW, GW, RW, and CRW. The radargram in Figure 4 is a dataset of scans taken in Lizama, near Bucaramanga (Colombia) using the GSSI-SIR 3000 equipment, and a shielded antenna at 400 MHz.

The interest in FWI of GPR data has been growing in recent years. Initially, the research is

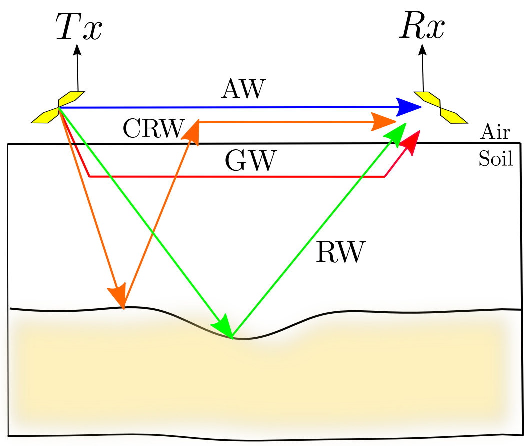

Figure 3. GPR trajectories between transmitter and receiver antennas: air-wave (AW), ground-wave (GW), reflected-wave (RW) and critical refracted wave (CRW), adapted from Jol (2008) Figure 1.7.

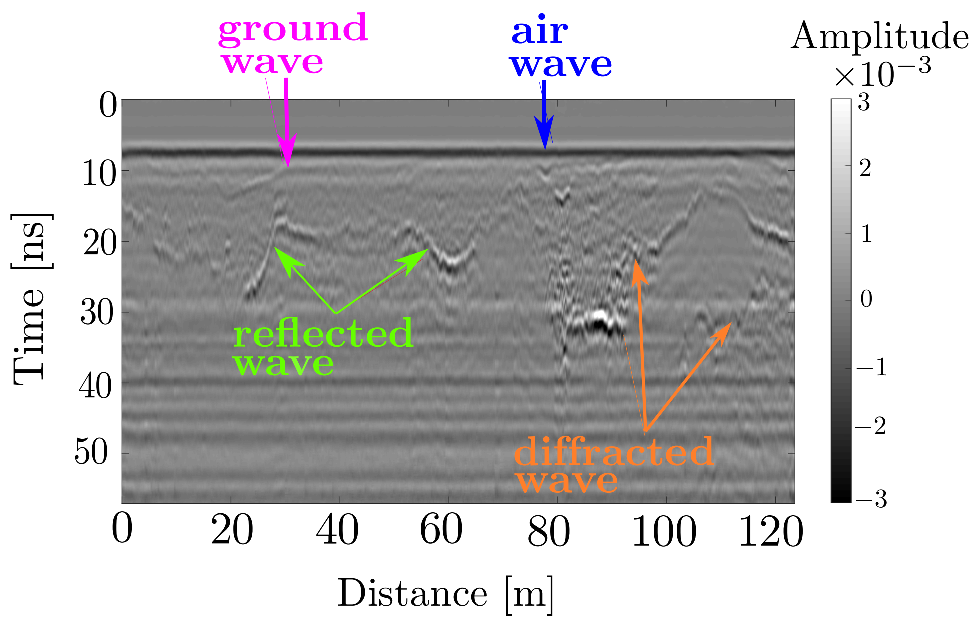

Figure 4. Ground penetration radar data (radargram).

oriented to agriculture using high frequency to characterize the first centimeters of the subsurface, Lambot et al. (2004b). Later, the research is oriented to the estimation of permittivity and the conductivity to characterize layered structures of concrete, Kalogeropoulos et al. (2011). These studies are focused on the frequency domain and restricted to 1D geometries.

The first work in which FWI 2D is performed in the time domain to find the parameters of permittivity and conductivity, it is developed by Ernst et al. (2007). First, the permittivity is set when the conductivity is estimated. Then, the conductivity is set when the permittivity is estimated from the data. This algorithm is improved by Meles et al. (2010) using Crosshole/Borehole-to-Surface GPR data for simultaneous inversion of permittivity and conductivity.

In the frequency domain, some authors have implemented the full waveform inversion method over GPR data using parallel programming. Yang et al. (2012) adapted the algorithm of Meles et al. (2010) in the frequency domain using an alternative scheme of inversion. Lavoué et al. (2014) applied an inversion of GPR data in the frequency domain for multi-offset data obtaining models of permittivity and conductivity with synthetic models. Watson (2016) implemented a 3D FWI with GPR data in synthetic case using finite-elements and considering the permittivity parameter only.

2.2. Spatial resolution in GPR

The spatial resolution in GPR can be: vertical (Δr) (longitudinal or depth) and horizontal (Δl) (lateral, angular or plain);Rial et al. (2007). Figure 5 shows the principle of vertical and horizontal resolution. The vertical and horizontal resolution defines the capacity of the GPR method to distinguish two objects of interest in the subsurface.

The vertical resolution is expressed by the efective duration of the radar pulse (W_P). The vertical resolution is defined by:

$$\Delta r \geq \frac{c \cdot W_P}{4\sqrt{\varepsilon_r}}, \tag{1}$$

where $W_P = \frac{1}{f_c}$, f_c is the central frequency, c is the speed of the light and ε_r is the relative permittivity. Two events are distinguishable if they have a separation greater than $W_p/2$. The horizontal resolution indicates the minimum distance that must exist between two reflectors, such that the radar can identify both reflectors. This lateral resolution mainly depends on the spatial sampling and on the propagating wavelengths; that is, the number of scans per meter. The horizontal resolution is defined by:

$$\Delta l \geq \sqrt{\frac{c \cdot \mathfrak{r} \cdot W_P}{2\sqrt{\varepsilon_r}}}, \tag{2}$$

where $\mathfrak{r}$ is the distance to the target, Jol (2008).

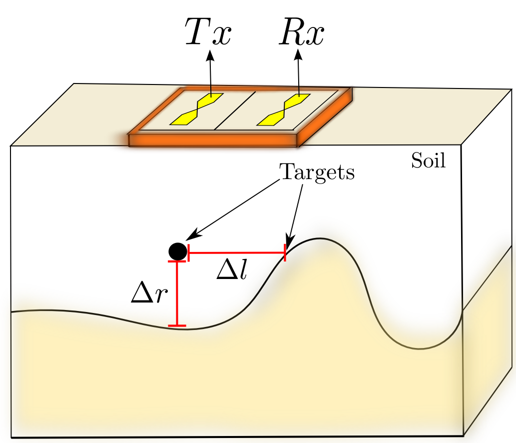

Figure 5. Ground Penetrating Radar resolution: Δl horizontal resolution and Δr vertical resolution.

2.3. Electric and magnetic properties

The GPR method is widely used in engineering, for its versatility and simplicity in the data collection. The GPR method can be considered as the electromagnetic response of the subsurface due to an injected electromagnetic pulse. The frequencies used in GPR ranges from 30 MHz to 2 GHz. Materials are characterized: dielectric, magnetic or conductive, depending on the predominant phenomenon.

2.3.1. Dielectrics. The dielectrics are composed of negative and positive charges that cannot travel freely. In the presence of an external field, the centroids of the charges can change position slightly creating electric dipoles. In GPR dipole polarization is the predominant mechanism, where the dipoles in the material align with the electric field (**E**). The constitutive relationship between the electric field and the electric flux density (**D**):

$$\mathbf{D} = \varepsilon_0 \mathbf{E}, \tag{3}$$

where the electric polarization (**P**) is given by

$$\begin{aligned} \mathbf{D} &= \varepsilon_0 \mathbf{E} + \mathbf{P}, \\ \mathbf{D} &= \varepsilon_0 \mathbf{E} + \varepsilon_0 \chi_e \mathbf{E}, \\ \mathbf{D} &= \varepsilon_0 (1 + \chi_e) \mathbf{E}, \end{aligned} \tag{4}$$

where ε_0 (F/m) is the the permittivity of free-space. The permittivity is the capacity to store and release electromagnetic energy in electric charges. The permittivity is a complex value where the real part is associated with the store, and the imaginary part is associated with lossy. Some books assign the real part of the permittivity parameter to the symbol κ, and it is called the dielectric constant. Materials with low saline conditions and low clay contents are the best choice to characterize in GPR because they have low attenuation. On the left side of Figure 6, the particles do not have displacement or polarization. On the right side of Figure 6, an external field is applied such that the cloud of electrons is moved from its original position (polarization) creating electric dipoles.

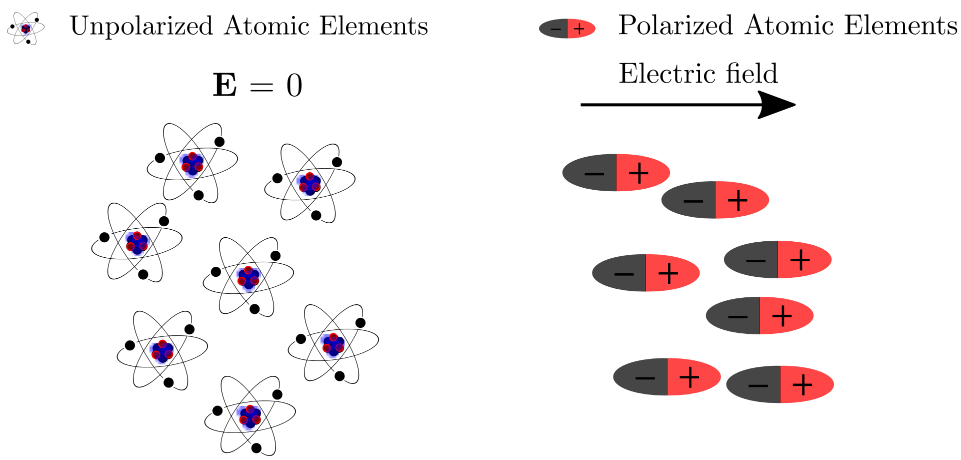

Figure 6. (left) non polarizate material ($\mathbf{E} = 0$); (right) Electrically polarized material with external electric field.

2.3.2. Conductors. Conductivity refers to the ability to pass the electric charges through the material. Materials with many free electrons are called conductors, which have a random movement that produce zero current. With an external field, electrons move in the opposite direction to the field, producing conduction current and the energy will be converted to heat due to collisions. The conductivity of a conductor (σ), characterizes the conducting properties of free electrons. Figure 7 shows the behavior of the charges when an external electric field is applied. Conductivity defines the ratio between the electric field (**E**) and the current density (**J**), the relation is given by

$$\mathbf{J} = \sigma \mathbf{E} \tag{5}$$

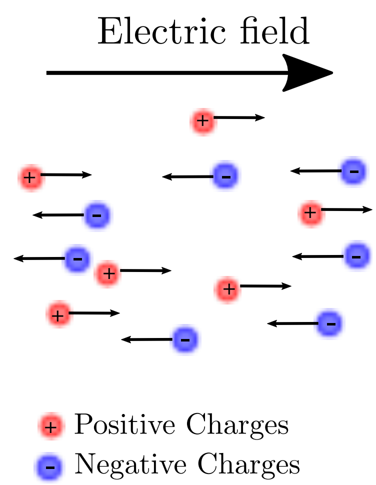

Figure 7. Concept of conductivity with external electric field.

Like the permittivity parameter, conductivity is also a complex measure, where its imaginary part produces a phase shift in the current conduction, produced by the electric field. The high values of conductivity (>20 m/s) are a problem in FWI because the energy is converted into heat, and the receiver antenna does not measure the energy. Table 1 shows some values of permittivity and conductivity of different materials that can be present in the subsurface at 100 MHz.

Medium	Static conductivity (mS/m) at 100 MHz	Relative permittivity at 100 (MHz)
Air	[0]	[1]
Clay - dry	[1,100]	[2,20]
Clay - wet	[100,1000]	[15,40]
Concrete - dry	[1,10]	[4,10]
Concrete - wet	[10,100]	[10,20]
Freshwater	[1×10^{-1},10]	[78,88]
Freshwater ice	[10^{-6}, 1]	[3]
Seawater	[4000]	[81,88]
Seawater ice	[10,100]	[4,8]
Permafrost	[10^{-1},10]	[2,8]
Granite - dry	[10^{-5}-10^{-3}]	[5,8]
Granite fractured and wet	[1,10]	[5,15]
Limestone - dry	[10^{-6}-10^{-3}]	[4,8]
Limestone - wet	[10,100]	[6,15]
Sandstone - dry	[10^{-7}-10^{-3}]	[4,7]
Sandstone - wet	[10^{-3}-10^{-2}]	[5,15]
Shale - saturated	[10,100]	[6,9]
Sand - dry	[10^{-4},1]	[3,6]
Sand - wet	[10^{-1},10]	[10,30]
Sand - coastal, dry	[10^{-2},1]	[5,10]
Soil - sandy, dry	[10^{-1},100]	[4,6]
Soil - sandy, wet	[10,100]	[15,30]
Soil - loamy, dry	[10^{-1},1]	[4,6]
Soil - loamy, wet	[10,100]	[10,20]
Soil - clayey, dry	[10^{-1},100]	[4,6]
Soil - clayey, wet	[100,1000]	[10,15]
Soil - average	[5]	[16]

Table 1. *Typical values of permittivity and conductivity for different materials. Taken from Jol (2008).*

3. The forward and inverse problem

The GPR method is based on the propagation of electromagnetic waves. In this chapter, a review of the Maxwell equations, the constitutive relations, the finite time difference method, the convolutional perfectly matched layer is described. The forward and inverse methods are introduced with their equations and discretization. The mathematical formulation to obtain the adjoint operator and the definition of gradients are shown in this chapter. In the last part of the chapter, the definitions of gradient descent, L-BFGS, lack of illumination in short-offset and single-channel acquisitions, and the selection of the number of scans, are introduced.

3.1. The forward problem

In the forward problem, the electric field at the surface is computed given that the source and electromagnetic parameters of the subsurface are known. Some alternatives to solve the forward problem is to use finite differences in the time domain, finite differences in the frequency domain, or finite elements. The section introduce the use of the electromagnetic wave equation described by Maxwell, its discretization using finite differences in the time domain, the use of CPML to avoid non-natural borders, numerical dispersion, and stability analysis.

3.1.1. The electromagnetic wave equations. The isotropic and non-dispersive electromagnetic wave equations are described by the Maxwell equations:

$$\nabla \times \mathbf{E}(\mathbf{r},t) = -\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t}, \tag{6}$$

$$\nabla \times \mathbf{H}(\mathbf{r},t) = \frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} + \mathbf{J}(\mathbf{r},t), \tag{7}$$

$$\nabla \cdot \mathbf{D}(\mathbf{r},t) = \mathbf{q}(\mathbf{r},t), \tag{8}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r},t) = 0, \tag{9}$$

where $\mathbf{E}(\mathbf{r})$ is the electric field (V/m), $\mathbf{H}(\mathbf{r},t)$ is the magnetic field intensity (A/m), $\mathbf{D}(\mathbf{r},t)$ is the displacement field or electric flux density(C/m^2), $\mathbf{B}(\mathbf{r},t)$ is the magnetic flux density (T), $\mathbf{J}(\mathbf{r},t)$ is the current density (A/m^2), $\mathbf{q}(\mathbf{r},t)$ is the charge density (C/m^3) and t is time. Equation (6) defines that the variation with respect to the time of the magnetic field ($\mathbf{B}(\mathbf{r},t)$) produces an electric field ($\mathbf{E}(\mathbf{r},t)$) that is located around of $\mathbf{B}(\mathbf{r},t)$ and $\mathbf{r}$ is the position vector in the spatial coordinates x, z. Taking the Equation (6) and solving the rotational:

$$\nabla \times \mathbf{E}(\mathbf{r},t) = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \mathbf{E}_x(\mathbf{r},t) & \mathbf{E}_y(\mathbf{r},t) & \mathbf{E}_z(\mathbf{r},t) \end{vmatrix} = -\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t},$$

$$-\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t} = \left(\frac{\partial \mathbf{E}_z(\mathbf{r},t)}{\partial y} - \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial z}\right)\hat{\imath} + \left(\frac{\partial \mathbf{E}_x(\mathbf{r},t)}{\partial z} - \frac{\partial \mathbf{E}_z(\mathbf{r},t)}{\partial x}\right)\hat{\jmath} + \left(\frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial x} - \frac{\partial \mathbf{E}_x(\mathbf{r},t)}{\partial y}\right)\hat{k}. \tag{10}$$

Taking the Equation (7) and solving the rotational, then

$$\nabla \times \mathbf{H}(\mathbf{r},t) = \frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} + \mathbf{J}(\mathbf{r},t) = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \mathbf{H}_x(\mathbf{r},t) & \mathbf{H}_y(\mathbf{r},t) & \mathbf{H}_z(\mathbf{r},t) \end{vmatrix},$$

therefore,

$$\frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} + \mathbf{J}(\mathbf{r},t) = \left(\frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial y} - \frac{\partial \mathbf{H}_y(\mathbf{r},t)}{\partial z}\right)\hat{\imath} + \left(\frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial z} - \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial x}\right)\hat{\jmath} + \left(\frac{\partial \mathbf{H}_y(\mathbf{r},t)}{\partial x} - \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial y}\right)\hat{k}. \tag{11}$$

Two propagation modes are defined to describe the Maxwell equation in 2D: transverse electric (TE) and transverse magnetic (TM). The GSSI GPR equipment considers only the TE mode because of the orientation of the antennas within the shielding. In TE mode, the fields $\mathbf{H}_z(\mathbf{r},t)$, $\mathbf{H}_x(\mathbf{r},t)$ and $\mathbf{E}_y(\mathbf{r},t)$ are different from zero.

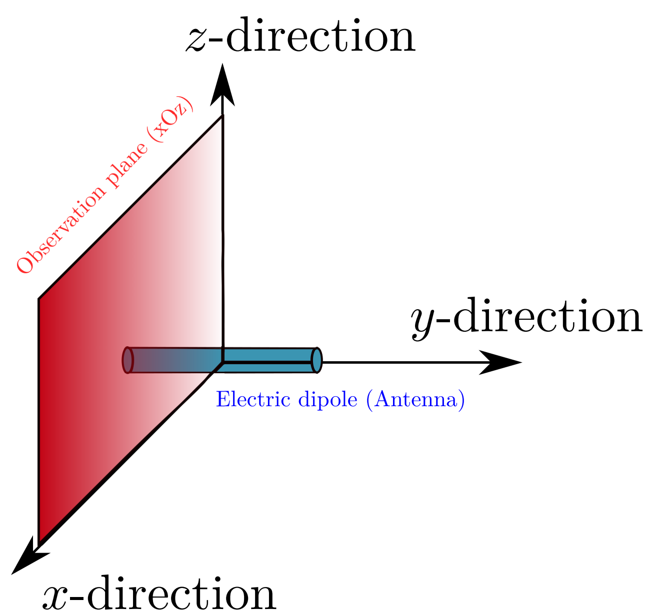

Figure 8. Transverse electric mode configuration.

From Equations (10) and (11) the propagation (TE) is obtained, where the fields ($\mathbf{E}_z(\mathbf{r},t)$ and $\mathbf{E}_x(\mathbf{r},t)$) are zero. Rewriting the Equation (10), then

$$-\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t} = \left(-\frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial z}\right)\hat{i} + \left(\frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial x}\right)\hat{k}. \tag{12}$$

Using the constitutive relation $\mathbf{B} = \mu\mathbf{H}$, then

$$\begin{aligned} \mu\frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial t} &= \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial z}, \\ -\mu\frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial t} &= \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial x}. \end{aligned} \tag{13}$$

The Equation (11) is different from zero in the y-direction. Therefore, rewriting the Equation (11)

$$\frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} + \mathbf{J}(\mathbf{r},t) = \left(\frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial z} - \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial x} \right), \tag{14}$$

and using the constitutive relation $\mathbf{D} = \varepsilon \mathbf{E}$

$$\varepsilon \frac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t} + \mathbf{J}(\mathbf{r},t) = \left(\frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial z} - \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial x} \right). \tag{15}$$

3.1.2. Modeling the electromagnetic wave propagation. GPR is based on electromagnetic wave propagation described by Maxwell. The Maxwell's equations in a medium non-dispersive and isotropic are described by

$$\varepsilon(\mathbf{r}) \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial t} = \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial z} - \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial x} - \sigma(\mathbf{r})\mathbf{E}_y(\mathbf{r},t) + \mathbf{J}_s(\mathbf{r},\mathbf{t}), \tag{16}$$

$$\mu(\mathbf{r}) \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial t} = \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial z}, \tag{17}$$

$$-\mu(\mathbf{r}) \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial t} = \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial x}, \tag{18}$$

where $\mathbf{E}_y(\mathbf{r},t)$ is the electric field in the y-direction (V/m); $\mathbf{H}_x(\mathbf{r},t)$ is the magnetic field in the x-direction (A/m); $\mathbf{H}_z(\mathbf{r},t)$ is the magnetic field in the z-direction (A/m) ; $\mathbf{J}_s(\mathbf{r},t)$ is the electromagnetic pulse (A/m^2); $\mu(\mathbf{r})$ is the permeability (H/m); $\varepsilon(\mathbf{r})$ is the permittivity (F/m); $\sigma(\mathbf{r})$ is the conductivity (S/m). The permittivity, permeability and conductivity are defined as

$$
\begin{aligned}
\varepsilon(\mathbf{r}) &= \varepsilon_0 \varepsilon_r(\mathbf{r}), \\
\mu(\mathbf{r}) &= \mu_0 \mu_r(\mathbf{r}), \\
\sigma(\mathbf{r}) &= \sigma_0 \sigma_r(\mathbf{r}),
\end{aligned}
\tag{19}
$$

where ε_0 and μ_0 are the permittivity and permeability in the vacuum and σ_0 is a regularization value. $\varepsilon_r(\mathbf{r})$, $\mu_r(\mathbf{r})$, and $\sigma_r(\mathbf{r})$ are the relative permittivity, permeability, and conductivity, respectively. The parameters ε_0 and μ_0 are defined as

$$
\begin{aligned}
\varepsilon_0 &\simeq 8.85 \times 10^{-12} \text{ (F/m)}, \\
\mu_0 &\simeq 4\pi \times 10^{-7} \text{ (H/m)}.
\end{aligned}
\tag{20}
$$

The Equations (16), (17) and (18) are discretized using the 3D Yee's scheme, as it is shown in Figure 9.

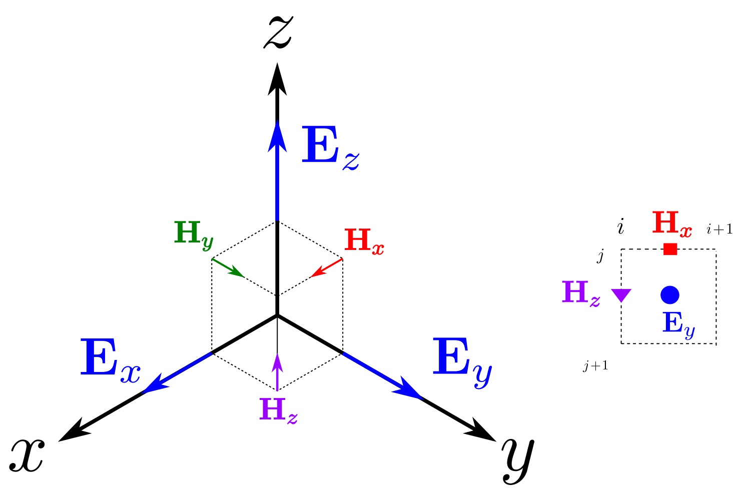

Figure 9. Discretization of the electromagnetic equation using FDTD method (Yee, 1966).

Finally, the electromagnetic wave equation (Equations (16), (17) and (18)) is discretized using eighth order in space and second order in time. The electrical field $\mathbf{E}_y$ and the magnetic fields $\mathbf{H}_x$, $\mathbf{H}_z$ are computed as

$$\begin{aligned} &\mathbf{E}_{y_{i+1/2,j+1/2}} \\ &\mathbf{H}_{x_{i+1/2,j}} \\ &\mathbf{H}_{z_{i,j+1/2}} \end{aligned} \tag{21}$$

$$\begin{aligned} \varepsilon_{i+1/2,j+1/2}\frac{\mathbf{E}_{y_{i+1/2,j+1/2}}^{n+1/2}-\mathbf{E}_{y_{i+1/2,j+1/2}}^{n-1/2}}{\Delta_t} &= \sum_{l=0}^{3}\frac{C_{FD}[l]\cdot(\mathbf{H}_{x_{i+1/2,j+l}}^{n}-\mathbf{H}_{x_{i+1/2,j-1-l}}^{n})}{\Delta_z} - \\ \sum_{l=0}^{3}\frac{C_{FD}[l](\mathbf{H}_{z_{i+l,j+1/2}}^{n}-\mathbf{H}_{z_{i-1-l,j+1/2}}^{n})}{\Delta_x} &- \sigma_{i+1/2,j+1/2}\mathbf{E}_{y_{i+1/2,j+1/2}}^{n}, \end{aligned} \tag{22}$$

where $C_{FD} = \left[\frac{1225}{1024}, -\frac{245}{3072}, \frac{49}{5120}, -\frac{5}{7168}\right]$; Δ_x, Δ_z are the length step in x-direction and z-direction, respectively; and Δ_t is the time step; and i, j and n are the dicretized version of x, z and t, respectively. According with Sullivan (2013), the electrical field in the n time is computed as $\mathbf{E}_{y_{i+1/2,j+1/2}}^{n} = 0.5\mathbf{E}_{y_{i+1/2,j+1/2}}^{n+1/2} + 0.5\mathbf{E}_{y_{i+1/2,j+1/2}}^{n-1/2}$. The magnetic fields are computed using the following equations

$$\mu_{i+1/2,j}\frac{\mathbf{H}_{x_{i+1/2,j}}^{n+1} - \mathbf{H}_{x_{i+1/2,j}}^{n}}{\Delta_t} = \sum_{l=0}^{3} \frac{C_{FD}[l](\mathbf{E}_{y_{i+1/2,j+3/2+l}}^{n+1/2} - \mathbf{E}_{y_{i+1/2,j+1/2-l}}^{n+1/2})}{\Delta_z}, \tag{23}$$

$$-\mu_{i,j+1/2}\frac{\mathbf{H}_{z_{i,j+1/2}}^{n+1} - \mathbf{H}_{z_{i,j+1/2}}^{n}}{\Delta_t} = \sum_{l=0}^{3} \frac{C_{FD}[l](\mathbf{E}_{y_{i+3/2+l,j+1/2}}^{n+1/2} - \mathbf{E}_{y_{i+1/2-l,j+1/2}}^{n+1/2})}{\Delta_x}. \tag{24}$$

Replacing $i+1/2 = i$, $i-1/2 = i-1, j+1/2 = j$ and $j-1/2 = j-1$ then:

$$\begin{aligned}\varepsilon_{i,j}\frac{\mathbf{E}_{y_{i,j}}^{n+1/2} - \mathbf{E}_{y_{i,j}}^{n-1/2}}{\Delta_t} &= \sum_{l=0}^{3} \frac{C_{FD}[l]\cdot(\mathbf{H}_{x_{i,j+l}}^{n} - \mathbf{H}_{x_{i,j-1-l}}^{n})}{\Delta_z} - \\ &\sum_{l=0}^{3} \frac{C_{FD}[l](\mathbf{H}_{z_{i+l,j}}^{n} - \mathbf{H}_{z_{i-1-l,j}}^{n})}{\Delta_x} - \sigma_{i,j}\mathbf{E}_{y_{i,j}}^{n},\end{aligned} \tag{25}$$

$$\mu_{i,j}\frac{\mathbf{H}_{x_{i,j}}^{n+1} - \mathbf{H}_{x_{i,j}}^{n}}{\Delta_t} = \sum_{l=0}^{3} \frac{C_{FD}[l](\mathbf{E}_{y_{i,j+1+l}}^{n+1/2} - \mathbf{E}_{y_{i,j-l}}^{n+1/2})}{\Delta_z}, \tag{26}$$

$$-\mu_{i,j}\frac{\mathbf{H}_{z_{i,j}}^{n+1}-\mathbf{H}_{z_{i,j}}^{n}}{\Delta_t}=\sum_{l=0}^{3}\frac{C_{FD}[l](\mathbf{E}_{y_{i+1+l,j}}^{n+1/2}-\mathbf{E}_{y_{i-l,j}}^{n+1/2})}{\Delta_x}. \tag{27}$$

3.1.3. Convolutional perfectly matched layer in electromagnetic waves. The simulation space for the propagation of electromagnetic waves is limited. These fixed dimensions mean that when the energy reaches the bounders, they generate reflections that do not exist in the observed data. Convolutional Perfectly Matched Layer (CPML) is a method that allows the energy reaching the edges to be attenuated and thus mitigates the effect of unnatural bounders created by computational limitations. According to Pasalic and McGarry (2010), four auxiliary fields are produced in the spatial derivatives for the 2D electromagnetic wave equation: $\Psi_{Eyx},\Psi_{Eyz},\Psi_{Hxz},\Psi_{Hzx}$. The partial derivative concerning i is defined by

$$\frac{\partial}{\partial\tilde{i}}=\frac{\partial}{\partial i}+\Psi_i, \tag{28}$$

where $\Psi_i^n=a_i\Psi_i^{n-1}+b_i\left(\frac{\partial}{\partial i}\right)^n$. Therefore, the electromagnetic wave equations are reformulated as:

$$\varepsilon(\mathbf{r})\frac{\partial\mathbf{E}_y(\mathbf{r},t)}{\partial t}=\frac{\partial\mathbf{H}_x(\mathbf{r},t)}{\partial z}-\frac{\partial\mathbf{H}_z(\mathbf{r},t)}{\partial x}-\sigma(\mathbf{r})\mathbf{E}_y(\mathbf{r},t)+\mathbf{J}_s(\mathbf{r},\mathbf{t})+\Psi_{Hxz}-\Psi_{Hzx}, \tag{29}$$

$$\mu(\mathbf{r})\frac{\partial\mathbf{H}_x(\mathbf{r},t)}{\partial t}=\frac{\partial\mathbf{E}_y(\mathbf{r},t)}{\partial z}+\Psi_{Eyz}, \tag{30}$$

$$-\mu(\mathbf{r})\frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial t} = \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial x} + \Psi_{Eyx}. \tag{31}$$

The initial condition in Ψ_i^{n-1} is zero. The parameters a_i and b_i are set to:

$$a_i = \frac{d_i}{d_i + \alpha_i}(b_i - 1), \tag{32}$$

$$b_i = e^{-(d_i+\alpha_i)\Delta_t}. \tag{33}$$

The parameters used in a_i and b_i are defined by:

$$R = 10^{-6},\ L_i = P_{CPML} \cdot \Delta_i,\ d_0 = \frac{-3}{2L_i}\ln(R). \tag{34}$$

$$\mathbf{F}_i = \begin{cases} (P_{CPML} - i)\Delta_i, & 0 \leq i < P_{CPML} \\ 0, & P_{CPML} \leq i < N_i - P_{CPML} \\ (i - (N_i - P_{CPML}))\Delta_i, & N_i - P_{CPML} \leq i < N_i, \end{cases} \tag{35}$$

$$\begin{aligned} d_i &= d_0 \cdot c \cdot \left(\frac{\mathbf{F}_i}{L_i}\right)^2, \\ \alpha_i &= \pi f_c \left(\frac{L_i - \mathbf{F}_i}{L_i}\right), \end{aligned} \tag{36}$$

where P_{CPML} is the total points used in the CPML boundary, c is the speed of the light in air, f_c is the central frequency of the source, and i is x-direction or z-direction. Equations (29), (30), and (31) shows that the Ψ parameter is expressed as an auxiliary field of dissipation very similar to the behavior of σ_r in the electromagnetic wave. The vectors a_i and b_i have an exponentially form that allows a gradual attenuation and avoid reflections when the energy reaches the CPML zone. The parameters d_i and $\mathbf{F}_i$ influence the behavior of the vectors a_i, and b_i; if they increase, they cause the exponential to quickly drop to zero, thus generating an effect of an unnatural barrier. On the contrary, if they decrease, it causes the exponential to decay slowly; therefore, the energy is not attenuated. Finally, the α_i parameter generates a shift in the vectors a_i and b_i, so if it is very large, it removes the CPML area. In our implementation, the CPML is used in top, bottom, left and rigth borders. The behavior of CPML is shown in Figure 10, where 40 points are selected in the implementation. We have produced an electromagnetic pulse in the vacuum. Figure 10 shows the electric field $\mathbf{E}_y$ for three snapshots of time: 250 (2.5 ns), 400 (4 ns) and 700 (7 ns). The first column in Figure 10 shows the electric field $\mathbf{E}_y$ without CPML, the second column shows the electric field $\mathbf{E}_y$ with CPML, and the third column shows the difference between the first and second column.

The efficiency of the CPML is study using the total energy density. The total energy density is computed according to:

$$\nu = \frac{1}{2}\varepsilon_0 \cdot \mathbf{E}_y^2 + \frac{1}{2}\mu_0 \cdot \mathbf{H}_x^2 + \frac{1}{2}\mu_0 \cdot \mathbf{H}_z^2. \tag{37}$$

Figure 11 shows the energy density for the electric and magnetic fields. Figure 11-a) shows

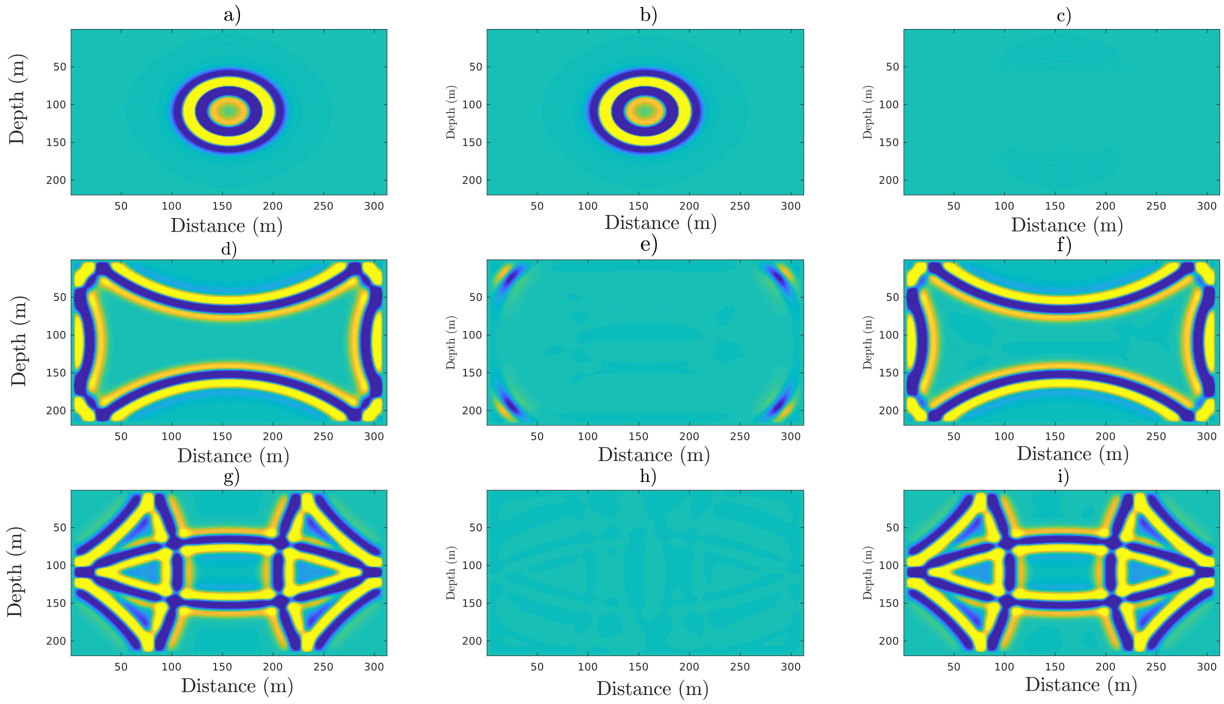

Figure 10. Electric field $\mathbf{E}_y$ in three differents snapshots: 250, 400, 700. First column without CPML, second column with CPML and third column shows the difference between first and second column.

the electric energy density $\mathbf{E}_y$, Figure 11-b) and Figure 11-c) shows the magnetic energy density for $\mathbf{H}_x$ and $\mathbf{H}_z$, respectively; and Figure 11-d) shows the total energy density (Ec. 37) where in blue shows the energy density without CPML and in orange the energy density with CPML. During the first 200 samples, the energy is injected and then gradually the energy is absorbed in the CPML layers. The energy is attenuated in 413 dB and it is enough for the modeled data (Berenger, 1994), (Komatitsch and Martin, 2007).

Figure 12 shows the behavior of the total energy density through the iterations in the two scenarios: with CPML and without CPML.

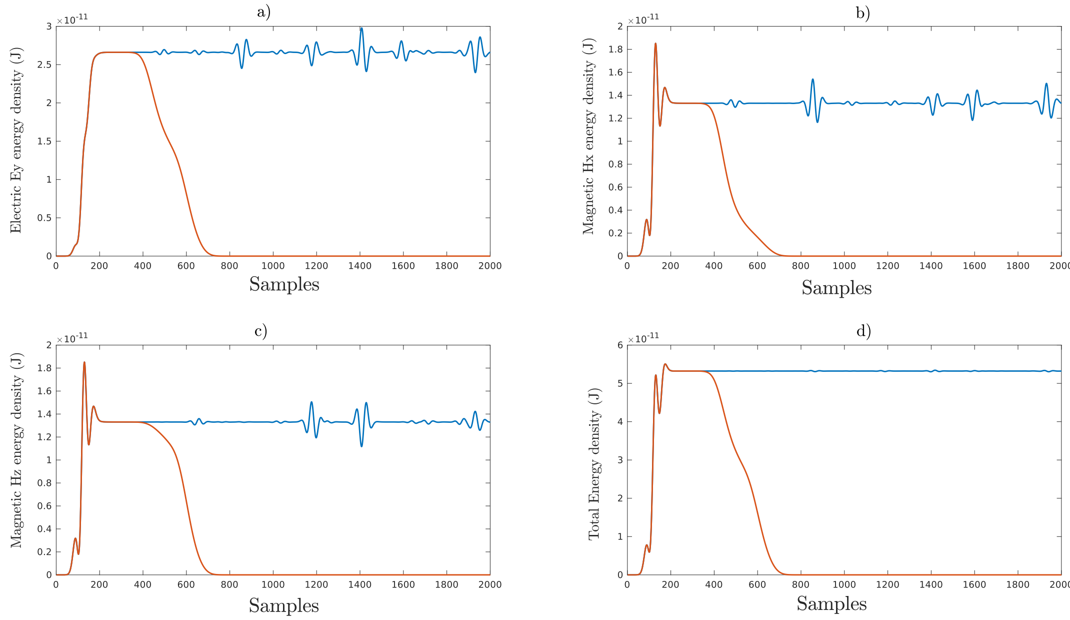

Figure 11. The behavior of the total energy density through the iteration, in blue the energy density without CPML and in orange the energy density with CPML: a) Electric $\mathbf{E}_y$ energy; b) Magnetic $\mathbf{H}_x$ energy , c) Magnetic $\mathbf{H}_z$ energy and d) Total energy density.

3.1.4. Numerical dispersion and numerical stability. According to the Courant-Friedrichs-Lewy (CFL) condition for numerical stability (De Moura and Kubrusly, 2013), the time-step (Δ_t) is defined for the cell size and the maximum velocity of the electromagnetic waves; it is given as

$$\Delta_t \leqslant \frac{1}{c\sqrt{\frac{1}{\Delta_x}+\frac{1}{\Delta_z}}}. \tag{38}$$

The FDTD method has some drawbacks related to the cartesian grid used. When the geometry in the objects does not fit the grid, the objects suffer a staircase effect. The staircase problem

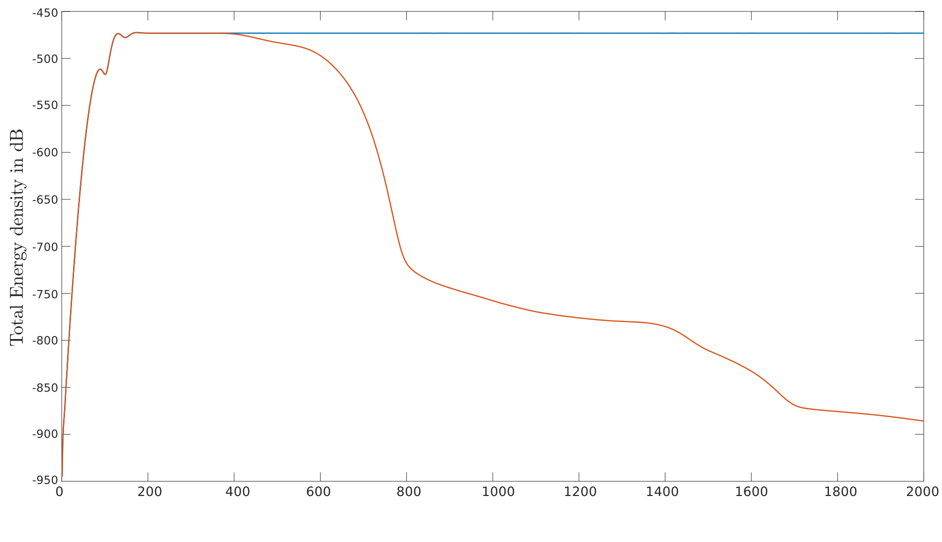

Figure 12. The behavior of the total energy density in the electric field through the iterations.

causes that the propagation of the electromagnetic waves suffers a delay, numerical dispersion and local spatial errors, Cangellaris and Wright (1991) and Holland (1993), (Dridi et al., 2001). Some alternatives are mentioned by Warren (Warren, 2009) to solve the staircase problems: an unstructured global grid to model non-conformal objects, a fine global grid, fine sub-grids within the global grid, and local sub-cell methods to model non-conformal objects. Despite the problems of the FDTD method, it is selected due to its simplicity, efficiency, and numerical stability. Besides, in many models, the FDTD approximation offers a reasonable estimation of parameters in FWI with less computational cost than finite elements.

3.1.4.1. Numerical dispersion of the electromagnetic wave equation in an isotropic and non-dispersive medium. The numerical dispersion of the electromagnetic wave equation is studied using the solution of the electromagnetic fields as planar waves. For each field, the electromagnetic field is defined by

$$
\begin{aligned}
\mathbf{E}_y &= \mathbf{E}_{oy} e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)},\\
\mathbf{H}_x &= \mathbf{H}_{ox} e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)},\\
\mathbf{H}_z &= \mathbf{H}_{oz} e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)},
\end{aligned}
\tag{39}
$$

where $\mathbf{E}_{oy}$, $\mathbf{H}_{ox}$, $\mathbf{H}_{oz}$ are the maximum amplitude, ω is the angular frequency, k_x and k_z are the wavenumber in x and z directions, repectively. In the following equations, the spatial coordinates i and j are replaced by r_x and r_z, respectively. The Equation (40) is obtained after replacing the expressions of Equation (39) on the left side of Equation (25).

$$
\begin{aligned}
\varepsilon \frac{\mathbf{E}_{y r_x, r_z}^{n+\frac{1}{2}} - \mathbf{E}_{y r_x, r_z}^{n-\frac{1}{2}}}{\Delta_t} &= \frac{\varepsilon}{\Delta_t} \left(\mathbf{E}_{oy} e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)} \right) \left(e^{\left(\frac{j\omega\Delta_t}{2}\right)} - e^{\left(\frac{-j\omega\Delta_t}{2}\right)} \right)\\
&= \frac{\varepsilon}{\Delta_t} \left(\mathbf{E}_{oy} e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)} \right) \left(2jsin\left(\frac{\omega\Delta_t}{2} \right) \right).
\end{aligned}
\tag{40}
$$

Equation (41) is obtained after applying the same methodology on the right side of Equation (25).

$$
\sum_{l=0}^{3} \frac{C_{FD}[l] \cdot (\mathbf{H}_{x r_x, r_z+l}^{n} - \mathbf{H}_{x r_x, r_z-1-l}^{n})}{\Delta_z} - \sum_{l=0}^{3} \frac{C_{FD}[l] \cdot (\mathbf{H}_{z r_x+l, r_z}^{n} - \mathbf{H}_{z r_x-1-l, r_z}^{n})}{\Delta_x}. \tag{41}
$$

After factoring terms, then

$$
\begin{aligned}
&\frac{\mathbf{H}_{ox}e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)}}{\Delta_z}\sum_{l=0}^{3} C_{FD}[l]\left(e^{-j(l)k_z\Delta_z} - e^{-j(-1-l)k_z\Delta_z}\right)\\
&-\frac{\mathbf{H}_{oz}e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)}}{\Delta_x}\sum_{l=0}^{3} C_{FD}[l]\left(e^{-j(l)k_x\Delta_x} - e^{-j(-1-l)k_x\Delta_x}\right).
\end{aligned}
\tag{42}
$$

The Equation (43) is obtained by replacing the expressions of Equation (39) on the left side of Equation (26)

$$
\frac{\mu}{\Delta_t}\left(\mathbf{H}_{x\,r_x,r_z}^{\,n+1} - \mathbf{H}_{x\,r_x,r_z}^{\,n}\right) = \frac{\mu}{\Delta_t}\left(\mathbf{H}_{ox}e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)}\right)\left(e^{j\omega\Delta_t} - 1\right).
\tag{43}
$$

The Equation (44) is obtained by replacing the expressions of Equation (39) in the right side of Equation (26)

$$
\begin{aligned}
&\sum_{l=0}^{3}\frac{C_{FD}[l](\mathbf{E}_{y\,r_x,r_z+1+l}^{\,n+1/2} - \mathbf{E}_{y\,r_x,r_z-l}^{\,n+1/2})}{\Delta_z}\\
&= \frac{1}{\Delta_z}\left(\mathbf{E}_{oy}e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)}\right)\left(\sum_{l=0}^{3} C_{FD}[l]e^{\frac{j\omega\Delta_t}{2}}\left(e^{-j(1+l)k_z\Delta_z} - e^{jlk_z\Delta_z}\right)\right).
\end{aligned}
\tag{44}
$$

Putting the Equations (43) and (44) together, then:

$$
\begin{aligned}
&\mathbf{H}_{ox}e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)} =\\
&\frac{\Delta_t}{\mu\Delta_z}\left(\mathbf{E}_{oy}e^{j(n\omega\Delta_t - r_x k_x \Delta_x - r_z k_z \Delta_z)}\right)\left(\sum_{l=0}^{3} C_{FD}[l]\frac{\left(e^{\frac{j\omega\Delta_t}{2}}\left(e^{-j(1+l)k_z\Delta_z} - e^{jlk_z\Delta_z}\right)\right)}{(e^{j\omega\Delta_t} - 1)}\right).
\end{aligned}
\tag{45}
$$

Replacing the definition of the magnetic field of Equation (39) in Equation (27) from the left side

$$\frac{-\mu}{\Delta_t}\left(\mathbf{H}_{z_{r_x,r_z}}^{n+1}-\mathbf{H}_{z_{r_x,r_z}}^{n}\right)=\frac{-\mu}{\Delta_t}\left(\mathbf{H}_{oz}e^{j(n\omega\Delta_t-r_xk_x\Delta_x-r_zk_z\Delta_z)}\right)\left(e^{j\omega\Delta_t}-1\right), \tag{46}$$

and from the right side

$$\begin{aligned}&\sum_{l=0}^{3}\frac{C_{FD}[l](\mathbf{E}_{y_{r_x+1+l,r_z}}^{n+1/2}-\mathbf{E}_{y_{r_x-l,r_z}}^{n+1/2})}{\Delta_x}\\ &=\frac{1}{\Delta_z}\left(\mathbf{E}_{oy}e^{j(n\omega\Delta_t-r_xk_x\Delta_x-r_zk_z\Delta_z)}\right)\left(\sum_{l=0}^{3}C_{FD}[l]e^{\frac{j\omega\Delta_t}{2}}\left(e^{-j(1+l)k_x\Delta_x}-e^{jlk_x\Delta_x}\right)\right).\end{aligned} \tag{47}$$

Putting the Equations (46) and (47) together, then

$$\begin{aligned}&\mathbf{H}_{oz}e^{j(n\omega\Delta_t-r_xk_x\Delta_x-r_zk_z\Delta_z)}=\\ &\frac{-\Delta_t}{\mu\Delta_x}\left(\mathbf{E}_{oy}e^{j(n\omega\Delta_t-r_xk_x\Delta_x-r_zk_z\Delta_z)}\right)\left(\sum_{l=0}^{3}C_{FD}[l]\frac{\left(e^{\frac{j\omega\Delta_t}{2}}\left(e^{-j(1+l)k_x\Delta_x}-e^{jlk_x\Delta_x}\right)\right)}{(e^{j\omega\Delta_t}-1)}\right).\end{aligned} \tag{48}$$

Taking the Equations (45) and (48) and replacing in Equation (42), where the term $\mathbf{E}_{oy}e^{j(n\omega\Delta_t-r_xk_x\Delta_x-r_zk_z\Delta_z)}$ is simplyfied, then

$$\begin{aligned}&\frac{\varepsilon}{\Delta_t}\left(2jsin\left(\frac{\omega\Delta_t}{2}\right)\right)=\\ &\frac{\Delta_t}{\mu\Delta_z^2}\left(\sum_{l=0}^{3}C_{FD}[l]\frac{\left(e^{\frac{j\omega\Delta_t}{2}}\left(e^{-j(1+l)k_z\Delta_z}-e^{jlk_z\Delta_z}\right)\right)}{(e^{j\omega\Delta_t}-1)}\sum_{l=0}^{3}C_{FD}[l]\left(e^{-jlk_z\Delta_z}-e^{j(l+1)k_z\Delta_z}\right)\right)\\ &+\frac{\Delta_t}{\mu\Delta_x^2}\left(\sum_{l=0}^{3}C_{FD}[l]\frac{\left(e^{\frac{j\omega\Delta_t}{2}}\left(e^{-j(1+l)k_x\Delta_x}-e^{jlk_x\Delta_x}\right)\right)}{(e^{j\omega\Delta_t}-1)}\sum_{l=0}^{3}C_{FD}[l]\left(e^{-jlk_x\Delta_x}-e^{j(l+1)k_x\Delta_x}\right)\right),\end{aligned} \tag{49}$$

Reagruping terms and defining $\Delta_x = \Delta_z = \Delta_h$, the following expresion is obtained

$$
\begin{aligned}
&2j\sin\left(\frac{\omega\Delta_t}{2}\right) = \\
&\frac{\Delta_t^2}{\mu\varepsilon\Delta_h^2}\left(\sum_{l=0}^{3} C_{FD}[l]\frac{e^{-jk_z\Delta_h/2}\left(-e^{j(1/2+l)k_z\Delta_h}+e^{-j(1/2+l)k_z\Delta_h}\right)}{\left(e^{\frac{j\omega\Delta_t}{2}}-e^{\frac{-j\omega\Delta_t}{2}}\right)}\right) \\
&\left(\sum_{l=0}^{3} C_{FD}[l]e^{jk_z\Delta_h/2}\left(-e^{j(1/2+l)k_z\Delta_h}+e^{-j(1/2+l)k_z\Delta_h}\right)\right) \\
&+\frac{\Delta_t^2}{\mu\varepsilon\Delta_h^2}\left(\sum_{l=0}^{3} C_{FD}[l]\frac{e^{-jk_x\Delta_h/2}\left(-e^{j(1/2+l)k_x\Delta_h}+e^{-j(1/2+l)k_x\Delta_h}\right)}{\left(e^{\frac{j\omega\Delta_t}{2}}-e^{\frac{-j\omega\Delta_t}{2}}\right)}\right) \\
&\left(\sum_{l=0}^{3} C_{FD}[l]e^{jk_x\Delta_h/2}\left(-e^{j(1/2+l)k_x\Delta_h}+e^{-j(1/2+l)k_x\Delta_h}\right)\right).
\end{aligned}
\tag{50}
$$

Using Euler's expansion, $e^{j\theta}+e^{-j\theta} = 2\cos(\theta)$, $e^{j\theta}-e^{-j\theta} = 2j\sin(\theta)$, then

$$
\begin{aligned}
&2j\sin\left(\frac{\omega\Delta_t}{2}\right) = \\
&\frac{\Delta_t^2}{\mu\varepsilon\Delta_h^2}\left(\sum_{l=0}^{3} C_{FD}[l]\frac{(-2j\sin((1/2+l)k_z\Delta_h))}{\left(2j\sin(\frac{\omega\Delta_t}{2})\right)}\right)\left(\sum_{l=0}^{3} C_{FD}[l]\,(-2j\sin((1/2+l)k_z\Delta_h))\right) \\
&\frac{\Delta_t^2}{\mu\varepsilon\Delta_h^2}\left(\sum_{l=0}^{3} C_{FD}[l]\frac{(-2j\sin((1/2+l)k_x\Delta_h))}{\left(2j\sin(\frac{\omega\Delta_t}{2})\right)}\right)\left(\sum_{l=0}^{3} C_{FD}[l]\,(-2j\sin((1/2+l)k_x\Delta_h))\right).
\end{aligned}
\tag{51}
$$

simplyfing terms

$$\sin^2\left(\frac{\omega\Delta_t}{2}\right) = \frac{\Delta_t^2}{\mu\varepsilon\Delta_h^2}\left(\left(\sum_{l=0}^{3} C_{FD}[l]\sin\left(\left(\frac{1}{2}+l\right)k_z\Delta_h\right)\right)^2 + \left(\sum_{l=0}^{3} C_{FD}[l]\sin\left(\left(\frac{1}{2}+l\right)k_x\Delta_h\right)\right)^2\right). \tag{52}$$

Finally, ω is given by

$$\omega = \frac{2}{\Delta_t}\sin^{-1}\left(\frac{\Delta_t}{\sqrt{\mu\varepsilon}\Delta_h}\sqrt{\left(\sum_{l=0}^{3} C_{FD}[l]\sin\left(\left(\frac{1}{2}+l\right)k_z\Delta_h\right)\right)^2 + \left(\sum_{l=0}^{3} C_{FD}[l]\sin\left(\left(\frac{1}{2}+l\right)k_x\Delta_h\right)\right)^2}\right). \tag{53}$$

If $k_x = k\cos(\theta)$, $k_z = k\sin(\theta)$, $k = \sqrt{k_x^2 + k_z^2}$ and $\mathfrak{B} = \frac{\Delta_t}{\sqrt{\varepsilon\mu}\Delta_h}$, then

$$\omega = \frac{2}{\Delta_t}\sin^{-1}\left(\mathfrak{B}\sqrt{\left(\sum_{l=0}^{3} C_{FD}[l]\sin\left(\left(\frac{1}{2}+l\right)k\cos(\theta)\Delta_h\right)\right)^2 + \left(\sum_{l=0}^{3} C_{FD}[l]\sin\left(\left(\frac{1}{2}+l\right)k\sin(\theta)\Delta_h\right)\right)^2}\right). \tag{54}$$

The stability condition is obtained when the Equation (54) reaches its maximum value, that is

$$\omega = \frac{2}{\Delta_t}\sin^{-1}\left(\mathfrak{B}\sqrt{3.3092}\right). \tag{55}$$

When $\mathfrak{B}\sqrt{3.3092} \leq 1$, the electromagnetic field is real, therefore the stability condition is given by

$$\mathfrak{B} \leq \frac{1}{\sqrt{3.3092}} \leq 0.5497. \tag{56}$$

The non-dispersive medium is defined when the relation $\frac{\mathbf{v}_p}{c} = 1$ or $\frac{\omega}{kc}$, where $\mathbf{v}_p$ is the phase velocity and c is the speed light velocity in air. Therefore

$$\frac{\mathbf{v}_p}{c} = \frac{2}{\mathfrak{B} k \Delta_h} \sin^{-1}\left(\mathfrak{B}\sqrt{\left(\sum_{l=0}^{3} C_{FD}[l] \sin\left(\left(\frac{1}{2}+l\right) k \cos(\theta)\Delta_h\right)\right)^2 + \left(\sum_{l=0}^{3} C_{FD}[l] \sin\left(\left(\frac{1}{2}+l\right) k \sin(\theta)\Delta_h\right)\right)^2}\right). \tag{57}$$

Figures 13, 14 and 15 shows the behavior of the relation $\frac{\mathbf{v}_p}{c}$ with $\mathfrak{B}$ equal to 0.1 for 2^{nd}, 4^{th} and 8^{th} order for derivatives in space, respectively. Non-centered finite difference are used (FDN). The numerical dispersion is avoided if the relation $\frac{\mathbf{v}_p}{c}$ is close to one. Therefore, we recommend select $k\Delta_h \leqslant 1$. According to Figures 13, 14 and 15, the 8^{th} order space derivative aproximation reach more frequencies with the relation $\frac{\mathbf{v}_p}{c} = 1$ and the propagation is less dispersive. 8^{th} order is selected in our implementation for the derivatives in space. Although 8^{th} order discretization in space, increases the computational cost and execution time, these issues are solved using GPU architectures. The time derivatives is selected in 2^{nd} order, and an increase of the discretization time does not affect the final results of FWI.

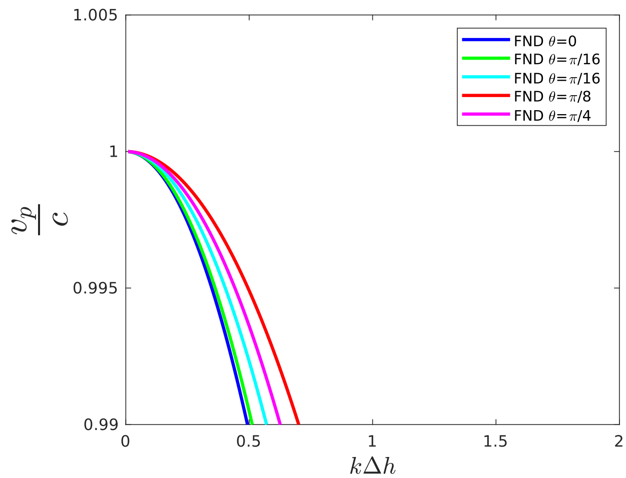

Figure 13. Numerical dispersion for $\theta = \pi/16, \pi/8, \pi/4, \pi/3$ with $\mathfrak{B} = 0.1$ and 2^{nd} order in space.

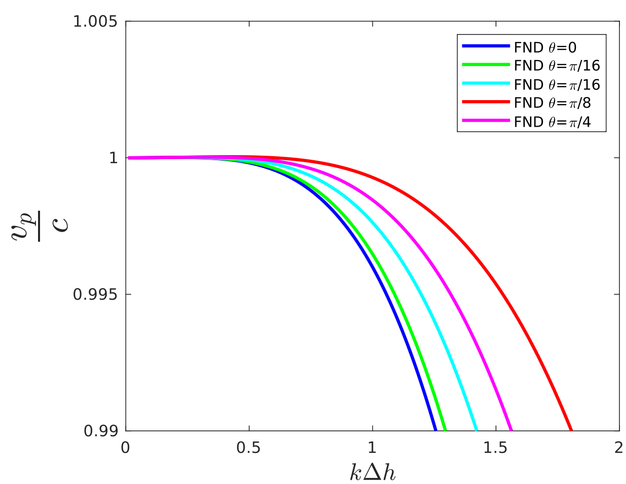

Figure 14. Numerical dispersion for $\theta = \pi/16, \pi/8, \pi/4, \pi/3$ with $\mathfrak{B} = 0.1$ and 4^{th} order in space.

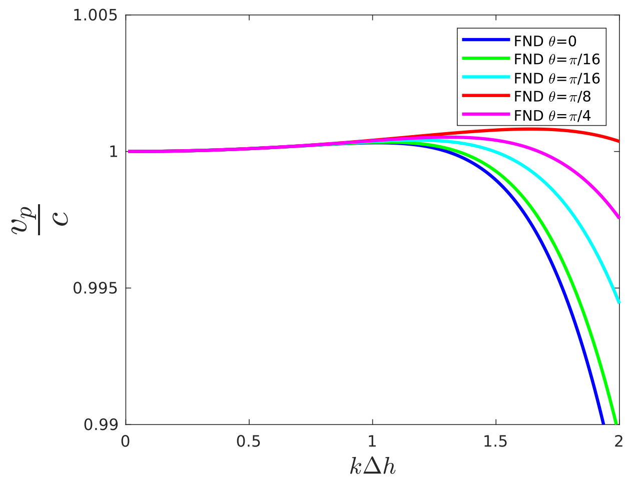

Figure 15. Numerical dispersion for $\theta = \pi/16, \pi/8, \pi/4, \pi/3$ with $\mathfrak{B} = 0.1$ and 8^{th} order in space.

3.2. The inverse problem

The inverse problem consists on estimating the parameters of the subsurface $\mathbf{m}$ ($\varepsilon_r(\mathbf{r})$, $\mu_r(\mathbf{r})$ and $\sigma_r(\mathbf{r})$) from the observations on the surface $\mathbf{d}(\mathbf{r},t)$ (radargram), and a known source ($\mathbf{J}_s(\mathbf{r},t)$). Since the number of independent observations usually is lower than the number of parameters to be estimated, then the inverse problem is an ill-posed problem (Snieder and Trampert, 1999) [1]. Standard Full Waveform Inversion (FWI) minimizes the ℓ^2-norm between the modeled data $\mathbf{d}_{mod}(\mathbf{m}^k)$ and observed data $\mathbf{d}_{obs}$ (Virieux and Operto, 2009), as

$$\Phi(m_k) = \sum_{i}^{Ns} \frac{1}{2} ||\mathbf{d}^i_{mod}(\mathbf{m}^k) - \mathbf{d}^i_{obs}||^2_2, \tag{58}$$

[1] A well posed problem has three characteristics (Hadamard, 1902): a solution exists; the solution is stable, small changes in the input do not produce large changes in the output, and the solution is unique

with

$$\mathbf{d}_{mod}^{i}(\mathbf{m}^{\mathbf{k}}) = R\mathbf{E}_{y}^{i}(\mathbf{r},t), \tag{59}$$

where $\mathbf{m}^k$ are the unknown parameters in the k-th iteration, Ns the number of sources used and the operator R extracts the electromagnetic wavefield $\mathbf{E}_y(\mathbf{r},t)$ at the receiver positions. The inverse problem is solved iteratively by selecting an initial model $\mathbf{m}^0$, and updating it by using Newton-like methods (Goldstein, 1965), as

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \alpha_k \Delta\mathbf{m}^k, \tag{60}$$

where α_k is the step size and $\Delta\mathbf{m}^k$ at the k^{th} iteration is given by

$$\Delta\mathbf{m}^k = -[\mathfrak{h}(\mathbf{m}^k)]^{-1}\mathbf{g}(\mathbf{m}^k). \tag{61}$$

The inverse of the Hessian matrix $[\mathfrak{h}(\mathbf{m}^k]^{-1}$ and the gradient $\mathbf{g}(\mathbf{m}^k)$ are both evaluated at $\mathbf{m}^k$. As the inverse problem is ill-posed, the cost function given in Equation (58) has several minimums and an inadequate starting point ($\mathbf{m}^0$) will produce convergence to a local minimum.

3.2.1. Mathematical formulation for the FWI. The adjoint method (Plessix, 2006) is an alternative method to obtain the gradients avoiding computing the Fréchet derivatives. The minimization problem can be formulated using the Lagrange operators where the general minimization problem is given by

$$\min_{\mathbf{m}} \chi(s,\mathbf{m}) = \int_0^{Tend} \Phi(s,\mathbf{m})dt, \tag{62}$$

subject to

$$W(\ddot{s}, \dot{s}, s, \mathbf{m}) = 0, \tag{63}$$

$$u(s(0), \mathbf{m}) = 0, \tag{64}$$

$$du(\dot{s}(0), \mathbf{m}) = 0, \tag{65}$$

where $\Phi(s, \mathbf{m})$ is the cost function given in Equation (58), W is the forward operator (Equations (16), (17), (18)), s is a discretized vector of the fields, the dots denote derivatives in time and $Tend$ is the final time. The initial conditions are denoted as $u(s(0), \mathbf{m})$ and $du(\dot{s}(0), \mathbf{m})$ for the field vectors s and $\dot{s}$. The lagrangian ($\mathscr{L}$) is defined as

$$\mathscr{L}(\mathbf{m}, \Phi, W, u, du) = \int_0^{Tend} \left[\Phi(s, \mathbf{m}) + \lambda^T W(\ddot{s}, \dot{s}, s, \mathbf{m})\right] dt + \mathfrak{u}^T u(s(0), \mathbf{m}) + \mathfrak{n}^T du(\dot{s}(0), \mathbf{m}); \tag{66}$$

where ^{T} is the transpose operator and λ^T, $\mathfrak{u}^T$, $\mathfrak{n}^T$ are the Lagrange multipliers. Applying chain rule $\frac{\partial \mathscr{L}}{\partial \mathbf{m}}$ and assumming that $\lambda(T)$, $\dot{\lambda}(T)$, $s(0)$ and $\dot{s}(0)$ are zero for initial conditions, then

$$\begin{aligned}\frac{\partial \mathscr{L}}{\partial \mathbf{m}} = \int_0^{Tend} [\frac{\partial s}{\partial \mathbf{m}} \left(\frac{\partial \Phi}{\partial s} + \ddot{\lambda}^T \frac{\partial W}{\partial \ddot{s}} + \dot{\lambda}^T \left(2\frac{\partial}{\partial t}\frac{\partial W}{\partial \ddot{s}} - \frac{\partial W}{\partial \dot{s}}\right) + \lambda^T \left(\frac{\partial^2}{\partial t^2}\frac{\partial W}{\partial \ddot{s}} + \frac{\partial W}{\partial s} - \frac{\partial}{\partial t}\frac{\partial W}{\partial \dot{s}}\right)\right) \\ + \frac{\partial \Phi}{\partial \mathbf{m}} + \lambda^T \frac{\partial W}{\partial \mathbf{m}} + \mathfrak{m}^T \frac{\partial u}{\partial \mathbf{m}} + \mathfrak{n}^T \frac{\partial du}{\partial \mathbf{m}}]dt,\end{aligned} \tag{67}$$

however, the initial condition are independent from the model paremeters. Therefore, $\frac{\partial u}{\partial \mathbf{m}}$ and $n^T \frac{\partial du}{\partial \mathbf{m}}$ are zero. Note that $\frac{\partial s}{\partial \mathbf{m}}$ is never computed and therefore the first term in Equation (67) is cancelled using the operator

$$\ddot{\lambda}^T \frac{\partial W}{\partial \ddot{s}} + \dot{\lambda}^T \left(2 \frac{\partial}{\partial t} \frac{\partial W}{\partial \ddot{s}} - \frac{\partial W}{\partial \dot{s}} \right) + \lambda^T \left(\frac{\partial^2}{\partial t^2} \frac{\partial W}{\partial \ddot{s}} + \frac{\partial W}{\partial s} - \frac{\partial}{\partial t} \frac{\partial W}{\partial \dot{s}} \right) = -\frac{\partial \Phi}{\partial s}. \tag{68}$$

The Equation (68) is known as the adjoint operator (Plessix, 2006) (Jaap, 2018). Finally, the gradients are given by

$$\frac{\partial \mathscr{L}}{\partial \mathbf{m}} = \int_0^{Tend} \left[\frac{\partial \Phi}{\partial \mathbf{m}} + \lambda^T \frac{\partial W}{\partial \mathbf{m}} \right] dt. \tag{69}$$

3.3. Adjoint operator for the inverse problem

The electromagnetic wave equations in a non-dispersive and isotropic medium are considered where the fields $\mathbf{E}_y(\mathbf{r},t)$, $\mathbf{H}_x(\mathbf{r},t)$ and $\mathbf{H}_z(\mathbf{r},t)$ are different from zero. The forward operator is described by $W(\ddot{s}, \dot{s}, s, \mathbf{m})$, where

Equation (63) is rewritten in matrix form as

$$W(\ddot{s}, \dot{s}, s, \mathbf{m}) = F_1(\mathbf{m})\dot{s} + F_2(\mathbf{m})s - \mathbf{J}_s(\mathbf{r},t) = 0 \tag{70}$$

or

$$W(\ddot{s},\dot{s},s,\mathbf{m}) = \begin{cases} \varepsilon_r(\mathbf{r})\varepsilon_0 \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial t} - \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial z} + \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial x} + \sigma_r(\mathbf{r})\sigma_0 \mathbf{E}_y(\mathbf{r},t) - \mathbf{J}_s(\mathbf{r},t), \\ \mu_r(\mathbf{r})\mu_0 \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial t} - \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial z}, \\ \mu_r(\mathbf{r})\mu_0 \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial t} + \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial x}. \end{cases} \tag{71}$$

where

$$F_1(\mathbf{m}) = \begin{bmatrix} \varepsilon_r(\mathbf{r})\varepsilon_0 & 0 & 0 \\ 0 & \mu_r(\mathbf{r})\mu_0 & 0 \\ 0 & 0 & \mu_r(\mathbf{r})\mu_0 \end{bmatrix}, \tag{72}$$

$$F_2(\mathbf{m}) = \begin{bmatrix} \sigma_r(\mathbf{r}) & -\frac{\partial}{\partial z} & \frac{\partial}{\partial x} \\ -\frac{\partial}{\partial z} & 0 & 0 \\ \frac{\partial}{\partial x} & 0 & 0 \end{bmatrix}, \tag{73}$$

$$s = \begin{bmatrix} \mathbf{E}_y(\mathbf{r},t) \\ \mathbf{H}_x(\mathbf{r},t) \\ \mathbf{H}_z(\mathbf{r},t) \end{bmatrix}, \tag{74}$$

$$\mathbf{m} = \begin{bmatrix} \varepsilon_r(\mathbf{r}) \\ \\ \mu_r(\mathbf{r}) \\ \\ \sigma_r(\mathbf{r}) \end{bmatrix}. \tag{75}$$

Using the Equations (68) and (138) the adjoint operator in matrix form is

$$-F_1^T(\mathbf{m})\dot{\lambda} + F_3(\mathbf{m})\lambda + \frac{\partial \Phi}{\partial s} = 0, \tag{76}$$

$$\lambda^T = \left[\lambda_{\mathbf{E}_y}(\mathbf{r},t), \lambda_{\mathbf{H}_x}(\mathbf{r},t), \lambda_{\mathbf{H}_z}(\mathbf{r},t)\right], \tag{77}$$

where $F_1^T(\mathbf{m}) = F_1(\mathbf{m})$ and

$$F_3(\mathbf{m}) = \begin{bmatrix} \sigma_r(\mathbf{r})\sigma_0 & \frac{\partial}{\partial z} & -\frac{\partial}{\partial x} \\ \\ \frac{\partial}{\partial z} & 0 & 0 \\ \\ -\frac{\partial}{\partial x} & 0 & 0 \end{bmatrix}. \tag{78}$$

Replacing $F_1^T(m)$ and $F_3(m)$ in Equation (76) , the adjoint operator for the traditional cost function (Φ) is defined as

$$\varepsilon_r(\mathbf{r})\varepsilon_0 \frac{\partial \lambda_{\mathbf{E}_y}(\mathbf{r},t)}{\partial t} - \frac{\partial \lambda_{\mathbf{H}_x}(\mathbf{r},t)}{\partial z} + \frac{\partial \lambda_{\mathbf{H}_z}(\mathbf{r},t)}{\partial x} - \sigma_r(\mathbf{r})\sigma_0 \lambda_{\mathbf{E}_y}(\mathbf{r},t) + \frac{\partial \Phi}{\partial s} = 0; \tag{79}$$

where $\frac{\partial \Phi}{\partial s} = \mathbf{d}_{mod}(s) - \mathbf{d}_{obs}$.

$$\mu_r(\mathbf{r})\mu_0 \frac{\partial \lambda_{\mathbf{H}_x}(\mathbf{r},t)}{\partial t} - \frac{\partial \lambda_{\mathbf{E}_y}(\mathbf{r},t)}{\partial z} = 0; \tag{80}$$

$$\mu_r(\mathbf{r})\mu_0 \frac{\partial \lambda_{\mathbf{H}_z}(\mathbf{r},t)}{\partial t} + \frac{\partial \lambda_{\mathbf{E}_y}(\mathbf{r},t)}{\partial x} = 0. \tag{81}$$

The Equations (79), (80), and (81) are discretized using the same scheme described in the Figure 9. The equations (79), (80), and (81) are very similar with the Maxwel equations (16), (17) and (18), with the difference in some signs and the source that is used. The fields $\lambda_{\mathbf{E}_y}(\mathbf{r},t)$, $\lambda_{\mathbf{H}_z}(\mathbf{r},t)$, $\lambda_{\mathbf{H}_x}(\mathbf{r},t)$ are computed using the same FDTD scheme that simulates electromagnetic waves propagation but in reverse time. In the adjoint operator, the source $\frac{\partial \Phi}{\partial s}$ is the difference between the modeled data and the observed data and it depends on the cost function used.

3.3.1. Gradients for the inverse problem for GPR. Rewriting Equation (69) in matrix form, then

$$\frac{d\mathscr{L}}{d\mathbf{m}} = \int_0^{Tend} \lambda^T \left(\frac{\partial F_1(\mathbf{m})}{\partial \mathbf{m}} \dot{s} + \frac{\partial F_2(\mathbf{m})}{\partial \mathbf{m}} s - \frac{\partial \mathbf{J}_s(\mathbf{r},t)}{\partial \mathbf{m}} \right) dt. \tag{82}$$

The following expressions are defined to obtain the gradients for each parameter

$$g(\varepsilon_r(\mathbf{r})) = \frac{d\mathscr{L}}{d\varepsilon_r(\mathbf{r})} = \int_0^{Tend} [\lambda_{\mathbf{E}_y}(\mathbf{r},t), \lambda_{\mathbf{H}_x}(\mathbf{r},t), \lambda_{\mathbf{H}_z}(\mathbf{r},t)] \begin{bmatrix} \varepsilon_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial t} \\ \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial t} \\ \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial t} \end{bmatrix} dt, \tag{83}$$

$$g(\varepsilon_r(\mathbf{r})) = \int_0^{Tend} \varepsilon_0 \lambda_{\mathbf{E}_y}(\mathbf{r},t) \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial t} dt, \tag{84}$$

$$g(\mu_r(\mathbf{r})) = \frac{d\mathscr{L}}{d\mu_r(\mathbf{r})} = \int_0^{Tend} [\lambda_{\mathbf{E}_y}(\mathbf{r},t), \lambda_{\mathbf{H}_x}(\mathbf{r},t), \lambda_{\mathbf{H}_z}(\mathbf{r},t)] \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mu_0 & 0 \\ 0 & 0 & \mu_0 \end{bmatrix} \begin{bmatrix} \frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial t} \\ \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial t} \\ \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial t} \end{bmatrix} dt, \tag{85}$$

$$g(\mu_r(\mathbf{r})) = \int_0^{Tend} \mu_0 \left(\lambda_{\mathbf{H}_x}(\mathbf{r},t) \frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial t} + \lambda_{\mathbf{H}_z}(\mathbf{r},t) \frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial t} \right) dt, \tag{86}$$

$$g(\sigma_r(\mathbf{r})) = \frac{d\mathscr{L}}{d\sigma_r(\mathbf{r})} = \int_0^{Tend} [\lambda_{\mathbf{E}_y}(\mathbf{r},t), \lambda_{\mathbf{H}_x}(\mathbf{r},t), \lambda_{\mathbf{H}_z}(\mathbf{r},t)] \begin{bmatrix} \sigma_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_y(\mathbf{r},t) \\ \mathbf{H}_x(\mathbf{r},t) \\ \mathbf{H}_z(\mathbf{r},t) \end{bmatrix} dt, \tag{87}$$

$$g(\sigma_r(\mathbf{r})) = \int_0^{Tend} \sigma_0 \lambda_{\mathbf{E}_y}(\mathbf{r},t)\mathbf{E}_y(\mathbf{r},t)dt. \tag{88}$$

3.3.2. Gradient descent method. The gradient descent method is a method to find the local minimum in a differentiable function. The gradient descent method takes a constant (α) to scale the gradient where the parameter is updated in the opposite directions of the gradient. Figure (16) shows an example of a cost function, and the update direction using the gradient. Some authors refer to the gradient descent method like how to descent a hill covered in mist, where the gradient is point to a local minimum, which may or may not coincide with the global minimum, and α is how fast to go down the hill. Following this analogy when the frequency increases, the cost function contains more rocks and trees that makes visibility to the global minimum difficult. The starting point defines the location on the hill to start the search for the global minimum. If the starting point does not have enough low-frequency information, the location on the hill is too high, therefore more likely to be trapped in a minimum local during the parameters inversion.

In this section, the results of FWI over a synthetic model are shown. The synthetic case is a realistic model called SEAM Foothills-Phase II (Regone et al., 2017) that is built to represent the complex geological underground of a Foothills area. The SEAM model's selected area is rescaled around 80 times from its initial dimensions to have a section of 12.5 (m) in depth and 30 m in the dip direction. The spatial discretization is 5 (cm), such that we generate 600 points in the x-direction (dip) and 301 points in z-direction (depth). A total of 320 scans are simulated using the equations for the forward electromagnetic modeling. Transmitter and receiver antennas are located in the air-layer at 25 (cm) of the ground layer, according to the topography. The time sampling of the data is 0.08 (ns), and the number of time samples is 3000; thus, the total acquisition time for

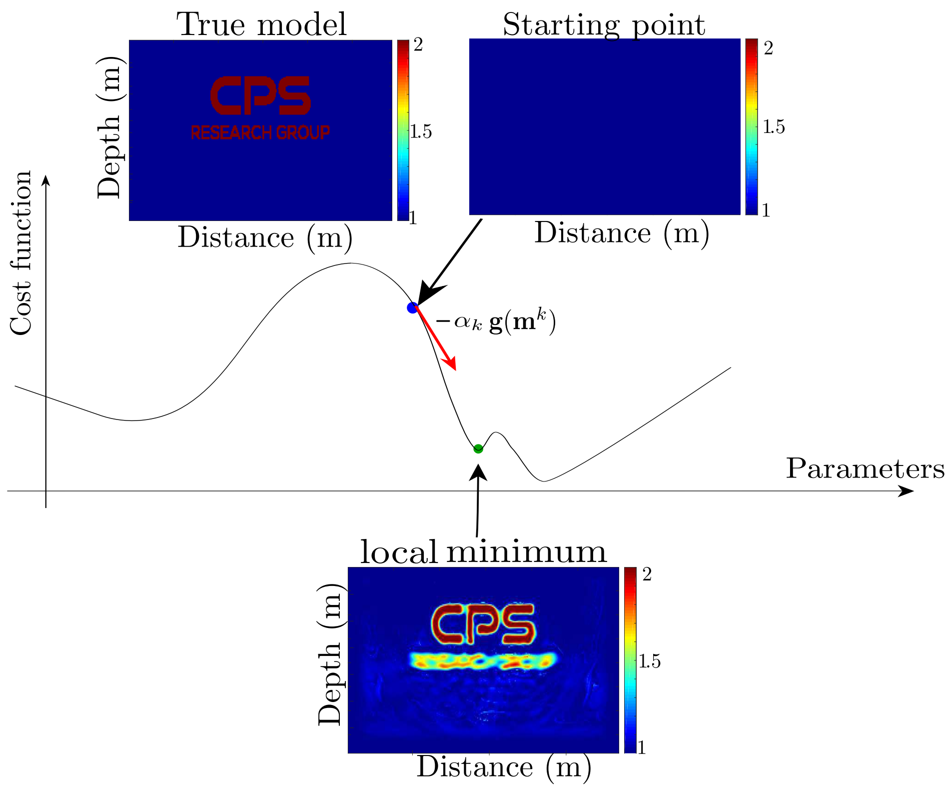

Figure 16. Example of starting point with gradient descent method

each scan is 240 (ns). The scheme of FDTD with 8^{th} and 2^{nd} order of approximation is used to discretize the spatial and time derivatives, respectively.

Three frequencies are selected in the inversion process: 30 (MHz), 50 (MHz) and 100 (MHz). The multiscale methodology proposed by (Bunks et al., 1995) is used, where the FWI process started from low frequency to high frequency, and the final model obtained in each frequency is used as a starting point for the next frequency step. Figure 17-a) and 17-b) show the starting points used in the FWI methodology for the relative permittivity and conductivity, respectively.

Figure 20 shows the final models of FWI with the gradient descent method. The experiments use the condition $(\Phi(\mathbf{m}^k - \alpha_k \mathbf{g}(\mathbf{m}^k)) < \Phi(\mathbf{m}^k)$ where forward propagation and backward propaga-

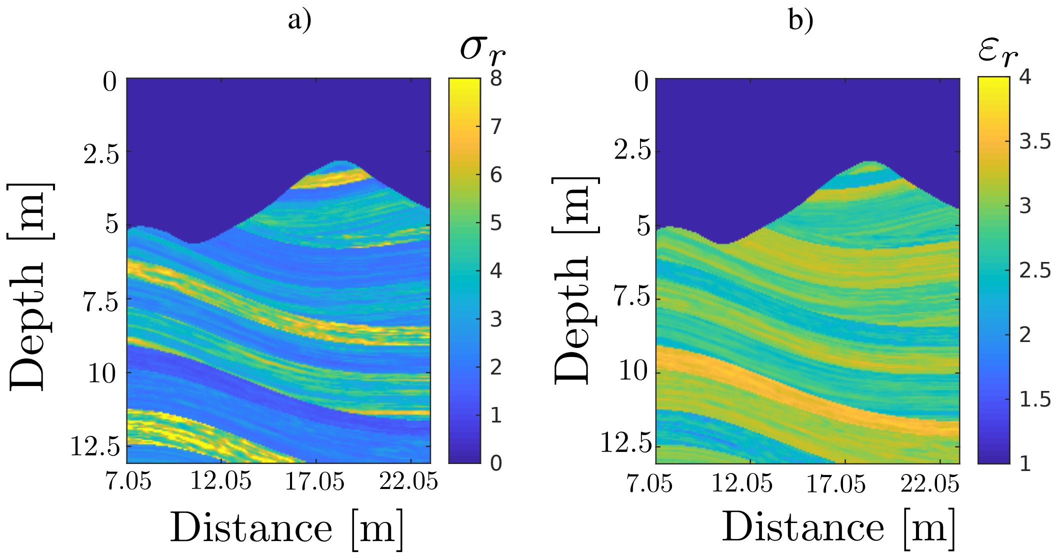

Figure 17. a) Relative conductivity adapted from SEAM, b) Relative permittivity adapted from SEAM.

tion are used to calculate the gradient and determine if the step length is accepted. In the gradient descent experiments or in the first iteration of L-BFGS the α_k is obtained as $1/||(\mathbf{g}(m^k))||_2$. The results obtained using the descending gradient method are still far from the global minimum and require more iterations, which implies more computational cost. To speed up this convergence, Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) is used in the following subsection, where the Hessian matrix information is incorporated allows improving the estimation in both parameters.

3.3.3. L-BFGS method. The method L-BFGS proposed by (Liu and Nocedal, 1989) is a modified version of the BFGS, where the total number of gradients and models that are stored in memory is reduced. The L-BFGS is a Quasi-Newton method that computes an approximation of the product between the inverse of the Hessian matrix and the gradient. The main

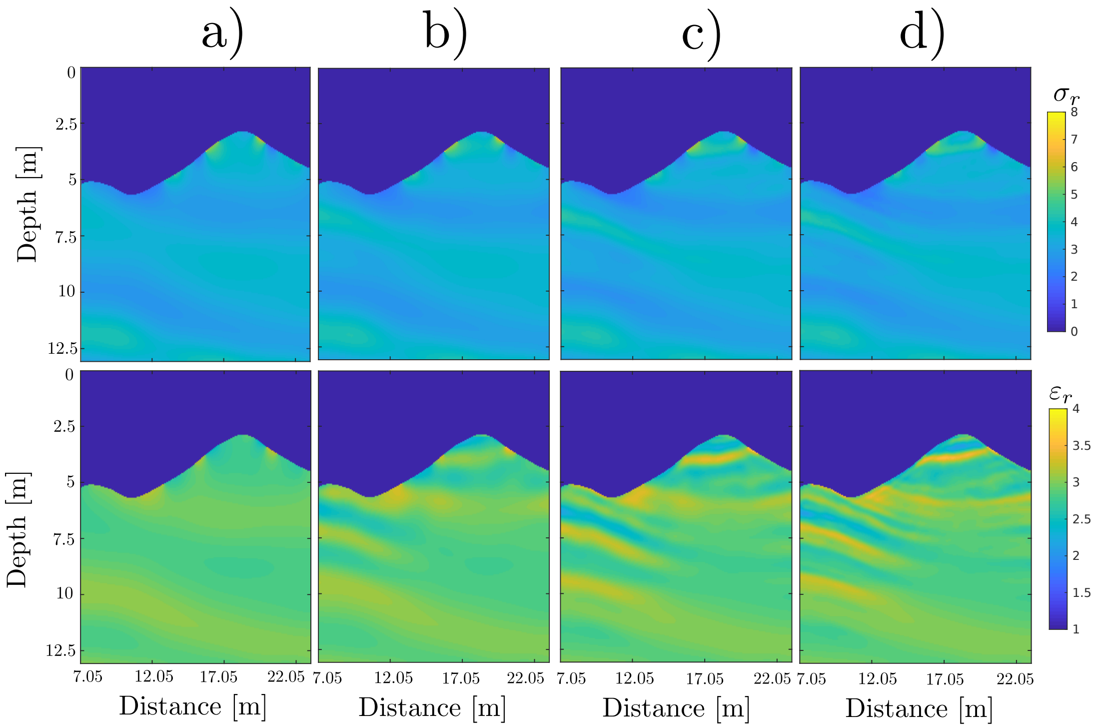

Figure 18. FWI models using gradient descent, where the first and second row shows the conductivity and permittivity, respectively: a) initial model, b) FWI Models after 20 iterations at 30 (MHz), c) FWI Models after 20 iterations at 50 (MHz), d) FWI Models after 20 iterations at 100 (MHz).

advantage of the L-BFGS method is to avoid computing the inverse of the Hessian matrix because of order increase from $\mathscr{O}(n)$ to $\mathscr{O}(n^2)$, where n is the number of elements to be estimated. The L-BFGS method only requests a history of gradients and models on the last ℓ iterations. Usually, the parameter ℓ is selected to be 10. The L-BFGS method computes the parameters s_k as the difference between models, and y_k as the difference between gradients. The L-BFGS method is a better approximation than gradient descent because the information of the Hessian matrix related to the cost function's concavity is included. The parameters are updated according to:

$$\mathbf{m}^{k+1} = \mathbf{m}^{k} - \alpha \cdot r_{opt}, \tag{89}$$

where the parameter r_{opt} is computed using the algorithm 1. The L-BFGS algorithm is tested using the same number of sources, frequencies, starting point, and geometry that the proposed in the section **3.3.2** (Serrano et al., 2019). Figure 20 shows the final models of FWI with the L-BFGS method. The Peak Signal to Noise Ratio (PSNR) metric is computed by

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{\sqrt{MSE}}\right), \tag{90}$$

where MAX_I is the maximum value in the true model and MSE is computed as follow:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[\mathbf{m}^{true}(i,j) - \mathbf{m}^{k}(i,j)]^2. \tag{91}$$

The PSNR metric is compute in both experiments Gradient descent and L-BFGS; the results are shown in Figure 19. The L-BFGS method, which includes the product of the inverse of the Hessian matrix with the gradient in the inversion process, provides better PSNR values than the Gradient descent method. However, the L-BFGS method requires an additional propagation in each iteration because r_{opt} is evaluated in the cost function to accept the advance. The inverse of the Hessian matrix includes concavity information in the cost function, allowing each dimension in the gradient to be scaled appropriately and not as a global value as in the Gradient descent.

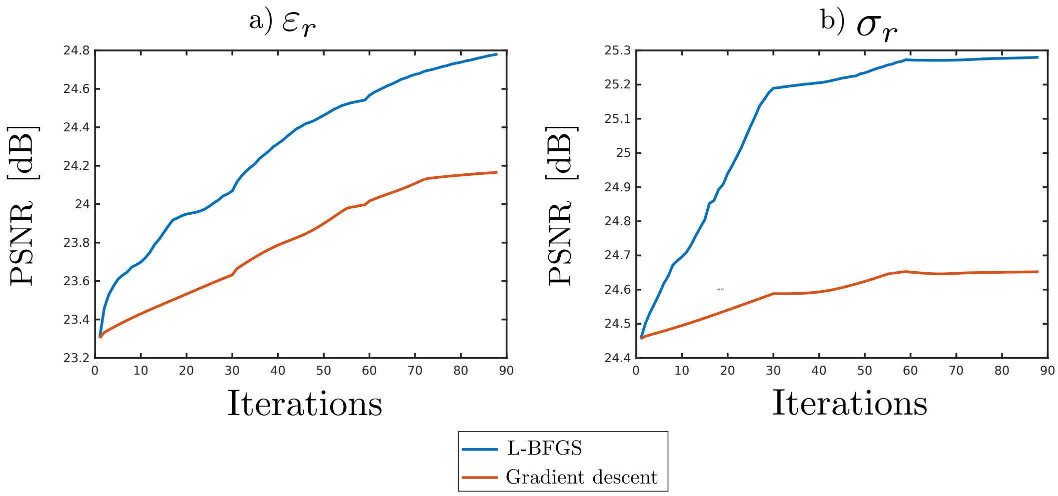

Figure 19. PSNR between the original model and the FWI models for: Gradient descent method and L-BFGS method: a) relative permittivity and b) relative conductivity.

Algorithm 1 L-BFGS

$\mathbf{s}_k = \mathbf{m}_{k+1} - \mathbf{m}_k$
$\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$
$\sigma_k = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}$
$\mathbf{q} \leftarrow \mathbf{g}_k$
for i=k-1:-1:k-m **do**
    $\varepsilon_i \leftarrow \sigma_i \mathbf{s}_i^T \mathbf{q}$
    $\mathbf{q} \leftarrow \mathbf{q} - \varepsilon_i \mathbf{y}_i$
end for
$\gamma_k = \frac{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}{\mathbf{y}_{k-1}^T \mathbf{y}_{k-1}}$
$H_k^0 = \gamma_k \mathbf{I}$
$r_{opt} \leftarrow H_k^0 \mathbf{q}$
for i=k-m:1:k-1 **do**
    $\beta_i \leftarrow \sigma_i \mathbf{y}_i^T r_{opt}$
    $r_{opt} \leftarrow r_{opt} + \mathbf{s}_i(\varepsilon_i - \beta_i)$
end for

3.3.4. Lack of illumination in short-offset acquisition for FWI. According to Sirgue and Pratt (2004), the configuration of source and receivers influences the inversion results of FWI. For a particular 1D case, with a fixed offset between source and receiver and a given

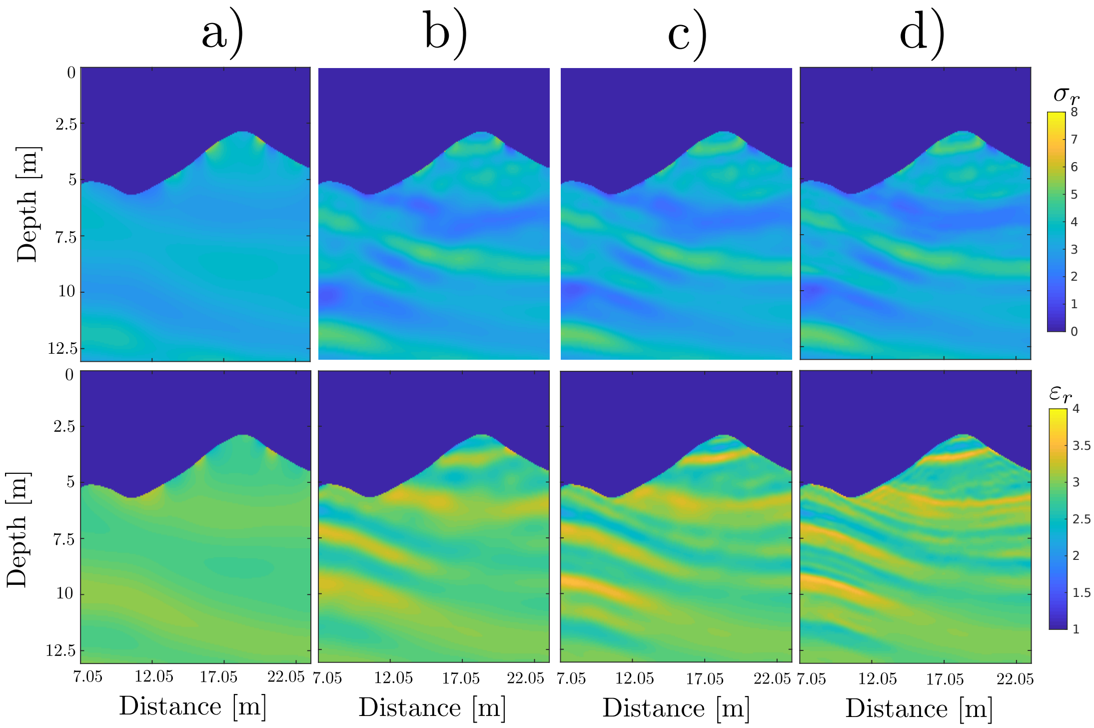

Figure 20. FWI models using L-BFGS, where the first and second row shows the relative conductivity and relative permittivity, respectively: a) initial model, b) FWI Models after 20 iteration at 30 (MHz), c) FWI Models after 20 iteration at 50 (MHz), d) FWI Models after 20 iteration at 100 (MHz).

frequency, the inversion only solves a vertical wavenumber k_z. Consider the following example, an interface is located at z (m) in-depth, the ray propagates the electromagnetic wave at a specific angle θ, and the distance between source and receiver is $2h$ (m). The trajectory from source to the interface is $\check{s}$, and from the interface to the receiver is $\check{r}$, as it is shown in Figure 21. The wavenumber in the common midpoint (CMP) is related with $k_0(\check{s}+\check{r})$, where k_0 is the background wave number and it is defined as

$$k_0 = \frac{f_c}{c}, \tag{92}$$

where f_c is the central frequency, and c is the speed of the light in air. The incident and scattering angles in 1D case are symmetric, then $k_0\check{s} = k_0 \cos\theta + k_0 \sin\theta$ and $k_0\check{r} = -k_0 \sin\theta + k_0 \cos\theta$ (Sirgue and Pratt, 2004). From Figure 21 then the following Equations are defined

$$\begin{aligned} \cos\theta &= \frac{z}{\sqrt{h^2+z^2}}, \\ \sin\theta &= \frac{h}{\sqrt{h^2+z^2}}. \end{aligned} \tag{93}$$

Using the definition in Equation (93) then the wavenumber located at CMP position is

$$\begin{aligned} k_x &= 0, \\ k_z &= \frac{2k_0 z}{\sqrt{h^2+z^2}}. \end{aligned} \tag{94}$$

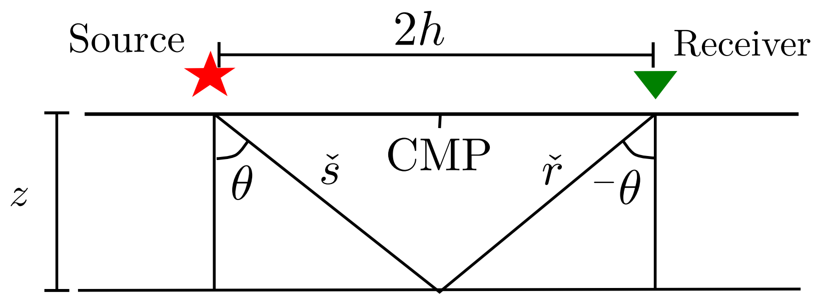

Figure 21. Explanation about FWI wavenumbers when changing the offset.

A synthetic case is implemented to understand the lack of illumination in the gradient for single-channel and short-offset acquisitions. A model with a single interface is used, as it is shown in Figure 22.

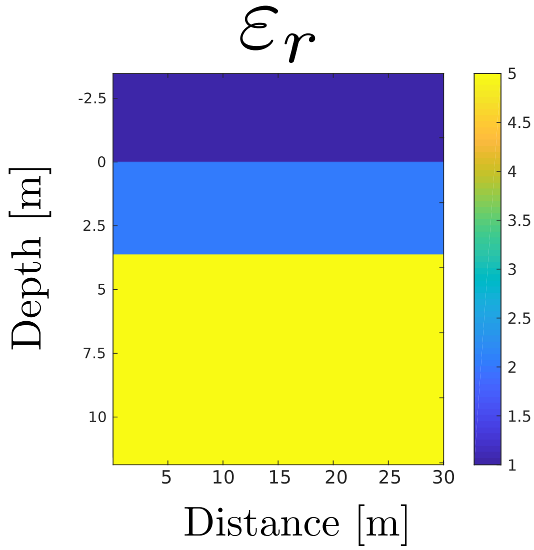

Figure 22. Permittivity parameter model used to test the lack of illumination problems, in the gradients..

The source position is fixed and it is located at $x = 2$ (m) and $z = 3.25$ (m). 520 scans are acquired where the receiver position changes, its position goes in x from 2 (m) to 28 (m) with a step 0.05 (m), and in z it is always $z = 3.25$ (m). Three frequencies are used in this synthetic case at 30 (MHz), 50 (MHz) and 100 (MHz). The results of the first gradient (CMP 0) and the last gradient (CMP 519) for 30 (MHz) are presented in Figure 23-a) and Figure 23-b), respectively. The CMP column is taken of each pair transmitter-receiver to make a new Figure that related Depth vs CMP as it is shown in Figures 24-a), 24-b) and 24-c) for frequencies 30 (MHz), 50 (MHz), and 100 (MHz), respectively. Remember that in a CMP position for the gradient, only the vertical component in the wavenumber k_z is different from zero.The Fourier transform is applied in each CMP-gradient of Figure 24 to obtain the Figure 25 that relates the wavenumbers k_z vs offset.

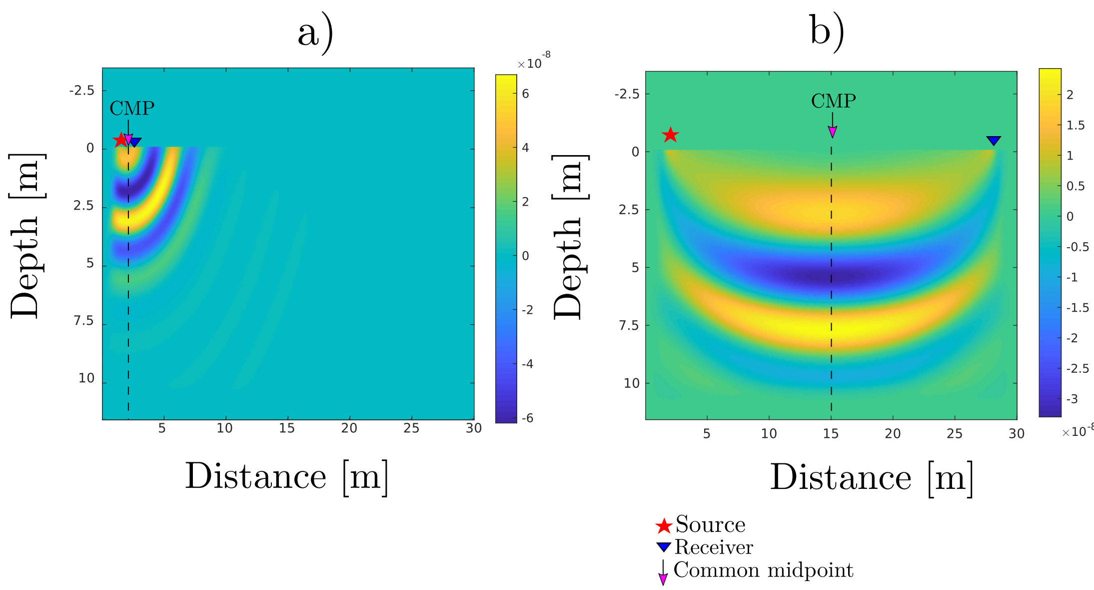

Figure 23. The gradient for different offsets values for 30 (MHz): a) First CMP, zero offsets; b) Last CMP, offset 27.95 (m).

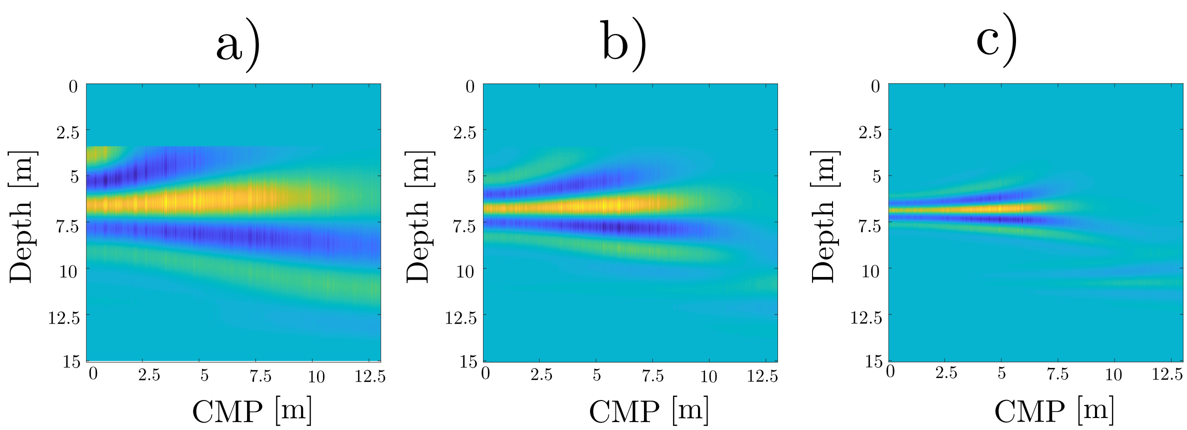

Figure 24. Depth vs CMP of each frequency: a) 30 (MHz), b) 50 (MHz) and c) 100 (MHz).

Figure 25 shows that when the offset increase, then the inversion takes into account smaller wavenumbers. This behavior is observed for all frequencies (30, 50, and 100 (MHz)). In other words, the single-channel and short-offset inversion is more sensitive to the starting point than a multi-channel acquisition. This is a critical point for the Full Waveform Inversion.

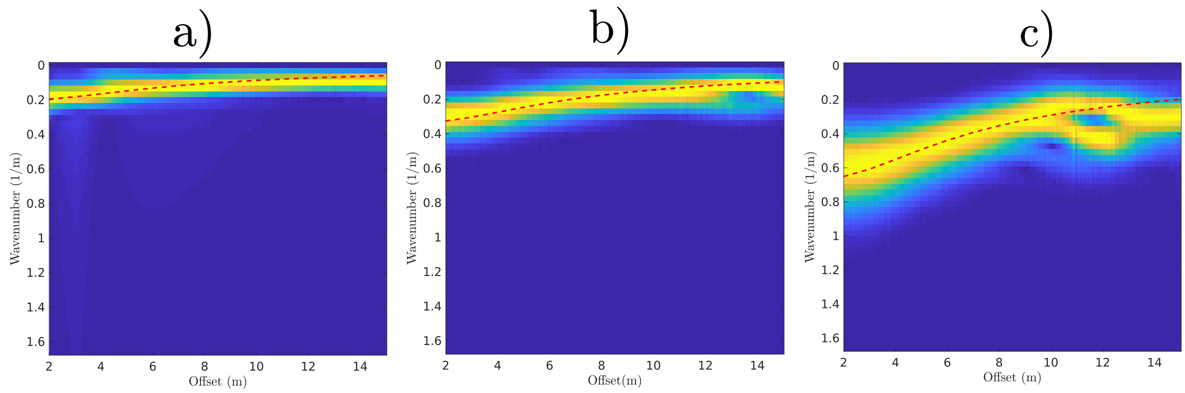

Figure 25. k-offset domain for each CMP gradient: a) 30 (MHz), b) 50 (MHz) and c) 100 (MHz). The red dash line is the theoretical wavenumber, Equation (94)

Three synthetic experiments are proposed to show the critical point concerning the starting point. In all the experiments the true conductivity and permittivity are selected according to Figures 17-a) and 17-b), respectively. The spatial and time step are selected in 0.05 (m) and 0.08 (ns), respectively. The propagation time is 240 (ns). A multi-scale methodology is used with frequencies of 30 (MHz), 50 (MHz) and 100 (MHz). Two acquisitions are used: single-short-offset acquisition and multi-channel acquisition. The single-short-offset acquisition uses 260 sources located each 10 (cm), and in-depth the sources are located in the air layer at 25 (cm) of the ground layer, according to the topography. The receiver antenna is located at the same location as the transmitter. In the multi-channel acquisition, only one source is used with 260 receivers. According to the topography, the receivers are located in the air layer at 25 (cm) of the ground layer. In the first experiment, the initial guess is selected at 2 for the relative permittivity and 1.11×10^{-3} (S/m) for the conductivity. A blurring filter is applied 100 times over the true models to obtain the starting point in the second experiment.

Figure 26 shows the results for the two scenarios proposed in the two configurations. Column a) shows the correct parameters that also are used to obtain the PSNR metric. Column b) presents the starting points used to start the inversion methodology. Column c) shows the FWI models after the FWI methodology in multi-scale configuration; in this column, the single-channel and short-offset configuration are used. Column d) refers to the models after FWI methodology; in this column, the multi-channel scheme is selected. The first and second rows present the results of using the starting point with a homogeneous layer. Note that in the results of the short-offset configuration contains high wave numbers and oscillating artefacts, however the mains structures of permittivity are visible. The lack of low wavenumbers produce a mislocate in the high-permittivity layers as it is shown in column-c) of Figure 26.

The multi-offset configuration has more wavenumbers, and the solution managed to get closer to the correct model. When the starting point with enough low wavenumbers information is used, rows three and four, the results of the two inversions have a behavior similar. The PSNR values for the conductivity in the multi-offset configuration are lower than for the short-offset acquisition, as it is shown in Figure 27. However, the results of permittivity and conductivity are very close. Again, this verifies the importance of selecting a starting point with enough low wavenumbers information in the single-channel and short acquisition, being an important point to obtain good results in FWI.

The same number of observations to solve the unknown parameters is used. The single-channel and short-offset acquisition could reach up to 520 sources, this being its maximum number of sources in the model. However, the single-channel and short-offset acquisition still have low illumination compared to multi-channel acquisition. In a multi-channel acquisition, we have only used one source, and it is possible to increase the number of the sources and improve the illumination, which allows a better estimation in both parameters when using a starting point without enough low-frequency information.

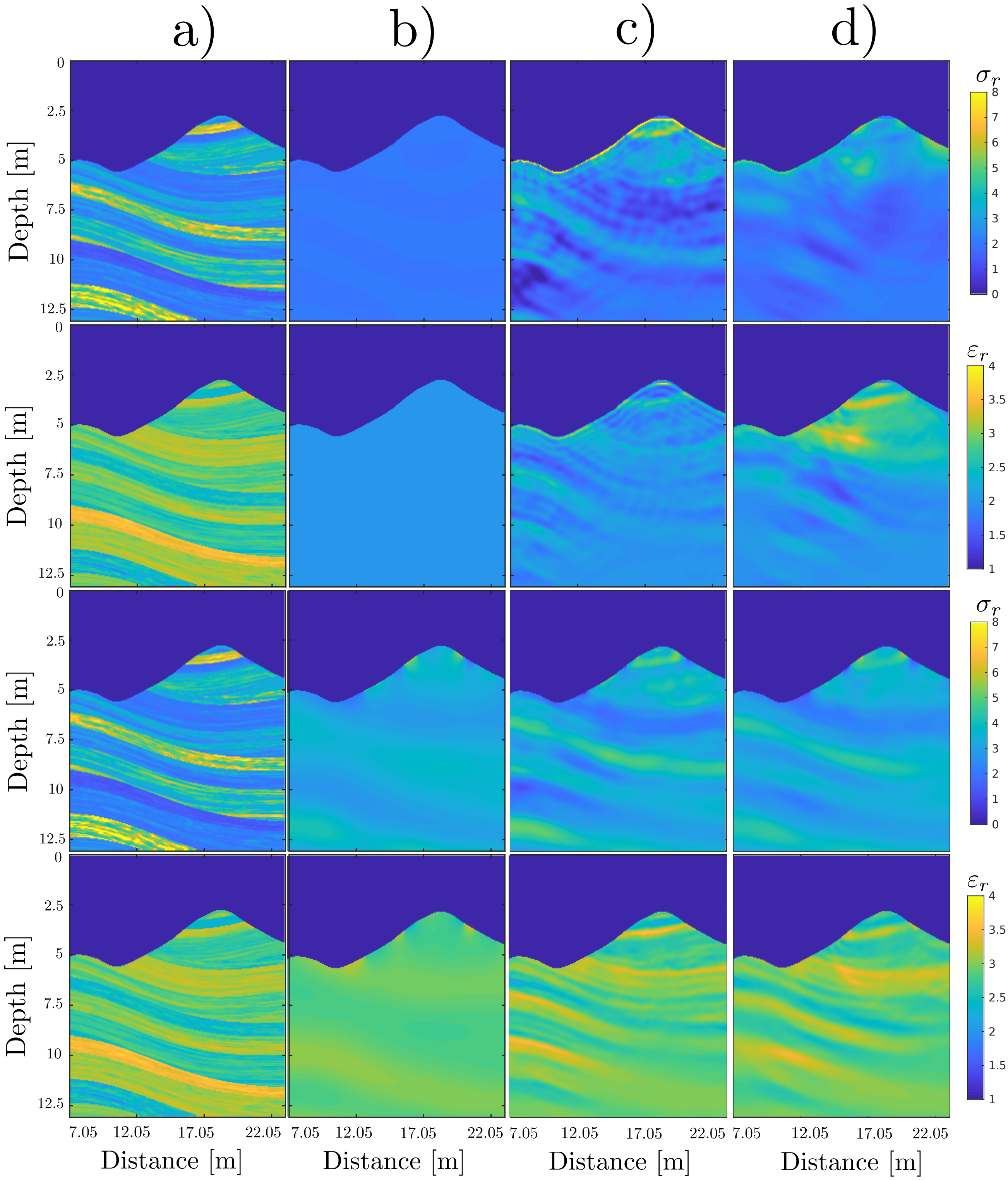

Figure 26. The behavior of the FWI in short-offset and single-channel vs multi-offset/single-source: the first and third rows are the relative conductivity, and the second and fourth rows are the relative permittivity. a) True models, b) starting points c) Results of FWI in single-channel and short-offset configuration; d) multi-channel configuration.

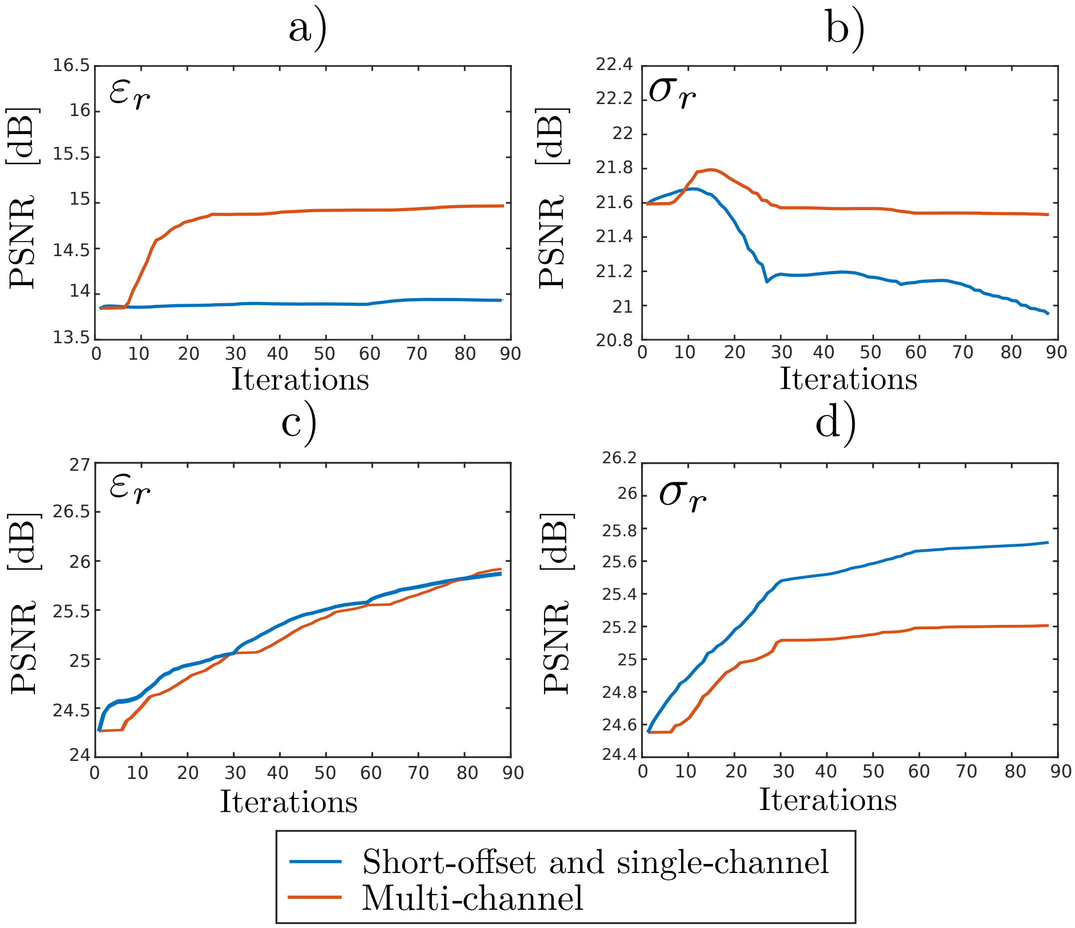

Figure 27. PSNR metric between the correct models and the models obtained after the FWI methodology: a) and b) relative permittivity and conductivity using as the starting point a constant value of 2 for the relative permittivity and conductivity; c) and d) relative permittivity and conductivity using as the starting point a smoothing version of the correct model.

A final experiment is proposed to compare a dense single-offset configuration vs a less dense multi-offset and multi-source acquisition. The number of observations is balanced between both configurations: in multi-offset and multi-source 20 sources and 26 offsets are used and in single-offset configuration 520 sources. The source and receivers' location in multi-offset/multi-sources configuration are equidistant over the model. Figure 28 show the relative conductivity and permittivity results and Figure 29 shows the achieved PSNR values. Although in both parameters the less-dense multi-offset and multi-source configuration reach better PSNR values than the dense single-channel acquisition, note that the estimated values in both parameters for the first reflectors are largely comparable and present a correct location for the first 7.5m. The multi-offset and multi-source configuration improves the estimation of the deepest reflector locations as well as improves the estimation of the permittivity and conductivity parameters.

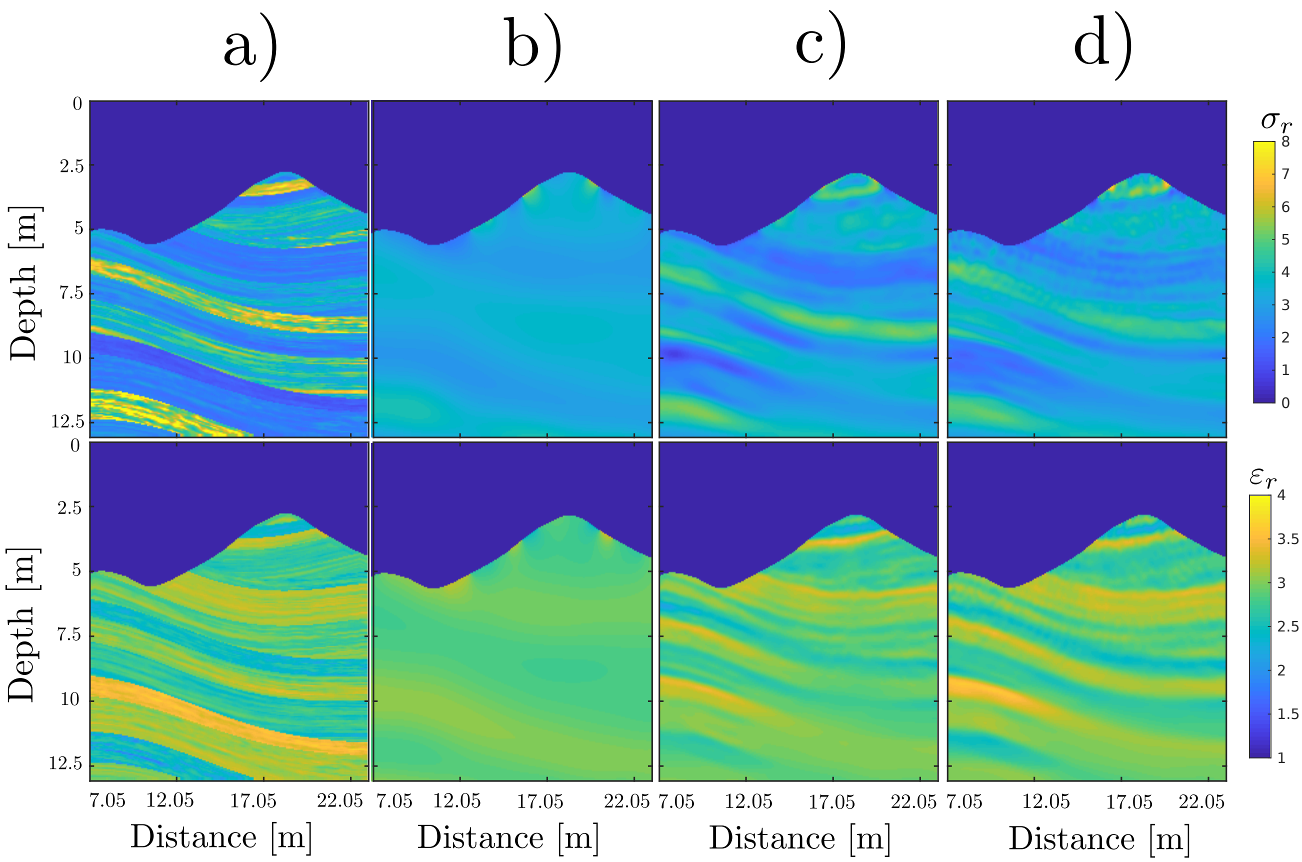

Figure 28. The behavior of the FWI in short-offset /single-channel vs multi-offset/multi-source: the first is the relative conductivity, and the second is the relative permittivity. a) True models, b) starting points c) Results of FWI in single-channel and short-offset configuration; d) multi-channel configuration.

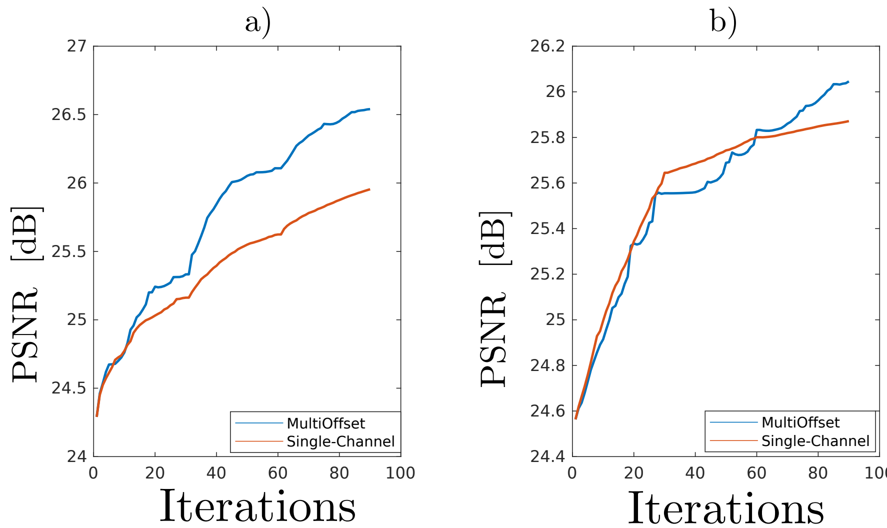

Figure 29. PSNR metric in the FWI models in short-offset and single-channel vs multi-offset/multi-source: a) relative permittivity and b) relative conductivity.

According to Feng et al. (2020), our multi-scale inversion strategy is sequential, which means that the FWI model obtained in one frequency is used as a starting point for the next frequency. The sequential strategy help to mitigate the nonlinearity and the ill-posed problem. Feng et al. (2020) proposed the Bunks strategy, where the lack of illumination problem is mitigated in short-offset acquisitions and the strategy uses a frequency hierarchy and broad frequency bandwidth.

3.3.5. Selecting scans-sampling in single short-offset acquisition. FWI in single-channel and short-offset acquisition has a severe problem because the high wavenumbers are solved during inversion, such that it is susceptible to the starting point. This type of acquisition requires many scans on the exploration-line to improve the horizontal resolution. Each additional scan requires one propagation and one backpropagation to obtain the gradient. The execution time, and

RAM memory in a GPU to obtain a gradient in a model with 600 points in the horizontal direction, 301 points in the vertical direction, and 3000 step times is presented in Table 2. The implementation has been developed using a hybrid architecture GPU-CPU using a cluster PE ProLiant XL270d Gen10 with two Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz and eight NVIDIA Tesla V100, 16 GB. The execution times are very challenging for a Full Waveform inversion in a 2D domain in single-channel and short offset configuration. As presented in the table 2, a gradinet takes about 2.8 (s) on GPU (1.4 (s) per scan). For a gradient of 260 scans, 520 scans are needed, 260 for the forward fields and 260 for the backward fields. Additionally, we need another 260 propagations when L-BFGS is used because it is necessary to evaluate if the new model is better than the previous one, so in total, 780 propagations are needed per iteration. Leaving 30 iterations per frequency and using three frequencies, this could take 21.4 hours. We used a shot decomposition with 8 GPUs, and our implementation reduced to 2.6 hours per experiment. In a three-dimensional case, the scenario is not very pleasant, a 3D propagation can take around 220 (s) in GPUs in a small model of 380 in x, by 380 in y, and by 360 in z, and a number of time samples of 6233, this could take approximately 4290 hours. Due to this high computational cost, we have decided to carry out our experiments in a 2D scenario. For more details on the implementation of FWI, you can refer to the pseudocode introduced in the annexes section A. In this subsection, we propose a new rule in single channel and short offset configuration that can reduce the execution time. In the following experiments, the number of scans used in the gradient is changed.

The results of FWI with six different scans-sampling are tested: 16, 32, 65, 130, 260, 520. The scans are distributed equidistant, and the receiver antenna is located at the same location at

# scans	Execution time (s)	RAM CPU (GB)	RAM GPU (GiB)
1	2.8	2	0.8
260	728	2	0.8

Table 2. *Execution time in a GPU to obtain a gradient with Nx=600, Nz=301 and Nt=3000.*

the transmitter. The SEAM model is again used with a spatial step of 0.05 and a time sampling of 800 ns with 3000 time samples. Three frequencies are used with 30 iterations per frequency: 30 (MHz), 50 (MHz) and 100 (MHz).

The Figures 30 and 31 show the FWI results and the PSNR when the scans change for single short-offset acquisition. Figure 30 shows that when there are very few scans, the gradients create gaps causing the inversion to converge to incorrect values. According to the PSNR values (Figure 31), the conductivity is highly affected by the gaps where the PSNR decrease for the scans sampling: 16, 32, 65, and 130. As the number of scans increases, the horizontal resolution improves, and there is a greater correlation in the events. However, as it is presented in Table 2, each additional scan increases the execution time of the algorithm. If the models have n points in x-directions, FWI in single-channel and short-offset acquisition must have at least $n/2$ scans to guarantee a correct convergence and avoid the gaps. In these tests, no regularization is being used, and the stopping criteriums are: when L-BFGS tries ten times without reducing the cost function or by the number of iterations established by frequency, which is 30 iterations. For this reason, the PSNR in Figure 31 for 260 scans shows that although the permittivity parameter continues to improve, the PSNR in the conductivity parameter decrease. Using the horizontal resolution described in the previous chapter we can establish that:

$$\Delta l = \sqrt{\frac{c \cdot \mathfrak{r}}{2 \cdot 4 \cdot f_c \cdot \sqrt{\varepsilon_r}}} = \sqrt{\frac{3e8 \cdot 0.05}{2 \cdot 4 \cdot 100e6 \cdot \sqrt{4}}}, \tag{95}$$

where ε_r is the maximum value in the relative permittivity, it is taken from the initial model; f_c is the maximum central frequency used in FWI; and $\mathfrak{r}$ is the distance from the subsurface to the object, it is assumed as d_h according to with the discretization. Replacing in the previous equation, we found that at least one scan is needed for every 0.0968 (m), so according to our discretization, at least 260 scans are needed to avoid losing horizontal resolution. With this rule we reduce the computational cost by a factor of 2x.

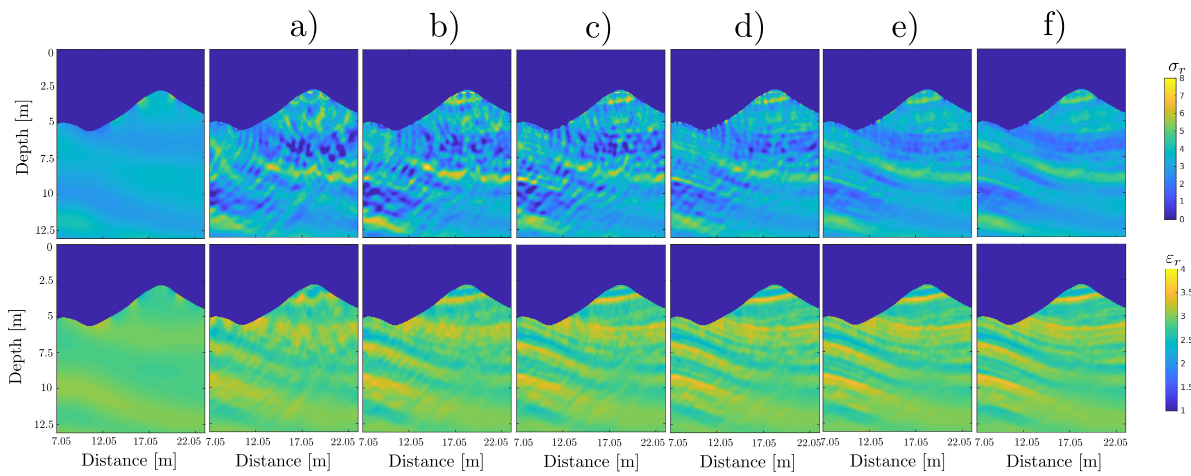

Figure 30. FWI models for different number of scans in single short-offset acquisition: a) 16, b) 32, c) 65 , d) 130, e) 260, and f) 520.

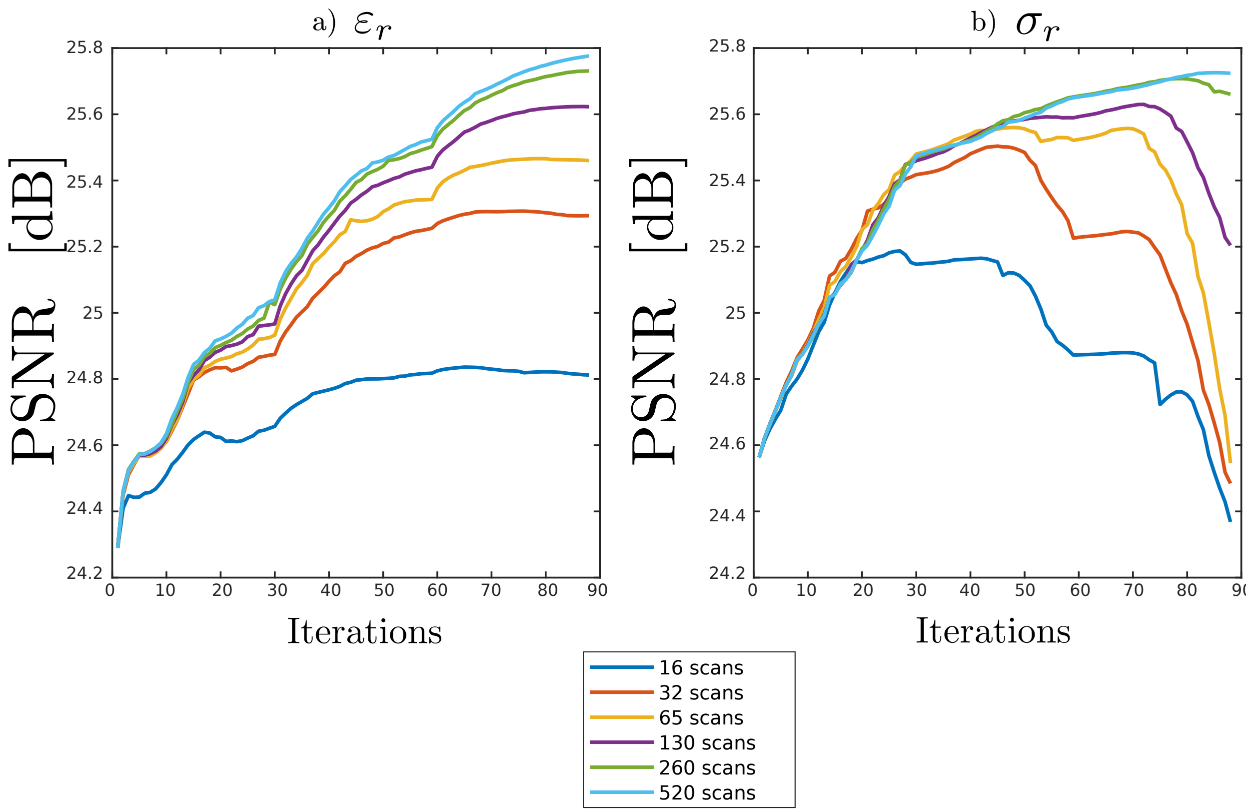

Figure 31. PSNR metric for different number of scans: 16, 32, 65, 130, 260, and 520: a) relative permittivity and b) relative conductivity.

4. Constrains and alternative cost functions

In this chapter, two regularization are applied in the model domain, and one preprocessing is applied in the data domain. The objective of these regularizations and preprocessing is to prevent that FWI reaches unwanted values. We have explored a cost function that normalizes the observed and modeled data and three regularizations that provide a smooth version of the model and improve the fit of the data between the modeled and observed data. These regularizations require many experiments for tuning the regularization parameters. However, this section shares some values that are useful as starting criteria for future research. In this chapter, four alternative cost functions are explored: (1) The normalization seeks to compensate the amplitude between the observed and modeled data; (2) the alternative cost function with Gaussian preprocessing produces convexity in the cost function and ignores the high-frequency information. It allows reaching better electromagnetic parameters with noising data and starting point without low-wavenumbers information; (3) Two alternative cost functions are proposed to obtain a smoothed solution of the parameters. These cost functions mitigate the effect of noise from the data and serve as a new starting point in FWI.

4.1. Alternative cost function based on data normalization in the FWI methodology

In a real scenario, the electromagnetic source is usually estimated from selected measured data through linear inversion (Pratt, 1999). However, the scan depends on the antenna inclination, the source amplitude, and the radiation patterns in each plane. The amplitude difference between observed and modeled data could be compensated in the source estimations or, in our case is, solved using a normalized cost function. When the amplitude difference is not taken into account, the FWI process can produce artifacts at the near-surface, and the parameters converge into a local minimum. The alternative cost function normalize the amplitude of the predicted data $\mathbf{d}_{mod}(s)$ with respect to the amplitude of the observed data $\mathbf{d}_{obs}$.The alternative cost function is used because the AVO (amplitude vs offset) is not important in single and short offset data and the data can be normalized trace by trace. However, in multi-offset data, the data can not be normalize trace by trace, and it should be normalized using all the shot-gather or scaling the source estimation. Mathematically, the alternative cost function is given by

$$\Phi_A = \sum_{i=1}^{Ns} \frac{1}{2} \left|\left| \frac{\mathbf{d}_{mod}(s) \cdot ||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2} - \mathbf{d}_{obs} \right|\right|_2^2 ; \tag{96}$$

In the traditional case, the adjoint source is defined by:

$$\frac{\partial \Phi}{\partial s} = \mathbf{d}_{mod}(s) - \mathbf{d}_{obs}. \tag{97}$$

However, a new adjoint source is defined for the alternative cost function, and it is given by

$$\frac{\partial \Phi_A}{\partial s} = \left(\frac{\mathbf{d}_{mod}(s) \cdot ||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2} - \mathbf{d}_{obs} \right) \left(\frac{||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2} - \frac{\mathbf{d}_{mod}(s) \cdot \mathbf{d}_{mod}^T(s) \cdot ||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2^3} \right). \tag{98}$$

By simplifying Equation (98), the new adjoint source is given by

$$\frac{\partial \Phi_A}{\partial s} = \frac{\mathbf{d}_{mod}(s) \cdot \mathbf{d}_{mod}^T(s) \cdot \mathbf{d}_{obs} \cdot ||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2^3} - \frac{\mathbf{d}_{obs} \cdot ||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2}. \tag{99}$$

Note that if the observed data is defined as a scaled version in the amplitude of the predicted data that also includes noise, *i.e.,* $\mathbf{d}_{obs} = \kappa \cdot \mathbf{d}_{mod}(\mathbf{m}_{ori}) + \eta$. Where η is the noise, therefore the new adjoint source is given by

$$\frac{\partial \Phi_A}{\partial s} = \frac{\mathbf{d}_{mod}(s) \cdot (\mathbf{d}_{mod}^T(s) \cdot (\kappa \cdot \mathbf{d}_{mod}(\mathbf{m}_{ori}) + \eta)) \cdot \|\mathbf{d}_{obs}\|_2}{\|\mathbf{d}_{mod}(s)\|_2^3} - \frac{\mathbf{d}_{obs} \cdot \|\mathbf{d}_{obs}\|_2}{\|\mathbf{d}_{mod}(s)\|_2}; \tag{100}$$

expanding terms in Equation (100), it is rewritten as:

$$\frac{\partial \Phi_A}{\partial s} = \frac{\kappa \cdot \mathbf{d}_{mod}(s) \cdot (\mathbf{d}_{mod}^T(s) \cdot \mathbf{d}_{mod}(\mathbf{m}_{ori}))\ ||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2^3} + \frac{\mathbf{d}_{mod}(s) \cdot (\mathbf{d}_{mod}^T(s) \cdot \eta) \cdot \|\mathbf{d}_{obs}\|_2}{\|\mathbf{d}_{mod}(s)\|_2^3} - \frac{\mathbf{d}_{obs} \cdot \|\mathbf{d}_{obs}\|_2}{\|\mathbf{d}_{mod}(s)\|_2}. \tag{101}$$

In the real scenario, the observed data has noise in a specific bandwidth. The inverse crime occurs when the same elements are used in an inverse problem to invert data(Wirgin, 2004). The inverse crime is avoided by including white gaussian noise (WGN). The WGN has uniform power across all the band frequencies. The samples generated with the normal distribution has average zero and standard deviation equal to 1. In a real scenario, the observed data is filtered in a specific bandwidth of the electronic instruments. Table 3 summarizes the bandwidth used to filter the WGN in each frequency.

When two variables are uncorrelated, $E[XY] = E[X]E[Y]$, and one of them has an expected value different from zero (orthogonal), the second term on the right side in Equation (99) is relevant

f1 (MHz)	fc (MHz)	f2 (MHz)
15	30	45
25	50	75
50	100	150

Table 3. *Cut-off frequencies to filter the white gaussian noise.* f_1 : *low cut-off* , f_c : *central cut-off,* f_2 : *upper cut-off.*

Serrano et al. (2020). An expected value is selected in each observed data for frequency: $E(\eta) = 0.0022$ (Signal to Noise Ratio, 33 dB), $E(\eta) = 0.0034$ (Signal to Noise Ratio, 31 dB) and $E(\eta) = 0.0062$ (Signal to Noise Ratio, 28 dB) for 30 MHz, 50 MHz and 100 MHz, respectively. Figure 32 shows the observed data adding the WGN.

The results of the FWI, after 30 iterations per frequency are shown in Figure 33. The first column in Figure 33-a) refers to the initial models; second column Figure 33-b) refers to the alternative cost function Φ_A with adjoint source $\frac{\partial \Phi_A}{\partial s}$; the third column Figure 33-c) refers to the cost function Φ with adjoint source $\frac{\partial \Phi}{\partial s}$.

Figure 34 shows the zoom-in zones A and B, corresponding to Figure 33-a) and Figure 33-b), respectively. The alternative cost function allows obtaining a better estimation of the parameters because the adjoint source compensates the norm value of the samples acquired.

The norm value of the samples acquired produces artifacts in the estimation of the near-surface parameters. The PSNR metric is used to measure the difference between the right model and the result of FWI, as it is shown in Figure 35. Both parameters are highly affected by norm value in the adjoint source.

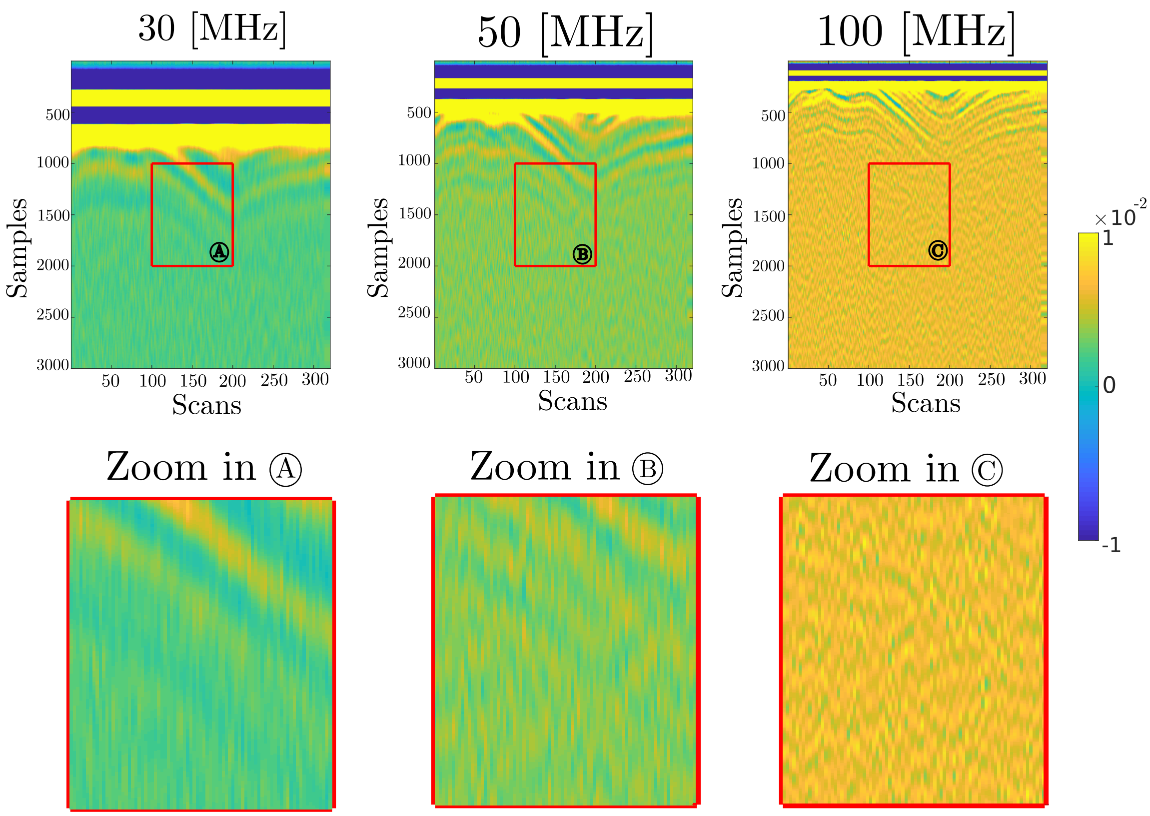

Figure 32. Observed data adding WGN with $E(\eta) = 0.0022$, $E(\eta) = 0.0034$ and $E(\eta) = 0.0062$ for 30 MHz, 50 MHz and 100 MHz, respectively.

4.2. Alternative cost function with Gaussian preprocessing for FWI.

This section shows an alternative cost function using Gaussian preprocessing. According to (Xue et al., 2016), a smoothing parameter is defined and allows to improve the convexity of the cost function. In the strategy, the smoothing parameter starts from high values to low values. A high value in the smoothing parameter produces convexity in the cost function and ignores the high frequency information. On the other hand, a low value in the smoothing parameter incorporates a multimodal function and high frequency information. The method proposed in (Xue et al., 2016) uses a small number of misfit functions with smoothing kernels of decreasing strengths, this allows

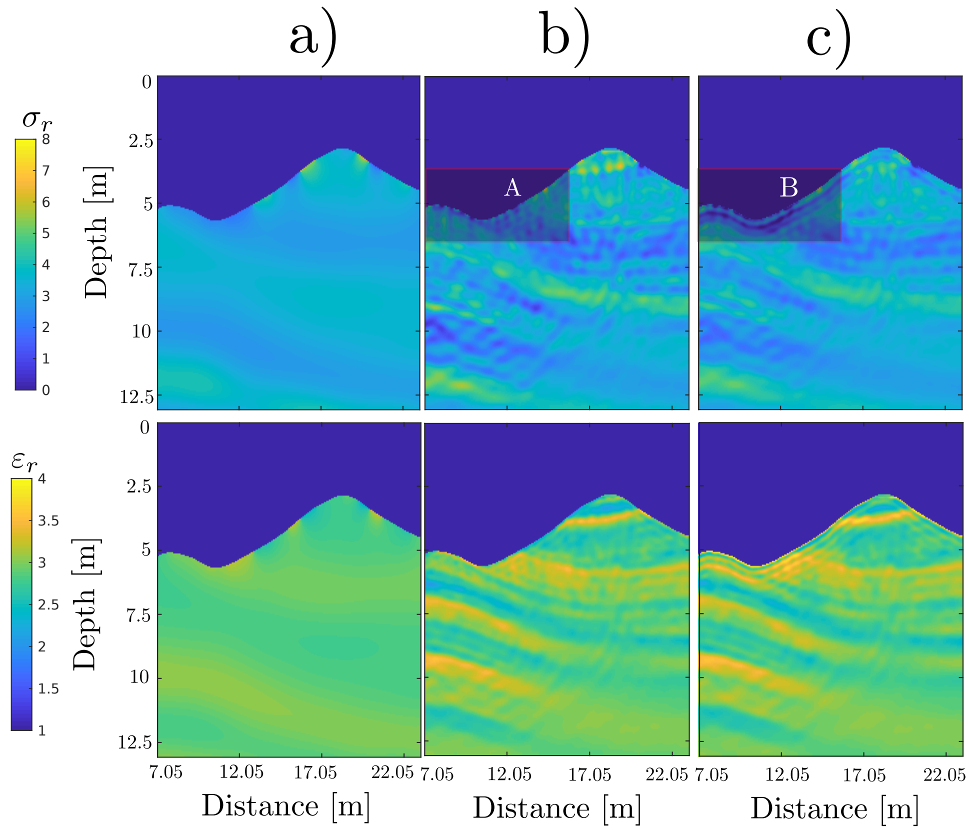

Figure 33. FWI models using L-BFGS, where the first and second row shows the conductivity and permittivity, respectively: a) initial model, b) FWI Models after multiscale (30, 50 and 100 MHz) with the cost function Φ_A and adjoint source $\frac{\partial \Phi_A}{\partial s}$, c) FWI Models after multiscale (30, 50 and 100 MHz) with the cost function Φ and adjoint source $\frac{\partial \Phi}{\partial s}$.

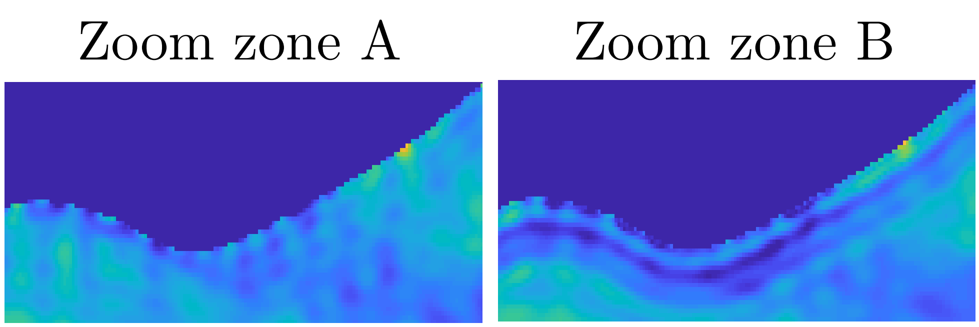

Figure 34. FWI results with zoom in zone A and B from Figure 33.

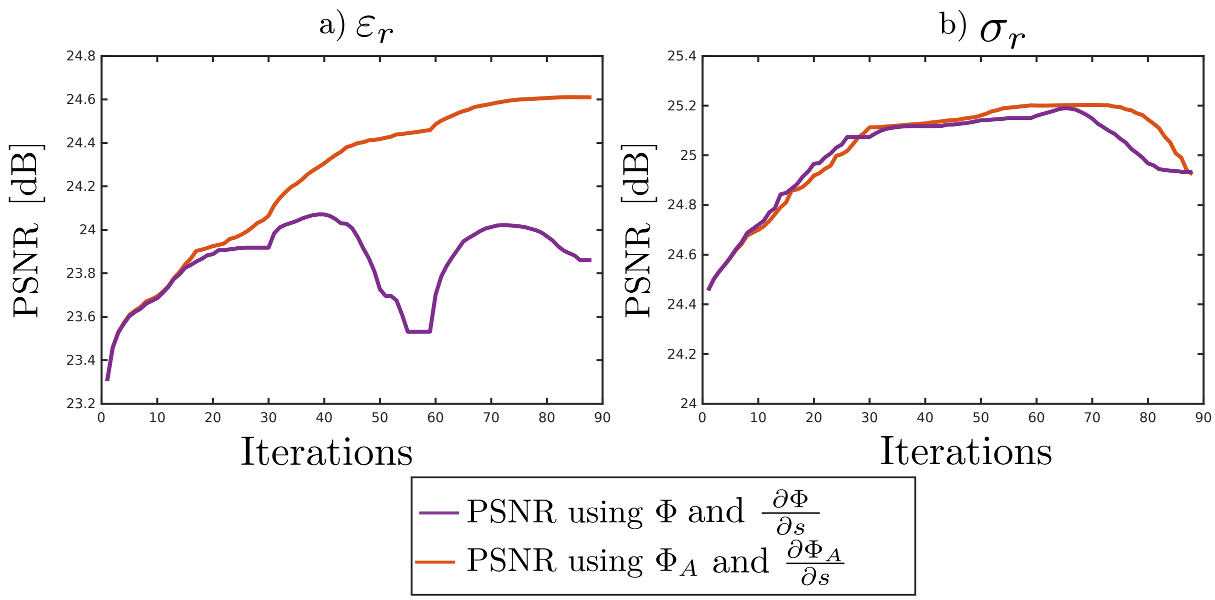

Figure 35. PSNR metric between the true model and the results of FWI in each frequency using Φ_A with $\frac{\partial \Phi_A}{\partial s}$ and Φ with $\frac{\partial \Phi}{\partial s}$: a) relative permittivity and b) relative conductivity.

estimating in the FWI process high-quality models with starting points that do not have enough low-frequency information. The cost function is given as

$$\Phi_{GS}(s,\mathbf{m},\rho) = \frac{1}{2}\left|\left|\nu(\rho,t) * \left(\frac{\mathbf{d}_{mod}(s,t)\cdot||\mathbf{d}_{obs}(t)||_2}{||\mathbf{d}_{mod}(s,t)||_2} - \mathbf{d}_{obs}(t)\right)\right|\right|_2^2, \tag{102}$$

where $\nu(\rho,t)$ is a Gaussian function and ρ is associated with the standard deviation of a Gaussian function. $\nu(\rho,t)$ is convolved with the difference between the modeled and observed data. The Gaussian function is defined as:

$$\nu(\rho,t) = \frac{1}{\sqrt{2\pi}\rho} e^{\frac{-(t-T_{aux})^2}{2\rho^2}}, \tag{103}$$

where $T_{aux} = (T - dt)/2$, t is the time and change from dt to T, and T is the time of the Gaussian function. Figure 36 shows the Gaussian function with the width fixed $T = 80 \cdot dt$ and four different values of variance (ρ^2): 200, 500, 1000, 2000. According to the cost function presented in the Equation (102), the adjoint source is defined as

$$\frac{\partial \Phi_{GS}}{\partial s} = \nu(\rho,t) * \left(\nu(\rho,t) * \left(\frac{\mathbf{d}_{mod}(s,t) \cdot \mathbf{d}_{mod}^{T}(s,t) \cdot \mathbf{d}_{obs}(t) \cdot \|\mathbf{d}_{obs}(t)\|_2}{\|\mathbf{d}_{mod}(s,t)\|_2^3} - \frac{\mathbf{d}_{obs}(t) \cdot \|\mathbf{d}_{obs}(t)\|_2}{\|\mathbf{d}_{mod}(s,t)\|_2} \right) \right), \tag{104}$$

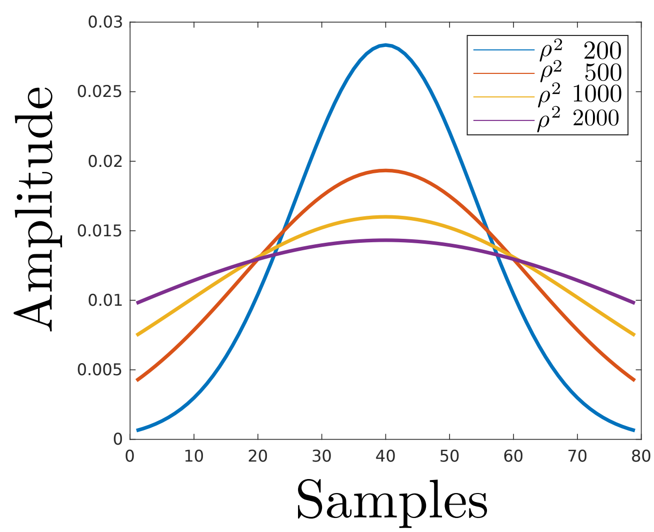

Figure 36. Gaussian function with variance: 200, 500, 1000, 2000.

If the parameter ρ^2 is huge, it causes the cost function is not multimodal, which allows ignoring the high-frequency information and approaching the global minimum. On the contrary, if the parameter ρ^2 is very small, the cost function is multimodal, and the noisy data together with an initial model without sufficient low-frequency information produces that the FWI parameters estimation is noisy. Gaussian preprocessing seeks to converge to a solution of low wavenumbers that depend on the width of the window and the parameter ρ. We have proposed applying the preprocessing in two directions: samples and scans. In both cases, the window size in the Gaussian function is chosen according to the FWI gradient resolution. Equation 94, relates the resolution of the FWI gradient in the CMP, where only the vertical component is different from zero. In short-offset acquisition, the parameter h is equal to zero; therefore, the expression of k_z is defined as:

$$k_z = \frac{2 \cdot f_c}{c}. \tag{105}$$

According to the Equation (105) for the frequencies of 30 (MHz), 50 (MHz) and 100 (MHz), the resolution k_z in the gradients are $0.2(1/m)$, $0.33(1/m)$ and $0.66(1/m)$, respectively. The width of the Gaussian filter window is selected in 100, 60, and 30 for the frequencies of 30 (MHz), 50 (MHz), and 100 (MHz), respectively. If the values of the window size are greater than those presented, the horizontal/vertical resolution is lost, which can incur few iterations of FWI for the highest frequencies.

Therefore, several experiments have been carried out, changing the parameter ρ^2: 500, 1000, 4000, and 8000; and applying the preprocessing in the t-direction. Figure 37 shows the results of FWI when the parameter ρ^2 increases. FWI with Gaussian preprocessing converges to a solution with lower wavenumber values when a huge length window is selected together with high values of ρ. The results that are shown in Figure 37 does not reach a better fit in the data than the

FWI without preprocessing as it is shown in Figure 38. Figure 39 shows the PSNR values for these same experiments. When the preprocessing is performed in t-direction, there is a loss of resolution of the residual, which deteriorates the fit of the data and the estimated parameters. Therefore, we do not recommend performing this Gaussian preprocessing in single-channel and short-offset acquisition in the direction of t.

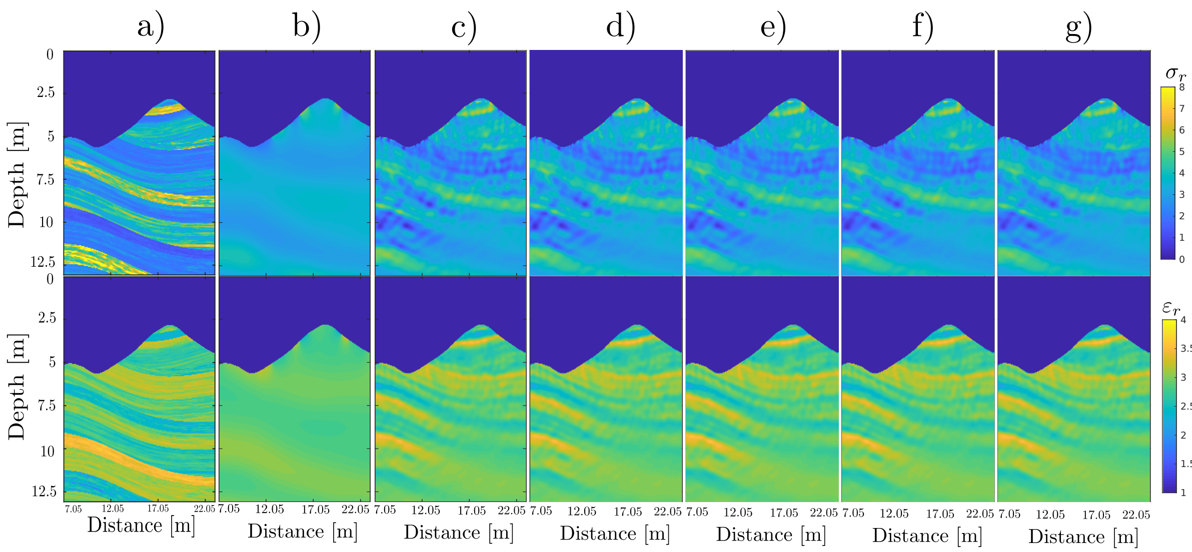

Figure 37. FWI results using Gaussian preprocessing in *samples*-directions and changing ρ^2: a) true models, b) initial models, c) no preprocessing, d) $\rho^2 = 500$, e) $\rho^2 = 1000$, f) $\rho^2 = 4000$, g) $\rho^2 = 8000$. The first row is the relative conductivity and the second rows is the relative permittivity.

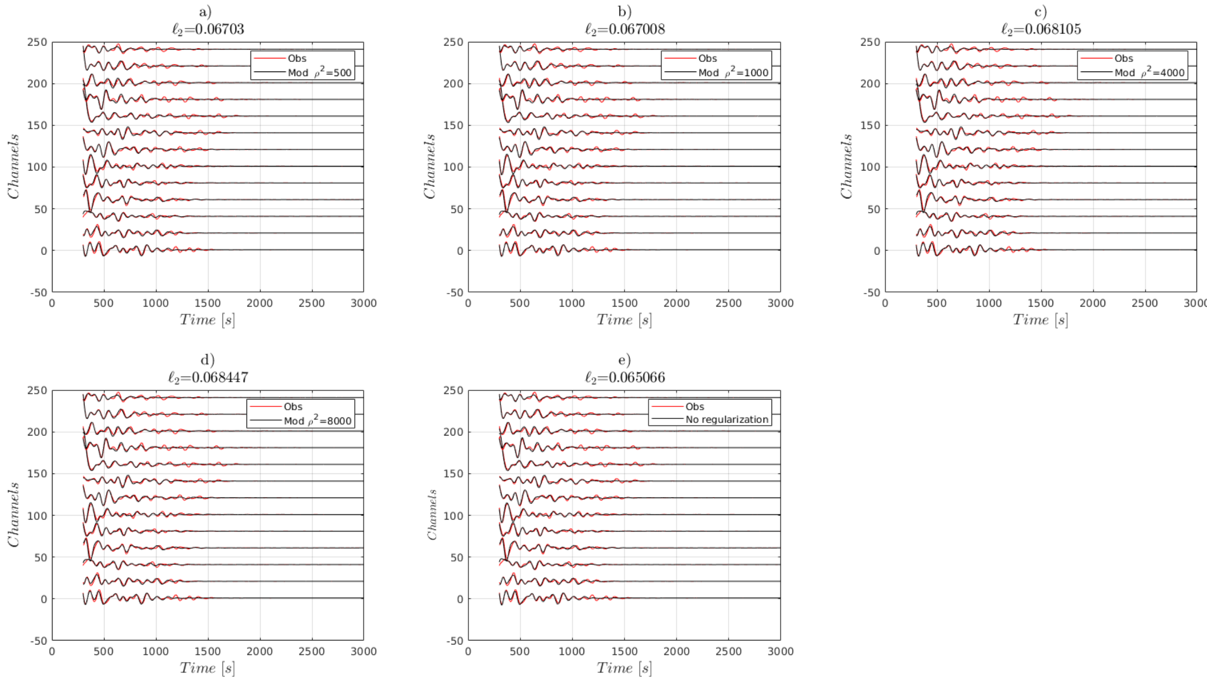

Figure 38. The fit between the observed and modeled data is presented in each panel using Gaussian preprocessing in t direction. Each panel shows 13 observed and modeled scans; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) $\rho^2 = 500$, b) $\rho^2 = 1000$, c) $\rho^2 = 4000$, d) $\rho^2 = 8000$, and e) No Gaussian preprocessing.

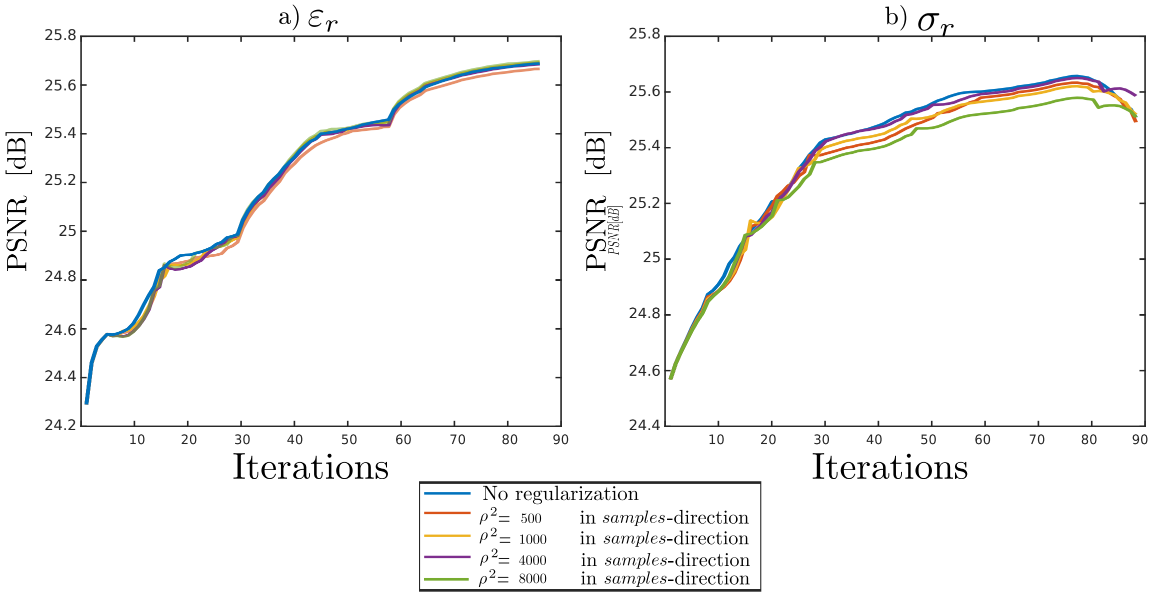

Figure 39. PSNR results using Gaussian constrain function in *samples*-direction for different ρ^2 values $500, 1000, 4000, 8000$. In blue no regulariztion, in orange $\rho^2 = 500$, in yellow $\rho^2 = 1000$, in purple $\rho^2 = 4000$ and in green $\rho^2 = 8000$: a) relative permittivity and b) relates conductivity.

The Gaussian preprocessing is applied by changing the direction from *samples*-direction to *scans*-direction. Figure 40 shows the residuals in the direction of scans changing ρ^2. If the preprocessing is applied in the scan direction, it implies losing horizontal resolution when ρ^2 increases. If we do not tune the parameter ρ^2 the FWI can lose continuity in the events that are not horizontal as it is in the case of the SEAM Foothills. Some experiments are carried out when the parameter ρ^2 changed; the experiments use the following values: 50, 25, 12, 6, and 2. The recommendation when applying the regularizer is to take low values of ρ^2 between 1-3 and gradually increase if more resolution is required. The window size is defined again according to the resolution of the gradient presented in the equation (105).

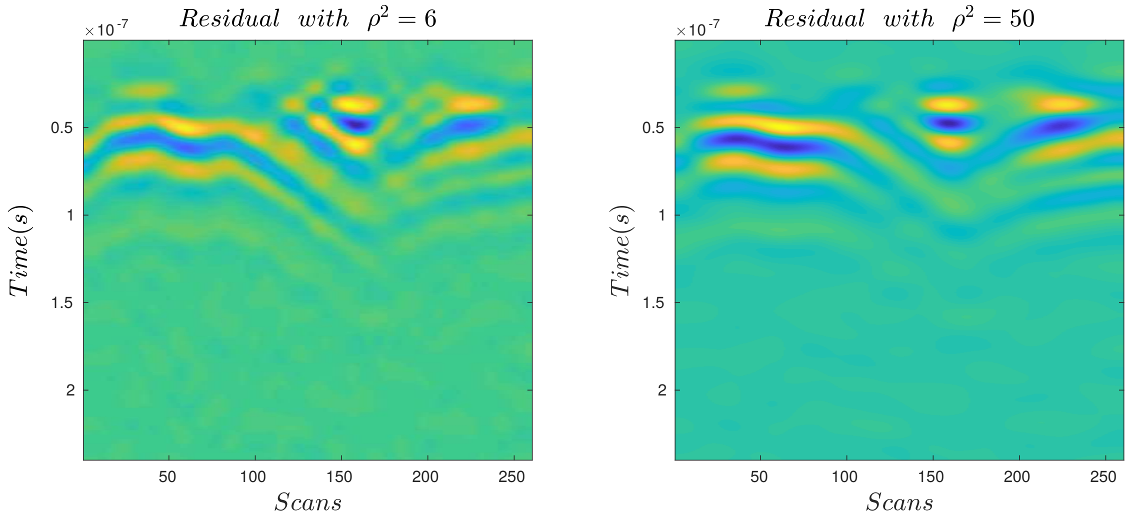

Figure 40. Difference between the modeled and observed data (residual) using a Gaussian preprocessing in scans direction with (left) $\rho^2 = 6$. (Right) $\rho^2 = 50$.

A new inversion experiment is carried out using the preprocessing in the *scans* -directions. The cost function is given as

$$\Phi_{GS2}(s, \mathbf{m}, \rho) = \frac{1}{2} \left|\left| \nu(\rho, scans) * \left(\frac{\mathbf{d}_{mod}(s,t) \cdot ||\mathbf{d}_{obs}(t)||_2}{||\mathbf{d}_{mod}(s,t)||_2} - \mathbf{d}_{obs}(t) \right) \right|\right|_2^2, \tag{106}$$

and the adjoint source is given by

$$\frac{\partial \Phi_{GS2}}{\partial s} = \nu(\rho, scans) * \left(\nu(\rho, scans) * \left(\frac{\mathbf{d}_{mod}(s,t) \cdot \mathbf{d}_{mod}^T(s,t) \cdot \mathbf{d}_{obs}(t) \cdot \|\mathbf{d}_{obs}(t)\|_2}{\|\mathbf{d}_{mod}(s,t) \cdot\|_2^3} - \frac{\mathbf{d}_{obs}(t) \cdot \|\mathbf{d}_{obs}(t)\|_2}{\|\mathbf{d}_{mod}(s,t)\|_2} \right) \right). \tag{107}$$

The results are presented in Figure 41. The PSNR results are shown in Figure 42, where the results show that the inversion using the Gaussian preprocessing is more robust to WGN. Figure 43 shows a matching between the traces obtained with the preprocessing in the direction of scans together with the observed data. As you can see, the fit of the data deteriorates as ρ^2 increases since it is smoothing the non-horizontal events, and coherence is lost. As shown by the values of ℓ_2 norm, the fit of the data improves for the case of $\rho^2 = 2$. Our recommendation is to start at low values of $\rho(> 1)$ and gradually increase the value to the desired resolution that best fits the data. The annexes section B introduces some results using FWI with the alternative and the traditional cost functions in a real data.

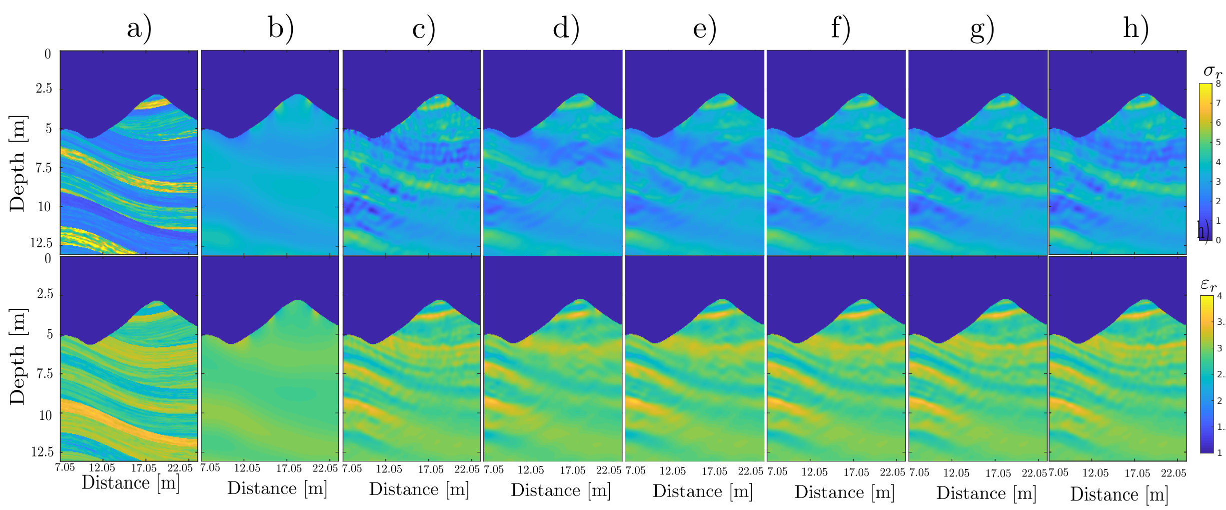

Figure 41. FWI results using Gaussian preprocessing: a) true models, b) initial models, c) No preprocessing, d) $\rho^2 = 50$ e) $\rho^2 = 25$, f) $\rho^2 = 12$, g) $\rho^2 = 6$ and h) $\rho^2 = 2$. The first row is the relative conductivity and the second-row is the relative permittivity.

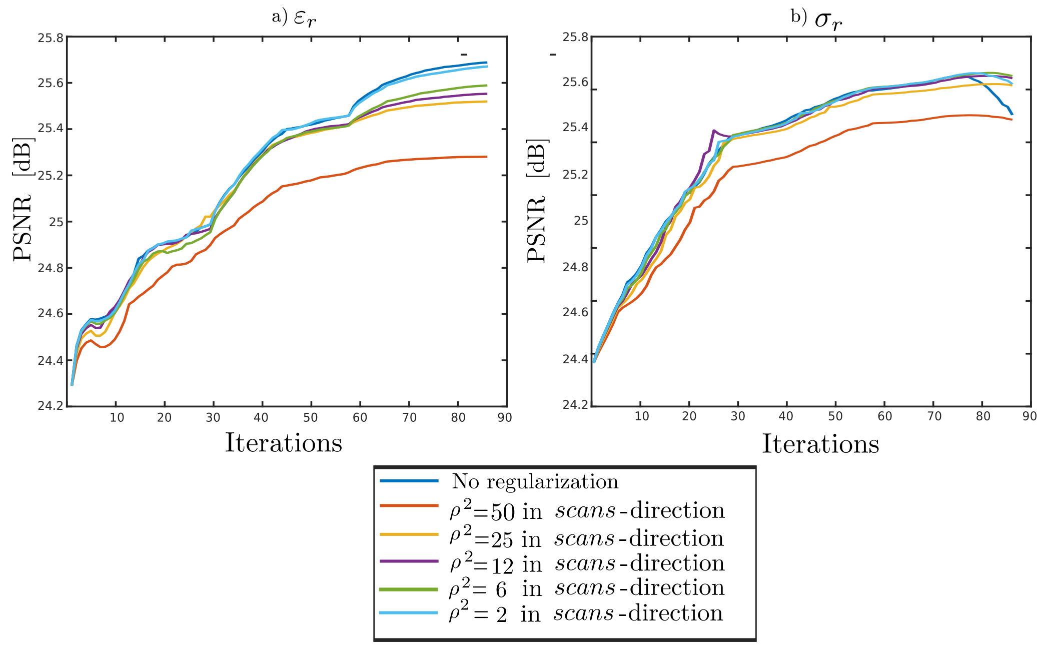

Figure 42. PSNR results using Gaussian preprocessing in the *scans*-direction. In blue no regularization, in orange $\rho^2 = 50$, in yellow $\rho^2 = 25$, in purple $\rho^2 = 12$, in green $\rho^2 = 6$ and in cyan $\rho^2 = 2$: a) relative permittivity and b) relative conductivity.

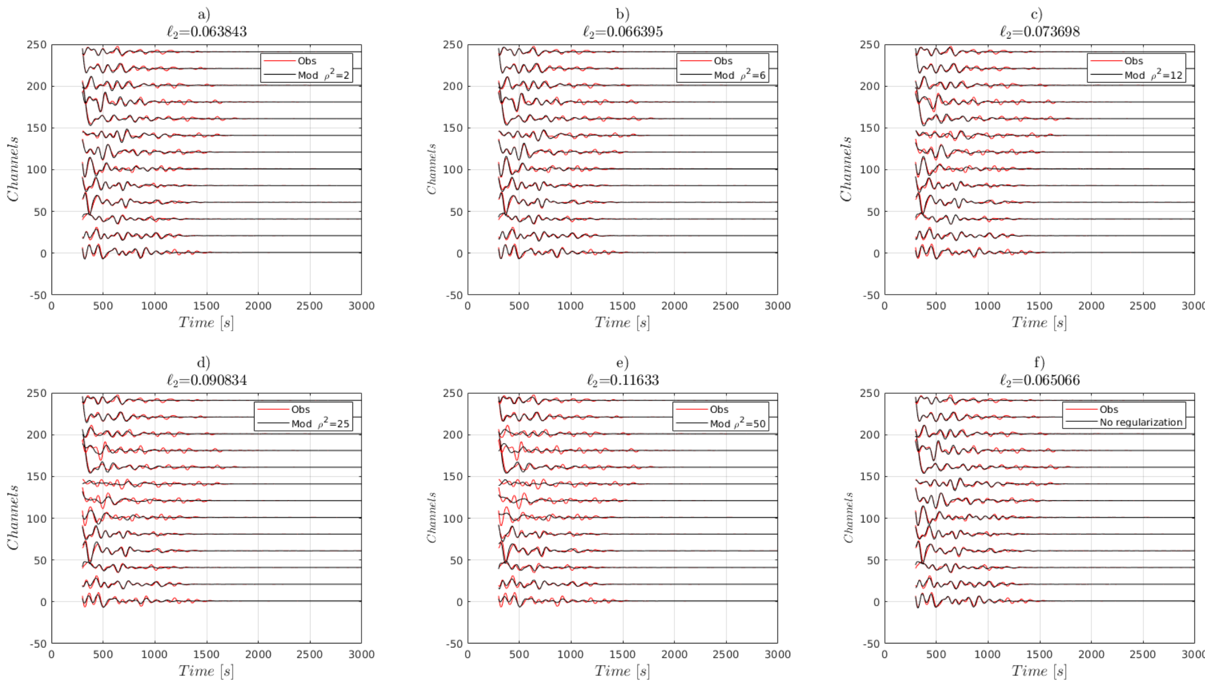

Figure 43. The fit between the observed and modeled data is presented in each panel using Gaussian preprocessing in *scans* direction. Each panel shows 13 observed and modeled scans; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) $\rho^2 = 2$, b) $\rho^2 = 6$, c) $\rho^2 = 12$, d) $\rho^2 = 25$, e) $\rho^2 = 50$, and e) No preprocessing.

4.3. FWI with TV-regularization.

The regularization of total variation (TV) is used to reduce the noise contamination in the inverse problem i.e., salt and pepper, Gaussian, Poisson, and speckle (Rudin et al. (1992) and Rodríguez (2013)). TV regularization seeks to obtain a smoothed estimate while preserving the main structures. Section **3.3.4** shows that the inversion results of a single-channel and short-offset acquisition are limited to high wavenumbers of the image. For that reason, when the TV regularization is included, the FWI reaches a smoothed version of the parameters with low wavenumbers, and it can be used as a new starting point. Besides, It is possible to attenuate the noise of the data using TV regularization. In this section, the TV regularization is applied on the SEAM model and the fit between the modeled and observed data is used to sintonized the TV regularization parameters.

The cost function for FWI using TV regularization is given by (Anagaw and Sacchi, 2012)

$$\Phi_{TV}(s,\mathbf{m}) = \frac{1}{2}\left|\left|\frac{\mathbf{d}_{mod}(s)\cdot||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2} - \mathbf{d}_{obs}\right|\right|_2^2 + \lambda_{TV}\sum_j^{Nx}\sum_i^{Nz}\sqrt{(D_x\mathbf{m}_{i,j})^2+(D_z\mathbf{m}_{i,j})^2+\alpha_{TV}^2}), \tag{108}$$

where $D_x m_{i,j} = \mathbf{m}_{i+1,j} - \mathbf{m}_{i,j}$, and $D_z\mathbf{m}_{i,j} = \mathbf{m}_{i,j+1} - \mathbf{m}_{i,j}$; α_{TV} is small to avoid the numerical division by zero in the gradient. The first term in Equation (108) is computed using the formulation given by Equation (69). Defining $||\overrightarrow{\delta}_{\mathbf{m}}|| = \sqrt{(D_x\mathbf{m}_{i,j})^2+(D_z\mathbf{m}_{i,j})^2+\alpha_{TV}^2}$ and taking the derivative with respect to $\mathbf{m}_{i,j}^k$, the gradient of TV is given by

$$\frac{\partial\mathscr{L}}{\partial\mathbf{m}_{i,j}^k} = \int_0^{Tend}\left[\lambda^T\frac{\partial W}{\partial\mathbf{m}}\right]dt + \lambda_{TV}\left(D_x^T\left(\frac{D_x\mathbf{m}_{i,j}}{||\overrightarrow{\delta}_{\mathbf{m}}||}\right)+D_z^T\left(\frac{D_z\mathbf{m}_{i,j}}{||\overrightarrow{\delta}_{\mathbf{m}}||}\right)\right), \tag{109}$$

where the differential operator $D_x^T = -D_x$, and $D_z^T = -D_z$. The gradient for relative permittivity and conductivity are given by

$$\frac{\partial\mathscr{L}}{\partial\varepsilon_{\mathbf{r}_{i,j}}^k} = \mathbf{g}_{TV}(\varepsilon_{\mathbf{r}i,j}) = \int_0^{Tend}\varepsilon_0\lambda_{\mathbf{E}_y}(\mathbf{r},t)\frac{\partial\mathbf{E}_y(\mathbf{r},t)}{\partial t}dt - \lambda_{TV}\left(D_x\left(\frac{D_x\varepsilon_{\mathbf{r}i,j}}{||\overrightarrow{\delta}_{\varepsilon_\mathbf{r}}||}\right)+D_z\left(\frac{D_z\varepsilon_{\mathbf{r}i,j}}{||\overrightarrow{\delta}_{\varepsilon_\mathbf{r}}||}\right)\right), \tag{110}$$

$$\frac{\partial \mathscr{L}}{\partial \sigma_{\mathbf{r}_{i,j}}^{k}} = \mathbf{g}_{TV}(\sigma_{\mathbf{r}_{i,j}}) = \int_{0}^{Tend} \sigma_0 \lambda_{\mathbf{E}_y}(\mathbf{r},t)\mathbf{E}_y(\mathbf{r},t)dt - \lambda_{TV}\left(D_x\left(\frac{D_x \sigma_{\mathbf{r}_{i,j}}}{\| \overrightarrow{\delta}_{\sigma_{\mathbf{r}}} \|}\right) + D_z\left(\frac{D_z \sigma_{\mathbf{r}_{i,j}}}{\| \overrightarrow{\delta}_{\sigma_{\mathbf{r}}} \|}\right)\right). \tag{111}$$

The parameter λ_{TV} is computed according to the following equation:

$$\lambda_{TV} = \gamma_{TV} \left|\left| g_{TV}(\mathbf{m}^k) \right|\right|_2, \tag{112}$$

where γ_{TV} is a constant value and in each iteration the regularization parameter is computed for permittivity and conductivity concerning the gradient energy, and over smoothing is avoid. The parameters γ_{TV} should be tuned in the TV regularization. A large value of γ_{TV} tends to produce a very smooth result. On the other hand, a small value of γ_{TV} does not significantly change the FWI results. The TV regularization is applied in the relative permittivity and conductivity. However, the relative conductivity is the more sensitive parameter in the inversion process. Four experiments with different γ_{TV} values have been performed: 1×10^{-3}, 5×10^{-3}, 7.5×10^{-3}, and 1×10^{-2}; and the results are shown in Figure 44. A large value of γ_{TV} can produce a smooth solution Figure 44-f). On the other hand, a small value of γ_{TV} does not produce a significant change in the solution (Figure 44-c)). The ℓ_2 norm is used to identify which of the models displayed in Figure 44 presents the best fit between the data. Figure 45 presents a 13-scans comparison between modeled and observed data. We can see that the data with the best fit occurs when γ_{TV} is $5e-3$ where the regularization reached a value of 0.5899. This tuning process is totally empirical and requires several experiments. However, the unique parameter that must be tuned for this regularization is γ_{TV}. We found that values between $1e-3$ and $7e-3$ generate good tuning results where the fit of the data is better than not having the regularization. Table 4 shows: the PSNR values for each parameter between the true model and the FWI model in the inversion process and the ℓ_2 norm using the residual between modeled and observed data. In collected data, it is not possible to compute the PSNR values because the true model is unknown therefore, our choice is highlighted in green according to the best fit in the data. Although, The PSNR values are lower with regularization than without regularization; it is important to mention that the inversion process with regularization is more stable when the data is highly noisy. The results with noisy data using TV regularization are presented in annexes sections C.

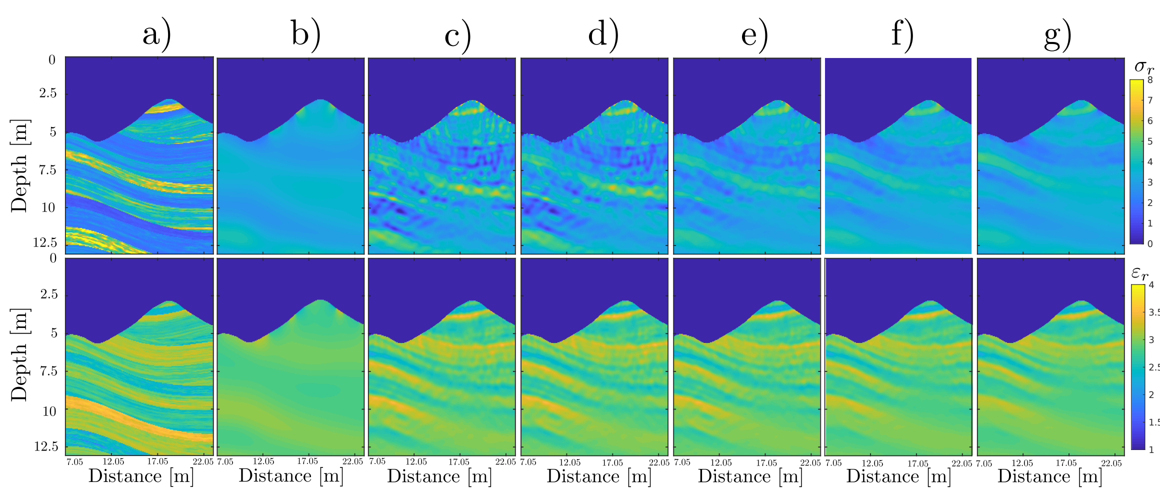

Figure 44. FWI results for different values of γ_{TV} in the TV regularization: a) true models, b) initial models, c) $\gamma_{TV} = 0.0$, d) $\gamma_{TV} = 1 \times 10^{-3}$, e) $\gamma_{TV} = 5 \times 10^{-3}$, f) $\gamma_{TV} = 7.5 \times 10^{-3}$, and g) $\gamma_{TV} = 1 \times 10^{-2}$. The first row is the relative conductivity and the second rows is the relative permittivity.

γ_{TV}	ε_r PSNR (dB)	σ_r PSNR (dB)	$\ell 2$ norm
0.0	25.97	25.43	0.0624
1.0e-3	25.95	25.52	0.0616
5.0e-3	25.80	25.37	0.0589
7.5e-3	25.60	25.31	0.0614
1.0e-2	25.44	25.24	0.0641

Table 4. *PSNR and ℓ_2 norm when the parameter γ_{TV} change in the FWI process and TV regularization is included. ur choice is highlighted in green according to the best fit in the data.*

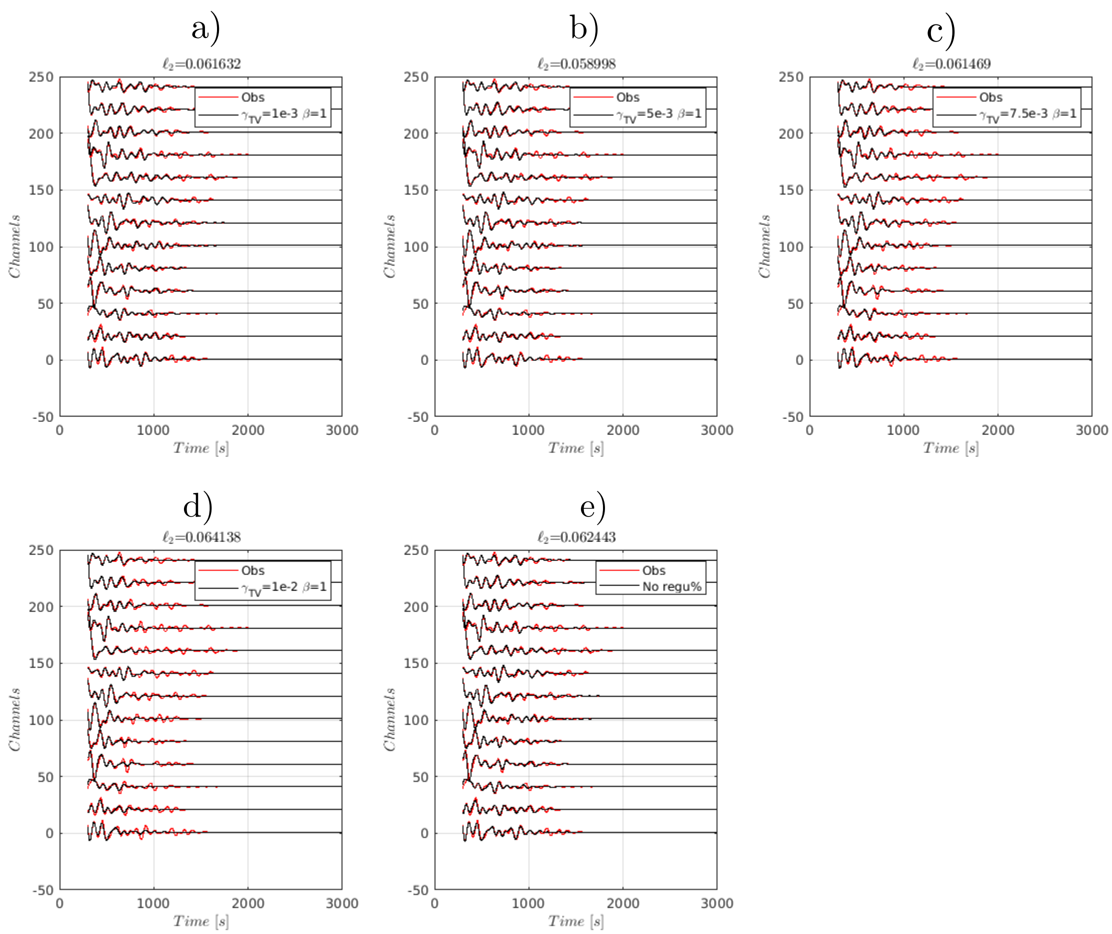

Figure 45. The fit between the observed and modeled data is presented in each panel using TV regularization. Each panel shows 13 observed and modeled scans; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) $\gamma_{TV} = 1e-3$, b) $\gamma_{TV} = 5e-3$, c) $\gamma_{TV} = 7.5e-3$, d) $\gamma_{TV} = 1e-2$, and e) No regularization

4.3.1. The scale parameter β. According to state-of-the-art, using the parameter β as a scale parameter is an alternative to improve the conductivity estimation. This parameter β can be included together with the TV constrain as it is presented in this subsection. A suitable scale between the parameters can produce a natural behavior in both parameters' sensitivity in the cost function. The parameter β is introduced in the inversion process to control the weights between the relative permittivity and the relative conductivity Lavoué et al. (2014). The scaled conductivity parameter is given by

$$\sigma = \frac{\sigma_0 \cdot \sigma_r}{\beta}, \tag{113}$$

where $\sigma_0 = 5.56 \times 10^{-4}$, and β is the scale parameter. The relative conductivity gradient is then given by

$$g(\sigma_r) = \int_0^{Tend} \beta \cdot \sigma_0 \cdot \lambda_{E_y} E_y dt. \tag{114}$$

The parameter is fixed in $\gamma_{TV} = 5e-3$, and the parameter β is changed. The results obtained from the inversion process are presented in Figure 46. Table 5 presents the PSNR results achieved for relative permittivity and relative conductivity, as can be seen in Figure 46 when β increase, the image quality in the conductivity parameter increase too and the image quality in the permittivity parameter decreases.

Figure 47 shows again a the fit between the modeled and observed data. The best fit between the data is reached when $\beta = 4.0$ with a $\ell_2 = 0.057$. Our recommendation is to use a parameter $\gamma_{TV} = 5e-3$ with $\beta = 4.0$. With the TV regularization and the scaling parameter β, it is possible to improve the conductivity parameter estimation even with noisy data in single-channel and short-offset acquisitions.

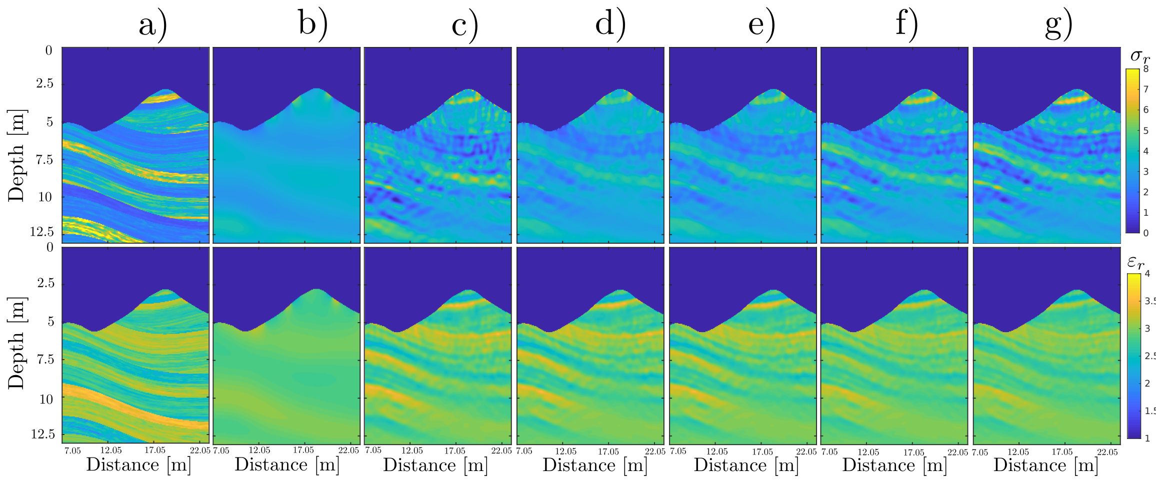

Figure 46. FWI results for different values of β in the TV regularization and fixed $\gamma_{TV} = 5e-3$: a) true models, b) initial models, c) 0.0, d) 0.5, e) 1.0, f) 4.0, and g) 8.0. The first row is the relative conductivity and the second rows is the relative permittivity.

β	ε_r PSNR (dB)	σ_r PSNR (dB)	$\ell 2$ norm
0.0	25.97	25.43	0.062
0.5	25.44	25.15	0.060
1.0	25.45	25.24	0.059
4.0	25.38	25.40	0.057
8.0	25.28	25.41	0.059

Table 5. *PSNR and ℓ_2 norm when the parameter β change in the FWI process and TV regularization is included with $\gamma_{TV} = 5e-3$. Our choice is highlighted in green according to the best fit in the data.*

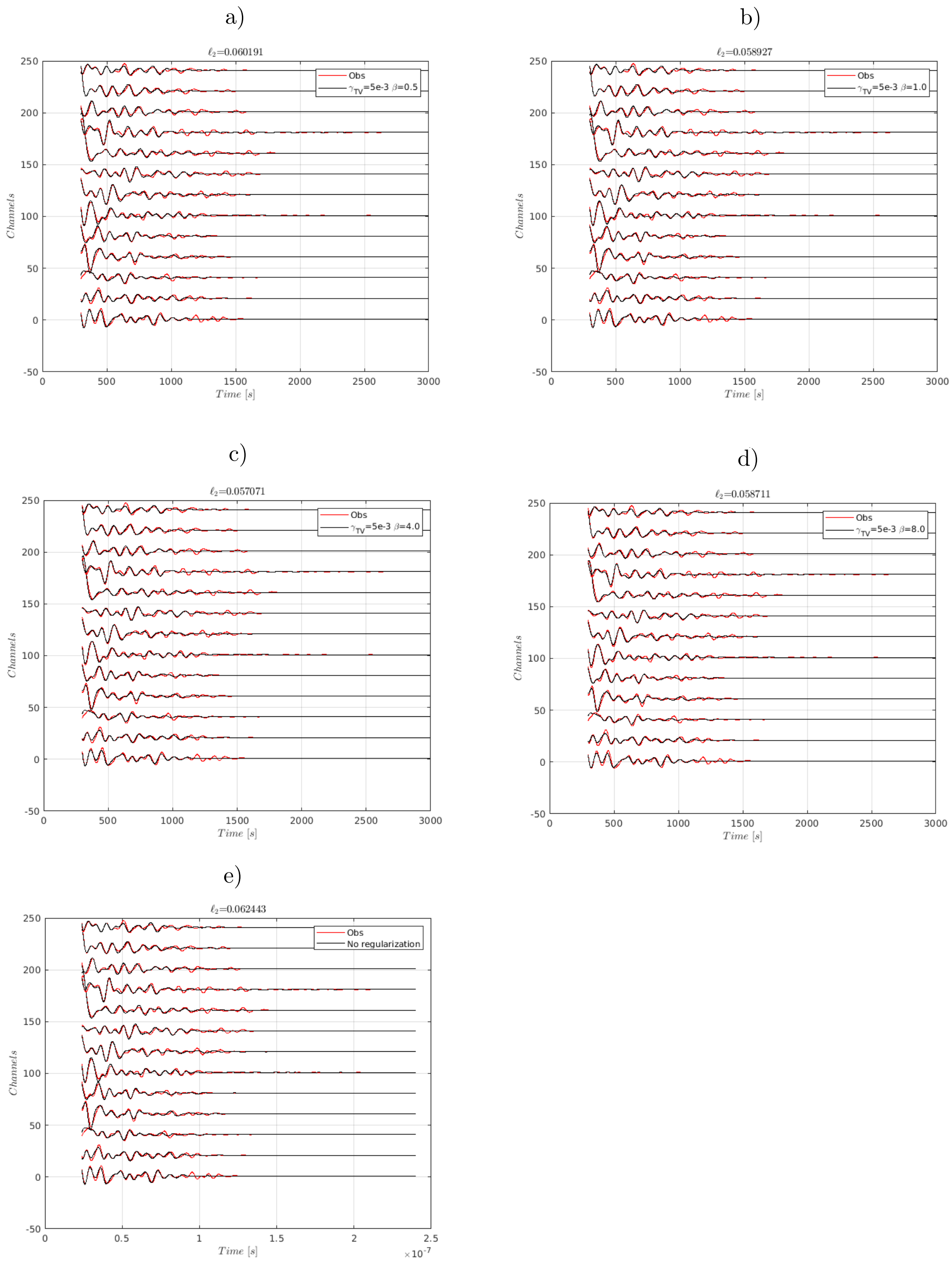

Figure 47. The fit between the observed and modeled data is presented in each panel using TV regularization and β scale parameter. Each panel shows 13 observed and modeled scans; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) $\gamma_{TV} = 5e-3$ and $\beta = 0$, b) $\gamma_{TV} = 5e-3$ and $\beta = 0.5$, c) $\gamma_{TV} = 5e-3$ and $\beta = 1.0$, d) $\gamma_{TV} = 5e-3$ and $\beta = 4.0$, e) $\gamma_{TV} = 5e-3$ and $\beta = 8.0$, and e) No regularization.

4.4. FWI with MTV regularization.

The modified total variation (MTV), is an alternative regularization method used in the seismic case as a modification to the TV method Lin and Huang (2014). The MTV regularization is a more stable version of TV since it does not depend on the α_{TV} value included to avoid divisions by zero in the gradients (see Equation (108)). The MTV method is selected in GPR as for single-channel and short-offset acquisition as the first step to obtaining low wavenumbers in the image, which is difficult to obtain due to the lack of illumination. The MTV method uses the auxiliary variable $\mathbf{u}$, and an additional term is added, as it is shown in Equation (115).

$$\min_{\mathbf{m},\mathbf{u}}, \Phi_d(s,\mathbf{m},\mathbf{u}) = \frac{1}{2}\left|\left|\frac{\mathbf{d}_{mod}(s)\cdot||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2} - \mathbf{d}_{obs}\right|\right|_2^2 + \lambda_{MTV|1}||\mathbf{m}-\mathbf{u}||_2^2 + \lambda_{MTV|2}||\mathbf{u}||_{TV}. \quad (115)$$

The advantage of the MTV method compared with the TV method is that the parameter α_{TV} is not included in the gradient. α_{TV} avoids zero divisions in the gradient and has a high weight in the results of the inversion results. The main problem is separated into two subproblems, and the formulation is given in Equations (116) and (117).

$$\min_{\mathbf{m}} \mathbf{m}^k = \frac{1}{2}\left|\left|\frac{\mathbf{d}_{mod}(s)\cdot||\mathbf{d}_{obs}||_2}{||\mathbf{d}_{mod}(s)||_2} - \mathbf{d}_{obs}\right|\right|_2^2 + \lambda_{MTV|1}||\mathbf{m}-\mathbf{u}^{k-1}||_2^2, \quad (116)$$

$$\min_{\mathbf{u}} \mathbf{u}^k = ||\mathbf{m}^k-\mathbf{u}||_2^2 + \lambda_{MTV|2}||\mathbf{u}||_{TV}. \quad (117)$$

The first subproblem, defined in Equation (116), is solved using the L-BFGS method, where the gradient is defined as:

$$g_{MTV}(\mathbf{m}^k) = g(\mathbf{m}^k) + 2\lambda_{MTV|1}(\mathbf{m} - \mathbf{u}^{k-1}). \tag{118}$$

The gradient $g_{MTV}(\mathbf{m}^k)$ and the model $\mathbf{m}^k$ are the inputs in each iteration for the L-BFGS algorithm. The second subproblem is solved using the Split-Bregman iterative method, where the new problem is reformulated as follows:

$$\min_{\mathbf{u},\mathbf{d}_x,\mathbf{d}_z} \{||\mathbf{u} - \mathbf{m}^k||_2^2 + \lambda_{MTV|2}||\mathbf{u}||_{TV} + \alpha_{MTV}||\mathbf{d}_x - D_x\mathbf{u}||_2^2 + \alpha_{MTV}||\mathbf{d}_z - D_z\mathbf{u}||_2^2\}, \tag{119}$$

where $\mathbf{d}_x \approx D_x\mathbf{u}$, $\mathbf{d}_z \approx D_z\mathbf{u}$ and $\alpha_{MTV} = 2\lambda_{MTV|2}$, (Goldstein and Osher, 2009). Using the Bregman method the new formulation is given by:

$$\min_{\mathbf{u},\mathbf{d}_x,\mathbf{d}_z} \{||\mathbf{u} - \mathbf{m}^k||_2^2 + \lambda_{MTV|2}||\mathbf{u}||_{TV} + \alpha_{MTV}||\mathbf{d}_x - D_x\mathbf{u} - \mathbf{b}_x^k||_2^2 + \alpha_{MTV}||\mathbf{d}_z - D_z\mathbf{u} - \mathbf{b}_z^k||_2^2\}, \tag{120}$$

where $\mathbf{b}_x$ and $\mathbf{b}_z$ are updated as:

$$\begin{aligned} \mathbf{b}_x^{k+1} &= \mathbf{b}_x^k + (D_x u^{k+1} - \mathbf{d}_x^{k+1}), \\ \mathbf{b}_z^{k+1} &= \mathbf{b}_z^k + (D_z u^{k+1} - \mathbf{d}_z^{k+1}). \end{aligned} \tag{121}$$

The Equation (120) is separated in two new subproblems, the formulation is shown in Equations (122) and (123).

$$\min_{\mathbf{u}} \{||\mathbf{u} - \mathbf{m}^k||_2^2 + \alpha_{MTV}||\mathbf{d}_x^k - D_x\mathbf{u} - \mathbf{b}_x^k||_2^2 + \alpha_{MTV}||\mathbf{d}_z^k - D_z\mathbf{u} - \mathbf{b}_z^k||_2^2\}, \tag{122}$$

$$\min_{\mathbf{d}_x, \mathbf{d}_z} \{\lambda_{MTV|2}||\mathbf{u}||_{TV} + \alpha_{MTV}||\mathbf{d}_x - D_x\mathbf{u} - \mathbf{b}_x||_2^2 + \alpha_{MTV}||\mathbf{d}_z - D_z\mathbf{u} - \mathbf{b}_z||_2^2\}. \tag{123}$$

The solution of Equation (122) is computed by using the Gauss-Seidel iterative method:

$$\begin{aligned} \mathbf{u}_{i,j}^{k+1} = & \frac{\alpha_{MTV}}{1+4\alpha_{MTV}}(\mathbf{u}_{i+1,j}^k + \mathbf{u}_{i-1,j}^k + \mathbf{u}_{i,j+1}^k + \mathbf{u}_{i,j-1}^k \\ & + \mathbf{d}_{x,i-1,j}^k - \mathbf{d}_{x,i,j}^k + \mathbf{d}_{z,i,j-1}^k - \mathbf{d}_{z,i,j}^k \\ & - \mathbf{b}_{x,i-1,j}^k + \mathbf{b}_{x,i,j}^k - \mathbf{b}_{z,i,j-1}^k + \mathbf{b}_{z,i,j}^k) + \frac{1}{1+4\alpha_{MTV}}\mathbf{m}_{i,j}^k. \end{aligned} \tag{124}$$

The solution of Equation (123) is computed by using a generalized shrinkage formulation

(Wang et al., 2008):

$$\begin{aligned} \mathbf{d}_x^{k+1} &= \mathbf{max}(\mathbf{s}^k - \frac{\lambda_{MTV|2}}{2\alpha_{MTV}}, 0)\frac{D_x\mathbf{u}^k + \mathbf{b}_x^k}{\mathbf{s}^k}, \\ \mathbf{d}_z^{k+1} &= \mathbf{max}(\mathbf{s}^k - \frac{\lambda_{MTV|2}}{2\alpha_{MTV}}, 0)\frac{D_z\mathbf{u}^k + \mathbf{b}_z^k}{\mathbf{s}^k}, \end{aligned} \tag{125}$$

where $\mathbf{s}^k = \sqrt{\left(D_x\mathbf{u}^k + \mathbf{b}_x^k\right)^2 + \left(D_z\mathbf{u}^k + \mathbf{b}_z^k\right)^2}$. The parameters $\lambda_{MTV|1}$ and $\lambda_{MTV|2}$ are given as

$$\begin{aligned} \lambda_{MTV|1} &= \gamma_{MTV}\frac{||g(\mathbf{m}^k)||_2}{||\mathbf{m}^k - \mathbf{u}^{k-1}||_2}, \\ \lambda_{MTV|2} &= \frac{max(|\mathbf{m}^k - \mathbf{u}^{k-1}|)}{\kappa}, \end{aligned} \tag{126}$$

where γ_{MTV} can take values from 0.05 to 0.5 and κ can take values from 0 to 200 Gao and Huang (2019). A large value of γ_{MTV} does not generate changes in the initial estimated model's parameter. On the other hand, a small value of the γ_{MTV} regularization results in the traditional FWI solution. Small values of κ result in the traditional FWI solution, whereas a very high value for κ results in a very smoothed solution

The MTV regularization is applied in the relative permittivity and relative conductivity. Three experiments are performed by changing the parameter γ_{MTV} between [0-0.5]. In each experiment, the other parameters are kept fixed. The parameter $\lambda_{MTV|1}$ is directly associated with γ_{MTV} and it is the weight associated with the *apriori* information obtained from the parameters $\mathbf{m}^k - \mathbf{u}^{k-1}$. Table 6 shows the PSNR values where $\gamma_{MTV} = 0.05$ reach the highest PSNR in both parameters.

γ_{MTV}	ε_r PSNR (dB)	σ_r PSNR (dB)	ℓ_2 norm
0.0	25.97	25.43	0.0624
0.05	25.86	25.63	0.0630
0.1	25.80	25.46	0.0638
0.2	25.36	25.38	0.0724

Table 6. *PSNR when the parameter γ_{MTV} change in the FWI process and MTV regularization is included*

The fit between the modeled and observed data is used as a strategy of tunning the γ_{MTV} parameter, where the best fit is reached when $\gamma_{MTV} = 0.05$. Figure 48 shows the fit data results.

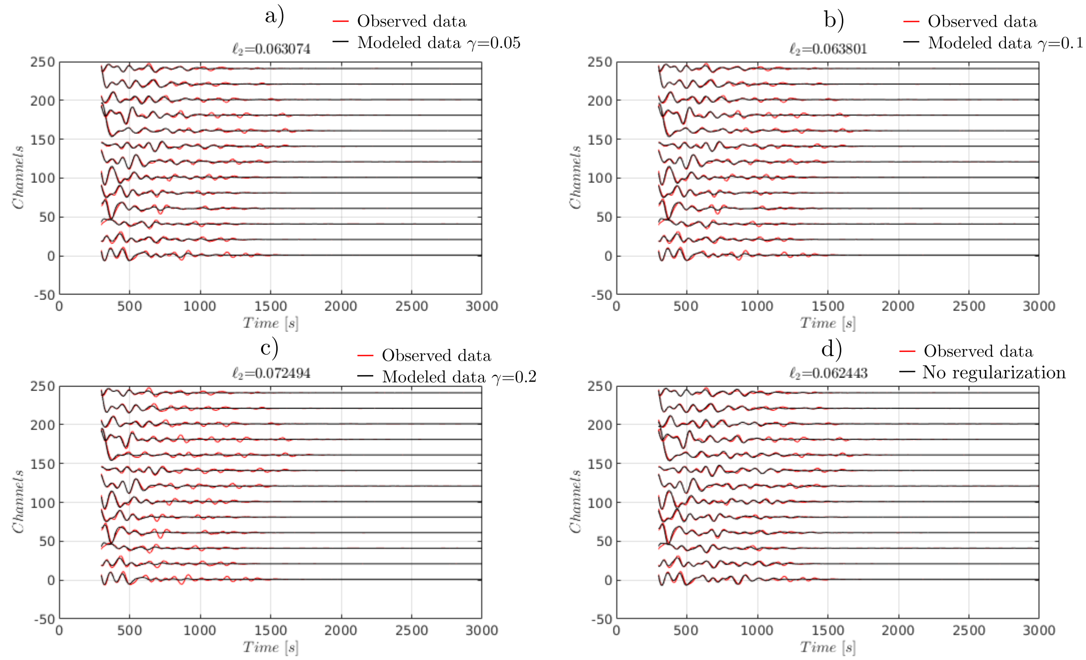

Figure 48. The fit between the observed and modeled data is presented in each panel using MTV regularization. Each panel shows 13 observed and modeled scans; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) $\gamma_{MTV} = 0.05$, b) $\gamma_{MTV} = 0.1$, c) $\gamma_{MTV} = 0.2$, and d) No regularization.

Once the γ_{MTV} is selected, the parameter κ is syntonized. The parameter κ changes with the following values: 100, 150, and 200. We again seek to find which value of κ achieves the best fit between the modeled and observed data. Figure 49 and Table 7 shows the results of FWI, PSNR and ℓ_2 reached by changing the values of κ. Figure 50 shows the fit between the modeled and observed data. The best fit between the data is reached when $\kappa = 200$ with a ℓ_2 norm = 0.0606. Our recommendation is to use a parameter $\gamma_{MTV} = 0.05$ with $\kappa = 200$.

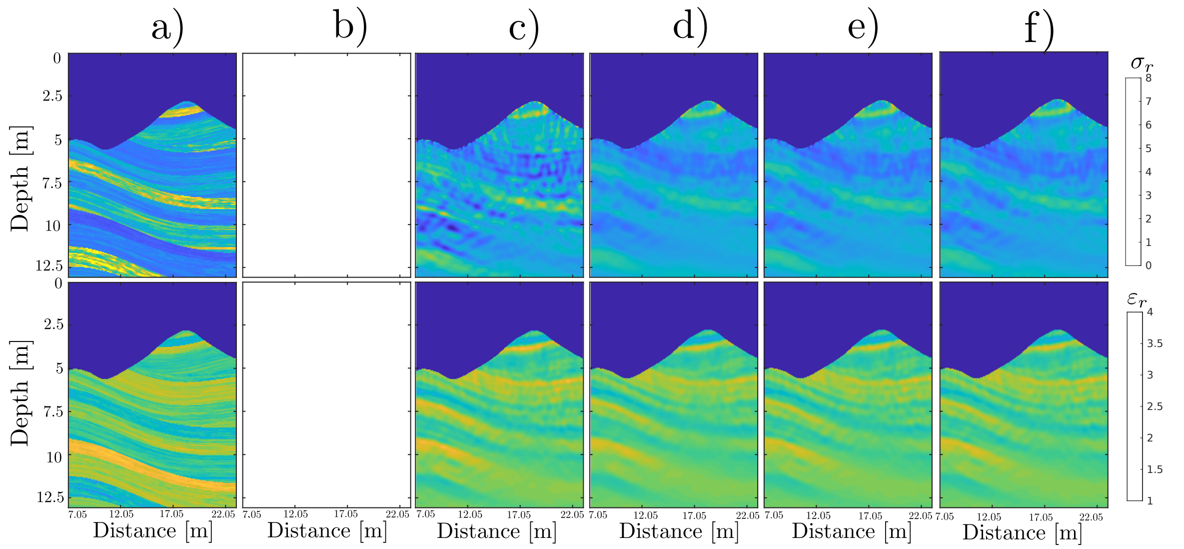

Figure 49. FWI results changing the κ parameter in the MTV regularization, $\gamma_{MTV} = 0.05$: a) true models, b) initial models, c) $\kappa = 0.0$, d) $\kappa = 100$, e) $\kappa = 150$, f) $\kappa = 200$. The first row is the relative conductivity and the second rows is the relative permittivity.

κ	ε_r PSNR (dB)	σ_r PSNR (dB)	$\ell 2$ norm
0	25.97	25.43	0.0619
100	25.87	25.27	0.0635
150	25.86	25.34	0.0616
200	25.91	25.35	0.0606

Table 7. *PSNR when the parameter κ change in the FWI process and MTV regularization is included*

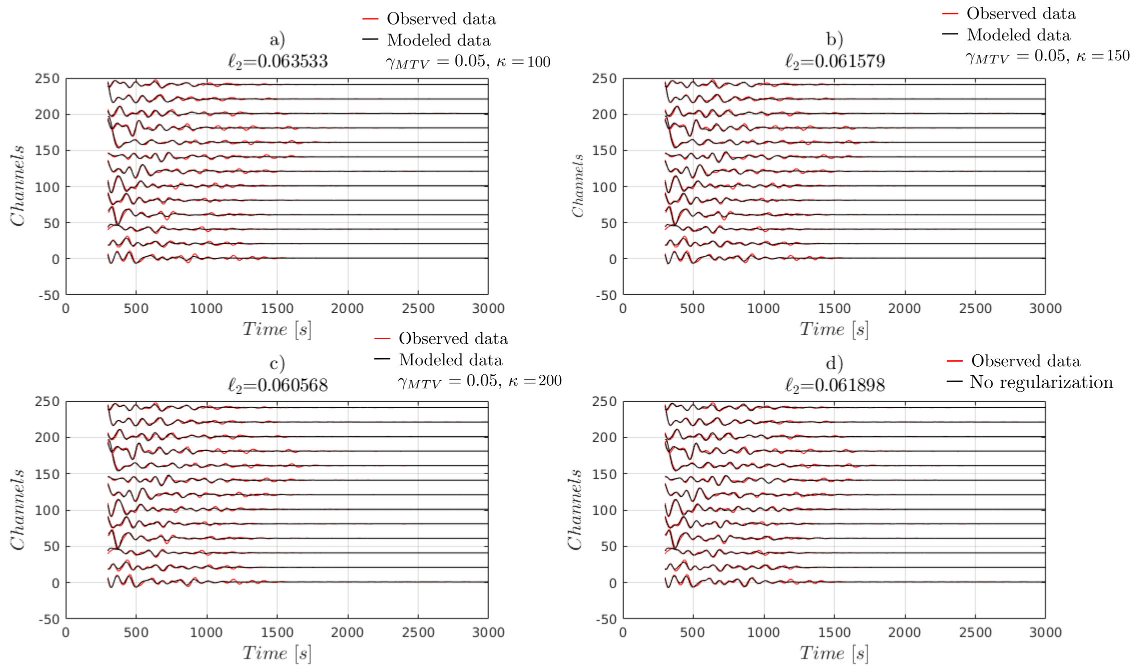

Figure 50. The fit between the observed and modeled data is presented in each panel using MTV regularization. The parameter γ_{MTV} is fixed in 0.05. Each panel shows 13 observed and modeled scans; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) $\gamma_{MTV} = 0.05$ and $\kappa = 100$, b) $\gamma_{MTV} = 0.05$ and $\kappa = 150$, c) $\gamma_{MTV} = 0.05$ and $\kappa = 200$, d) No regularization.

Finally, four experiments are performed where the parameter β is changed from 0.5 to 4.0, the results are shown in Figure 51 and the PSNR and ℓ_2 norm are shown in Table 8. Figure 52 shows again a the fit between the modeled and observed data. The best fit between the data is reached when $\beta = 4.0$ with a ℓ_2 norm = 0.0608. Our recommendation is to use a parameter $\gamma_{MTV} = 0.05$, $\kappa = 200$ with $\beta = 4.0$.

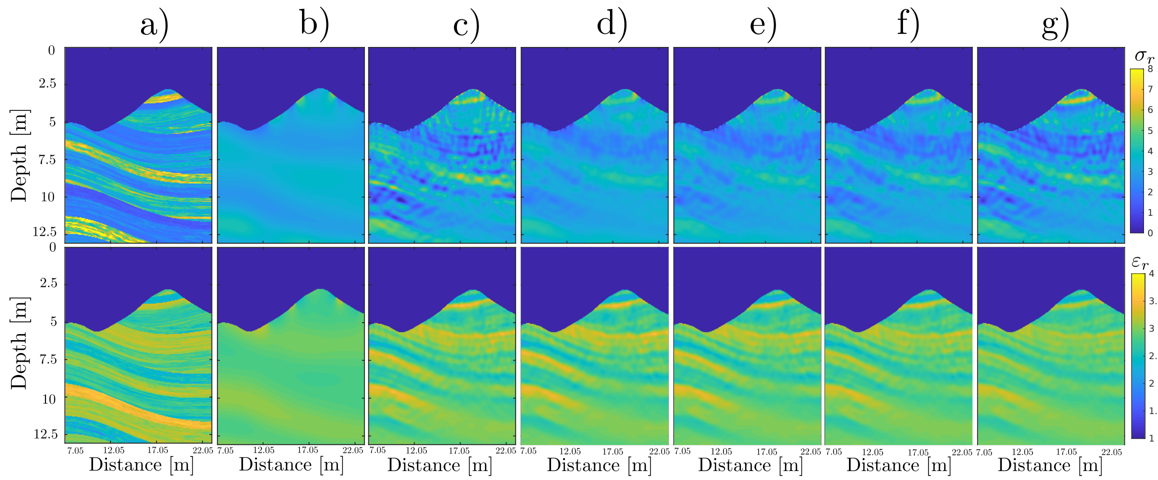

Figure 51. FWI results changing the β parameter in the MTV regularization $\gamma_{MTV} = 0.05$ and $\kappa = 200$: a) true models, b) initial models, c) without regularization, d) $\beta = 0.5$, e) $\beta = 1$, f) $\beta = 2$ and, g) $\beta = 4$. The first row is the relative conductivity and the second rows is the relative permittivity.

β	ε_r PSNR (dB)	σ_r PSNR (dB)	ℓ_2 norm
0.0	25.97	25.43	0.0624
0.5	25.86	25.12	0.0652
1.0	25.93	25.06	0.0630
2.0	25.91	25.28	0.0609
4.0	25.81	25.30	0.0608

Table 8. *PSNR when the parameter β change in the FWI process and MTV regularization is included*

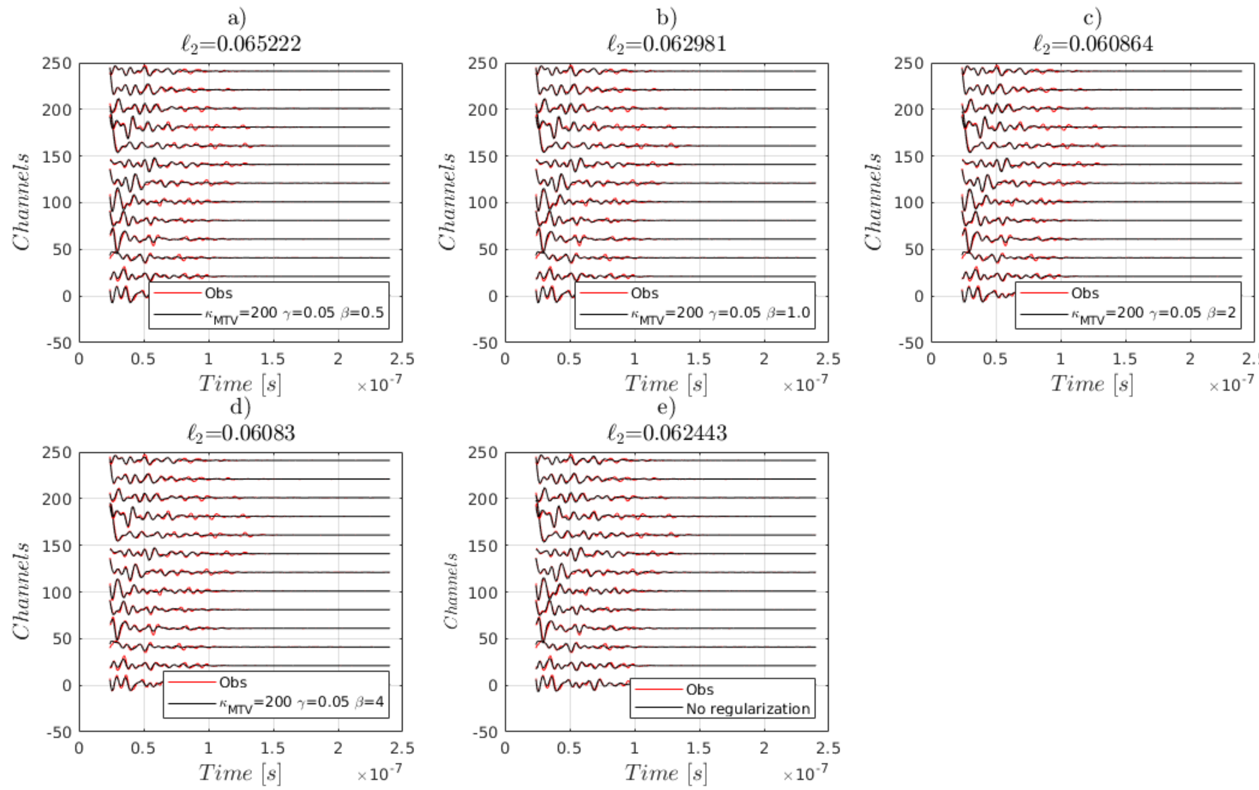

Figure 52. The fit between the observed and modeled data is presented in each panel using MTV regularization. The parameter $\gamma_{MTV} = 0.05$ and the paramter $\kappa = 200$. Each panel shows 13 observed and modeled scans; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) $\gamma_{MTV} = 0.05$, $\kappa = 200$ and $\beta = 0.5$, b) $\gamma_{MTV} = 0.05$, $\kappa = 200$ and $\beta = 1.0$, c) $\gamma_{MTV} = 0.05$, $\kappa = 200$ and $\beta = 2.0$, d) $\gamma_{MTV} = 0.05$, $\kappa = 200$ and $\beta = 4.0$, and e) No regularization.

From the results obtained from TV and MTV Figures 44-e) and 51-e), respectively, the regularizations allow to reduce the noise in the FWI results, it maintains the main subsurface layers, and the regularizations reach FWI stable process. However, the two regularizations require tuning processes that increase the computational cost and execution time. We recommend using the fit data as a selection criterion in the tuning of regularization parameters, such as γ_{TV}, γ_{MTV}, κ and β. Of the two regularizations used on the parameters in these synthetic data, we achieved a better fit with the TV regularization than the MTV regularization where the ℓ_2 norm=0.057 in TV and ℓ_2 norm=0.060 in MTV.

5. FWI of experimentally collected data

This chapter presents the estimation of the electromagnetic parameters of a small area in Mogotes, a local municipality located in the state of Santander, Colombia. The regularization introduced in chapter 4 is tested, showing the importance of each cost function in the inversion process. The collected data is obtained using the shielded single-channel, and short-offset antenna with a center frequency of 400 (MHz). Finally, the estimated parameters of permittivity and conductivity from FWI are shown.

5.1. Description of the experiment

A raw GPR data is studied in this section. The raw GPR data is obtained in Mogotes, a local municipality in the state of Santander, Colombia. Its geographic location is shown in Figure 53-a). The raw GPR data is acquired using a shielded antenna with central frequency at 400 (MHz). The acquisition line is shown in Figure 53-b). The acquisition line is selected near the Mogoticos river, where a visible outcrop is accessible and it shows two geological layers with different materials (see Figure 53-c)).

The GPR system uses a shielded antenna with central frequency at 400 (MHz), and it takes a scan each 0.1 m. In total, 213 scans are taken. The total propagation time of the scan is $99,9987$ (ns) with 512 samples, a time step of 0.1953 (ns), and an acquisition line with length of 23.4 (m). CPML is used in the top, bottom, left, and right edges to avoid non-natural reflections.

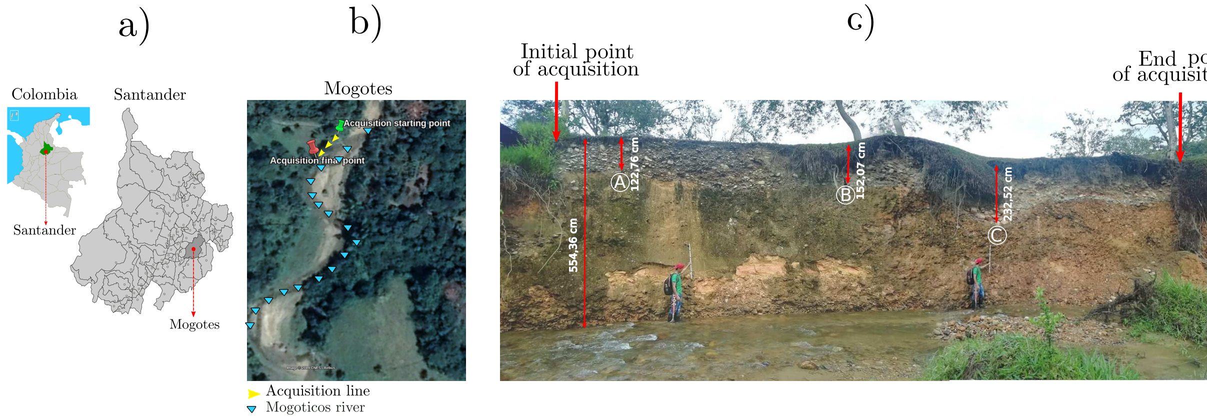

Figure 53. Acquisition area. a)Mogotes geographic location, b) acquisition line, c) outcrop in the acquisition zone

5.2. Initial permittivity and conductivity for the collected data

This section is focused on the initial parameters for FWI, where permittivity and conductivity are considered. The permeability is not taken into account in the inversion process because the materials in this zone are not magnetic. First, a picking in the raw data is performed to find the significant reflectors. Figure 54 shows the raw data and picking with three possibles layers. Next, the relative permittivity and conductivity values are selected taking into account the typical values of the subsurface materials (see Table 1), the field samples, and a visual inspection: in layer-A, the relative permittivity is 4 and conductivity 1 (mS/m), the material is considered a dry clayey soil. In layer-B, the relative permittivity is selected as dry-sand with a value equal to 6 and the conductivity is 1 (mS/m); and finally, the layer-C is chosen to be wet clay, and the relative permittivity value is 12 with the conductivity of 4 (mS/m). These parameters are discussed with geologists who know the area and determine those typical values according their knowledge of the area. In the zone, no laboratory tests are taken in the materials, and it could produce inaccuracies in the estimation of electromagnetic parameters.

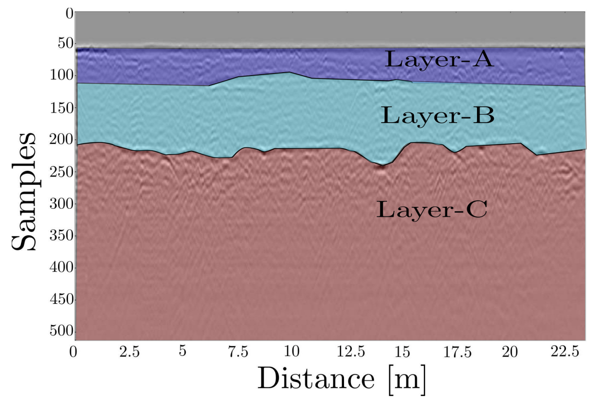

Figure 54. Raw GPR data interpretation with three possibles layers based on main reflectors: layer-A soil-clayey,dry, layer-B sand-dry and layer-C clay-wet.

The picking time and the relative permittivity of each layer allow obtaining the depth of each layer to build the initial model for permittivity and conductivity. The first depth layer is computed using Equation (127) and for the second and third layer is computed using Equation (128).

$$D_l = \frac{(t_l - t_{air}) \cdot c}{2\sqrt{\varepsilon_{r,l}(\mathbf{r})}}, \tag{127}$$

$$D_l = D_{l-1} + c \cdot \frac{(t_l - t_{l-1})}{2\sqrt{\varepsilon_{r,l}(\mathbf{r})}}, \tag{128}$$

where l is the layer index, t_l is the picking time with high amplitude (see Figure 54), t_{air} is the time for the air wave, it is approximately 0.9375 ns, c is the velocity in the vacuum which is approximately 3×10^8 m/s, and $\varepsilon_{r,l}$ is the relative permittivity in the layer l.

Figures 55-a) and 55-c) illustrates the permittivity and conductivity models for the layers.

The spatial step is 0.025 (m) in x and z-directions; and Figures 55-b) and 55-d) show the initial models using a bluring filter of size 5×5 and all its elements equal to 0.04; the filter is applied 500 times in each parameter.

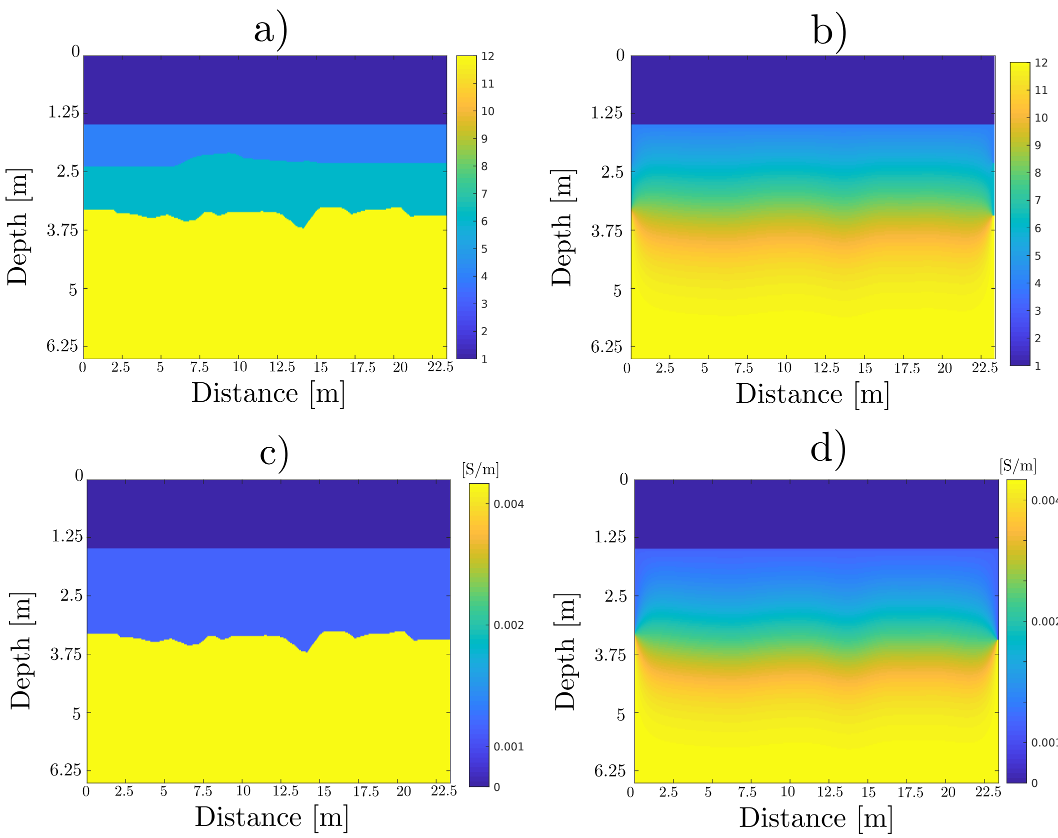

Figure 55. Initial parameters used in the FWI methodology: a) relative permittivity model using the picking time layers; b) Initial relative permittivity model obtained after apply the smoothing filter 500 times over the model illustrated in Figure (55)-a, c) conductivity model using the picking time layers; d) Initial conductivity model obtained after apply the smoothing filter 500 times over the model illustrated in Figure (55)-c).

5.3. Source wavelet estimation using the impulse response system

In the GPR scenario of single-channel and short-offset acquisition, the antenna is located in the air, and the sign source can be estimated with the air-wave events. The relationship of the source with the data in the airwave event is linear; therefore, the observed data can be obtained as the convolution between the source ($\mathbf{J}_s$) and the impulse response of the earth ($\mathbf{h}_{earth}$).

$$\mathbf{d}_{obs} = \mathbf{J}_s * \mathbf{h}_{earth}. \tag{129}$$

However, both the source and the impulse response of the earth are unknown.Belina et al. (2012) proposed that the impulse response of the earth can be estimated using a synthetic source and a modeled data that is obtained using the synthetic source. In the source estimation, the modeled data can be obtained modeling on any model. In the synthetic and collected data, the background filter is used to obtain the air wave events; the details of this filter are presented in the section 5.3.3. Mathematically, it is given by

$$\mathbf{d}_{mod}^{synt} = \mathbf{J}_s^{synt} * \mathbf{h}_{earth}. \tag{130}$$

If the Fast Fourier Transform (FFT) is applied in both sides of the Equation (130), then:

$$\mathbf{D}_{mod}^{synt}(f) = \mathbf{J}_s^{synt}(f) \cdot \mathbf{H}_{earth}(f), \tag{131}$$

from the Equation (131), the impulse response of the earth is obtained as

$$\mathbf{H}_{earth}(f) = \mathbf{D}_{mod}^{synt}(f) \cdot [\mathbf{J}_s^{synt}(f)]^{-1}. \tag{132}$$

Using the impulse response of the earth in the frequency domain ($\mathbf{H}_{earth}(f)$), the estimated source is obtained from the observed data as

$$\mathbf{J}_s = IFFT(\mathbf{D}_{obs}(f) \cdot [\mathbf{H}_{earth}(f)]^{-1}) = IFFT(\mathbf{D}_{obs}(f) \cdot [\mathbf{D}_{mod}^{synt}(f) \cdot [\mathbf{J}_s^{synt}(f)]^{-1}]^{-1}), \quad (133)$$

where $IFFT(\cdot)$ is the Inverse Fourier Transform. According to Equation 133 the source estimation is computed with the following inputs: a synthetic source with sufficiently broad frequency bandwidth in order to avoid divisions by zero in the deconvolution (see Equation 133), the modeled data using the synthetic source, and the observed data. In chapter 7, the estimation of the radiation patterns for planes E and H are presented for the shielded antenna at 400 MHz. However, it is essential to clarify that these radiation patterns are frequency-dependence therefore, it is necessary a source estimation when the observed data is filtered for FWI.

5.3.1. Source estimation in a synthetic data.. A synthetic case modified from the SEAM model is developed with the original conductivity and permittivity models presented in Figure 17-a) and 17-b), respectively. The initial model is a smoothed version of the original models. The spatial discretization is 5 (cm) in both dimension, such that 600 points are x-direction (distance) and 301 in z-direction (depth). A total of 520 scans are simulated using the equations for the forward electromagnetic modeling. Transmitter and receiver antennas are located in the air-layer at 25 (cm) of the ground layer, according to the topography. The time sampling of the data is 0.08 (ns), and the number of time samples is 3000. Thus, the total acquisition time for each scan is 240 (ns). The original and synthetic sources are shown in Figure 56. The original source is the source emitted by the transmitter antenna and produces the observed data, in this example both modeled and original are synthetic and known, but in the real context, the signature original source is unknown.

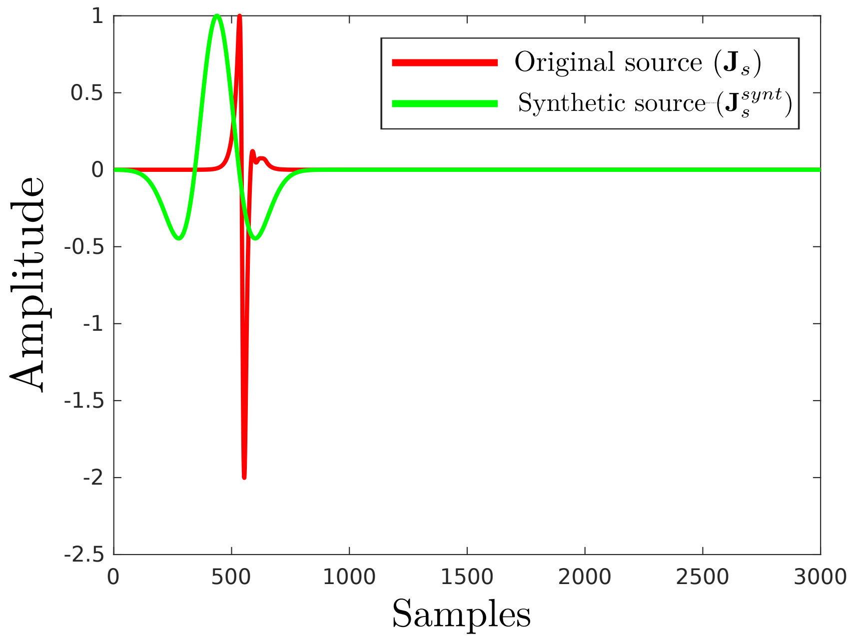

Figure 56. The synthetic source used to obtain the modeled data is presented in green and the original source is presented in red. The original source is the source emitted by the transmitter antenna and produces the observed data.

In single-channel and short-offset acquisitions, the values of permittivity and conductivity are known in the air-wave ($\varepsilon_r = 1$ and $\sigma_r = 0$). The air-wave event is separated using the *background* filter (Rashed, 2015). The *background* filter removes the horizontal events of the radargram. The horizontal events are obtained with the difference between the data without background filter and the data with background filter. The observed an modeled data are shown in Figure 57-a) and Figure 57-b), respectively; the horizontal events for the observed and modeled data are presented in Figure 57-c) and Figure 57-d), respectively. In total, 520 sources estimates are obtained, as it is shown in Figure 58 -a). All these approximations are averaged to have the estimated source. Figure 58 -b) shows the estimated source, the original source and the synthetic source. The correlation between the estimated source and the original source is 0.99. The correlation is high because this is an scenario where the exact permittivity and conductivity values of the air-wave are known and besides, the observed data is noise free.

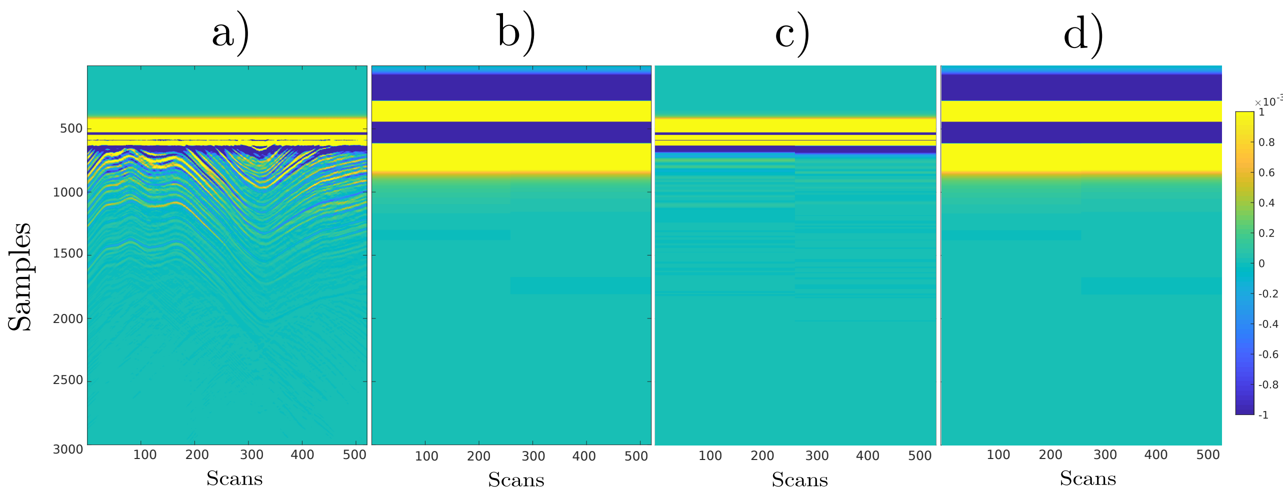

Figure 57. a) Observed data, b) Modeled data, c) horizontal events in the observed data and d) horizontal events in the modeled data.

5.3.2. Source estimation for real data. The source estimation methodology described in the previous section is now used with the collected data. The GPR system takes a scan each 0.1 (m). Two hundred thirteen scans are taken with 512 samples, and the sampling rate is 0.1953 (ns). The acquisition line has a length of 21.4 (m). CPML is used in top, bottom, left and right edges to avoid non-natural reflections.

The observed data is presented in Figure 59 -a). However, as in the synthetic case, only the air-wave is selected. According to (Lavoué, 2014) the direct arrivals are selected using an

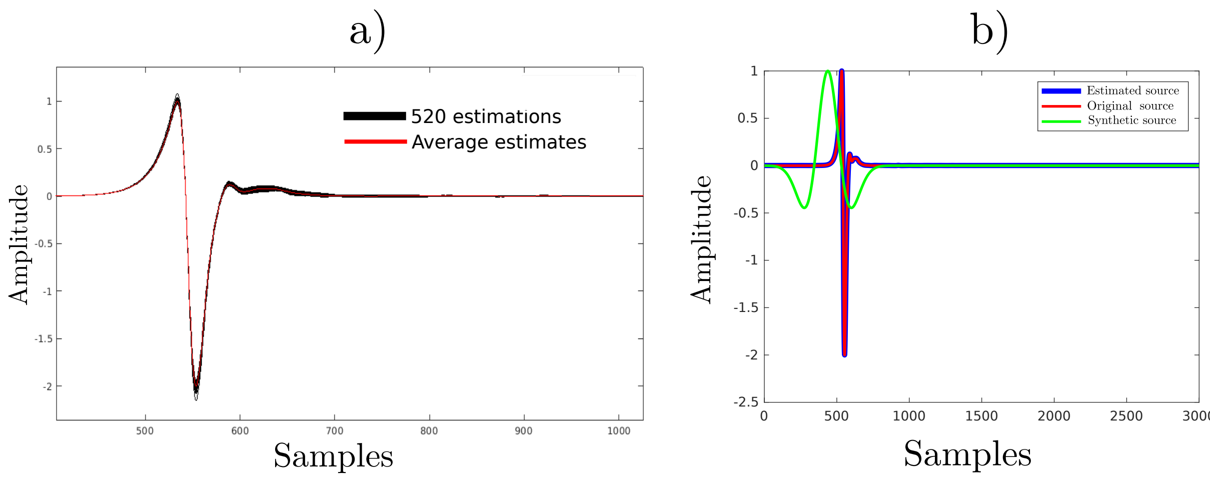

Figure 58. a) Estimations are presented in black, and the average of these estimations is presented in red color, b) the estimated source is presented in blue, the original source used to obtain the observed data is shown in red, and the synthetic source is shown in green.

exponential time damping, the exponential damping function is defined as follow:

$$d_{obs}(scan,t) = e^{-(t-t_{AW}(scan))/\tau} d_{obs}(scan,t), \tag{134}$$

The result of applying this exponential damping function is shown in Figure 59 -b). A wavelet ricker with the same time sampling and the same number of samples, as in the observed data, is selected for the synthetic source. For each scan, a source estimate is performed (see Equation (133)), such that in total, 213 estimations are obtained. The 213 estimations are averaged, resulting in the signal in blue of Figure 60.

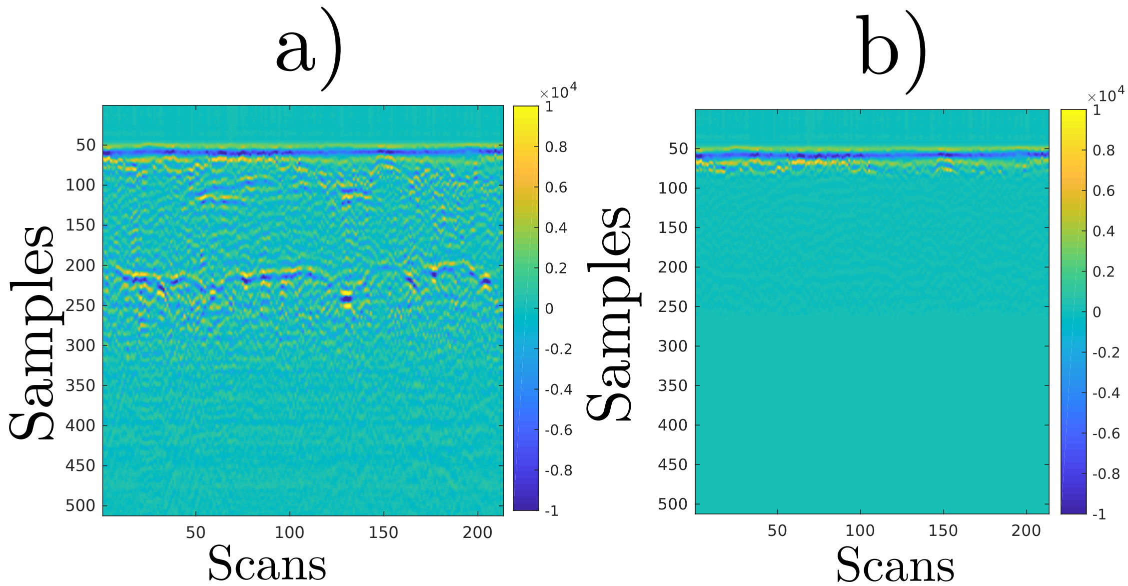

Figure 59. a) Observed data using a shielded antenna at 400 (MHz), b) Observed data with exponential time damping function.

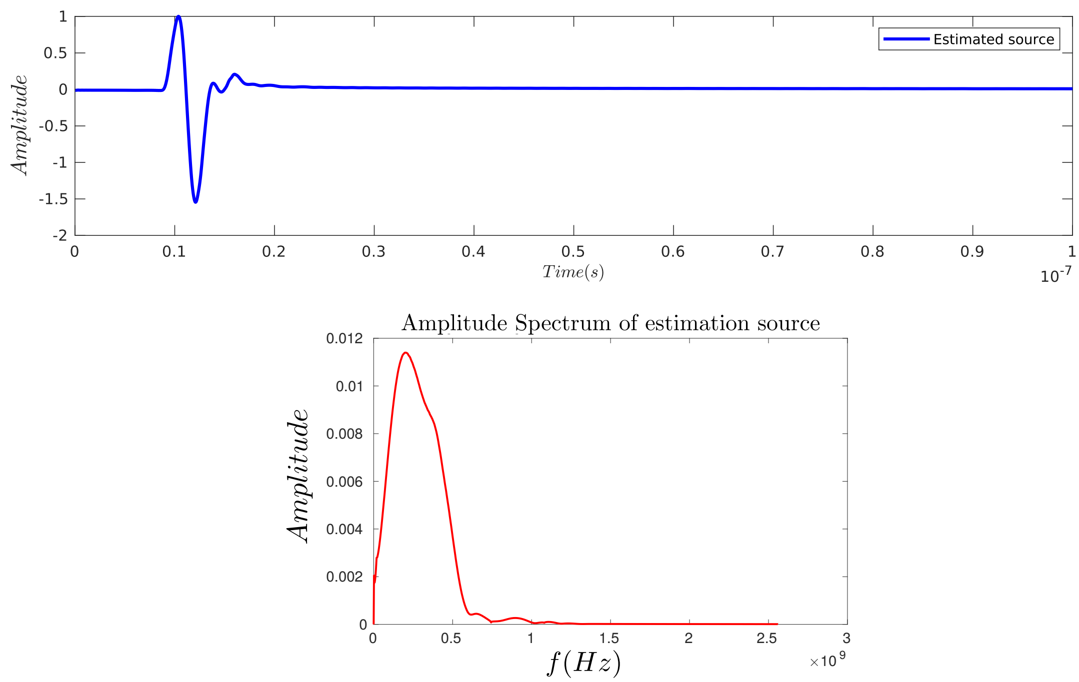

Figure 60. Estimated source and amplitude spectrum for the collected data in Mogotes.

The Antenna distance to the subsurface is tuned, taking values from 0.025 (m) to 0.1 (m), in steps of 0.025 (m). The norm ℓ_2 is computed between the observed and modeled data. The observed and modeled data of the best fit are presented in Figure 61 -a) and the best fit for the distance of the antenna to the surface is 0.05 (m) as it is shown in Figure 61 -b). Finally, the parameter $\sigma_0 = 5.0e-4$, is only for having a reference value and using a relative conductivity in the inversion, but this parameter does not make any physical sense as it is described in (Lavoué et al., 2014).

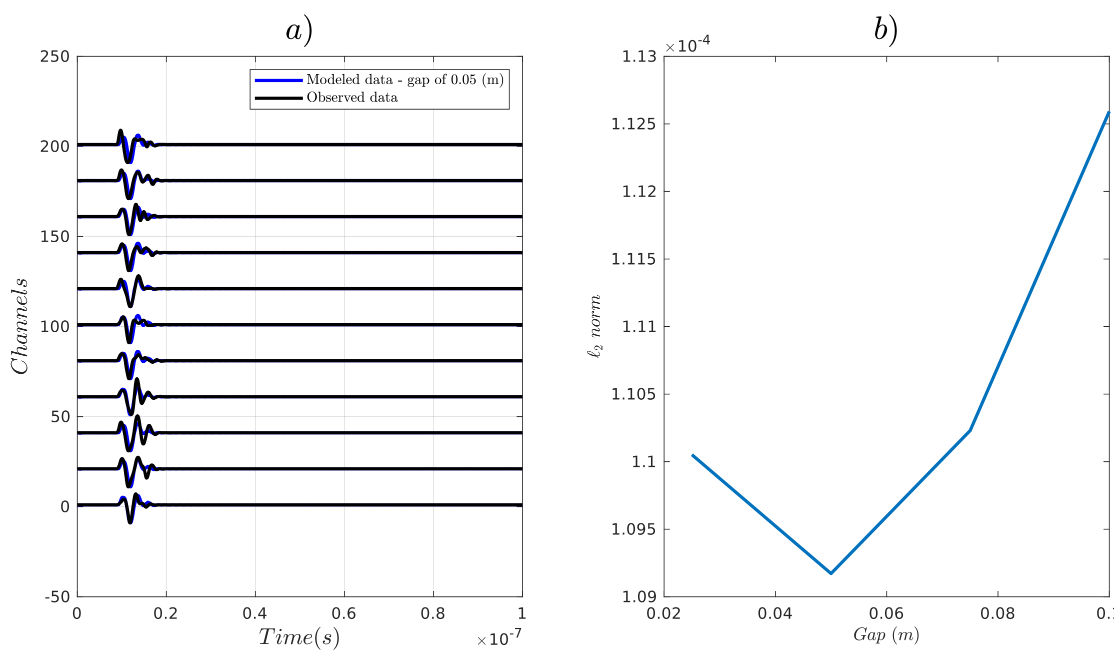

Figure 61. a) The fit between modeled and observed data when the distance between the source and subsurface is 0.05 (m). b) ℓ_2 norm between modeled and observed data when the distance between the transmitter antenna and subsurface changes between 0.02 (m) and 0.1 (m).

5.3.3. Preprocessing the collected data. The following processing steps are applied over the raw data:

- Dewow filter: The Dewow filter selects a window and takes the mean value in the window to remove the DC components in each sample. An alternative to implement this filter is to apply the FFT on the selected window and put to zero the first coefficient (this coefficient is associated with the DC level), then perform the IFFT to obtain the time domain signal.

- Lowpass filter (LP): FWI is a local optimization technique that requires starts from low frequencies to high frequencies to get a better parameter estimation. Since the data is collected using a 400 (MHz) antenna, the radargram has been filtered using a low pass filter at 250 (MHz) (see Figure 62). Furthermore, the source used is also filtered at this frequency range. Figure 63 shows the filtered source.

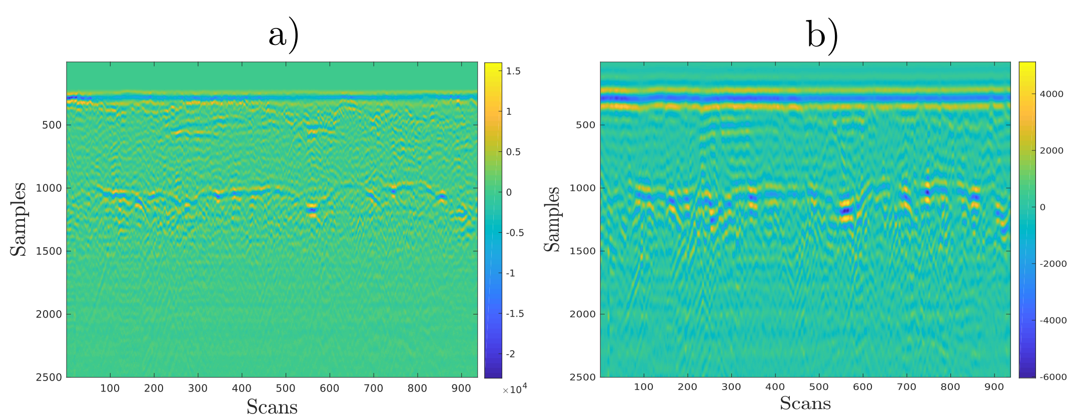

Figure 62. Collected data in the acquisition zone: a) unfiltered data b) data filtered using a low pass filter between 0-250 (MHz).

- Background filter: The background filter is used to mitigate the horizontal and periodic events in the observed data, and it is called ringing noise. The ringing noise is removed using two steps, according to Khan and Al-Nuaimy (2010). The first step is to compute the

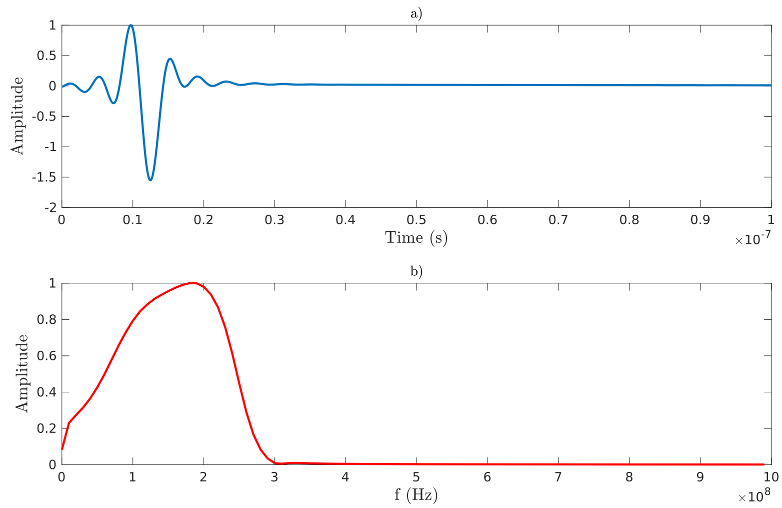

Figure 63. Source used during the inversion process for the collected data: a) filtered source in the time domain at 250 (MHz), b) spectrum of the filtered source.

eigenimages containing the highest and lowest values of the eigenvalues of the covariance matrix. The second step is to subtract the average value of the radargram. Figure 64 shows the results when the background filter is applied to the data. The background filter is only applied in the first 500 samples where the air-wave is observed, and it avoids attenuate some flat events of interest. The air-wave is removed in the observed and modeled data. Algorithm 2 summarizes the methodology to apply the background filter in $\mathbf{d}_{obs}$.

- Resampling (optional): using the stability condition proposed by Courant-Friedrichs-Lewy (CFL), (see Equation (135)) the time step necessary to comply with numerical stability can be obtained. As the spatial step is $\Delta_h = 0.025$ (m), we obtain that $\Delta t \leqslant 5.8966 \times 10^{-11}$ (s). In compliance with this condition, a time step equal to 0.04 (ns) is selected. The collected

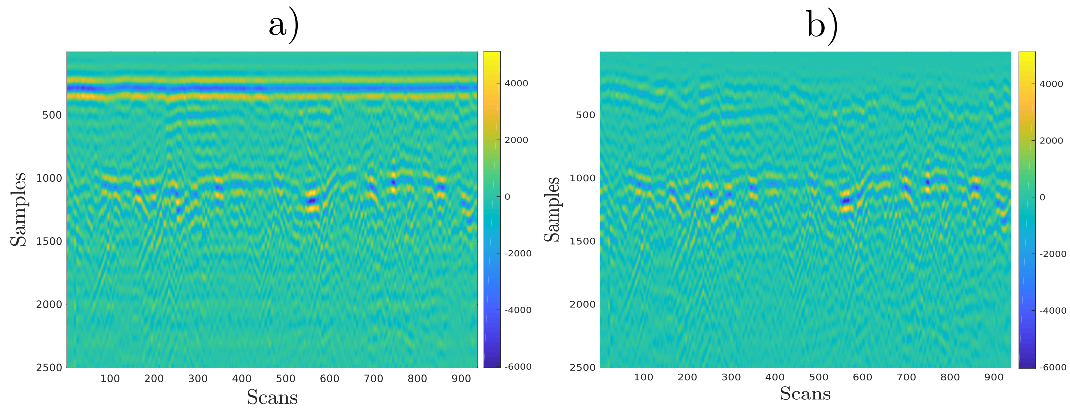

Figure 64. Background filter on the collected data: a) data collected with low pass filter and without background filter b) collected data filtered with low pass filter and background filter.

Algorithm 2 Background filter

C=$\mathbf{cov}(\mathbf{d}_{obs})$
L=$\mathbf{eig}$(C)
M=$L^T \cdot \mathbf{d}_{obs} \cdot L$ ▷ The diagonal values of M denote the eigenvalues of C
$[u_1, d_1]$=$\mathbf{eig}(\mathbf{d}_{obs}^T \cdot \mathbf{d}_{obs})$ ▷$\mathbf{eig}(\cdot)$ is the eigenvalues function and ^{T} denote transpose
$[u_2, d_2]$=$\mathbf{eig}(\mathbf{d}_{obs} \cdot \mathbf{d}_{obs}^T)$
$S = \mathbf{svd}(\mathbf{d}_{obs})$ ▷$\mathbf{svd}(\cdot)$ is the singular value descomposition function
$val_1 =$ Row corresponding to maximum eigen value of M
$val_2 =$ Row corresponding to mimimum eigen value of M
$P1 = S(val_1) * u_1(:, val1) * u_2(:, val_1)^T$ ▷$*$ denote matrix multiplication
$P2 = S(val_2) * u_1(:, val2) * u_2(:, val_2)^T$
$O_1 = \mathbf{norm}(\mathbf{d}_{obs}) - \mathbf{norm}(\mathbf{P}_1)$ ▷$\mathbf{norm}(\cdot)$ is the normalized function
$O_2 = \mathbf{norm}(\mathbf{d}_{obs}) - \mathbf{norm}(\mathbf{P}_2)$
$out = \mathbf{norm}(O_1). * \mathbf{norm}(O_2)$
$output = out - \mathbf{mean}(\mathbf{d}_{obs})$ ▷$\mathbf{mean}(\cdot)$ is the mean function

data is resampled from 0.1953 (ns) to 0.04 (ns).

$$\Delta t \leqslant \frac{\Delta_h}{c\sqrt{2}} \tag{135}$$

- Amplitude compensation: The difference in amplitude between observed and modeled data could be solved in the source estimation or in our tests using the normalized cost function (see Equation (96)). As is presented in section **4.1** when the amplitudes are not compensated, this could produce artifacts in the first layers of the parameters and cause the FWI to converge quickly to a local minimum.

5.4. FWI using TV and MTV regularizations.

TV and MTV regularizations allow obtaining smoother solutions for the electromagnetic parameters in the inversion process. TV regularization has been included in the inversion process to analyze its behavior in the real collected data. The definitions of gradients and the procedures described in the section **4.3** are used in this section. The regularization parameter for relative permittivity is fixed at $\lambda_{TV}(\varepsilon) = 5.0e-3$. Figure 65 -c), d) and e), show the results of the inversion process with TV using the values $\lambda_{TV}(\sigma)$: 2.5×10^{-3}, 5.0×10^{-3} and 7.5×10^{-3}, respectively. Figure 66 shows the behavior of the cost function for 20 iterations. The worst data misfit is reached with no regularization and the better results are with $\lambda_{TV}(\sigma) = 2.5\times10^{-3}$ or $\lambda_{TV}(\sigma) = 5.0\times10^{-3}$. Note that in the Figure 65 -g) having no regularization produces undesired values in the conductivity parameter, which does not have a physical sense. In addition, to generate a smooth version of the parameters, the TV regularization does not allow extreme changes in the permittivity and conductivity values, so the inversion is more stable. Figure 65-f) presents a very smooth solution for the conductivity parameter, so it is discarded. Figure 67 shows the fit achieved between the observed and modeled data for 11 scans. The inversion is stopped for the case of the FWI with TV and MTV when a flat area is reached in the cost function; although the algorithm can perform more iterations, the variation in the cost function is very small and therefore the algorithm stops. It can be noticed that the modeled data that our modeled data do not correctly estimate the first events of the observed data. This could be due to a poor estimate of the initial values for the permittivity and conductivity. In the case of not having the regularization, it manages to adjust these regularization, the method adjusts the first events better than including regularization, but the parameter solution is not correct. The electromagnetic parameters estimations with TV are better than a traditional FWI without regularization. Figure 68 presents the results by including the scale parameter β taking values of 0.25, 0.5, 0.75 and 1.0. The behavior of the cost function taking differentes values of β is show in Figure 69. Details about β parameter is presented in the section 4.3.1. From Figure 67 the best fit of the data is reached with $\beta = 1.0$.

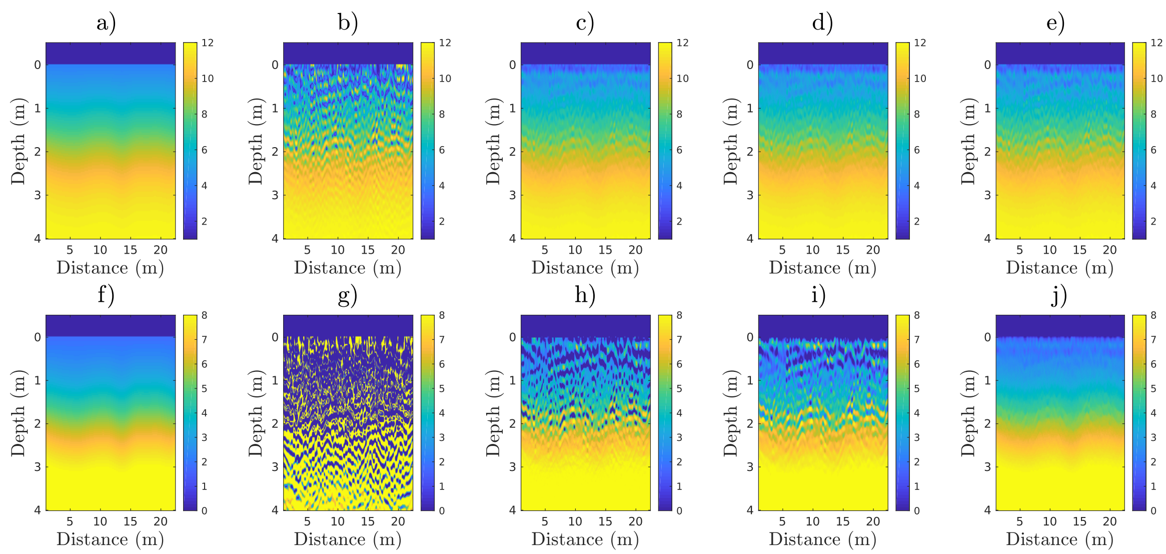

Figure 65. FWI results with TV regularization in both parameters, $\lambda_{TV}(\varepsilon) = 5.0e-3$ and $\lambda_{TV}(\sigma)$ takes values of $2.5e-3$, $5.0e-3$ and $7.5e-3$: a) and f) initial models; b) and g) No regularization; c) and h) $\lambda_{TV}(\varepsilon) = 5.0e-3$, $\lambda_{TV}(\sigma) = 2.5\times10^{-3}$; d) and i) $\lambda_{TV}(\varepsilon) = 5.0e-3$, $\lambda_{TV}(\sigma) = 5.0\times10^{-3}$ and e) and j) $\lambda_{TV}(\varepsilon) = 5.0e-3$, $\lambda_{TV}(\sigma) = 7.5\times10^{-3}$. The first row shows the relative permittivity and the second row shows the relative conductivity.

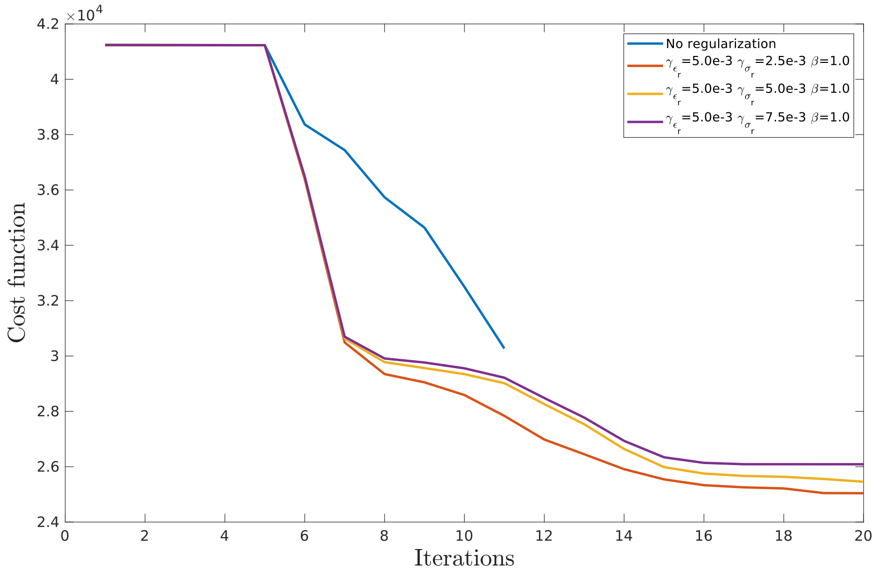

Figure 66. The behavior of the cost function for the data collected with TV regularization.

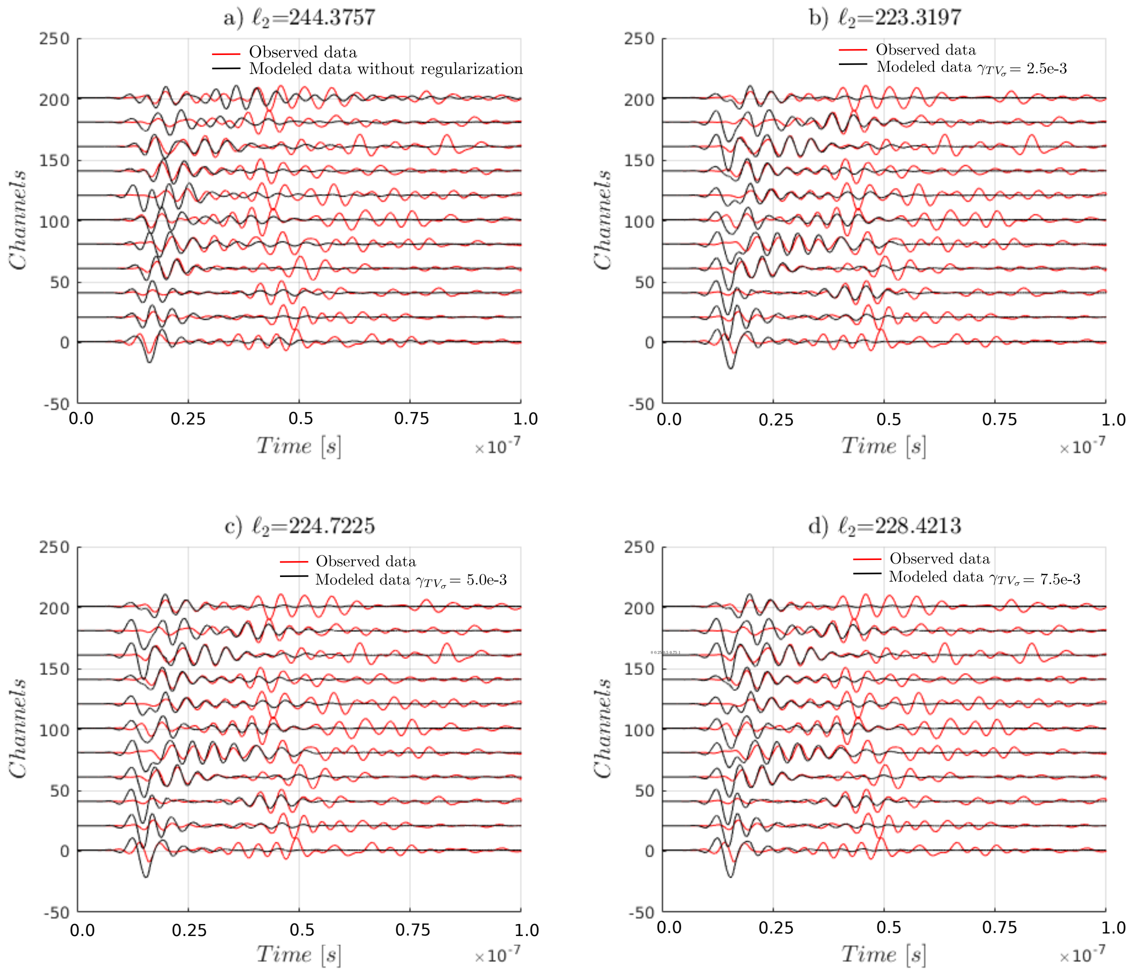

Figure 67. The fit between the observed and modeled data is presented in each panel using TV regularization. Each panel shows 11 scans observed and modeled; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) without regularization, b)$\gamma_{TV}(\varepsilon_r) = 5.0e-3$ and $\gamma_{TV}(\sigma_r) = 2.5e-3$, c) $\gamma_{TV}(\varepsilon_r) = 5.0e-3$ and $\gamma_{TV}(\sigma_r) = 5.0e-3$, d) $\gamma_{TV}(\varepsilon_r) = 5.0e-3$ and $\gamma_{TV}(\sigma_r) = 7.5e-3$.

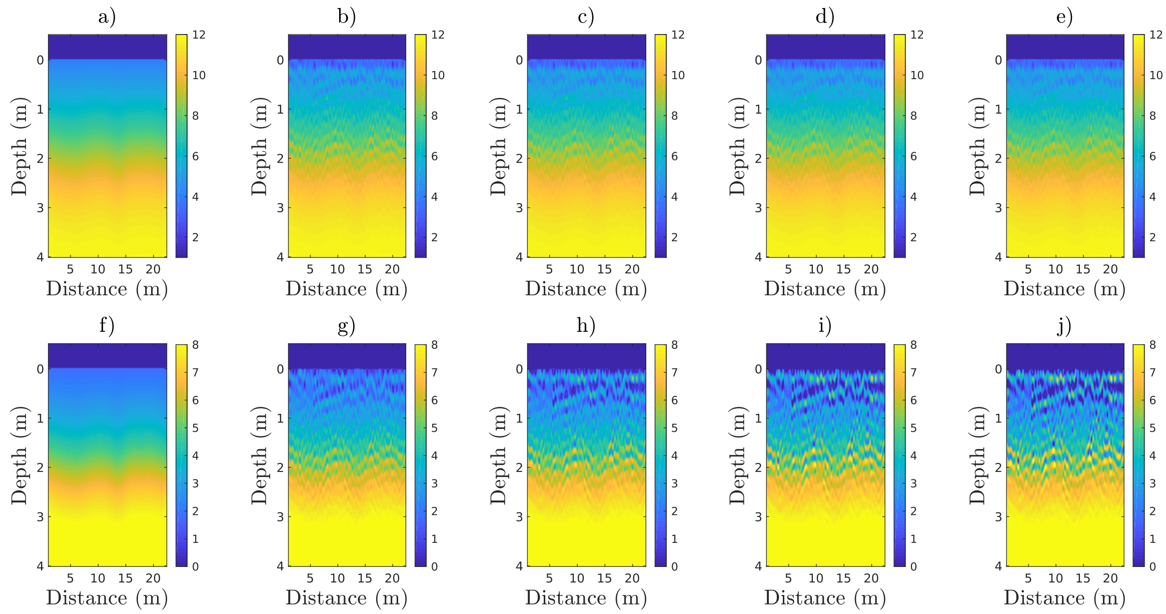

Figure 68. FWI results with TV regularization in both parameters, $\lambda_{TV}(\varepsilon) = \lambda_{TV}(\sigma) = 5.0e-3$ and β takes values of: a) and f) initial model, b) and g) 0.25, c) and h) 0.5 , d) and i) 0.75, e) and j) 1.0. The first row shows the relative permittivity and the second row shows the relative conductivity.

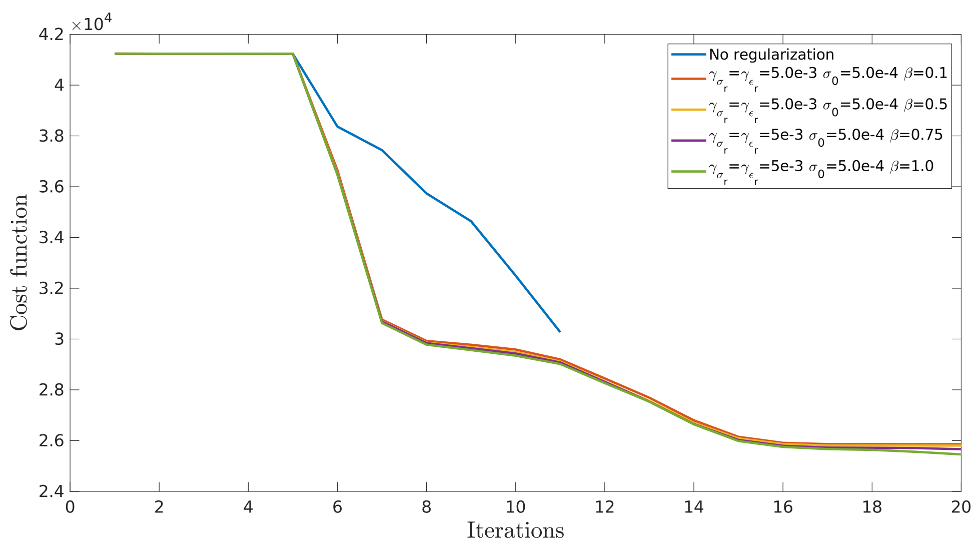

Figure 69. The behavior of the cost function for the data collected with TV regularization and the scale parameter β.

Similarly to the TV regularization, the MTV regularization is included in the inversion process, where the definitions of gradients and the procedure described in section **4.4** have been used. The regularization γ is tunned to 0.05 according to section **4.4**. However, the parameter κ must be tuned. Figures 70-b), c) and d), shows the inversion results with MTV regularization using the values $\kappa(\sigma)$= 100, 150 and 200, respectively. Figure 71 shows the behavior of the cost function for 20 iterations where the flat area is reached. The MTV regularization converges to a less smooth model than the TV regularization. Both regularizations allow controlling the permittivity and conductivity parameters, preventing the conductivity from converging to undesired values. Figure 72 shows the models obtained from FWI when the scale parameter β takes values of 0.25, 0.5, 0.75 and 1.0. The results of the cost function are presented in Figure 73. Note that the best fit is reached with $\beta = 0.75$. The cost function is better on MTV than TV as presented in trace comparisonshown in Figure 74. Finally, our TV and MTV regularizations only manage to fit the events that arrive before 1.0 $\times 10^{-7}$ (s). As described above, the inversion is stopped when it reaches the flat zone.

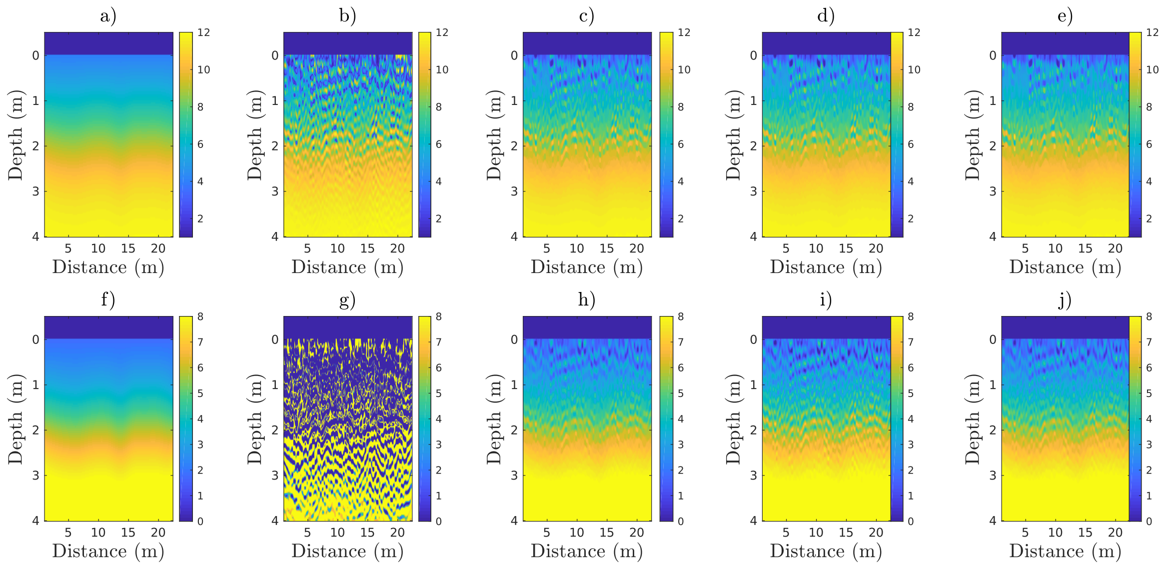

Figure 70. FWI results with MTV regularization in both parameters, κ takes values of 100, 150 and 200. a) and f) initial model, b) and g) no regularization, c) and h) $\kappa = 100$, d) and i) $\kappa = 150$, e) and j) $\kappa = 200$. The first row shows the relative permittivity and the second row shows the relative conductivity.

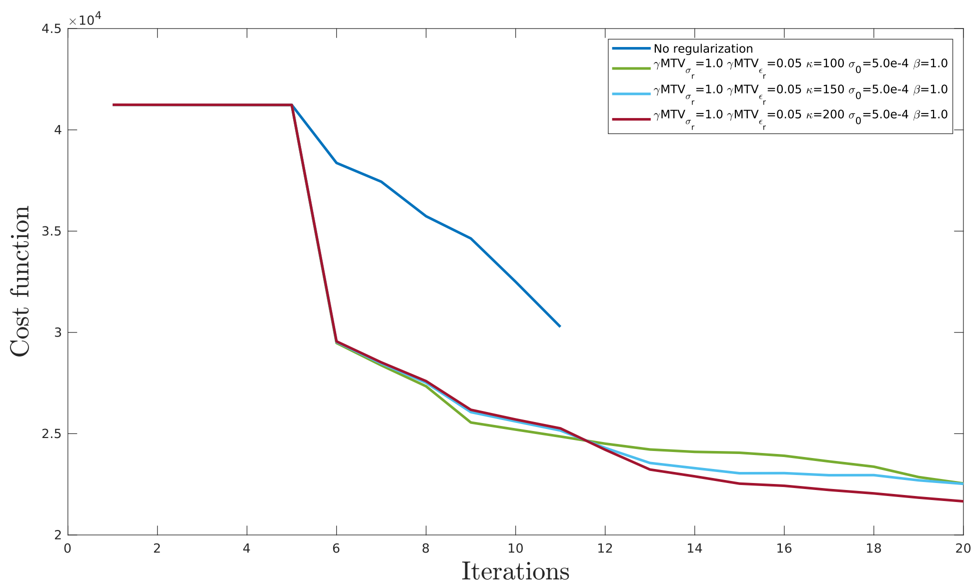

Figure 71. The behavior of the cost function for the data collected with MTV regularization and changing κ on the parameter σ.

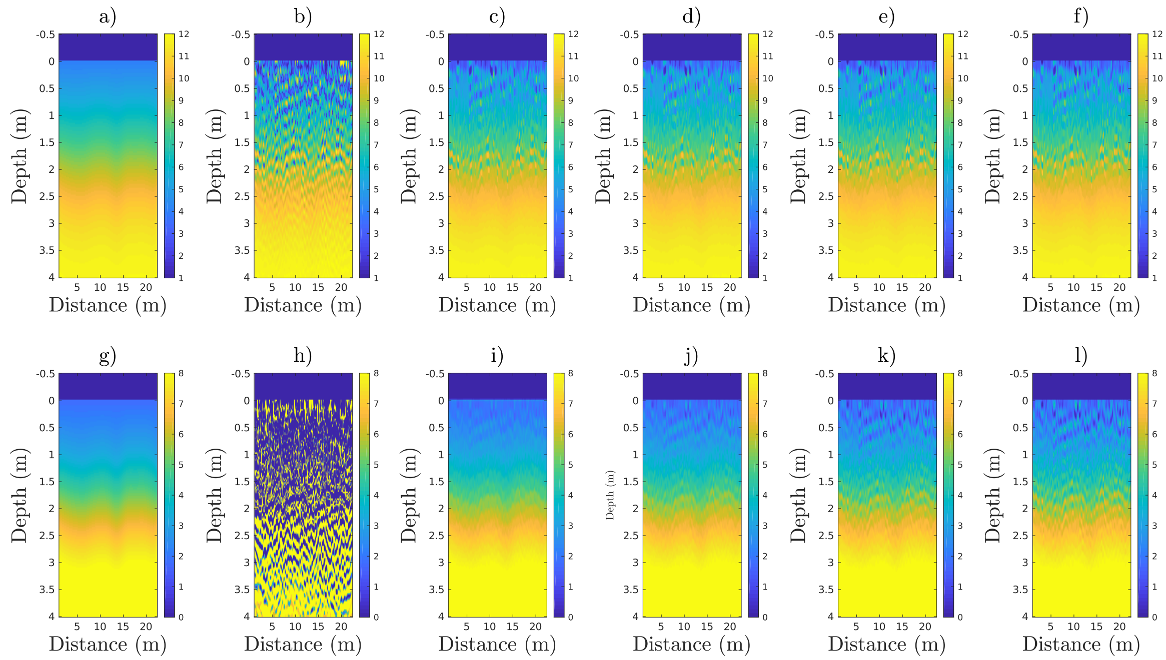

Figure 72. FWI results with MTV regularization when the parameter $\kappa = 200$ and β takes values of 0.25, 0.5, 0.75 and 1.0: a) and g) initial model, b) and h) no regularization, c) and i) $\beta = 0.25$, d) and j) $\beta = 0.5$, e) and k) $\beta = 0.75$ and f) and l) $\beta = 1.0$. The first row shows the relative permittivity and the second row shows the relative conductivity.

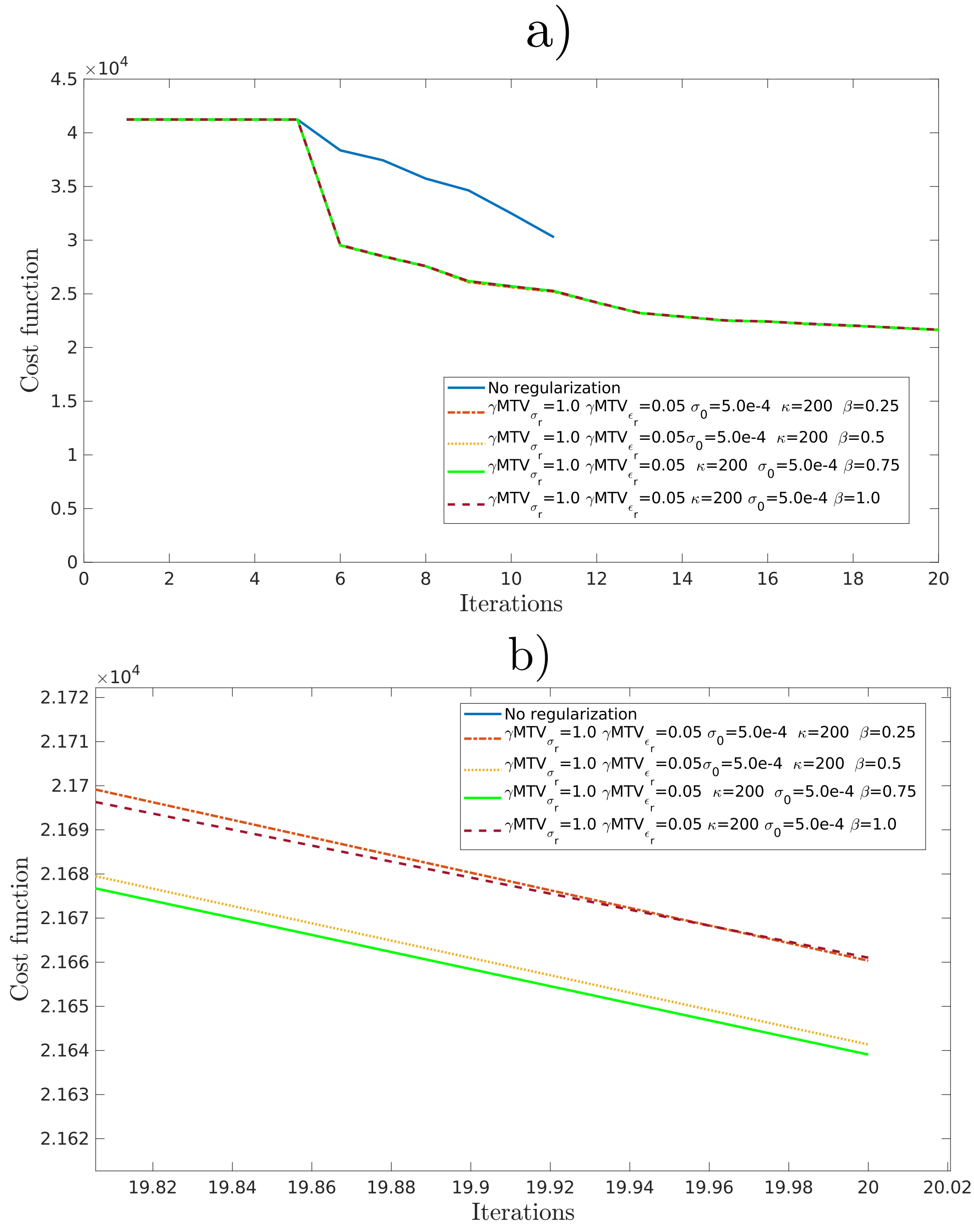

Figure 73. The behavior of the cost function for the data collected with MTV regularization and β takes values of 0.25, 0.5, 0.75 and 1.0.

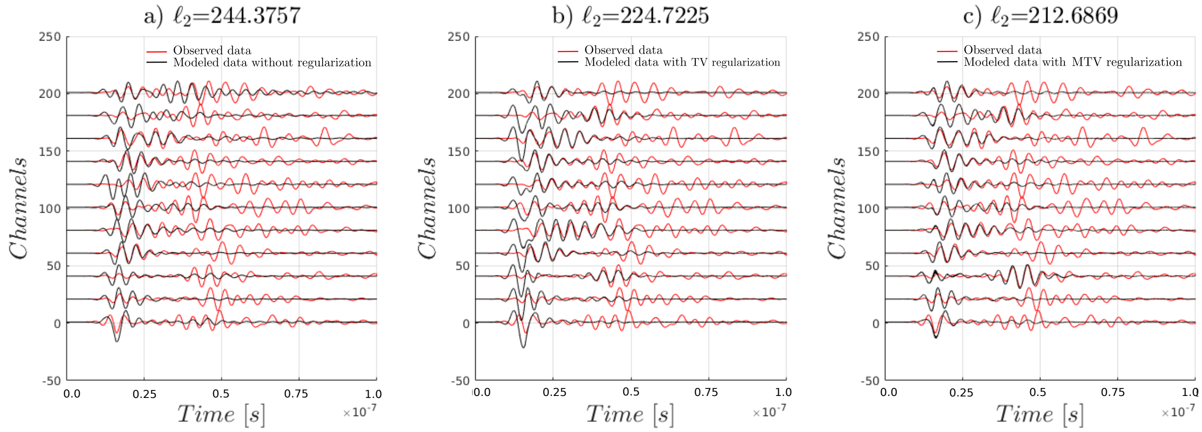

Figure 74. . The fit between the observed and modeled data is presented in each panel using TV and MTV regularization. Each panel shows 11 scans observed and modeled; in each panel, the airwave is removed. In addition, the ℓ_2 norm for the residual is shown in the upper part of each panel. a) no regularization, b) TV regularization and c) MTV regularization.

5.5. Gaussian preprocessing-TV regularization and Gaussian preprocessing-MTV regularization

TV and MTV regularization are combined in this section with the gaussian preprocessing. The methodology and definition of gradients presented in section **4.2** are used in this section. According to section **4.2**, the best estimation of the parameters is obtained using $\rho^2 = 2$ in *scans*-directions with a fixed window of length 66 using the FWI gradient resolution (see Equation 105). The results of the TV regularization with Gaussian preprocessing and the MTV with Gaussian preprocessing is shown in Figure 75. The Gaussian preprocessing reduces the incoherent noise in the scans-direction and the solutions reach low-wavenumber information. Likewise, Figure 76 shows the cost function values achieved in the two regularizations (TV + Gaussian and MTV + Gaussian). Gaussian preprocessing achieves the worse fit data on TV and MTV; this is due to the loss of horizontal resolution when the high frequency is used. Figure 75 shows a surface layer

(<0.25 (m)) where we interpret that it is the layer of soil with vegetation.

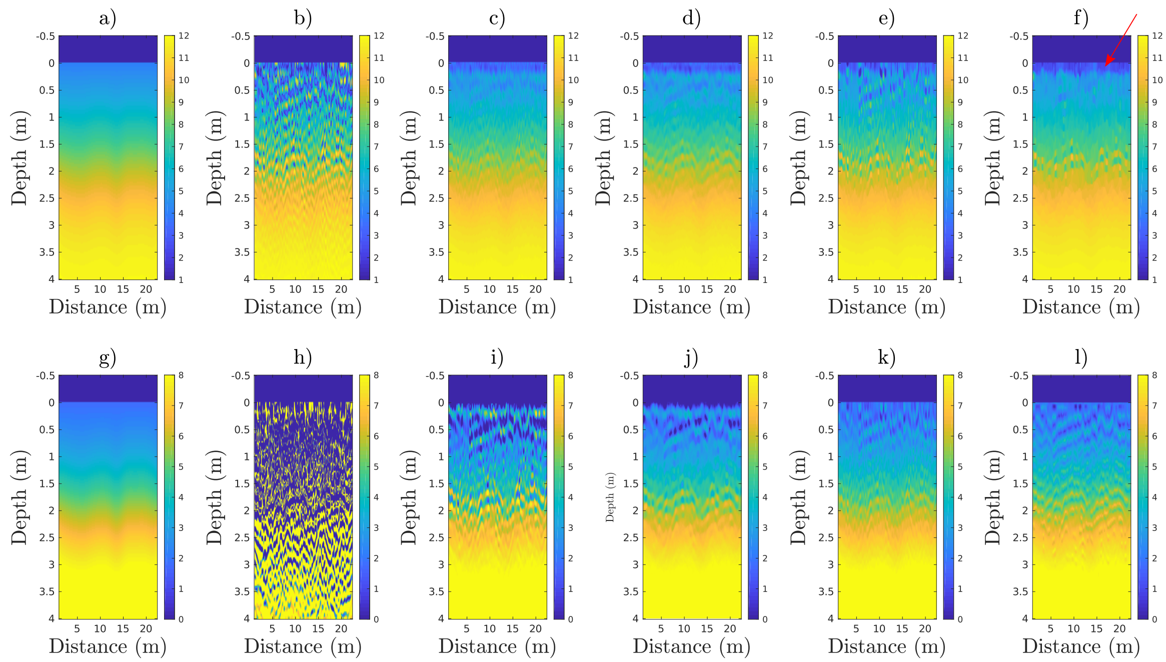

Figure 75. FWI results with TV, TV+Gaussian, MTV and MTV+Gaussian. a) and g) initial models, b) and h) no regularization, c) and i) TV regularization, d) and j) TV and Gaussian preprocessing, e) and k) MTV regularization, f) and l) MTV and Gaussian preprocessing. The first row shows the relative permittivity and the second row shows the relative conductivity.

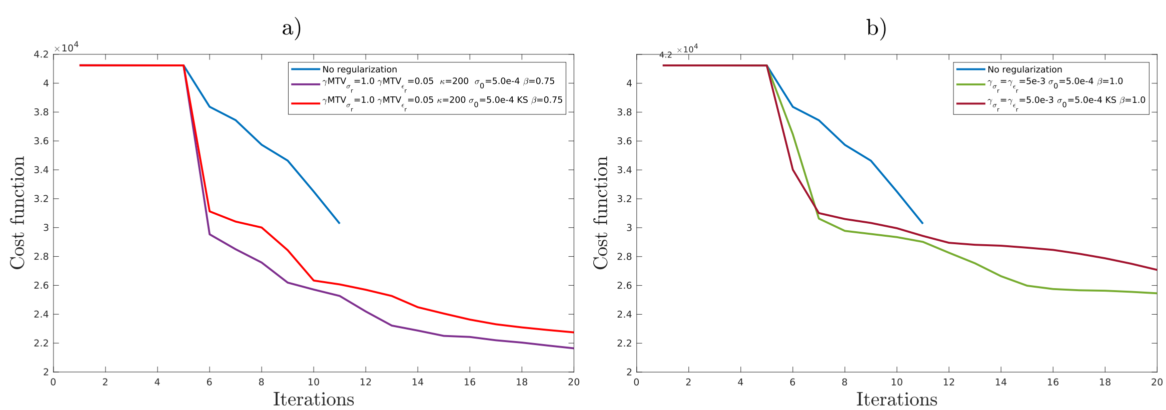

Figure 76. The behavior of the cost function for the data collected with : a) MTV and Gaussian preprocessing and b) TV and Gaussian preprocessing.

5.6. Quality control in FWI

The complex trace attributes allows measuring the coherence in the instantaneous phase between the events from a reflectivity map. The reflectivity map is obtained by the reverse time migration (RTM) method. RTM is a geophysical method developed by Baysal et al. (1983), and it uses the solution of the full-wave equation. The reflectivity map is computed in RTM using two wavefields: the first wavefield called source wavefield ($\mathbf{E}_s(\mathbf{r},t)$) is obtained from solving the wave equation with a source. The second wavefield called receiver wavefield ($\mathbf{E}_r(\mathbf{r},t)$) is obtained from the backpropagation of the wave equation using the information from the observed data as the source. The air-wave has been removed using the background filter presented in section **6.2.0.1**; it reduces the high contrast on the topography layer. The Equation 136 refers to the imaging condition used to obtain the reflectivity map by RTM using the source field and the receiver field.

$$I_{RTM}(\mathbf{r}) = \sum_{i=1}^{Ns} \sum_{t=0}^{T_{end}} \mathbf{E}_s(\mathbf{r},t)\mathbf{E}_r(\mathbf{r},t). \tag{136}$$

a Laplacian filter is applied in the imaging condition to remove the low-frequency artefacts due to back-scattering, and as the last step, a deconvolution is applied in the reflectivity map to remove the footprint of the source a find the location of the reflector. Two pairs of models of permittivity and conductivity have been used to obtain the reflectivity map: the FWI results models with the TV regularization using $\gamma_{TV}(\varepsilon) = 5.0e-3$ and $\gamma_{TV}(\sigma) = 5.0e-3$ (Figure 65-d), and the FWI results models using the MTV regularization with $\kappa_{MTV} = 200$ (Figure 72-e). The results of each pair of models previously described are presented in Figure 77 -a) and b) respectively.

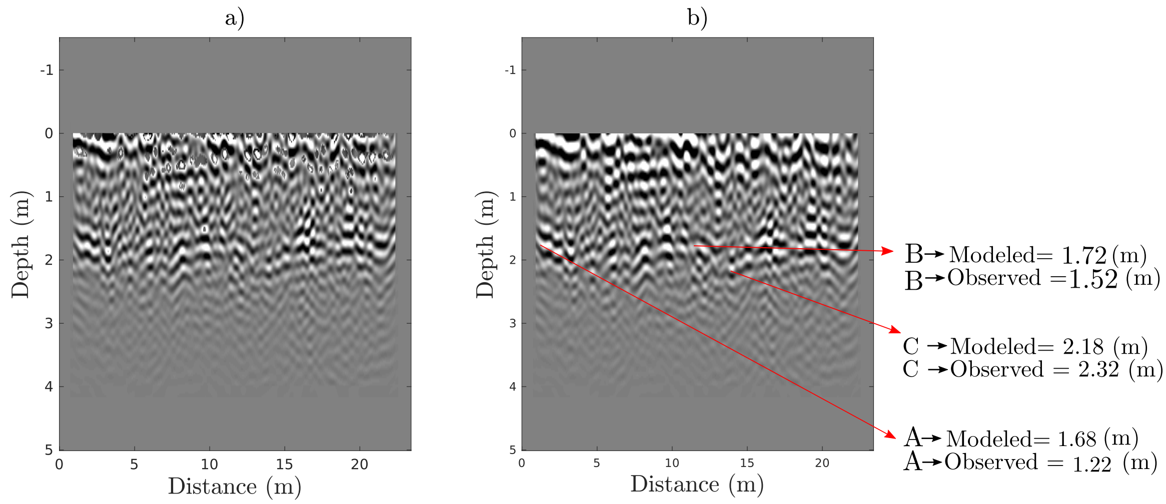

Figure 77. a) Reflectivity map for the TV models, and b) reflectivity map for the MTV models.

Figure 77 shows that the reflectivity map for the MTV model is better than for the TV model. This is because the TV model has conductivity values close to zero in the area near to the surface, which is not possible. The error between the estimate reflectivity map and the real location of the reflectors, measured in the field, are: 1. In point A, 46cm, 2. In point B, 14 cm. 3. In point C, 20 cm..

Jazayeri et al. (2018), proposed a method that could be applied when there is clear evidence of diffraction hyperbolas that indicate the presence of a pipe, but its diameter and internal content are unknown. The method is tested on synthetic data and two real data using a center frequency of 900 MHz. According to Jazayeri et al. (2018), the background parameters are known (permittivity and conductivity). The initial model is estimated using ray-based analysis of the hyperbolas previously identified by the interpreter's experience. The conductivity parameter always remains fixed in the inversion process. Jazayeri et al. (2018), an iterative source inversion problem is posed that is executed twice. In Jazayeri et al. (2019), the sparse blind deconvolution (SBD) method is included to estimate the source and obtain a sparse representation of the subsurface reflectivity series. With the reflectivity model and the ray-based model, an initial geometry model can be found to start FWI.

Unlike our estimates, these estimates have a higher frequency and the subsurface models in our case are complex (synthetic and collected). In our synthetic and collected data, the conductivity is updated using FWI. Our initial models are selected by geologist expertise as recommended by Jazayeri et al. (2018) when the diffraction hyperbolas identification is not clear. However, we consider as future work to use other classic methods (tomography) or to make measurements in the field that allow us to validate the starting point.

6. Full Waveform Inversion for 3D case

The 3D full waveform inversion of GPR data is computationally expensive, due to the number of hypercubes that should be stored in order to compute the gradient of the different parameters. This section presents a computational strategy to overcome the memory constrains and sucessfully implement a 3D FWI. The strategy consist of updating a set of 2D planes of the model instead of updating the entire gradient hypercube. The chapter compares our 3D electromagnetic implementation and the free software gprMax, where the correlation reached is 0.99. The experiments carried out in this chapter are synthetic and use the SEAM model inspired by the geology of the Colombian territory.

6.1. Forward and Inverse Problems for 3D case

Maxwell's equations for a non-dispersive and isotropic 3D medium are shown in Equation (137).

$$
\begin{aligned}
\frac{\partial E_x}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} + J_{sx} - \sigma E_x\right), \\
\frac{\partial E_y}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} + J_{sy} - \sigma E_y\right), \\
\frac{\partial E_z}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} + J_{sz} - \sigma E_z\right), \\
\frac{\partial H_x}{\partial t} &= \frac{1}{\mu}\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} - M_{sx} - \sigma^* H_x\right), \\
\frac{\partial H_y}{\partial t} &= \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} - M_{sy} - \sigma^* H_y\right), \\
\frac{\partial H_z}{\partial t} &= \frac{1}{\mu}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} - M_{sz} - \sigma^* H_z\right),
\end{aligned}
\tag{137}
$$

where ε, σ, μ are the permittivity, permeability and conductivity already described in previous sections, σ^* is the equivalent magnetic loss (Ω/m) ; E_x, E_y, E_z are the electric fields in the direction x, y and z, respectively; H_x, H_y, H_z are the magnetic fields in the direction x, y and z, respectively; J_{sx}, J_{sy} and J_{sz} are the densities of electric current in x, y and z , respectively and M_{sx}, M_{sy}, M_{sz} are the magnetic current densities at x, y and z respectively. In these experiments $\mu = 1$ since the materials are not magnetic and the parameter $\sigma^* = 0$ since no equivalent magnetic loss. The excitation and reception of the electric current density is only applied in the y-direction given the orientation of the antennas in this axis, therefore $J_{sx} = 0$ and $J_{sz} = 0$. In this 3D implementation, only the relative permittivity is considered in the inversion process, however the methodology could be implemented with conductivity parameter. The 3D forward equation in matrix form is given by:

$$W(\ddot{s},\dot{s},s,\mathbf{m}) = F_1(\mathbf{m})\dot{s} + F_2(\mathbf{m})s - \mathbf{J}_s(\mathbf{r},t) = 0 \tag{138}$$

where

$$F_1(\mathbf{m}) = \begin{bmatrix} \varepsilon_r(\mathbf{r})\varepsilon_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_r(\mathbf{r})\varepsilon_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_r(\mathbf{r})\varepsilon_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_r(\mathbf{r})\mu_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_r(\mathbf{r})\mu_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_r(\mathbf{r})\mu_0 \end{bmatrix}, \tag{139}$$

$$F_2(\mathbf{m}) = \begin{bmatrix} \sigma_r(\mathbf{r}) & 0 & 0 & 0 & \frac{\partial}{\partial z} & -\frac{\partial}{\partial y} \\ 0 & \sigma_r(\mathbf{r}) & 0 & -\frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} \\ 0 & 0 & \sigma_r(\mathbf{r}) & \frac{\partial}{\partial y} & -\frac{\partial}{\partial x} & 0 \\ 0 & -\frac{\partial}{\partial z} & \frac{\partial}{\partial y} & 0 & 0 & 0 \\ \frac{\partial}{\partial z} & 0 & -\frac{\partial}{\partial x} & 0 & 0 & 0 \\ -\frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{140}$$

$$s = \begin{bmatrix} \mathbf{E}_x(\mathbf{r},t) \\ \mathbf{E}_y(\mathbf{r},t) \\ \mathbf{E}_z(\mathbf{r},t) \\ \mathbf{H}_x(\mathbf{r},t) \\ \mathbf{H}_y(\mathbf{r},t) \\ \mathbf{H}_z(\mathbf{r},t) \end{bmatrix}, \tag{141}$$

$$\mathbf{m} = \begin{bmatrix} \varepsilon_r(\mathbf{r}) \\ \mu_r(\mathbf{r}) \\ \sigma_r(\mathbf{r}) \end{bmatrix}. \tag{142}$$

Using the Equations (68) and (138) the adjoint operator in matrix form is

$$-F_1^T(\mathbf{m})\dot{\lambda} + F_3(\mathbf{m})\lambda + \frac{\partial \Phi}{\partial s} = 0, \tag{143}$$

$$\lambda^T = \left[\lambda_{\mathbf{E}_x}(\mathbf{r},t), \lambda_{\mathbf{E}_y}(\mathbf{r},t), \lambda_{\mathbf{E}_z}(\mathbf{r},t), \lambda_{\mathbf{H}_x}(\mathbf{r},t), \lambda_{\mathbf{H}_y}(\mathbf{r},t), \lambda_{\mathbf{H}_z}(\mathbf{r},t)\right], \tag{144}$$

where $F_1^T(\mathbf{m}) = F_1(\mathbf{m})$ and

$$F_3(\mathbf{m}) = \begin{bmatrix} \sigma_r(\mathbf{r}) & 0 & 0 & 0 & -\frac{\partial}{\partial z} & \frac{\partial}{\partial y} \\ 0 & \sigma_r(\mathbf{r}) & 0 & \frac{\partial}{\partial z} & 0 & -\frac{\partial}{\partial x} \\ 0 & 0 & \sigma_r(\mathbf{r}) & -\frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial z} & -\frac{\partial}{\partial y} & 0 & 0 & 0 \\ -\frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} & 0 & 0 & 0 \\ \frac{\partial}{\partial y} & -\frac{\partial}{\partial x} & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{145}$$

The adjoint field are computed using the Equation (146).

$$
\begin{aligned}
\frac{\partial \lambda_{E_x}}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial \lambda_{H_z}}{\partial y} - \frac{\partial \lambda_{H_y}}{\partial z} + \sigma \lambda_{E_x}\right),\\
\frac{\partial \lambda_{E_y}}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial \lambda_{H_x}}{\partial z} - \frac{\partial \lambda_{H_z}}{\partial x} - \frac{\partial \Phi}{\partial E_y} + \sigma \lambda_{E_y}\right),\\
\frac{\partial \lambda_{E_z}}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial \lambda_{H_y}}{\partial x} - \frac{\partial \lambda_{H_x}}{\partial y} + \sigma \lambda_{E_z}\right),\\
\frac{\partial \lambda_{H_x}}{\partial t} &= \left(\frac{\partial \lambda_{E_y}}{\partial z} - \frac{\partial \lambda_{E_z}}{\partial y}\right),\\
\frac{\partial \lambda_{H_y}}{\partial t} &= \left(\frac{\partial \lambda_{E_z}}{\partial x} - \frac{\partial \lambda_{E_x}}{\partial z}\right),\\
\frac{\partial \lambda_{H_z}}{\partial t} &= \left(\frac{\partial \lambda_{E_x}}{\partial y} - \frac{\partial \lambda_{E_y}}{\partial x}\right),
\end{aligned}
\tag{146}
$$

where Φ is the cost function given in Eq.(58), λ_{E_x}, λ_{E_y}, λ_{E_z} are the adjoint fields of E_x, E_y, E_z, respectively; and λ_{H_x}, λ_{H_y}, λ_{H_z} are the adjoint fields of H_x, H_y, H_z. The gradients for the relative permittivity and relative conductivity are given by

$$
g(\varepsilon_r) = \int_0^{Tend} \varepsilon_0 \left(\lambda_{E_x}\frac{\partial E_x}{\partial t} + \lambda_{E_y}\frac{\partial E_y}{\partial t} + \lambda_{E_z}\frac{\partial E_z}{\partial t}\right) dt, \tag{147}
$$

$$
g(\sigma_r) = \int_0^{Tend} \sigma_0 \left(\lambda_{E_x}E_x + \lambda_{E_y}E_y + \lambda_{E_z}E_z\right) dt, \tag{148}
$$

6.1.1. Validatation gprCPS vs gprMax . In this subsection, the results of our implementation (gprCPS) are compared to gprMax are presented. Only for this validation, an antenna-like GSSI at 400 (MHz) is used (see Figure 78). The electromagnetic parameters used are described in the following table:

Parameters gprMAX	
Kind of source	Gaussian Voltage
Frequency	392.398 (MHz)
Source resistence	111.59927 (Ω)
Receiver resistence	111.59927 (Ω)
ε_r(abs)	2.35 (F/m)
ε_r(hdp)	2.35 (F/m)
ε_r(PCB)	1.1 (F/m)
σ_r(abs)	0.062034684 (S/m)

Table 9. *Parameters used in the gprMAX software to compare with gprCPS.*

The acquisition area is 380 points in x direction, 380 points in y direction, and 360 points in z direction. The space step is 0.01 (m), the time step is 0.96e-12 (s), and the antenna is located in the air. CPML is used for all edges to avoid unnatural bouncing. A GTX1070 GPU with 8 GiB VRAM and an AMD Ryzen 5 5600x 6-core processor $\times$ 12 is used in both tests. As it is shown in Figure 79, the correlation between both propagations is high (0.9994).

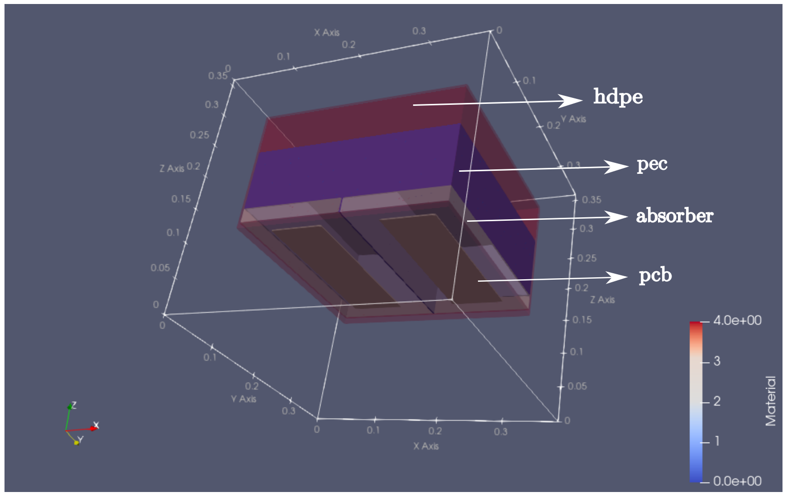

Figure 78. The internal materials used in the GSSI antenna with central frequency at 400 (MHz). The internal materials are used to compare both algorithms: gprCPS and gprMax.

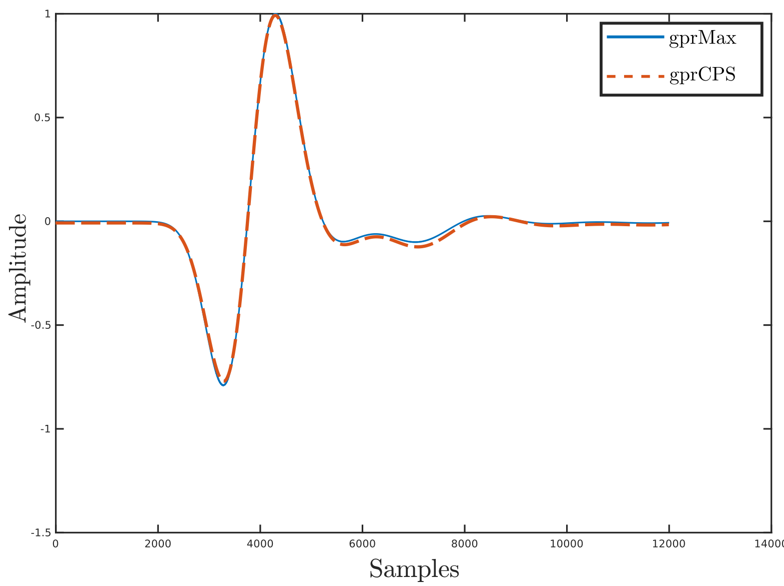

Figure 79. Electric field E_y using the antenna-like GSSI at 400 (MHz): in blue gprMax implementation and in orange gprCPS implementation.

6.2. Computational Strategy

FWI-3D is a great computational challenge and more so for a single-channel, short-offset acquisition. The Equation 147 shows that the forward (E_x, E_y, E_z) and adjoint $(\lambda_{E_x}, \lambda_{E_y}, \lambda_{E_z})$ fields are required to obtain the permittivity gradient. In a 3D implementation with float data, the following RAM resources are required to compute the gradient:

$$RAM(GB) = \frac{N_x \times N_y \times N_z \times N_t \times 4}{10^9} \tag{149}$$

where N_x, N_y, and N_z are the number of points in the dimensions x, y, and z, respectively. N_t is the number of samples in the field and 4 is the size in bytes occupied by the floating number data. As an example, using the antenna discretization and the SEAM model: $dh = 0.002$, $Nx = 7500$, $Nz = 5475$, $Ny = 5000$ and $Nt = 12000$, a field would occupy 9855 TB of RAM storage, this example is a rather crude and naive estimation.

The computational strategy seeks not to calculate the entire gradient volume used in the 3D inversion process. The energy is located near to source in a single and short offset acquisition (a radius for the 3D case). The upper part of Figure 80 presents the energy per column (the energy is estimated as the sum of the squared values) and the lower part of Figure 80 shows a gradient of the SEAM model on which the energy is computed. For compute the gradient, a forward propagation is made from time t=0 to time t=nt-1 and the fields E_x, E_y, E_z are stored. These fields are multiplied with the backpropagated fields that are obtained in the inverse time and where the residual of the observed and modeled data is used as a source to obtain the fields $\lambda_{E_x}, \lambda_{E_y}, \lambda_{E_z}$. At each instant of time, all the fields previously described have dimensions of $N_x \times N_y \times N_z$, this previously described procedure is presented in Figure 81-a). We propose to perform the multiplication of the backward and forward fields only in a specific gradient area. The FWI procedure is similar to the traditional

FWI with the difference that in our strategy only is necessary to store the fields in the zone to update. This zone contains the GPR acquisition line, and Figure 81-b) presents the cube at a specific time with the acquisition line in the zone to update. We have done several experiments changing the width of the zone to determine the recomended width for 3D inversion in zero offset and single-channel acquisition. Figure 81-c) presents the proposed methodology for computing the gradient, where it is highlighted in blue that only that part of the cube for each instant of time is considered for the gradient calculations.

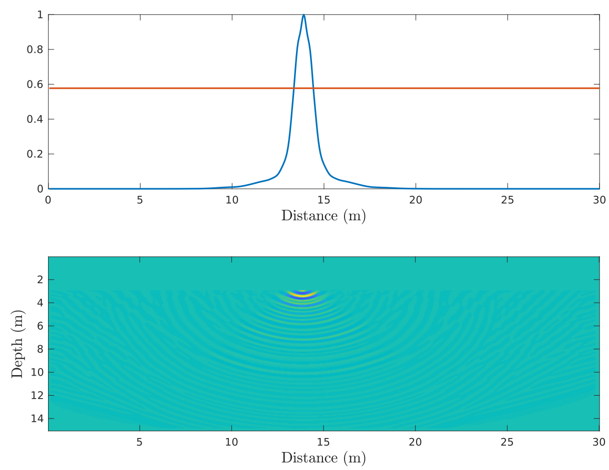

Figure 80. (Top) Energy for column in the gradient at 100 (MHz). (Bottom) Relative permittivity gradient for a single scan at 100 (MHz) on the SEAM foothill model.

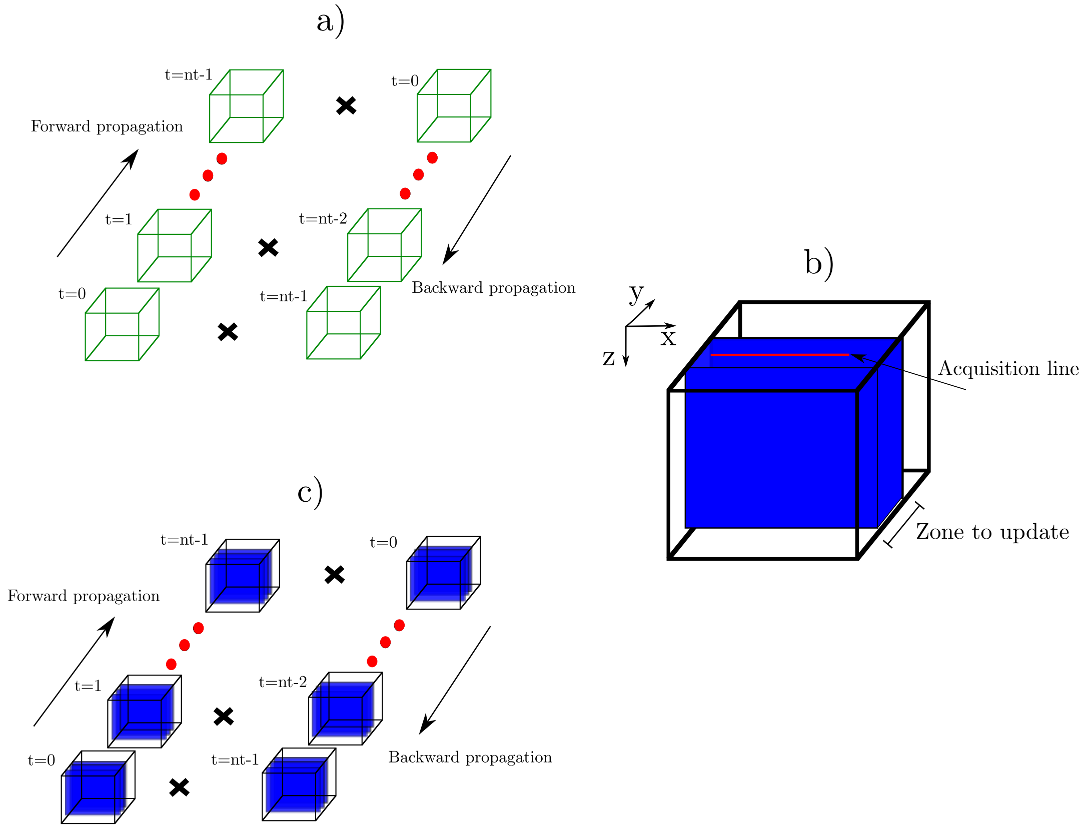

Figure 81. a) Gradient methodology used for the traditional FWI-3D, b) Description in an instant of time in the computational strategy for FWI-3D,c) Gradient methodology used with the proposed computational strategy for FWI-3D.

Some strategies like the one proposed in Tromp et al. (2008) can be used to reduce the RAM resources needed to compute the gradient. According to Tromp et al. (2008) the gradient is obtained during the adjoint field computation by recomputing the forward field. The strategy reduces the RAM resources and improves the approximation of the electromagnetic parameters. However, this strategy has a consequence on the execution time. An additional propagation is necessary per scan; in a 3D case for single-channel and short-offset acquisitions, this would imply, according to the same experiment presented in this section, 116 additional scans per iteration. In a multiscale scheme like the one shown with 3 frequencies for the experiment and with 30 iterations per frequency, this strategy would imply 10440 additional scans. This would imply an increase of 33 % per iteration in the execution times than the strategy proposed in this chapter. It should be noted that short-offset GPR 3D acquisitions are a great challenge (computational cost), for this reason our computational strategy seeks to obtain a 2D image of the electromagnetic parameters using 3D simulations.

6.3. FWI-3D Synthetic Results

In this section, the results of FWI over a synthetic model are shown. The synthetic case is a realistic model called SEAM-Phase II Regone et al. (2017) that is built to represent the geological underground of a foothill. The SEAM model's selected area is rescaled from its initial dimensions to have a 31.2 m section in-depth and 3 m in the dip direction. The spatial discretization is 5 cm, such that we generate 312 points in the x-direction (dip) and 219 points in z-direction (depth) and (200) point in y- direction (strike). A total of 116 scans are simulated using the equations for the forward electromagnetic modeling. The scans are located at $y = 99$, and only one line of sources and receivers is used, as it is shown in Figure 82. According to the topography, transmitter and receiver antennas are located in the air-layer at 10 (cm) of the ground layer. The time sampling of the data is 0.0963 ns, and the number of time samples is 2000; thus, the total acquisition time for each scan is 192.6 ns. The scheme of FDTD with 8^{th} and 2^{nd} order of approximation is used to discretize the spatial and time derivatives, respectively.

Three frequencies are selected in the inversion process: 30 MHz, 50 MHz and 100 MHz. The multiscale methodology proposed by Bunks et al. (1995) is used, where the FWI process started from low frequency to high frequency, and the final model obtained in each frequency is used as a starting point for the next frequency step. A full-wave dipole is used at the transmitter antenna. Figure 83-a), and 83-b) shows the true and initial models obtained in cut $y = 99$. Figure 83-c), Figure 83-d), Figure 83-e) and 83-f) shows the FWI models in cut $y = 99$ with 1, 5, 11 and 21 layers in the gradients zone. Note in Figure 83-c), d), e), and f) that as the number of planes used in the gradient increases, the estimation of the permittivity model in the plane improves. The PSNR values achieved are 19.84 (dB), 23.68 (dB), 25.77 (dB), 25.92 (dB), using 1, 5, 11 and 21 layers in the three-dimensional gradient, respectively. It is important to note that when a single layer is used as the gradient update zone (Figure 83 -c)) , FWI tries to fit all the data with that single plane, and as lot energy is out of the plane, therefore, it converges to incorrect values. We propose to use at least 2,1 (m), or 21 planes as a zone to update the gradient; this implies a 9.5x reduction of the volume capacity necessary to compute the gradient. Our synthetic tests show that we must keep the gradient update zone close to $\lambda/4$ by taking the lowest frequency in multi-scale methodologies.

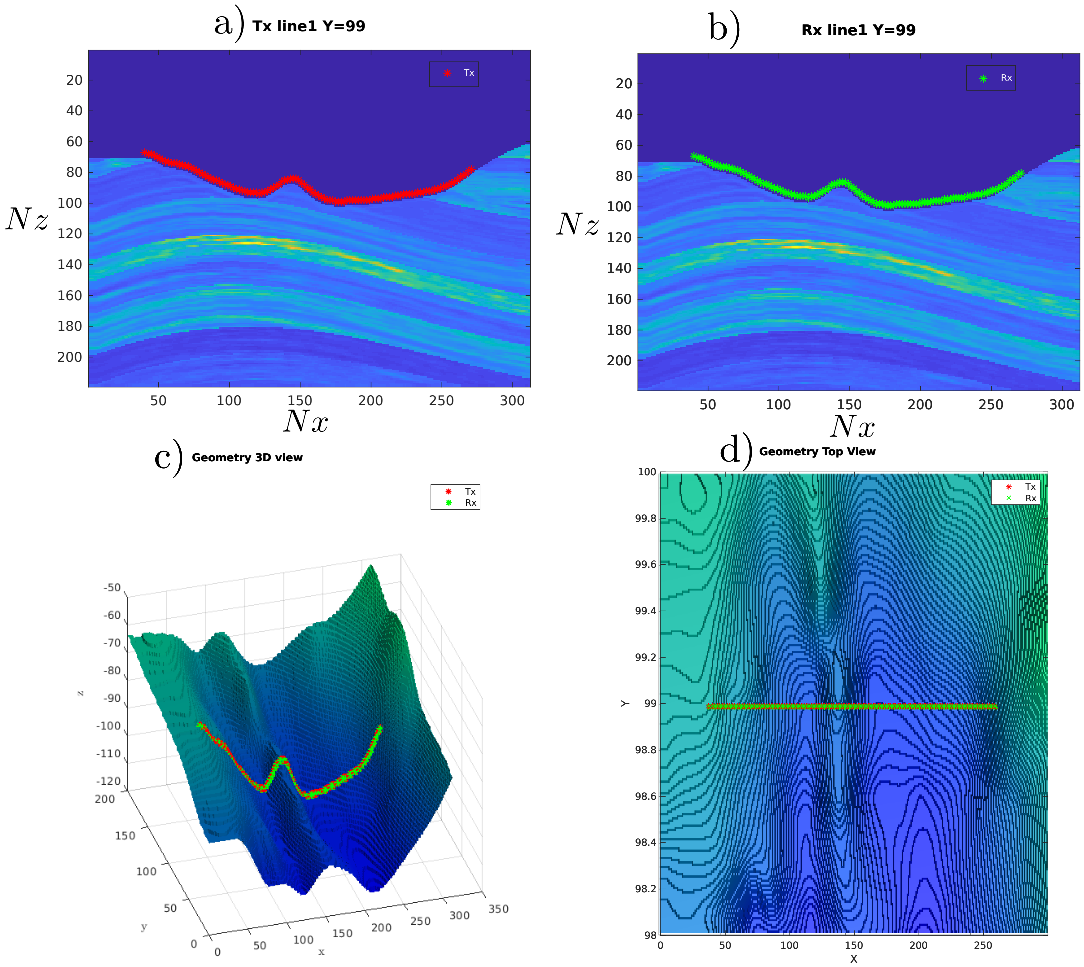

Figure 82. a) Sources positions, b) Receivers positions, c) 3D geometry, d) Top view geometry

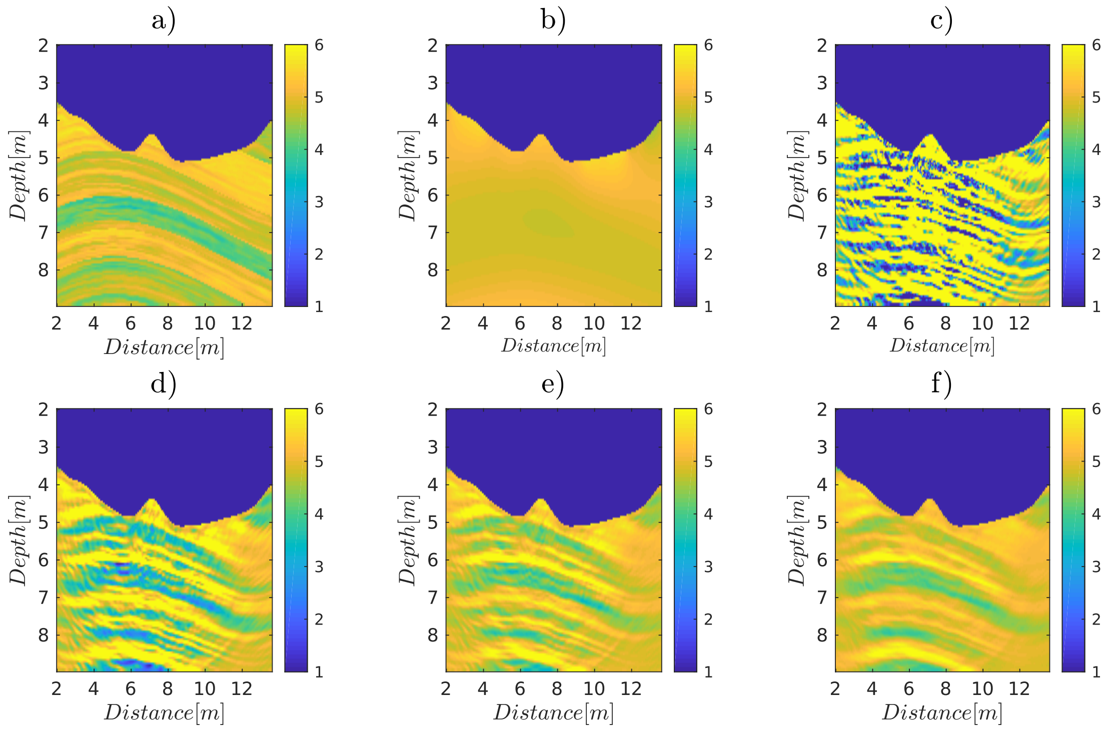

Figure 83. Results of FWI when the zone used for compute the gradient changes: a) true relative permittivity, b) initial relative permittivity, c) FWI with multiscale for relative permittivity using a zone of 0.1(m), d) FWI with multiscale for relative permittivity using a zone of 0.5(m), e) FWI with multiscale for relative permittivity using a zone of 1.1(m), and f) FWI with multiscale for relative permittivity using a zone of 2.1(m).

7. Characterization of a shielded antenna

The radiation pattern defines how the energy that is transmitted by the antenna is distributed throughout the medium. Usually, those patterns are assumed as an infinitesimal dipole in TE mode (transverse electric), the energy is distributed at the same for all directions. This chapter introduces the use of PSO as a global optimization technique for characterizing the shielded single-channel and short-offset antenna designed and built by the Geophysical Survey Systems. Inc (GSSI) at 400 MHz. Based on the antenna's internal parameters, the radiation patterns in planes E and H are obtained and then included in the modelling and inversion processes. In the first part of the chapter, the concepts of a shielded antenna are presented. Then, the proposed methodology for the estimation of the internal parameters of the antenna is presented and finally, the results obtained for the inversion process including the radiation patterns are presented.

7.1. The shielded antenna

In FWI, it is important to have an initial guess with enough low frequencies information and the source wavelet. If the initial guess contains a kinematically accurate (low-wavenumbers), FWI is capable of obtaining high-resolution models in the parameters. However, when the initial model is not able to describe the kinematics of the wavefield and the difference between the collected and modeled data is greater than half a cycle (cycle skipping problem), FWI usually reaches a local minimum where these parameters include artifacts Wu and Alkhalifah (2018).

Usually, in GPR applications the radiation patterns in TE mode are assumed as an infinitesimal dipole, where the energy is distributed uniform for any angle. There are different techniques to estimate the source signature: reverse-time propagation (Zhang et al., 2021), deconvolution of radar data with the parameters system (Ernst et al., 2007) or solving a linear inverse problem (Pratt, 1999).

The inverse problem is solved using a global optimization algorithm called Particle Swarm Optimization (PSO), which generates particles with the possible candidate values for the internal parameters of the antenna in a random way in a predefined search space (Shi et al., 2001). The global optimization algorithm compares R_{xobs} and R_{xmod} using two metrics: the correlation in the time domain and the spectrum in the frequency domain. PSO algorithm estimates the internal parameters of the antenna to obtain a simulated electromagnetic field that is compared with the measured electromagnetic field. The antenna used in this study is a commercial GSSI brand GPR with an operating frequency of 400 MHz.

The shielded antenna is a kind of antenna generally used in GPR applications where the transmitter and receiver antennas are located in a housing, and the distance between both antennas is fixed. One of the most recognized companies in the world, because of the production of antennas for GPR, is GSSI. A shielded antenna with a central frequency of 400 MHz is selected. Its internal geometry is depicted in Figure 84. The parameters W, L, and h, are 6 cm. The distance between both antennas, known as the offset, is 16.2 cm.

7.2. Methodology to characterize the shielded antenna

The following methodology is proposed for the characterization of the shielded antenna:

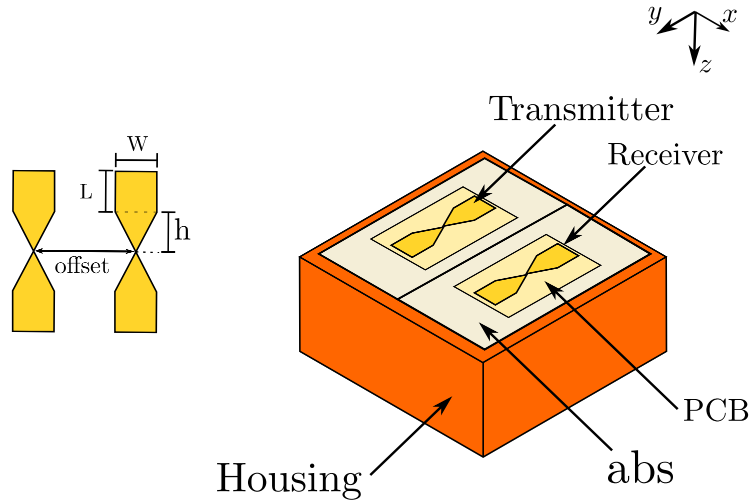

Figure 84. Internal geometry of the shielded GSSI antenna at 400 MHz.

- Measuring the electromagnetic pulse in the time domain.
- Generate modeled data that includes the internal geometry of the antenna.
- Definition of the global optimization process to estimate the given parameters.
- Selecting and tuning up the global optimization method (PSO).
- Performing statistical analysis of the estimated parameters.

7.2.1. Measuring the electromagnetic pulse in the time domain. The electromagnetic field is measured with the receiver antenna of the GSSI equipment and it is carried out in a controlled experiment in a free space avoiding unwanted reflections in the measurement. In this test, 600 scans are performed in the free space, and these scans are averaged, resulting in the source presented in Figure 85. The source measured is compared with the free software gprMAX using the proposed parameters in Warren et al. (2016) (see Table 10). The parameters selected

Parameters gprMAX	
Kind of source	Ricker Voltage
Frequency	392.398 MHz
Source resistence	111.59927 Ω
Receiver resistence	111.59927 Ω
ε_r(abs)	2.35 F/m
ε_r(hdp)	2.35 F/m
ε_r(PCB)	1.1 F/m
σ_r(abs)	0.062034684 S/m

Table 10. *Parameters used in the gprMAX software to compare with our observation.*

in the gprMAX software are: ε_r(abs) as the permittivity of the absorbent barrier; σ_r(abs) as the conductivity of the absorbent barrier, ε_r(PCB) as the permittivity of the PCB and $\varepsilon_r(hdp)$ as the permittivity of the coating. As shown in Figure 85, the two signals have a high correlation, 94.76 %. We improve the correlation in time and the difference in amplitude of the spectrum presented in gprMax using the global optimization technique.

7.2.2. Generate a modeled data that includes the internal geometry of the antenna. In this subsection, the free software gprMAX is used (Warren et al., 2016), which allows 3D modeling of the electromagnetic wave equation according to the selected antenna geometry. The gprMAX software uses finite differences in the time domain, and the electromagnetic wave

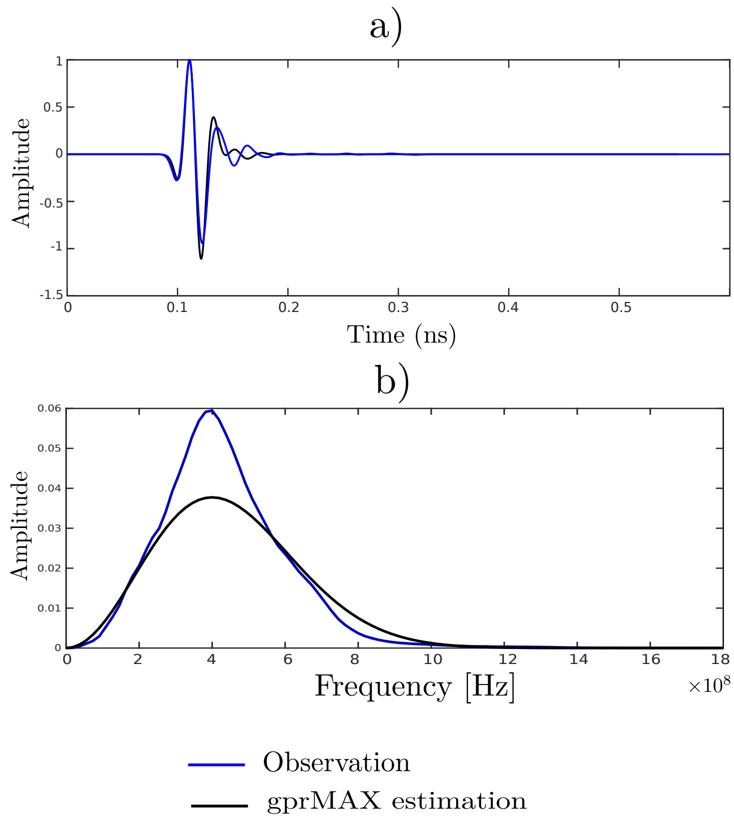

Figure 85. a) In blue and black the observed data and the electrical field computed with gprMAX, respectively, b) in blue and black the spectrum of the observed data and the electrical field computed gprMax, respectively.

equation is given by

$$
\begin{aligned}
\frac{\partial E_x}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - J_{sx} - \sigma E_x\right), \\
\frac{\partial E_y}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - J_{sy} - \sigma E_y\right), \\
\frac{\partial E_z}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - J_{sz} - \sigma E_z\right), \\
\frac{\partial H_x}{\partial t} &= \frac{1}{\mu}\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} - M_{sx} - \sigma^* H_x\right), \\
\frac{\partial H_y}{\partial t} &= \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} - M_{sy} - \sigma^* H_y\right), \\
\frac{\partial H_z}{\partial t} &= \frac{1}{\mu}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} - M_{sz} - \sigma^* H_z\right),
\end{aligned}
\tag{150}
$$

where ε, σ, μ are the permittivity, permeability and conductivity already described in previous sections, σ^* is the equivalent magnetic loss $[\Omega/m]$; $\mathbf{E}_x$, $\mathbf{E}_y$, $\mathbf{E}_z$ are the electric fields in the direction x, y and z, respectively; $\mathbf{H}_x$, $\mathbf{H}_y$, $\mathbf{H}_z$ are the magnetic fields in the direction x, y and z, respectively; J_{sx}, J_{sy} and J_{sz} are the densities of electric current in x, y and z , respectively and M_{sx}, M_{sy}, M_{sz} are the magnetic current densities at x, y and z respectively. In these experiments $\mu = 1$ since the materials are not magnetic and the parameter $\sigma^* = 0$ since no equivalent magnetic loss. The excitation and reception of the electric current density is only applied in the y-direction given the orientation of the antennas in this axis, therefore $J_{sx} = 0$ and $J_{sz} = 0$. Based on this modeling that includes the internal geometry of the antenna, the internal parameters are modified. The internal parameters are: resistance of the transmitter and receiver antenna, relative permittivity absorbent barrier, relative conductivity absorbent barrier, relative permittivity of the PCB and coating.

7.2.3. Cost function for the global optimization process. In this subsection, two metrics are introduced to measure the similarity between the measured electric field (R_{xobs}) and the modeled electric field (R_{xmod}) using the gprMAX software. Two metrics have been selected: one in time and the other in frequency. The time metric is the correlation that measures how closely these two signals are alike in phase. The correlation equation is described in the Equation (151).

$$R_{corr} = \frac{\sum_{i=1}^{n}(R_{xobs}(i) - \overline{R_{xobs}})(R_{xmod}(i) - \overline{R_{xmod}})}{\sqrt{(\sum_{i=1}^{n}(R_{xobs}(i) - \overline{R}_{xobs})^2)(\sum_{i=1}^{n}(R_{xmod}(i) - \overline{R_{xmod}})^2)}}, \tag{151}$$

where $\overline{R_{xobs}}$ and $\overline{R_{xmod}}$ are the mean value of R_{xobs} and R_{xmod}, respectively. R_{xobs} and R_{xmod} are measured in y-direction solving the electromagnetic wave equation described in the Equation (150).The second metric use the Fast Fourier Transform $FFT(\cdot)$ on R_{xobs} and R_{xmod} and the metric measure the difference between the maximum values of spectrum amplitude. The followig equation described the metric:

$$rf = |Max(abs(FFT(Rxmod))) - Max(abs(FFT(Rxobs)))|, \tag{152}$$

where $|\cdot|$ is the modulus, $Max(\cdot)$ is the maximum value, $abs(\cdot)$ is the amplitude spectrum and $FFT(\cdot)$ is the Fast Fourier Transform. The objective of combining these two metrics is to include more information to measure the similarity between these two signals, the correlation is a maximization problem and the magnitude of the FFT is a minimization problem.

7.2.4. Global optimization method. Particle swarm optimization (PSO) is a global optimization technique that uses a group of particles to explore the entire space of possible solutions by moving randomly (Shaw and Srivastava, 2007). The formulation used in PSO to find the position of the j^{th}-dimension of the i^{th} particle is given by

$$\mathbf{x}_{i,j}^{k+1} = \mathbf{x}_{i,j}^{k} + \mathbf{v}_{i,j}^{k+1} \cdot dt, \tag{153}$$

where $\mathbf{x}_{i,j}^{k+1}$ is the new position vector associated with the i^{th} internal parameter in the shielded antenna, $\mathbf{x}_{i,j}^{k}$ is the present position vector of the i^{th} particle, $\mathbf{v}_{i,j}^{k+1}$ is the new velocity vector and dt is the step size used by the particle to move inside the searching space. An update in the velocity vector is obtained as

$$\mathbf{v}_{i,j}^{k+1} = (w_{pso} \cdot \mathbf{v}_{i,j}^{k}) + c_1 \cdot rand() \cdot \frac{(\mathbf{p}_i^k - \mathbf{x}_{i,j}^k)}{dt} + c_2 \cdot rand() \cdot \frac{(\mathbf{g}_l^k - \mathbf{x}_{i,j}^k)}{dt}, \tag{154}$$

where w_{pso} is an inertial weight that controls the particle movement; c_1 and c_2 are the weights associated with the local and global behavior of the swarm, respectively; $rand()$ is a random number between $[0-1]$ with uniform distribution; $\mathbf{p}_i^k$ is the best local position; and $\mathbf{g}_l^k$ is the best global position. Each particle is moving in a j-dimensional space, which the internal parameters are $\varepsilon_r(abs)$, $\varepsilon_r(hdp)$, $\varepsilon_r(PCB)$, $\sigma r(PCB)$, source resistence and receiver resistence. The PSO solution at each iteration is accepted when a minimum error measure between the R_{xobs} and R_{xmod} is obtained. After, several experiments the parameters selected for PSO are: $w_{pso} = 0.3$, $c_1 = 0.5$ and $c_2 = 2.06$.

7.2.5. Statistical analysis of the estimated parameters. In this subsection, a global search is performed with the following search regions in the PSO algorithm: source and receiver resistance between [1-1000] Ω, $\varepsilon_r(abs)$ and $\varepsilon_r(PCB)$ between [1-30] F/m and finally, $\sigma r(abs)$ and $\sigma r(PCB)$ 0-20 S/m. In total, 50 experiments are performed as it is shown in Table 11.

In Table 11, the number of classes or subdivision (*Nclass*) is obtained using the rule of Sturges (Canavos et al., 1988), which describes that $Nclass = int(1 + log_2(n))$, where n is the total number of experiments using PSO, therefore $Nclass = 6$. Min_{cost} is the minimum error in the cost function (minimum value in column *LI*), Max_{cost} is the maximum error in the cost function (maximum value in column *LS*); R is the difference between Max_{cost} and Min_{cost}; K is the width of

Experiments using an antenna at 400 MHz					
	LI	*LS*	*X*	*F*	*FR*
	0,030664	0,081107	0,055885	18	36 %
	0,081107	0,131550	0,106328	1	2 %
	0,131550	0,181993	0,156772	1	2 %
	0,181993	0,232437	0,2072155	2	4 %
	0,232437	0,282880	0,257658	1	2 %
	0,282880	0,333323	0,308102	25	50 %
	0,333323	0,383767	0,358545	2	4 %

n	50
Min_{cost}	0,030664
Max_{cost}	0,383767
$Nclass$	7
R	0,353103
K	0,05044329

Table 11. *Frequency table of the experiments carried out to obtain the internal parameters of the antenna.*

the class ($R/Nclass$); LI and LS are the lower and upper limit in each class, respectively; X is the average value of each class, F is the absolute frequency or the total number of PSO experiments in the subdivision, FR is the relative frequency ($\frac{F}{n} \times 100$). In one of these class, the algorithm converges to small error values between 3.07 % to 8.11 %, and the other mode converges to large errors between 28.29 % to 33.33 %. Small errors indicate that the algorithm has a proper fit between R_{xobs} and R_{xmod}. Figure 86 shows the binomial behavior of the results.

The parameters where large errors occur are discarded. Adjusting these ranges, 15 new additional experiments reach an error of 2.65 % between R_{xobs} and R_{xmod}. Figures 87 -a) and b) show the signal in time, and the magnitude of the spectrum, respectively. The estimated parameters that model the materials of the internal structure of the antenna are: resistance of the source and receiver = 49.224 Ω, $\varepsilon_r(abs)$ = 3.896 F/m, $\sigma_r(abs) = 0.023$ S/m, $\varepsilon_r(PCB)$ = 7.445 F/m,$\varepsilon_r(hdp)$=1.000 F/m.

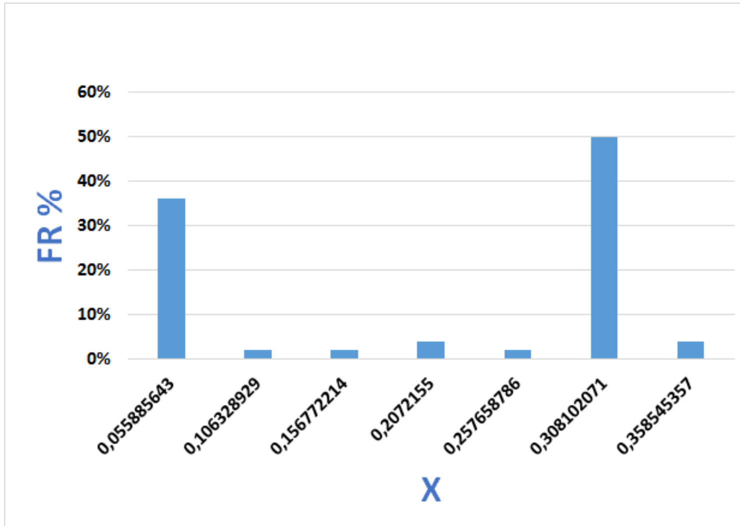

Figure 86. Frequency distribution for 50 experiments using PSO methodology.

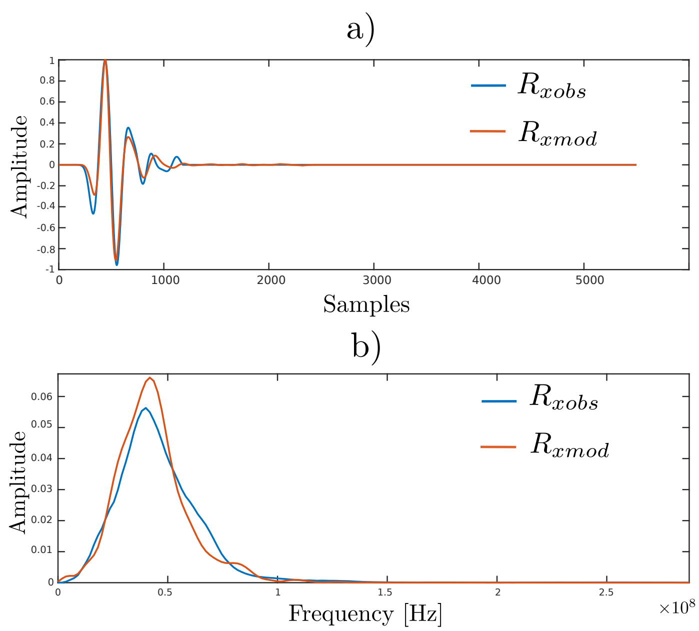

Figure 87. a) R_{xmod} and R_{xobs} in the time domain. b) Spectrum for R_{xmod} and R_{xobs}.

7.3. Radiation pattern

The radiation pattern and the inclination affect the information of the reflected waves than could be measure by the receiver antenna in short-offset acquisition. The radiation patterns allow analyzing the distribution of energy in the subsurface (Elliot, 1983). The far-field extends from radiating near-field to infinity. In the far-field region, the field's behavior is dependent of the distance where the electric and magnetic fields E and H change with the relation $\frac{1}{r}$, where r is the radius of the wavefront to the antenna ; therefore, it is the recommended area to measure them. However, the specific case of the antenna studied in this document is a short antenna because the length of the antenna (D) is shorter than half of the wavelength ($\lambda = 75cm$ in free space). If r is the distance from the radiation source to the wavelength λ of the radiation the short antenna define: far-field when $r \gg 2\lambda$, near-field when $r \ll \lambda$ and transition zone between $r = \lambda$ and $r = 2\lambda$. The distance between transmitter and receiver antenna is 16 cm for the GSSI antenna therefore we measure the electric field in the near-field. When the electric field is measured, the antennas are isolated by materials present in the construction, physically the antenna is a black box so it is not possible to uncover it to extract one of them and perform its characterization.

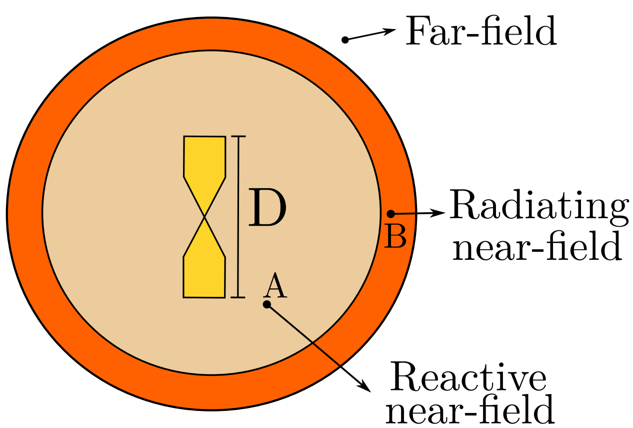

Figure 88. Regions usually studied in an antenna according to the distance from the origin: Far-field, Radiation near-field and Reactive near-field.

The bowtie antenna used in the antennas at 400 MHz of GSSI can be represented using a dipole as an approximation. In the hypothetical case of a *y*-axis oriented dipole, the radiation pattern should be a toroid as shown in Figure 89 -a) and their respective views in the *x*-*z* and *y*-*z* planes are presented in Figure 89-b) and Figure 89-c), respectively. With the characterization parameters obtained through the PSO process, the radiation patterns for the shielded antenna at 400 MHz can be estimated.

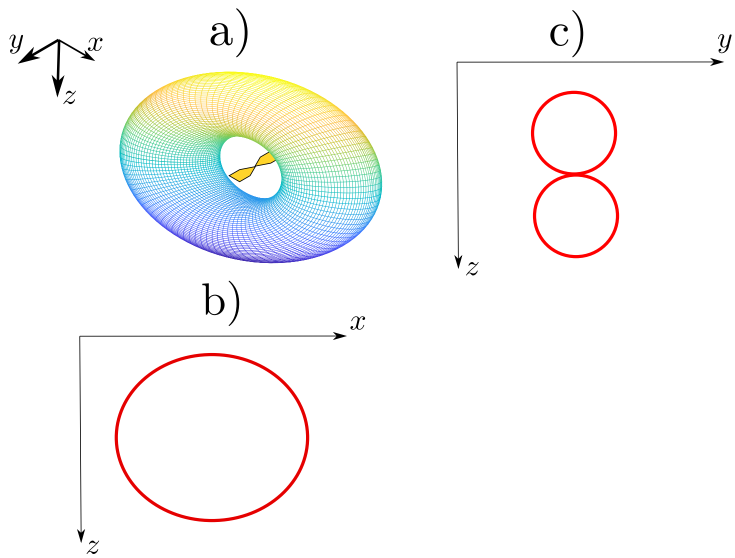

Figure 89. Radiation pattern in a dipole: a) 3D view,b) plane *x*-*z* and c) plane *y*-*z*.

The internal geometry of the antenna proposed in the gprMAX software is used with the following notation to measure each component from cartesian coordinates to spherical coordinates

$$
\begin{aligned}
&r = r_1\hat{i} + r_2\hat{j} + r_3\hat{k}, \\
&r_1 = \frac{x}{\sqrt{x^2+y^2+z^2}}, \quad \theta_1 = \frac{xz}{\sqrt{x^2+y^2+z^2}\sqrt{x^2+y^2}}, \quad \phi_1 = \frac{-y}{\sqrt{x^2+y^2}}, \\
&r_2 = \frac{y}{\sqrt{x^2+y^2+z^2}}, \quad \theta_2 = \frac{yz}{\sqrt{x^2+y^2+z^2}\sqrt{x^2+y^2}}, \quad \phi_2 = \frac{x}{\sqrt{x^2+y^2}}, \\
&r_3 = \frac{z}{\sqrt{x^2+y^2+z^2}}, \quad \theta_3 = \frac{-\sqrt{x^2+y^2}}{\sqrt{x^2+y^2+z^2}}, \quad \phi_3 = 0,
\end{aligned}
\tag{155}
$$

where x, y and z are the spatial coordinates in the direction of x, y and z. With these unit vectors the electric and the spherical coordiantes of Figure 90, the field in terms of r, θ and ϕ is defined as

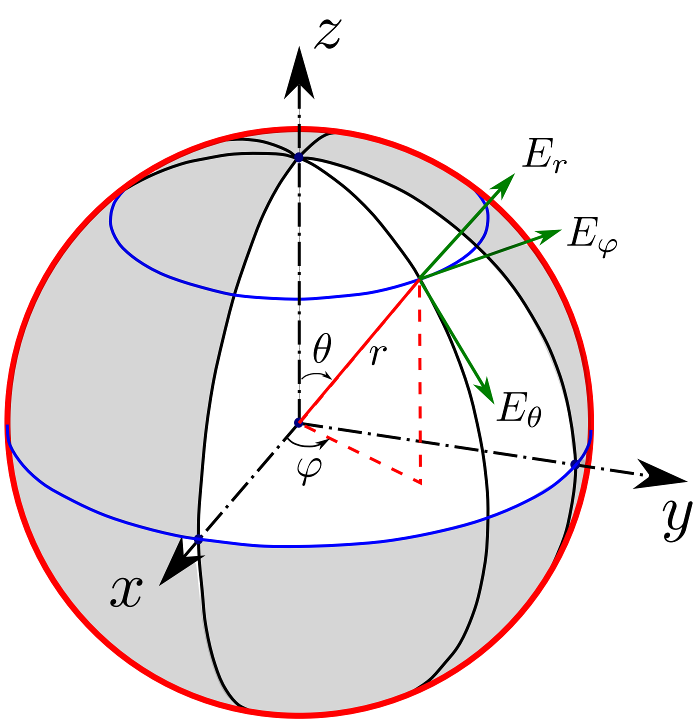

Figure 90. Spherical coordinates system used to measure the radiation pattern.

$$
\begin{aligned}
E_r &= E_x r_1 + E_y r_2 + E_z r_3, \qquad H_r = H_x r_1 + H_y r_2 + H_z r_3, \\
E_\theta &= E_x \theta_1 + E_y \theta_2 + E_z \theta_3, \quad H_\theta = H_x \theta_1 + H_y \theta_2 + H_z \theta_3, \\
E_\phi &= E_x \phi_1 + E_y \phi_2 + E_z \phi_3 \quad H_\phi = H_x \phi_1 + H_y \phi_2 + H_z \phi_3.
\end{aligned}
\tag{156}
$$

The sum of all the squared values of E_θ and H_θ are computed to obtain the radiation plane E and H, respectively. According to Warren in Warren and Giannopoulos (2017), radiation patterns are traditionally plotted at a specific frequency. However, this is a limited analysis for UWB (ultra-wideband) antennas. According to Warren, making the measurement on a single frequency can cause constructive or destructive interference on another frequency. For that reason, the Equation (157) presents the energy for each planeis considered to be a modification of Diamanti and Annan (2013) in a specific r and for all the angles θ.

$$
\Xi_E = \sum_{t=0}^{T} E_\theta^2, \qquad \Xi_H = \sum_{t=0}^{T} H_\theta^2. \tag{157}
$$

The internal parameters of the shielded antenna such as Source and receiver resistence, ε_r (abs), ε_r(hdp), ε_r (PCB), σ_r(abs) together with its geometry in 3D space, define the radiation pattern. The internal parameters of the shielded antenna such as source resistance, receiver resistance, ε_r (abs), ε_r(hdp), ε_r (PCB), σ_r(abs) together with its geometry in 3D space, define the radiation pattern. We estimate the internal parameters of the antenna using PSO. The electric and magnetic field is measured at a radius from 0.2 (m) to 0.9 (m) and a $dr = 0.116$ (m). The angle of observation changes from 0 (rad) to 2π (rad) with $d\theta = 0.1047$ (rad). Figure 91-a), and Figure 91-b) shows the radiation pattern E and H in spherical coordinates, respectively; Figure 91-c), and Figure 91-d) shows the normalized radiation pattern in amplitude vs θ. As it is shown in Figure 91-a) and Figure 91-b), the patterns are similar to a dipole pattern (see Figure 89-b) and Figure 89-c)). However, given the antenna's internal geometry, the energy is attenuated between -60° and 60°. Based on these obtained radiation patterns, we seek to include them in the propagation of the electromagnetic field, thus making it closer to reality and reducing the error between the modeled and observed data.

7.4. FWI using the radiation patterns

In this section, a Full Waveform inversion is performed, including the radiation pattern in E-plane. In the 3D case of gprMAX software, the radiation pattern is included taking into account the internal geometry of the antenna in the process of propagating the electromagnetic wave. However, this would substantially increase the computational resources required to model the propagation of the electromagnetic wave. For example, using a grid of 0.002 m the algorithm takes approximately 4 hours and 139 GB of RAM in a server with two processors Intel Xeon E5-2643 v3 3.4GHz,20M Cache, 9.60GT/s QPI, Turbo, HT, 6C/12T and 256 GB of memory RAM. In this section, the radiation patterns are included in the propagation of the electromagnetic wave in a 2D implementation. To achieve this implementation, we have developed the following methodology:

- To obtain the radiation pattern in spherical coordinates using the global optimization process (PSO) and the gprMAX software that includes the internal geometry of the antenna and the

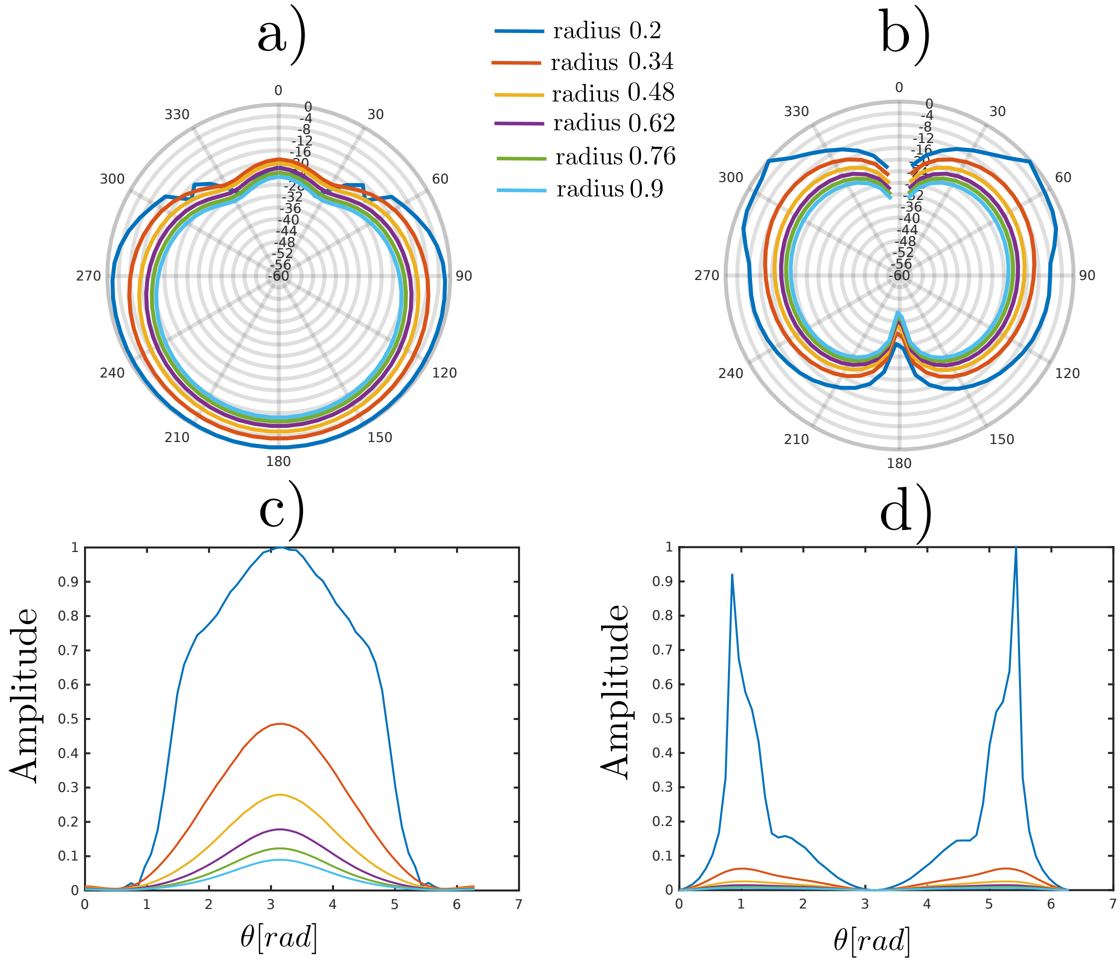

Figure 91. E and H plane radiation for the shielded antenna at 400 MHz: a) and b) show the radiation patterns in spherical coordinate using a plot of $r(dB) - \theta(grades)$ for planes E and H, respectively. The radius is changed from 0.2 (m) to 0.9 (m) with a step of 0.116 (m). c) and d) show the normalized radiation patterns with respect to the maximum of the highest energy pattern in planes E and H, respectively.

estimated parameters.

- To convert the radiation pattern from spherical coordinates to cartesian coordinates.
- To model the wave propagation of the electromagnetic wavefront taking into account the radiation pattern.

7.4.1. Radiation pattern in spherical coordinates. The radiation pattern is obtained using the global optimization process in search of the internal parameters of the GSSI antenna at 400 MHz. Based on these parameters, the gprMAX software uses these and the electromagnetic wave equations described in Equations (155) and (156) to obtain the fields in spherical coordinates. The radiation patterns in spherical coordinates for the E and H planes are presented in Figure 91.

7.4.2. Conversion of the radiation pattern from spherical coordinates to cartesian coordinates. An alternative to apply the radiation pattern in the transmission of the electromagnetic pulse is to transform the radiation pattern from spherical coordinates to cartesian coordinates, such that it can be applied at each instant of time on the fields $\mathbf{E}_y(\mathbf{r},t)$, $\mathbf{H}_x(\mathbf{r},t)$ and $\mathbf{H}_z(\mathbf{r},t)$. For this conversion, x and z are defined as: $x = r \cdot \cos(\theta)\sin(\varphi)$, $y = r \cdot \sin(\theta)\sin(\varphi)$ and $z = r \cdot \sin(\theta)$. The transformed points from spherical to cartesian coordinates are not exact values on the mesh and it requires an interpolation. In this particular case, a bilinear interpolation is performed where we want to know a value $f(x,z)$, with x and z being the locations of the point in distance-direction and depth-direction, respectively. The point [x, z] is known to lie within four known corners $f(x_0,z_0)$, $f(x_0,z_1)$, $f(x_1,z_0)$ and $f(x_1,z_1)$, fulfilling the conditions: $x_0 < x < x_1$, $z_0 < z < z_1$, $\Delta_x = x_1 - x_0$, $\Delta_z = z_1 - z_0$, $\delta_x = x - x_0$ and $\delta_z = z - z_0$. First the interpolation is done in the x-direction

$$
\begin{aligned}
f(x,z_0) &= \frac{\delta_x}{\Delta_x}(f(x_1,z_0) - f(x_0,z_0)) + f_(x_0,z_0), \\
f(x,z_1) &= \frac{\delta_x}{\Delta_x}(f(x_1,z_1) - f(x_0,z_1)) + f_(x_0,z_1),
\end{aligned}
\tag{158}
$$

and then the interpolation is done in the z-direction

$$f(x,z) = \frac{\delta_z}{\Delta_z}[f(x,z_1) - f_{x,z_0}] + f(x,z_0). \tag{159}$$

Figure 92 shows the concept of bilinear interpolation, where the order in the interpolation direction is irrelevant in the result since the bilinear interpolation is symmetric (Press et al., 1988).

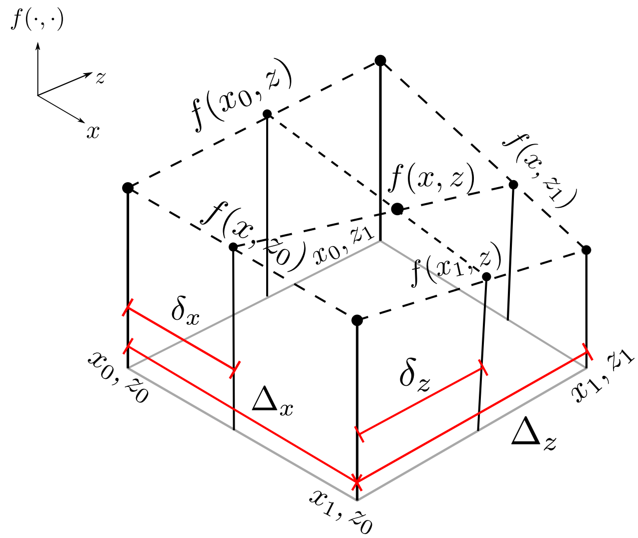

Figure 92. Explanation of the bilinear interpolation for the point $f(x,z)$ knowing the four points around it $f(x_0,z_0)$, $f(x_0,z_1)$, $f(x_1,z_0)$ and $f(x_1,z_1)$.

The result of the transformation of spherical coordinates to cartesian coordinates using the bilinear interpolation for the entire area of study is presented in Figure 93 -a). However, the radiation pattern does not apply over the entire study area, only to a radius lower than 0.2 m, where the radiation pattern has been measured. Figure 93 -b) shows the result of cutting the radiation pattern to a radius less than 0.2 m.

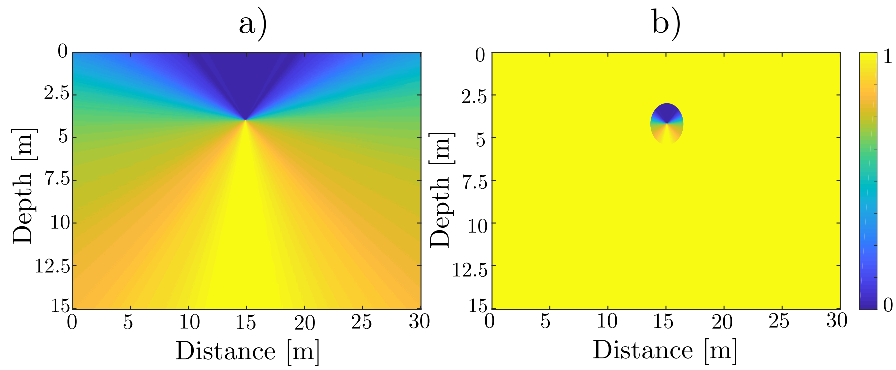

Figure 93. Radiation patterns: a) with 0° of inclination, and b) radiation pattern for radius less than 0.2 m and 0° of inclination.

7.4.3. Propagation of the electromagnetic wavefront taking into account the radiation pattern. As a last step in this methodology to include the radiation pattern, the estimated radiation pattern must be multiplied (Figure 93-b)) with the fields $\mathbf{E}_y(\mathbf{r},t)$, $\mathbf{H}_x(\mathbf{r},t)$ and $\mathbf{H}_z(\mathbf{r},t)$ in each iteration and the electromagnetic wave equation is discretized using the FDTD scheme again. The electromagnetic equation in a non-dispersive and isotropic medium is rewritten as:

$$\varepsilon(\mathbf{r})\frac{\mathbf{E}_y^{n+1}(\mathbf{r},t)-\mathbf{E}_y^{n}(\mathbf{r},t)}{\Delta t}=\frac{\partial \mathbf{H}_x(\mathbf{r},t)}{\partial z}-\frac{\partial \mathbf{H}_z(\mathbf{r},t)}{\partial x}-\sigma(\mathbf{r})\mathbf{E}_y^{n+1/2}(\mathbf{r},t)+\mathbf{J}_s(\mathbf{r,t}), \tag{160}$$

$$\mu(\mathbf{r})\frac{\mathbf{H}_x^{n+1}(\mathbf{r},t)-\mathbf{H}_x^{n}(\mathbf{r},t)}{\Delta t}=\frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial z}, \tag{161}$$

$$-\mu(\mathbf{r})\frac{\mathbf{H}_z^{n+1}(\mathbf{r},t)-\mathbf{H}_z^{n}(\mathbf{r},t)}{\Delta t}=\frac{\partial \mathbf{E}_y(\mathbf{r},t)}{\partial x}, \tag{162}$$

in each iteration the fields $\mathbf{E}_y$, $\mathbf{H}_x$ and $\mathbf{H}_z$ are modified by the radiation pattern $\mathbf{Pr}(\mathbf{r})$ as follow:

$$
\begin{aligned}
\mathbf{E}_y^{n+1}(\mathbf{r},t) &= \mathbf{Pr}(\mathbf{r}) \cdot \mathbf{E}_y^{n+1}(\mathbf{r},t),\\
\mathbf{H}_x^{n+1}(\mathbf{r},t) &= \mathbf{Pr}(\mathbf{r}) \cdot \mathbf{H}_x^{n+1}(\mathbf{r},t),\\
\mathbf{H}_z^{n+1}(\mathbf{r},t) &= \mathbf{Pr}(\mathbf{r}) \cdot \mathbf{H}_z^{n+1}(\mathbf{r},t).
\end{aligned} \tag{163}
$$

To propagate the electromagnetic wavefront, the SEAM model is used with a spatial step in both dimensions of 0.05 (m), a time step of 0.06 (ns) with a total of 3000 time samples. The model has 30 (m) in distance and 15.05 (m) in depth. The central frequency used for these sample of time is 100 MHz. The location of the source is 15 (m) in distance and 4 (m) in depth. Figure 94 presents the wave propagation in the SEAM model with different pattern configurations: a), and b), presents the snapshot at 30 (ns) c), and d) presents the snapshot at 60 (ns). Figure 94 -a) and Figure 94-c), shows the propagation including the radiation pattern at $0°$ of inclination (Radiation pattern in Figure 93-b)). Figure 94-b) and Figure 94-d), presents the propagation without taking into account the radiation pattern.

Analyzing the red ovals in the Figures 94-a), and b): in the white marks 1 and 3, there are notable changes in the air-wave energy distribution when the radiation pattern is included. It is important to remember that according to the radiation pattern little energy is distributed between the angles from $-60°$ to $60°$. At mark 2, the antenna inclination is associated with the energy distribution and, therefore, object detection. Due to the antenna inclination, it might not measure the energy that returns at subsurface. Figure 94-e) shows a comparison of the normalized scan, without airwave and at zero-offset acquisition. Figure 94-e) shows the scan with and without radiation pattern and although the events are similar, the amplitudes are slightly attenuated according to the radiation pattern.

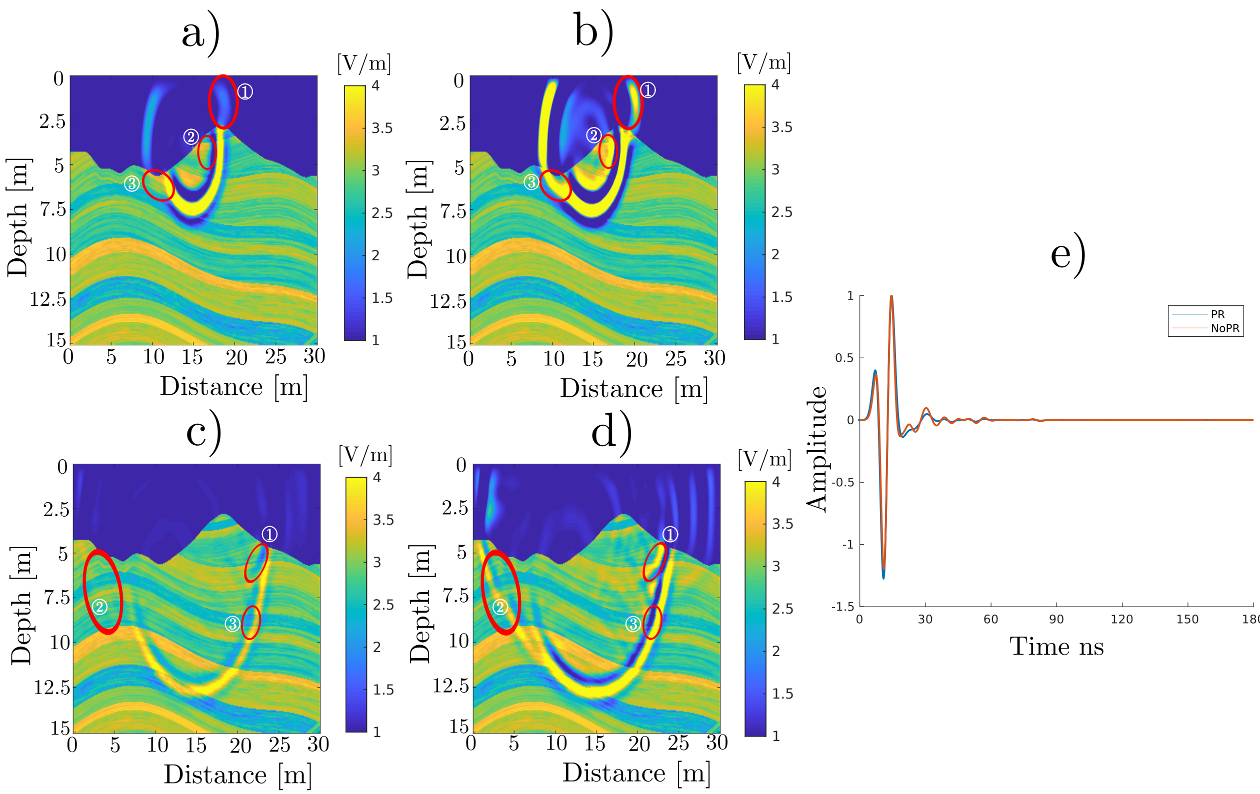

Figure 94. Propagation of the electromagnetic wavefront where the first row refers to the time at 30 (ns) and the second row at 60 (ns): a) and c) with radiation pattern at 0° of inclination, b) and d) without radiation pattern, and e) comparison of the modeled data with and without the radiation pattern (blue considering the radiation pattern and orange not considering the radiation pattern).

7.4.4. FWI including the radiation pattern. The same SEAM model is used here, and the forward and inversion model includes the radiation pattern of the antenna with an inclination angle of 0° is taking into account. We use the previously estimated radiation pattern of the antenna to simulate the acquisition of GRP data of a single-offset antenna on the surface. The shielded antenna at 400 MHz emits significant energy between 100 and 800 MHz. The radiation patterns depends on the frequency, but it is not easy to include them in the FDTD scheme in the time domain. Considering this limitation, we are assumed that the radiation patterns are frequency-independent for the antennas of 30, 50, and 100 MHz. The spatial step in both dimensions is 0.05 (m), a time step of 0.06 (ns) with a total of 3000-time samples. The model is 30 (m) in distance direction and 15.05 (m) in depth-direction. Three frequencies are used in a multi-scale methodology. 260 scans are distributed equispaced in the model and they are located at 25 (cm) from the surface. The receiving antenna is located at 15 (cm) from the transmitting antenna. The observed data is obtained using the radiation pattern presented in Figure 93-b), which is centered according to the position of each source. Figure 95 presents the results of performing the multi-scale approach on both scenarios: with and without the radiation pattern. The first row in Figure 95 is the relative conductivity parameter, and the second row is the relative permittivity parameter. In Figure 95, the first column presents the original models; the second column shows the initial models; the third column presents the results obtained using the radiation pattern, and the last column presents the results of the inversion process without taking into account the radiation pattern. When the radiation pattern is not taken into account, we can see that the estimation of the electromagnetic parameters is not correct and the conductivity parameter is the most affected in the inversion process.

Measuring the PSNR in the initial models, we have 24.29 (dB) for permittivity and 24.56 dB for conductivity. The results including the radiation pattern (Figure 95 -c)) reach PSNR values of 25.66 (dB) for permittivity and 25.59 (dB) for conductivity. The results where the radiation pattern is not included (Figure (95) -d)) reaches a PSNR of 25.59 (dB) for the relative permittivity and 22,87 (dB) for conductivity. The PSNR values obtained from including the radiation pattern are lower compared to the 260 scans experiment outlined in section **3.3.5** where PSNR values of 25.72 (dB) for permittivity and 25.65 (dB) for conductivity. The radiation pattern directs the energy between 120 degrees and 240 degrees, restricting a very short-offset of the energy that returns to the subsurface. Due to this short-offset, the wavenumbers are restricted to high wavenumbers, as it is shown in Figure 25. Note that in Figure 91-a) and Figure 91-b), the patterns are similar to a dipole pattern (see Figure 89-b) and Figure 89-c)). However, given the antenna's internal geometry, it produces a small amount of energy between -60° and 60°. Based on these obtained radiation patterns, we seek to include them in the propagation of the electromagnetic field, thus making it closer to reality and reducing the error between the modeled and observed data.

According to the manufacturer, the GPR acquisition line must be located in areas without topography for single-channel and short-offset acquisition. In the following experiment, the radiation pattern, topography, and antenna inclination behavior are considered in FWI. The results are presented in Figure 96. When the antenna inclination is considered, the FWI lost resolution in areas where the topography are high, as it is shown in zone A and B. When the antenna inclination is not considered (the antenna is parallel to horizontal), the PSNR values reached are 25.66 (dB) and 25.59 (dB) for permittivity and conductivity, respectively. The PSNR values when the antenna inclination is considered, are 25.41 (dB) and 25.44 (dB) for permittivity and conductivity, respectively. Therefore, in a real scenario with a stron topography, it is recommended to include the radiation patterns as well as the tilt of the antenna to have a better estimates of the electromagnetic parameters of the subsurface.

Figure 97-a) shows the maximum energy trajectory when the radiation pattern and the

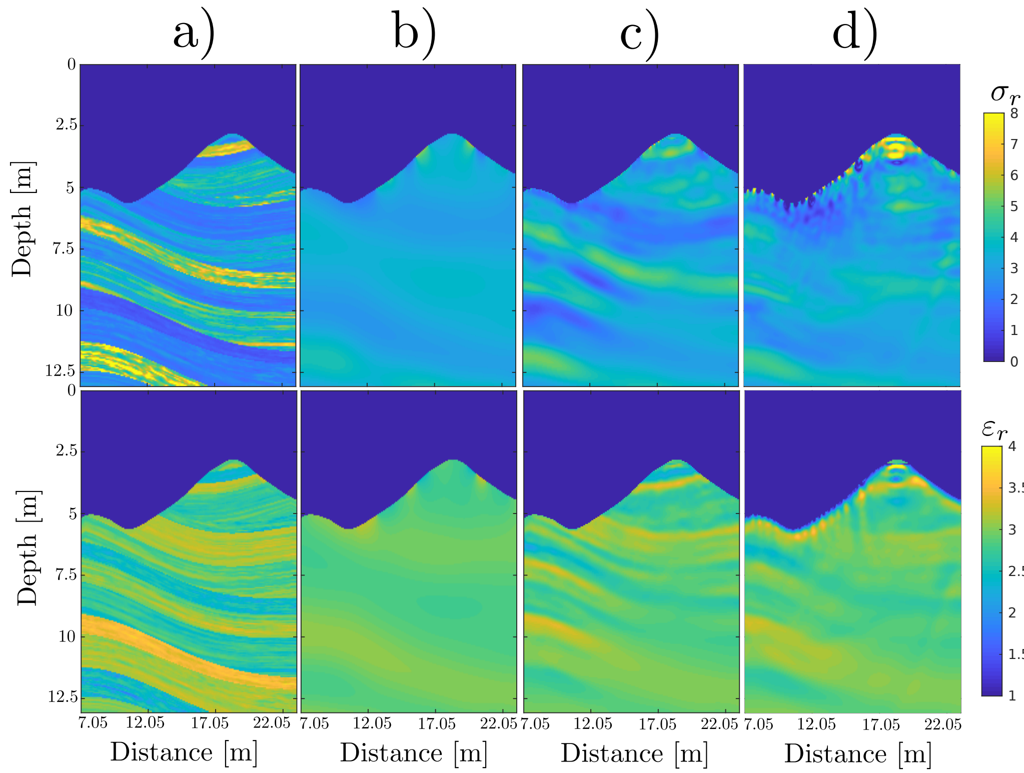

Figure 95. FWI results taking into account the radiation pattern. The first row refers to the relative conductivity and the second column refers to the relative permittivity: a) original models, b) initial models, c) Models obtained after FWI multi-scale using the radiation pattern and d) models obtained after of FWI multi-scale without the radiation pattern.

antenna inclination are considered according to the topography. The trajectory of the maximum energy is normal to the surface, and the reflected waves are not measured by the receiver antenna. Figure 97-b) shows the maximum energy trajectory with the radiation pattern but the inclination is not considered. When the antenna inclination is not considered, the receiver antenna sense more reflected waves without attenuation by the radiation pattern and it allows obtained better results in the electromagnetic parameters estimation of FWI. The results in the qualitative form show better results in the conductivity when the PR is included in the electromagnetic propagation than without

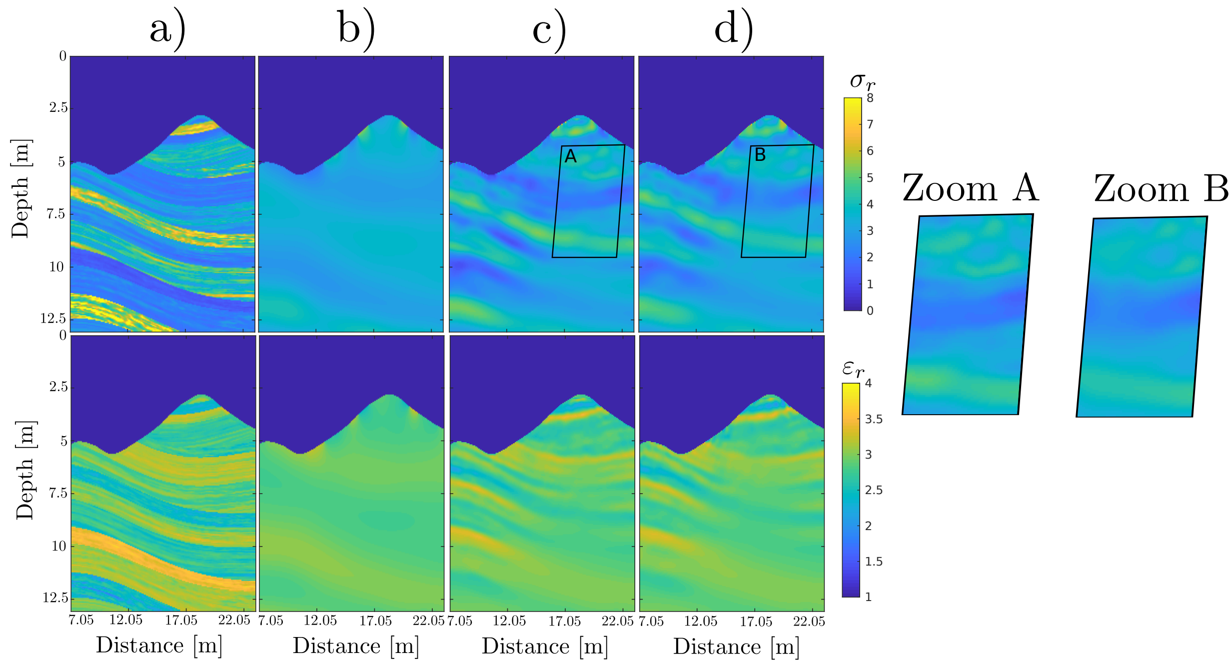

Figure 96. FWI results taking into account the radiation pattern and inclination. The first row refers to the relative conductivity and the second column refers to the relative permittivity: a) original models, b) initial models, c) Models obtained after FWI multi-scale using the radiation pattern without topography inclination and d) models obtained after of FWI multi-scale with the radiation pattern and topography inclination.

PR. The most affected electromagnetic parameter is the conductivity, which is associated with the phenomenon of dissipation or attenuation, and attenuation is associated with the process of the radiation pattern in electromagnetic propagation.

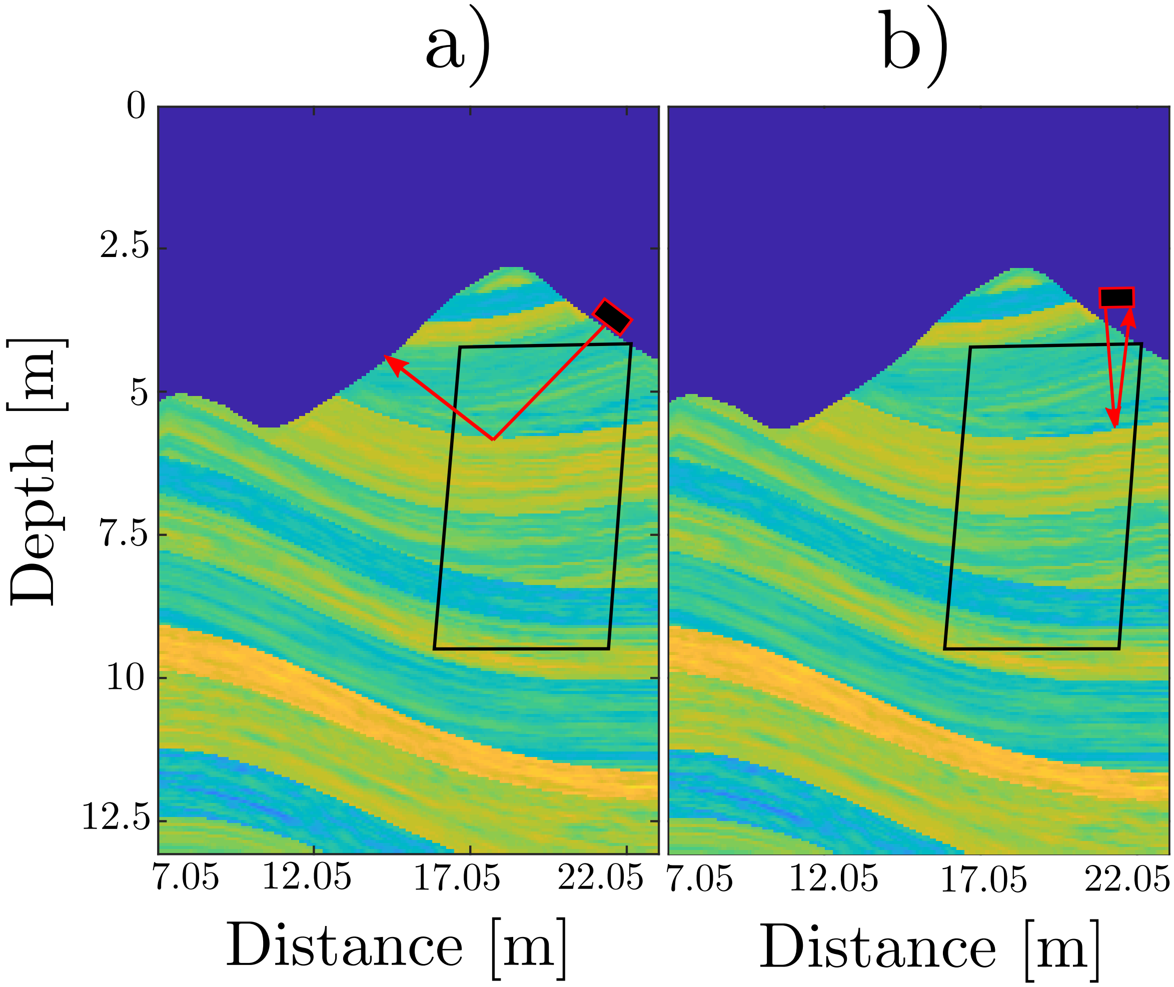

Figure 97. Maximum energy trajectory considering the radiation pattern: a) with inclination, b) without inclination

8. Future work

8.1. FWI-2D and Machine Learning

Colombia suffers from many post-conflict problems where civilians and soldiers lose their life or upper/lower extremities due to antipersonnel mines. In this section, we seek to present the possibility of using machine learning and FWI to accurately identify buried objects such as antipersonnel mines. Machine learning is a tool capable of "learning" and extracting the images characteristics and classified them according to some established labels.

Through a GPR with a drone, we can perform B-scans on a terrain where we suspect having antipersonnel mines and then apply machine learning algorithms to classify possible buried objects. The problem with applying FWI in a short-offset and single-channel scheme is that it has very little information from the subsurface, so Machine Learning can be a complementary tool, providing a starting point to FWI. With machine learning, we can also perform regression of the possible values of material like permittivity or conductivity of the subsurface or its geometry like width, height or depth. In this section, we use machine learning as a first approach to the buried object, and then FWI to improves the resolution of the buried objects object.

This section presents a future line of research to improve the results obtained from the traditional Full Waveform Inversion. As has been demonstrated, Full Waveform Inversion in a single-channel, short-offset scheme presents a great challenge due to the lack of low wavenumber information. In this project, FWI with regularizations such as TV or MTV allows mitigating the lack of information in addition to converging to more robust solutions to noise. Figure 98 presents a diagram with a target of interest. For this experiment, four parameters are expected to be identified using ML: height, width, depth, and relative permittivity. The background relative permittivity is known and it is selected in 2. Sources are located 5 cm from the surface. A total of 301 sources are uniformly distributed on the surface. The acquisition scheme is a single channel and short-offset (B-scans). Three thousand and eight hundred B-scans in total are used, where 70 % of the data is used in training, 15 % for validation and 15 % for test. The zone of interest has 30 (m) in x -direction and 15 (m) in z-direction. The time step is 0.08 (ns) and the step size is 0.05 (m). The central frequency is 100 (MHz). Figure 99 shows 20 B-scans, the air-wave is removed and a resize of 64×64 has been performed with a normalization.

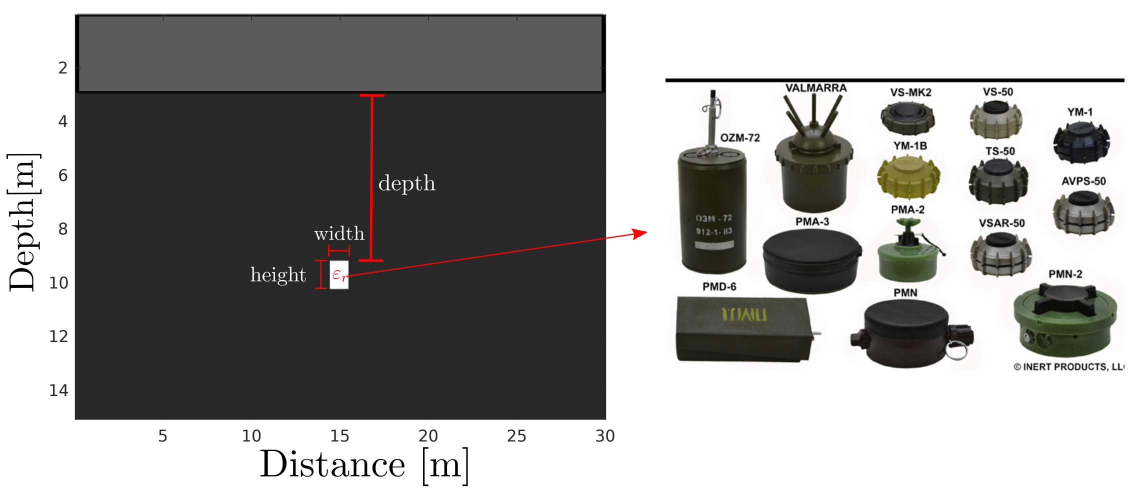

Figure 98. Parameters used for the ML estimation: depth,relative permittivity, height, and width.

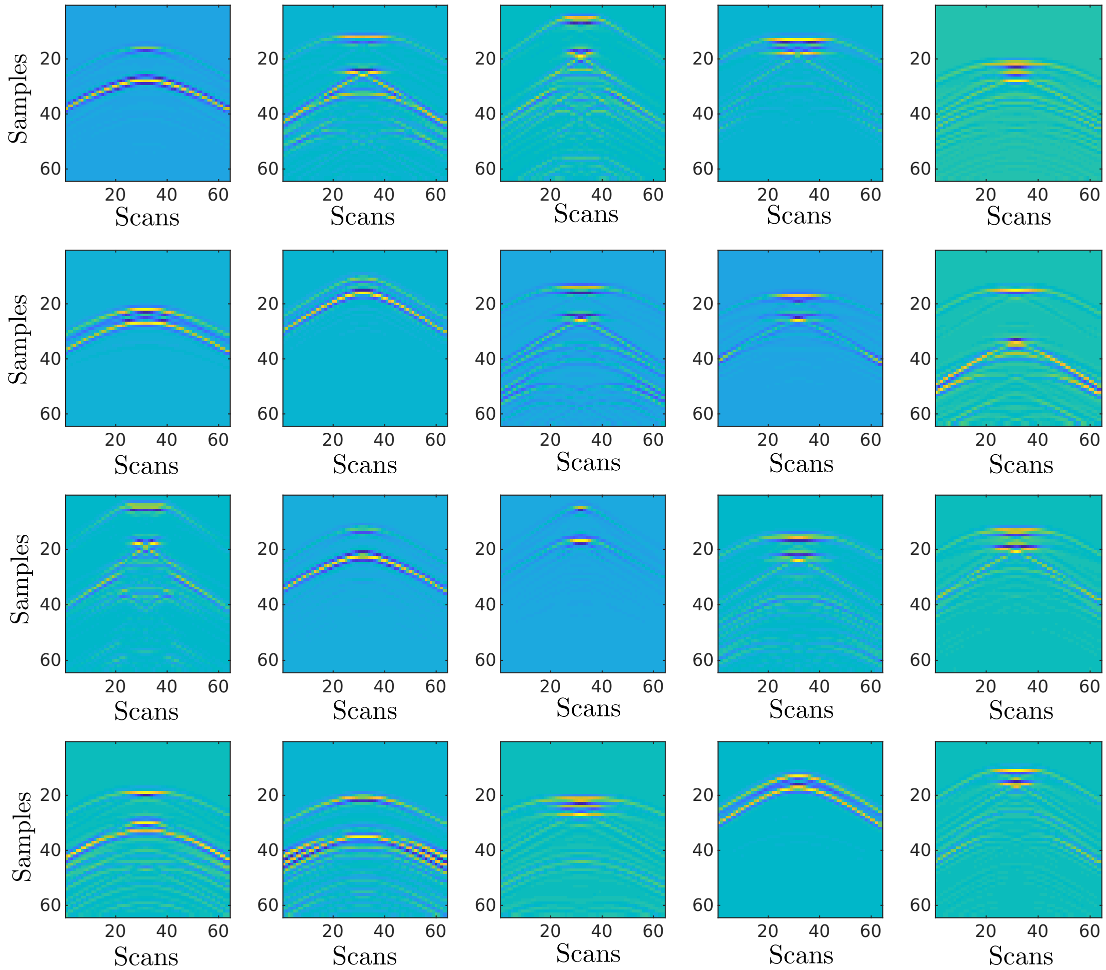

Figure 99. 20 B-scans used for neural network training.

Figure 100 shows that each parameter has a uniform distribution. A neural network is trained with the architecture shown in Figure 101.

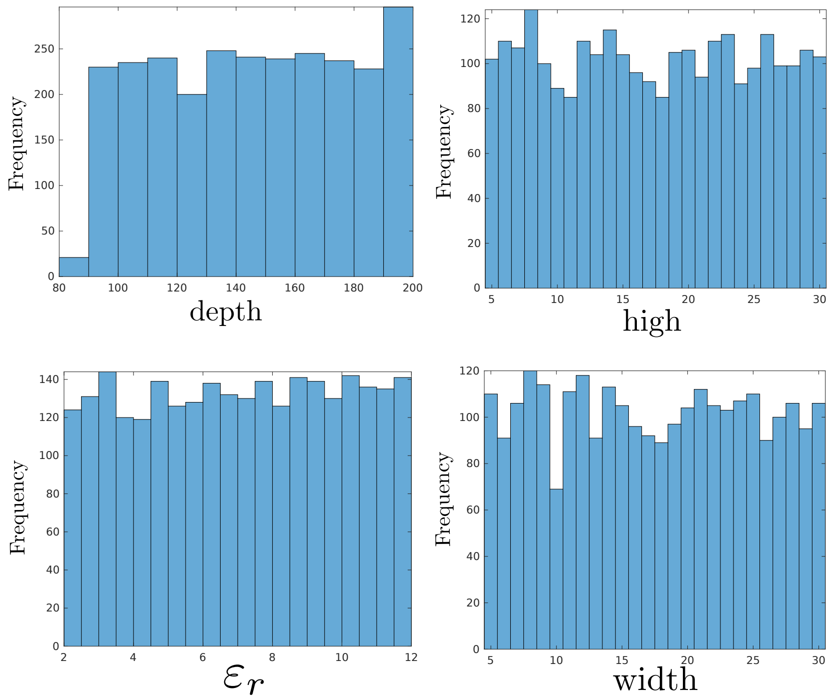

Figure 100. Histogram of each parameter used in ML

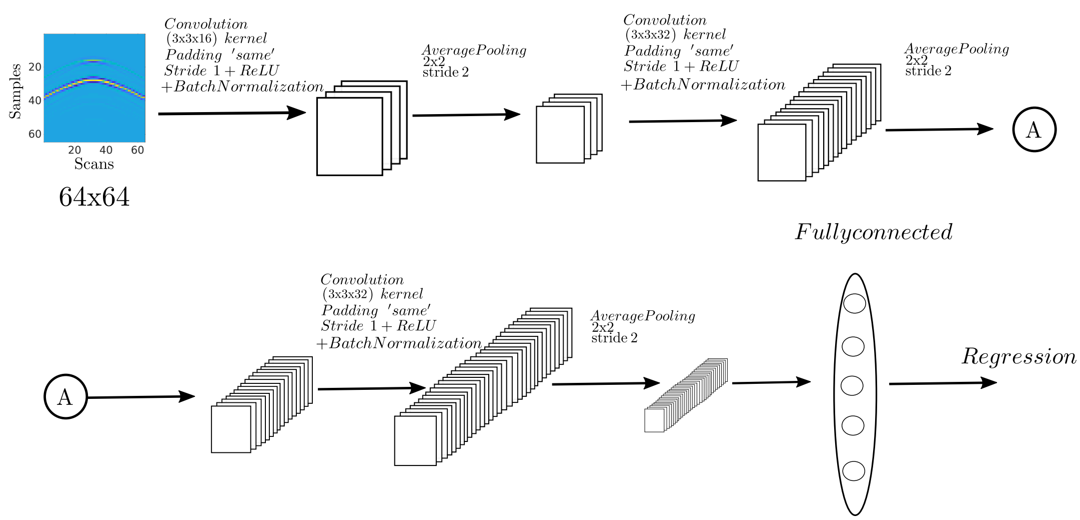

Figure 101. Convolutional Neural Network (CNN) architecture

This network allows performing a regression for each of the parameters. The training results are presented in Figure 102. The statistical measure (R^2) is used, which represents the proportion of a variance for a dependent variable, which is explained by an independent variable or by a regression model. The coefficient of determination is computed by:

$$R^2 = 1 - \frac{SSE}{SST}, \tag{164}$$

where SSE is the sum of squared error, and SST is the sum of the squared total. The coefficient of determination is 0.9836, 0.9551, 0.9642, and 0.9725 for the relative permittivity, depth, height, and width, respectively.

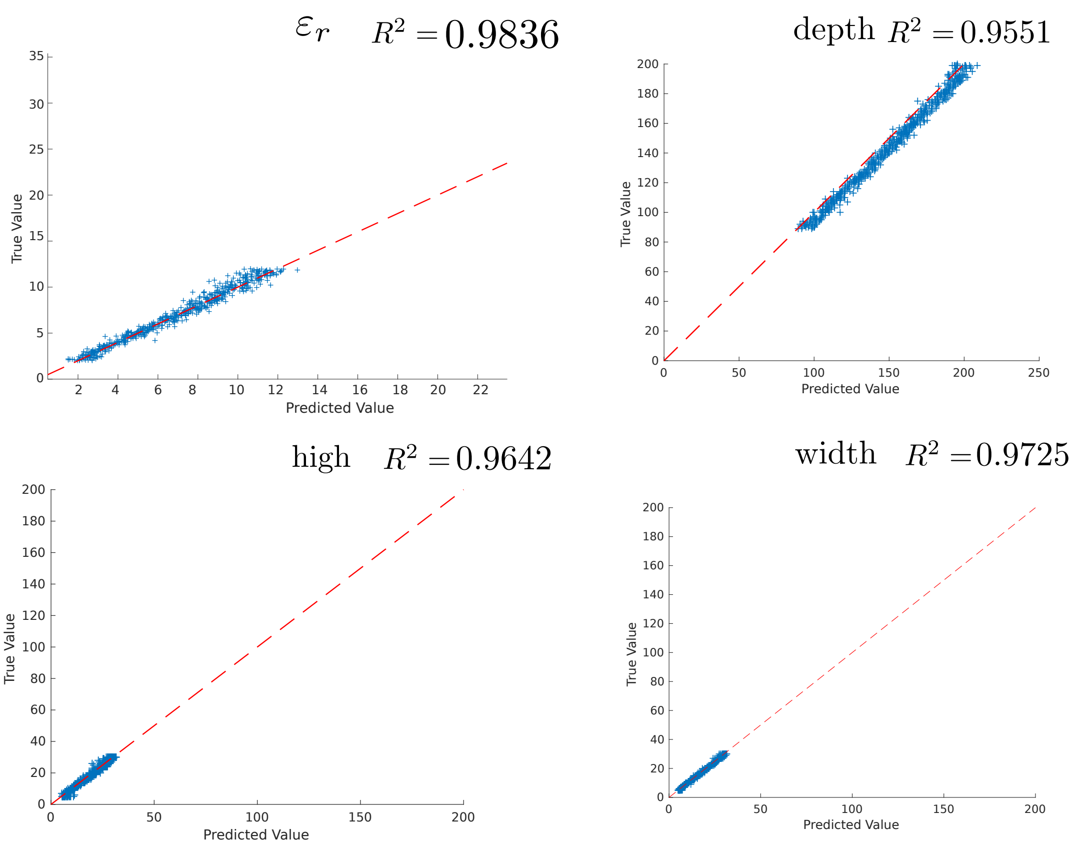

Figure 102. Linear regression for the parameters: relative permittivity , depth, height and, width.

Based on this network training, we have carried out two Full Waveform Inversion processes:

in the first, the *apriori* information delivered by the neural network is used to build the starting point, and in the second, we have not used any *apriori* information. Table 12 presents the true and obtained parameters using the neural network in the second and third row, respectively. Based on the parameters of the table 12, the starting point is built, where a smoothing filter is applied 40 times. Figures 103 -a) and 104 -a)show the true relative permittivity model. Figures 103 -b) and 104 -b) show the initial models used in FWI, using the ML solution and a uniform model, respectively. Figures 103 -c) and 104 -c) present the models obtained from applying the inversion process. When using the *apriori* information obtained from ML, the PSNR is 29.64 (dB), and without ML information, the PSNR is 23.83 (dB). It is evident how the *apriori* information obtained from the neural network significantly helps the estimation of the parameters. The depth object location has an error of 1.48 % (or 0.1m) when the initial model is obtained from ML. On the other hand, the error in the depth object location for a uniform initial model is 7.4 % (or 0.5m).

	ε_r	half-height	half-width	depth
True	6.32	29	8	163
ML	6.59	28	9	167

Table 12. *True and estimations parameters using ML for relative permittivity, height, width and depth.*

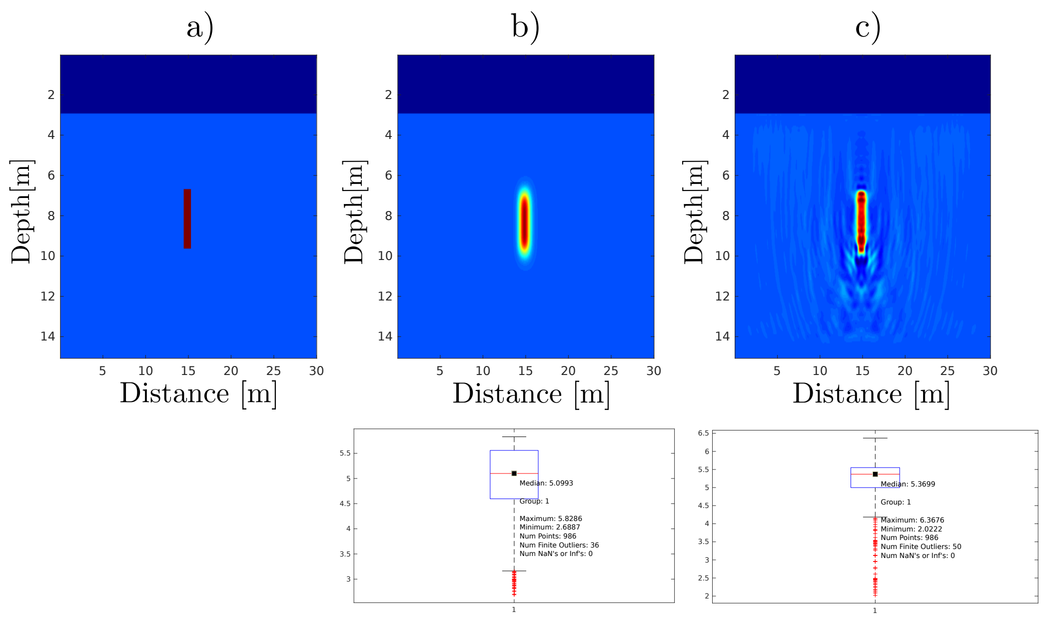

Figure 103. a) True model for the relative permittivity, b) Initial guess for relative permittivity, and c) FWI results with ML information

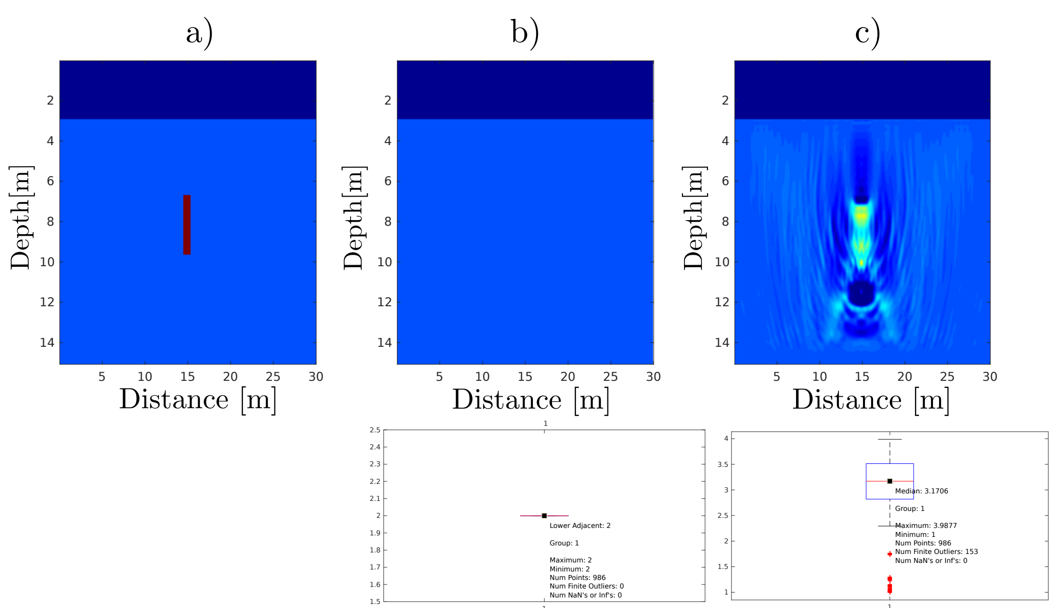

Figure 104. a) True model for the relative permittivity, b) Initial guess for relative permittivity, and c) FWI results without ML information.

9. Conclusions

- Single-channel and short-offset acquisitions are related to high wavenumbers in the solution, as it is shown in section **3.3.4**. This limitation requires better starting points than multi-offset acquisitions. TV and MTV allow converging in a smoother model that not only reduce the noise level while preserving the primary interfaces but also their regularization parameters can be tuned to obtain a better starting point before starting an inversion process again. We found that applying the Gaussian preprocessing in the direction of the samples does not affect. However, when it is applied in scans-direction, we have found that we can reduce the incoherent noise in the data, achieving a low-wavenumber version in the inversion parameters.

- The regularizations TV, MTV, and Gaussian preprocessing must have tuning stages of the regularization parameters such as γ_{TV} in TV, κ in MTV, or ρ^2 at Gaussian. In this thesis, we have found these parameters with many experiments in collected data and synthetic experiments for the electromagnetic case in short-offset and single-channel acquisitions. We propose to use data fitting as a protocol to tune these regularization parameters. In our synthetic and collected tests, we have achieved a better data fit than traditional FWI. Furthermore, as shown by the synthetic data in the annexes section or the real data, we found that the TV and MTV regularizations are more stable.

- In section **3.3.5**, we have analyzed the horizontal resolution in short-offset and single-channel acquisitions for different scans-samples. Based on many experiments, we have concluded that a scan should be carried out for every two times the desired spatial resolution in the model. According to our synthetic results, we adapt the horizontal resolution equation to find the following expression:

$$\Delta l = \sqrt{\frac{c \cdot dh}{2 \cdot 4 \cdot f_c \cdot \sqrt{\varepsilon_r}}} \tag{165}$$

With this expression we can reduce the execution time by a factor 2x.

- We have implemented a methodology that allows characterizing the internal parameters of a shielded antenna. The internal parameters are the permittivity and conductivity of the observant barrier, the permittivity of the PCB, and the resistance of the antennas (Transmitter and receiver). The PSO global optimization algorithm has been used together with the free software gprMAx for its characterization. In this optimization process, we have used two metrics: correlation and FFT. Based on the characterized parameters, we have obtained the radiation patterns in planes E and H. The radiation patterns are included to allow us to know the distribution of the energy in the medium and reduce the difference of energy between the observed and modeled data. Including the radiation patterns in the electromagnetic propagation, we avoid a quick converge and unwanted artifacts in the first layers of the subsurface. Figure 91 shows that the angles of the highest attenuation for the shielded antenna at 400 (MHz) are between -60° and 60°, this limits the information obtained from objects outside of these angles and also generates possible artifacts in the inversion process if it is not taken into account in the modeled data. We have included these patterns in a two-dimensional wavefront, and we have reached better electromagnetic parameters estimations.

- In section **4.1**, we have proposed an alternative cost function that compensates the amplitude of the modeled data making it comparable with the observed data. We show that when the data have noise (η) and with an expected value equal to zero, the alternative cost function has similar behavior with the traditional cost function. However, in the collected data or synthetic case with WGN and expected value different from zero, the alternative cost function reaches better results concerning to PSNR (synthetic data, see Figure 35) or cost function (collected data, see Figure 106).

- The 2D FWI algorithm has been implemented in short-offset and single-channel acquisitions for synthetic and collected data. The implementation has been developed using a hybrid architecture GPU-CPU using a cluster PE ProLiant XL270d Gen10 with two Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz and eight NVIDIA Tesla V100, 16 GB. In this doctoral thesis, we have taken advantage of the independence between scans to distribute the load in each GPU using the MPI protocol, which manages to reduce its execution time by 8x, given that it is the number of GPUs used. Our algorithm only stores the electric field's volume, and the magnetic field is computing from the electric field.

- We have implemented and developed a mathematical formulation for a new computational strategy in FWI-3D in single-channel and short-offset acquisitions. The strategy is based on that the gradient energy is located in a radius concerning the source location. Our synthetic tests show that we must keep the gradient update zone close to $\lambda/4$ by taking the lowest frequency in multi-scale methodologies. From our synthetic tests, we are reduced the computational cost by 9.5x compared with the traditional FWI. The execution time in a 3D-propagation increases 115 times compared to a 2D propagation, so implementing this type of inverse problem is still a significant challenge in single-channel and short-offset acquisitions.

- We have developed a methodology that allows joining ML with FWI for GPR data, which improves the estimation of electromagnetic parameters. The FWI results with ML information improve by 5.81 dB. Furthermore, the FWI with ML information reduces the target detection error by 5.92 %. We have found that removing the air-wave and normalizing the data improves ML efficiency by 15 %. We propose future work to use more complex scenarios, where 3D propagation includes their radiation pattern and the artifact geometry.

- As future work, we have proposed including artificial intelligence to relate electromagnetic parameters to seismic parameters. We suggest carrying out a 3D implementation in CPU-GPU of the inversion process using energy compensation to reduce the error from 3D to 2D conversion and include the near field effect.

Bibliography

Agudo, O., Guasch, L., Huthwaite, P., and Warner, M. (2018). 3d imaging of the breast using full-waveform inversion. In *Proceedings of the International Workshop on Medical Ultrasound Tomography*, pages 99–110.

Anagaw, A. Y. and Sacchi, M. D. (2012). Edge-preserving seismic imaging using the total variation method. *Journal of Geophysics and Engineering*, 9(2):138–146.

Baysal, E., Kosloff, D. D., and Sherwood, J. W. (1983). Reverse time migration. *Geophysics*, 48(11):1514–1524.

Belina, F. A., Irving, J., Ernst, J. R., and Holliger, K. (2012). Waveform inversion of crosshole georadar data: Influence of source wavelet variability and the suitability of a single wavelet assumption. *IEEE transactions on geoscience and remote sensing*, 50(11):4610–4625.

Berenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of computational physics*, 114(2):185–200.

Blom, N., Gokhberg, A., and Fichtner, A. (2020). Seismic waveform tomography of the central and eastern mediterranean upper mantle. *Solid Earth*, 11(2):669–690.

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

Canavos, G., Meyer, P., Spiegel, M., and Mendenhall, S. (1988). Probabilidad y estadística. *LICENCIATURA EN INGENIERÍA EN INFORMÁTICA*, 28.

Cangellaris, A. C. and Wright, D. B. (1991). Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in fdtd simulations of electromagnetic phenomena. *IEEE transactions on antennas and propagation*, 39(10):1518–1525.

Daniels, D. (1996). Surface-penetrating radar-ieee radar, sonar, navigation and avionics series 6. *The Institue of Electrical Engineers*.

Daniels, D. J. (2004). *Ground penetrating radar*, volume 1. Iet.

De Moura, C. A. and Kubrusly, C. S. (2013). The courant-friedrichs-lewy (cfl) condition. *AMC*, 10:12.

Diamanti, N. and Annan, A. P. (2013). Characterizing the energy distribution around gpr antennas. *Journal of Applied Geophysics*, 99:83–90.

Dridi, K., Hesthaven, J. S., and Ditkowski, A. (2001). Staircase-free finite-difference time-domain formulation for general materials in complex geometries. *IEEE Transactions on Antennas and Propagation*, 49(5):749–756.

Elliot, R. (1983). Ieee standard definitions of terms for antennas. *IEEE Transactions on Antennas and Propagation*, 31(6):1–29.

Ernst, J. R., Green, A. G., Maurer, H., and Holliger, K. (2007). Application of a new 2d time-domain full-waveform inversion scheme to crosshole radar data. *Geophysics*, 72(5):J53–J64.

Gao, K. and Huang, L. (2019). Acoustic-and elastic-waveform inversion with total generalized p-variation regularization. *Geophysical Journal International*, 218(2):933–957.

Goldstein, A. A. (1965). On newton's method. *Numerische Mathematik*, 7(5):391–393.

Goldstein, T. and Osher, S. (2009). The split bregman method for l1-regularized problems. *SIAM journal on imaging sciences*, 2(2):323–343.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52.

Holland, R. (1993). Pitfalls of staircase meshing. *IEEE Transactions on Electromagnetic Compatibility*, 35(4):434–439.

Jaap, W. (2018). A general recipe for obtaining adjoint equations and gradients and its application to full waveform inversion of 2d isotropic elastic media in finite difference time domain, internal report of CPS group.

Jazayeri, S., Klotzsche, A., and Kruse, S. (2018). Improving estimates of buried pipe diameter and infilling material from ground-penetrating radar profiles with full-waveform inversion. *Geophysics*, 83(4):H27–H41.

Jazayeri, S., Kruse, S., Hasan, I., and Yazdani, N. (2019). Reinforced concrete mapping using full-waveform inversion of gpr data. *Construction and Building Materials*, 229:117102.

Jol, H. M. (2008). *Ground penetrating radar theory and applications*. elsevier.

Kalogeropoulos, A., Kruk, J., Hugenschmidt, J., Busch, S., and Merz, K. (2011). Chlorides and moisture assessment in concrete by gpr full waveform inversion. *Near Surface Geophysics*, 9.

Khan, U. S. and Al-Nuaimy, W. (2010). Background removal from gpr data using eigenvalues. In *Proceedings of the XIII Internarional Conference on Ground Penetrating Radar*, pages 1–5. IEEE.

Klotzsche, A., van der Kruk, J., Meles, G. A., Doetsch, J., Maurer, H., and Linde, N. (2010). Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the thur river, switzerland. *Near surface geophysics*, 8(6):635–649.

Klotzsche, A., Vereecken, H., and van der Kruk, J. (2019). Review of crosshole ground-penetrating radar full-waveform inversion of experimental data: Recent developments, challenges, and pitfalls. *Geophysics*, 84(6):H13–H28.

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.

Lailly, P. and Bednar, J. (1983). The seismic inverse problem as a sequence of before stack migrations. In *Conference on inverse scattering: theory and application*, pages 206–220. Siam Philadelphia, PA.

Lambot, S., Slob, E., Van Den Bosch, I., Stockbroeckx, B., Scheers, B., and Vanclooster, M. (2004a). Estimating soil electric properties from monostatic ground-penetrating radar signal inversion in the frequency domain. *Water Resources Research*, 40(4).

Lambot, S., Slob, E. C., van den Bosch, I., Stockbroeckx, B., and Vanclooster, M. (2004b). Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties. *IEEE transactions on geoscience and remote sensing*, 42(11):2555–2568.

Lavoué, F. (2014). *2D full waveform inversion of ground penetrating radar data: towards multi-parameter imaging from surface data. PhD thesis, Université de Grenoble.*

Lavoué, F., Brossier, R., Métivier, L., Garambois, S., and Virieux, J. (2014). Two-dimensional permittivity and conductivity imaging by full waveform inversion of multioffset gpr data: A frequency-domain quasi-newton approach. *Geophysical Journal International*, 197(1):248–268.

Lin, Y. and Huang, L. (2014). Acoustic-and elastic-waveform inversion using a modified total-variation regularization scheme. *Geophysical Journal International*, 200(1):489–502.

Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.

Meles, G. A., Van der Kruk, J., Greenhalgh, S. A., Ernst, J. R., Maurer, H., and Green, A. G. (2010). A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination crosshole/borehole-to-surface gpr data. *IEEE Transactions on geoscience and remote sensing*, 48(9):3391–3407.

Miller, R. D., Bradford, J. H., and Holliger, K. (2010). *Advances in near-surface seismology and ground-penetrating radar*. Society of Exploration Geophysicists, American Geophysical Union.

Mozaffari, A., Klotzsche, A., Warren, C., He, G., Giannopoulos, A., Vereecken, H., and van der Kruk, J. (2020). 2.5 d crosshole gpr full-waveform inversion with synthetic and measured data. *Geophysics*, 85(4):H71–H82.

Pasalic, D. and McGarry, R. (2010). Convolutional perfectly matched layer for isotropic and anisotropic acoustic wave equations. In *SEG Technical Program Expanded Abstracts 2010*, pages 2925–2929. Society of Exploration Geophysicists.

Persico, R. (2014). *Introduction to ground penetrating radar: inverse scattering and data processing*. John Wiley & Sons.

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.

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3):888–901.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1988). Numerical recipes in c.

Rashed, E. A. (2015). Gpr background removal using a directional total variation minimisation approach. *Journal of Geophysics and Engineering*, 12(6):897–908.

Regone, C., Stefani, J., Wang, P., Gerea, C., Gonzalez, G., and Oristaglio, M. (2017). Geologic model building in seam phase ii-land seismic challenges. *The Leading Edge*, 36(9):738–749.

Rial, F. I., Pereira, M., Lorenzo, H., Arias, P., and Novo, A. (2007). Vertical and horizontal resolution of gpr bow-tie antennas. In *2007 4th International Workshop on, Advanced Ground Penetrating Radar*, pages 187–191. IEEE.

Rodríguez, P. (2013). Total variation regularization algorithms for images corrupted with different noise models: a review. *Journal of Electrical and Computer Engineering*, 2013.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268.

Serrano, J. O., Ramirez, A. B., Abreo, S. A., and Sadler, B. M. (2020). Alternative cost function for full waveform inversion of gpr data. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXV*, volume 11418, page 1141806. International Society for Optics and Photonics.

Serrano, J. O., Ramirez, A. B., Abreo Sr, S. A., and Sadler, B. M. (2018). Permittivity and conductivity parameter estimations using full waveform inversion. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIII*, volume 10628, page 106281K. International Society for Optics and Photonics.

Serrano, J. O., Ramirez, A. B., Abreo Sr, S. A., and Sadler Sr, B. M. (2019). Multiparametric full waveform inversion of 2d ground penetrating radar data using multiple gpus. In *Detection and*

Sensing of Mines, Explosive Objects, and Obscured Targets XXIV, volume 11012, page 1101219. International Society for Optics and Photonics.

Shaw, R. and Srivastava, S. (2007). Particle swarm optimization: A new tool to invert geophysical data. *Geophysics*, 72(2):F75–F83.

Shi, Y. et al. (2001). Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)*, volume 1, pages 81–86. IEEE.

Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.

Snieder, R. and Trampert, J. (1999). Inverse problems in geophysics. *Wavefield inversion*, pages 119–190.

Sullivan, D. M. (2013). *Electromagnetic simulation using the FDTD method*. John Wiley & Sons.

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

Tromp, J., Komatitsch, D., and Liu, Q. (2008). Spectral-element and adjoint methods in seismology. *Communications in Computational Physics*, 3(1):1–32.

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

Wang, Y., Yang, J., Yin, W., and Zhang, Y. (2008). A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272.

Warren, C. (2009). Numerical modelling of high-frequency ground-penetrating radar antennas. The University of Edinburgh.

Warren, C. and Giannopoulos, A. (2017). Characterisation of a ground penetrating radar antenna in lossless homogeneous and lossy heterogeneous environments. *Signal processing*, 132:221–226.

Warren, C., Giannopoulos, A., and Giannakis, I. (2016). gprmax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar. *Computer Physics Communications*, 209:163–170.

Watson, F. (2016). Towards 3d full-wave inversion for gpr. In *2016 IEEE Radar Conference (RadarConf)*, pages 1–6. IEEE.

Wirgin, A. (2004). The inverse crime. *arXiv preprint math-ph/0401050*.

Wu, Z. and Alkhalifah, T. (2018). Selective data extension for full-waveform inversion: An efficient solution for cycle skipping. *Geophysics*, 83(3):R201–R211.

Xue, Z., Alger, N., and Fomel, S. (2016). Full-waveform inversion using smoothing kernels. In *SEG Technical Program Expanded Abstracts 2016*, pages 1358–1363. Society of Exploration Geophysicists.

Yang, X., Van Der Kruk, J., Bikowski, J., Kumbhar, P., Vereecken, H., and Meles, G. (2012). Full-waveform inversion of gpr data in frequency-domain. In *2012 14th International Conference on Ground Penetrating Radar (GPR)*, pages 324–328. IEEE.

Yee, K. (1966). Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on antennas and propagation*, 14(3):302–307.

Zhang, P., Gao, R., Han, L., and Lu, Z. (2021). Refraction waves full waveform inversion of deep reflection seismic profiles in the central part of lhasa terrane. *Tectonophysics*, 803:228761.

Annexes

We have divided the annexes section into four subsections: FWI algorithm, FWI results in the collected data with the alternative cost function, FWI with Gaussian and TV regularization, and FWI results with the parameter of scale β and noisy data. We showed other experiments with combinations of regularization that we are eliminated in the main document for not saturated the reader with more results.

A. FWI algorithm

The following algorithm summarizes the use of FWI throughout the experiments in the thesis using synthetic and collected data. Please consider that the cost function and its respective gradient depend on the regularization used: no regularization, TV, MTV, and Gaussian.

Algorithm 3 FWI algorithm

$\mathbf{d}_{obs}^{raw}, J_s, \varepsilon_r^{(0)}, \mu_r^{(0)}, \sigma_r^{(0)}\ \varepsilon_r^{(k)}, \mu_r^{(k)}, \sigma_r^{(k)}$
1) Compute $[\mathbf{d}_{obs}] \Leftarrow$ **Peprocessing**$(\mathbf{d}_{obs}^{raw})$ **comment:** see subsection **Preprocessing**.
Initialization : $\alpha_k = 1$
For(From $k = 0$ to N_{ite}){ **comment:** N_{ite} is the number of iterations
Initialization: $f^{(k)} \Leftarrow 0,\ g^{(k)}(\varepsilon_r^{(k)}) \Leftarrow 0,\ g^{(k)}(\mu_r^{(k)}) \Leftarrow 0,\ g^{(k)}(\sigma_r^{(k)}) \Leftarrow 0;$
For(From $i = 0$ to N_s){ **comment:** N_s is the total scans
2) Compute $[\mathbf{E}_y, \mathbf{H}_x, \mathbf{H}_z, \mathbf{d}_{mod}^i] \Leftarrow$ **Forward_operator**$(J_s, \varepsilon_r^{(k)}, \mu_r^{(k)}, \sigma_r^{(k)})$ **comment:** see subsection **The forward problem**
3) Compute $[\lambda_{\mathbf{E}_y}, \lambda_{\mathbf{H}_x}, \lambda_{\mathbf{H}_z}] \Leftarrow$ **Adjoint_operator**$(\mathbf{d}_{obs}^i, \mathbf{d}_{mod}^i, \varepsilon_r^{(k)}, \mu_r^{(k)}, \sigma_r^{(k)})$ **comment:** see subsection **Adjoint operator for the inverse problem**
4) Compute $[g^i(\varepsilon_r^{(k)}), g^i(\mu_r^{(k)}), g^i(\sigma_r^{(k)})] \Leftarrow$ **gradients**$(\mathbf{E}_y, \mathbf{H}_x, \mathbf{H}_z, \lambda_{\mathbf{E}_y}, \lambda_{\mathbf{H}_x}, \lambda_{\mathbf{H}_z})$; **comment:** see subsection **Gradients for the inverse problem for GPR**
5) Compute $f^{(i)} \Leftarrow$ **cost_function**$(\mathbf{d}_{mod}^{(i)}, \mathbf{d}_{obs}^{(i)})$
$f^{(k)} \Leftarrow f^{(k)} + f^{(i)}$
$g^{(k)}(\varepsilon_r^{(k)}) \Leftarrow g^{(k)}(\varepsilon_r^{(k)}) + g^i(\varepsilon_r^{(k)})$ **comment:** Permittivity gradient
$g^{(k)}(\mu_r^{(k)}) \Leftarrow g^{(k)}(\mu_r^{(k)}) + g^i(\mu_r^{(k)})$ **comment:** Permeability gradient
$g^{(k)}(\sigma_r^{(k)}) \Leftarrow g^{(k)}(\sigma_r^{(k)}) + g^i(\sigma_r^{(k)})$ **comment:** Conductivity gradient
} **comment:** End For N_s
if(k=0){
$\varepsilon_r^{(k+1)} \Leftarrow \varepsilon_r^{(k)} - \alpha_{(k)} \cdot g^{(k)}(\varepsilon_r^{(k)})/||g^{(k)}(\varepsilon_r^{(k)})||_2^2$ **comment:** Update permittivity
$\mu_r^{(k+1)} \Leftarrow \mu_r^{(k)} - \alpha_{(k)} \cdot g^{(k)}(\mu_r^{(k)})/||g^{(k)}(\mu_r^{(k)})||_2^2$ **comment:** Update permeability
$\sigma_r^{(k+1)} \Leftarrow \sigma_r^{(k)} - \alpha_{(k)} \cdot g^{(k)}(\sigma_r^{(k)})/||g^{(k)}(\sigma_r^{(k)})||_2^2$ **comment:** Update conductivity }
else{
Compute $[\varepsilon_r^{(k+1)}, \mu_r^{(k+1)}, \sigma_r^{(k+1)}] \Leftarrow$ **L-BFGS**$(g^{(k)}(\varepsilon_r^{(k)}), g^{(k)}(\mu_r^{(k)}), g^{(k)}(\sigma_r^{(k)})\ , \varepsilon_r^{(k)}, \mu_r^{(k)}, \sigma_r^{(k)})$ }
} **comment:** End for N_{ite}

B. FWI results for the collected data with traditional and alternative cost functions

This subsection introduces the use of regularization using the traditional cost function and the alternative cost function. As in the synthetic data (see subsection 4.1), the alternative cost function better compensates for the differences in amplitudes between the observed and modeled data. The alternative costo function reaches better fits beteen the data than the traditional cost function. The results of FWI and the cost function are presented in the figures 105 and, 106, respectively.

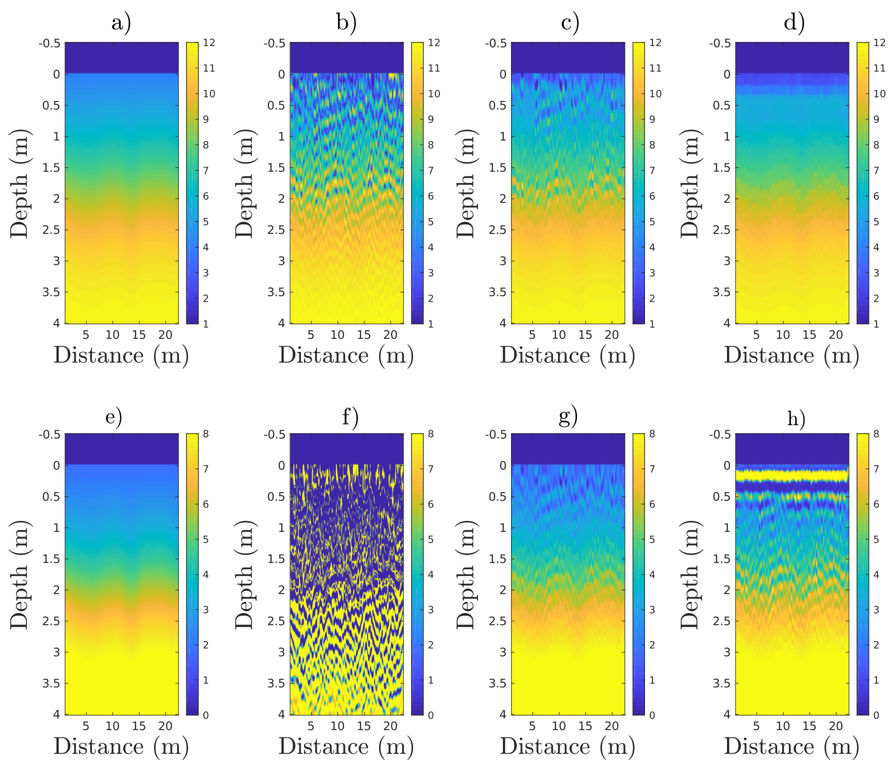

Figure 105. FWI results where the first row is the relative permittivity and the second row is the relative conductivity. MTV regularization is applied in both parameters (relative permittivity and conductivity). a) and e) initial parameters, b) and f) no regularization, c) and g) alternative cost function; d) and h) traditional cost function.

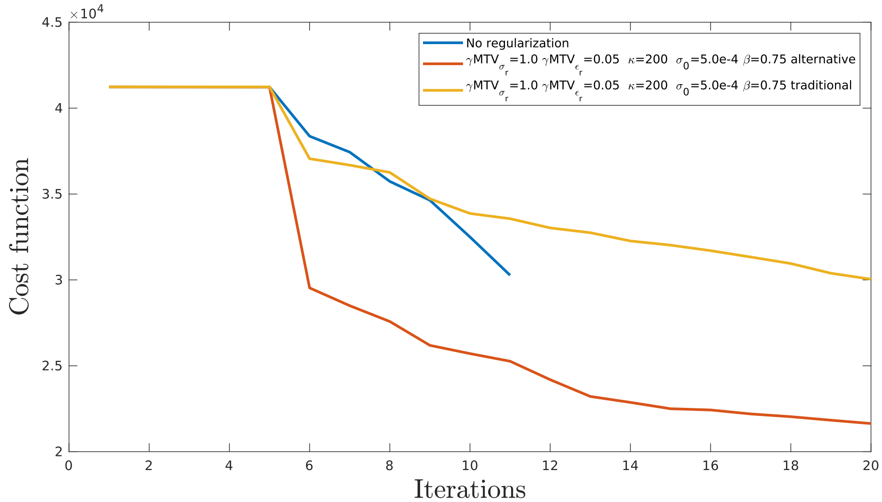

Figure 106. The behavior of the cost function for the data collected with MTV regularization with traditional cost function and alternative cost function.

C. β parameter with regularizations and noisy data

This section presents the synthetic results obtained from modifying the β parameter on a section of the SEAM model. For all the experiments, a multi-scale scheme with frequencies of 30, 50, and 100 MHz is used, where 30 iterations per frequency have been carried out. The β parameter control the weight in the gradient for the conductivity parameter. We identify that the parameter becomes important when it is included together with the restrictions. The β parameter is included to scale the σ parameter according with (Lavoué et al., 2014) and is defined by:

$$\sigma = \frac{\sigma_0 \cdot \sigma_r}{\beta}, \tag{166}$$

and the definition for the conductivity gradient according to this modification would be:

$$g(\sigma_r(\mathbf{r})) = \int_0^{Tend} \beta \cdot \sigma_0 \cdot \lambda_{\mathbf{E}_y}(\mathbf{r},t)\mathbf{E}_y(\mathbf{r},t)dt. \tag{167}$$

For all experiments, the regularizations are applied on both parameters: relative permittivity and relative conductivity. The observed data is contaminated with noise using a uniform distribution with an amplitude of 1 % of the maximum amplitude of the data. Figures 107 and 108 present for the permittivity and relative conductivity, the original model, initial model, and the results of FWI with and without being noise-constrained. As can be seen in Figure 108-d) when including noise in the data, the conductivity parameter is the most sensitive in the FWI. In the document, we have explored different regulations such as TV and MTV, but in this section, we include the regularizations together with the β parameter, which allows us to improve the achieved PSNR values. Figures 109 and 110 present the qualitative results achieved in each inversion process by including the regularizations and changing the β parameter. The Tables 13, 14 and 15 present the results achieved in PSNR, where the best PSNR value is obtained for σ_r and ε_r and using MTV and $\beta = 2.0$.

With both regularizations, the noise level is reduced and the FWI are stable. However, these regularizers require a tuning process, which can increase execution times. According to all the experiments carried out, it is recommended to use a $\beta = 0.5$ for TV and a $\beta = 2.0$ for MTV.

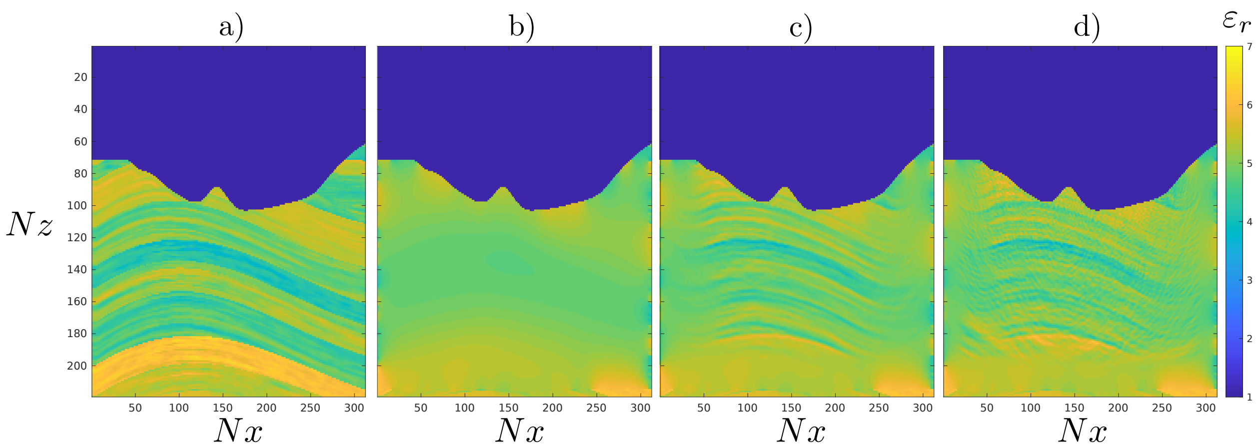

Figure 107. Relative permittivity: a) True model, b) Initial guess, c) FWI results without WGN and d) FWI results with WGN.

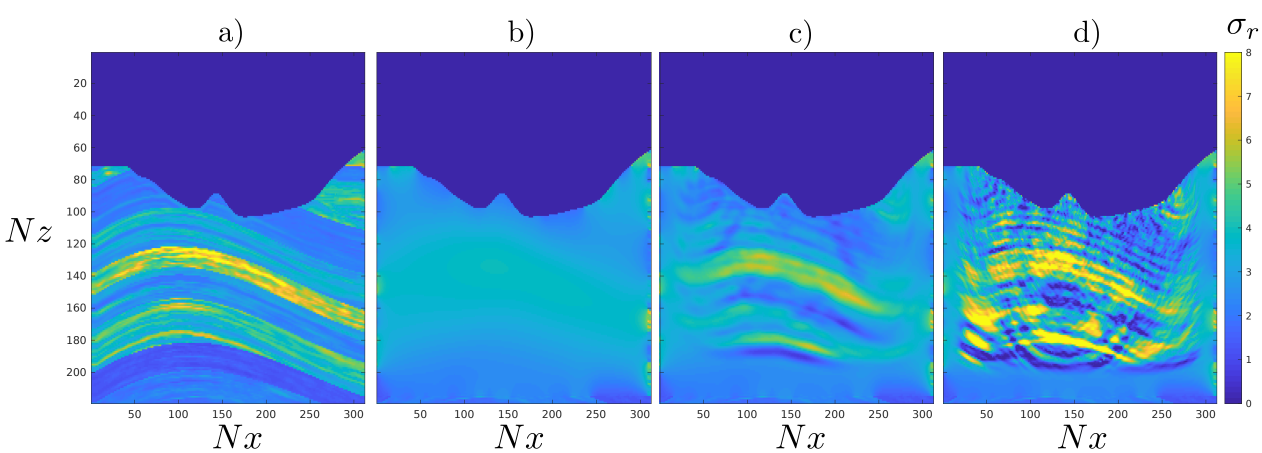

Figure 108. Relative conductivity: a) True model, b) Initial guess, c) FWI results without WGN and d) FWI results with WGN.

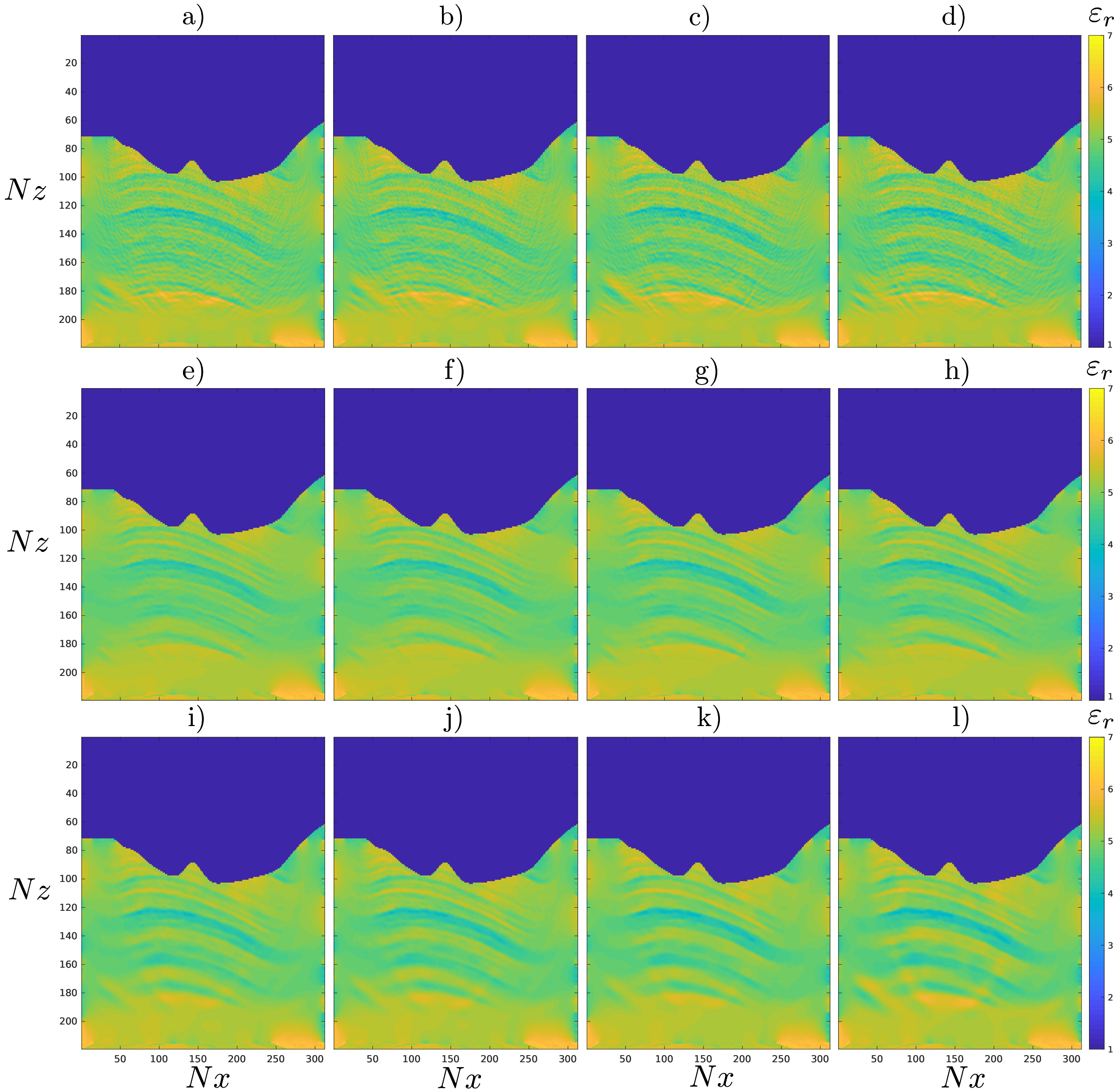

Figure 109. First row, FWI results for relative permittivity without regularizations: a) $\beta = 2.0$, b) $\beta = 1.0$, c) $\beta = 0.5$, d) $\beta = 0.25$. Second row, FWI results for relative permittivity with TV regularization: e) $\beta = 2.0$, f) $\beta = 1.0$, g) $\beta = 0.5$, h) $\beta = 0.25$. Third row, FWI results for relative permittivity with MTV regularization: i) $\beta = 2.0$, j) $\beta = 1.0$, k) $\beta = 0.5$, l) $\beta = 0.25$.

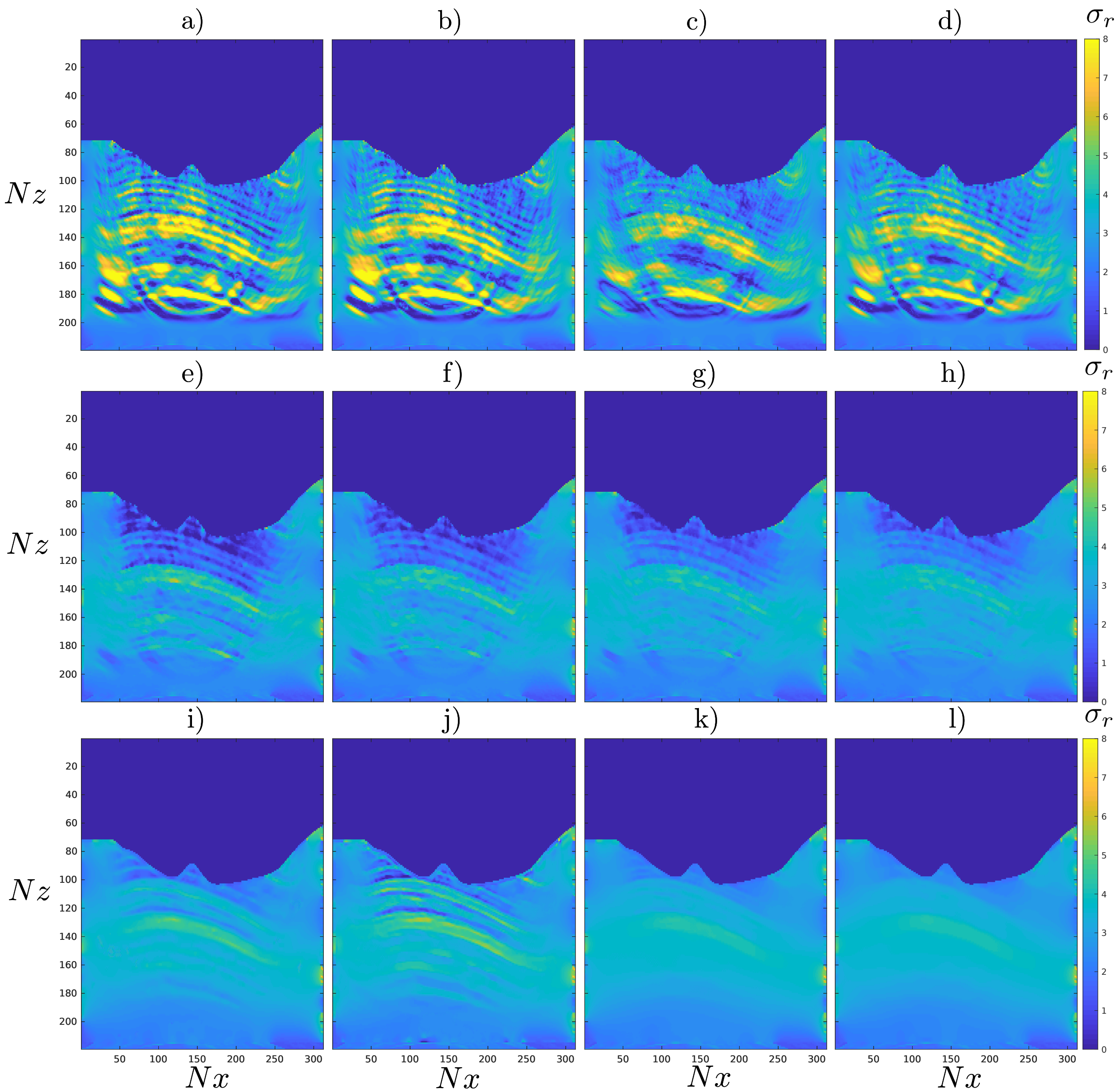

Figure 110. First row, FWI results for relative conductivity without regularizations: a) β=2.0, b) β=1.0, c) β=0.5, d) β=0.25. Second row, FWI results for relative conductivity with TV regularization: e) β=2.0, f) β=1.0, g) β=0.5, h) β=0.25. Third row, FWI results for relative conductivity with MTV regularization: i) β=2.0, j) β=1.0, k) β=0.5, l) β=0.25.

Not regularization	ε_r PSNR (dB)	σ_r PSNR (dB)
β=2.0	27.096103	19.269707
β=1.0	27.165063	19.151739
β=0.5	27.326458	20.925694
β=0.25	27.389384	20.913055

Table 13. *PSNR values reached for relative permittivity and realative conductivity without regularization changing the β parameter.*

TV Parameters	ε_r PSNR (dB)	σ_r PSNR (dB)
$\tau_{\sigma_r} = \tau_{\varepsilon_r} = 0.001$, $\gamma_{\varepsilon_r} = 0.01$, $\gamma_{\sigma_r} = 0.01$, β=2.0	27.882459	20.535970
$\tau_{\sigma_r} = \tau_{\varepsilon_r} = 0.001$, $\gamma_{\varepsilon_r} = 0.01$, $\gamma_{\sigma_r} = 0.01$, β=1.0	27.975319	20.963388
$\tau_{\sigma_r} = \tau_{\varepsilon_r} = 0.001$, $\gamma_{\varepsilon_r} = 0.01$, $\gamma_{\sigma_r} = 0.01$, β=0.5	28.090800	21.283203
$\tau_{\sigma_r} = \tau_{\varepsilon_r} = 0.001$, $\gamma_{\varepsilon_r} = 0.01$, $\gamma_{\sigma_r} = 0.01$, β=0.25	28.114301	21.604823

Table 14. *PSNR values reached for relative permittivity and realative conductivity with TV regularization changing the β parameter.*

MTV Parameters	ε_r PSNR (dB)	σ_r PSNR (dB)
$\mu_{\sigma_r} = \mu_{\varepsilon_r} = 200$, $\gamma_{\varepsilon_r} = 0.1$, $\gamma_{\sigma_r} = 1.0$, β=2.0	27.944586	22.616287
$\mu_{\sigma_r} = \mu_{\varepsilon_r} = 200$, $\gamma_{\varepsilon_r} = 0.1$, $\gamma_{\sigma_r} = 1.0$, β=1.0	28.101673	22.448459
$\mu_{\sigma_r} = \mu_{\varepsilon_r} = 200$, $\gamma_{\varepsilon_r} = 0.1$, $\gamma_{\sigma_r} = 1.0$, β=0.5	28.131485	22.513446
$\mu_{\sigma_r} = \mu_{\varepsilon_r} = 200$, $\gamma_{\varepsilon_r} = 0.1$, $\gamma_{\sigma_r} = 1.0$, β=0.25	28.248285	22.497362

Table 15. *PSNR values reached for relative permittivity and realative conductivity with MTV regularization changing the β parameter.*