Year : 2022

Quantitative interpretation approaches for GPR and seismic reflection data in complex environments

Liu Yu

Liu Yu, 2022, Quantitative interpretation approaches for GPR and seismic reflection data in complex environments

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_5488F9E20D846

UNIL | Université de Lausanne

Faculté des géosciences et de l'environnement
Institut des sciences de la Terre

Quantitative interpretation approaches for GPR and seismic reflection data in complex environments

Thèse de doctorat

Présentée à la
Faculté des géosciences et de l'environnementde
l'Université de Lausanne
par

Yu Liu

Master en ressources géologiques et ingénierie géologique
Université du Zhejiang, Chine

Jury

Prof. Dr. Klaus Holliger, Directeur de thèse
Dr. James Irving, Co-directeur de thèse
Prof. Dr. György Hetényi, Expert interne
Prof. Dr. Evert Slob, Expert externe

Lausanne, 2022

UNIL | Université de Lausanne
Faculté des géosciences et de l'environnement
bâtiment Géopolis bureau 4631

IMPRIMATUR

Vu le rapport présenté par le jury d'examen, composé de

Président de la séance publique :	M. le Professeur Pietro De Anna
Président du colloque :	M. le Professeur Pietro De Anna
Directeur de thèse :	M. le Professeur Klaus Holliger
Co-directeur de thèse :	M. le Docteur James Irving
Expert interne :	M. le Professeur Gyorgy Hetenyi
Expert externe :	M. le Professeur Evert Slob

Le Doyen de la Faculté des géosciences et de l'environnement autorise l'impression de la thèse de

Monsieur Yu LIU

Titulaire d'un
Master in Geological resources and Geological engineering
de Zhejiang University

intitulée

QUANTITATIVE INTERPRETATION APPROACHES FOR GPR AND SEISMIC REFLECTION DATA IN COMPLEX ENVIRONMENTS

Lausanne, le 6 décembre 2022

Pour le Doyen de la Faculté des géosciences et de l'environnement

Professeur Pietro De Anna

Contents

List of figures

List of tables

Résumé

Le sous-sol de la Terre présente une hétérogénéité à plusieurs échelles. Contraindre l'hétérogénéité est essentielle pour améliorer notre compréhension des conditions géologiques. Par rapport aux approches basées sur les sondages qui sont limitées à la dimension verticale, les mesures géophysiques basées sur la surface, notamment le géoradar (GPR) et la sismique, permettent de révéler efficacement les caractéristiques latéralement corrélées de la subsurface. Le GPR et la sismique sont basés sur la propagation des ondes électromagnétiques et sismiques, respectivement, qui ont de fortes analogies mathématiques. Les deux méthodes ont un bon potentiel pour contraindre l'hétérogénéité. Dans cette thèse, quatre nouvelles techniques d'interprétation quantitative sont proposées, dont trois pour explorer la vitesse des ondes radar ou la structure de corrélation du sous-sol peu profond en utilisant le GPR, et une pour caractériser les structures sub-verticales en milieu cristallin en utilisant la sismique. Pour le sous-sol peu profond, la connaissance détaillée des propriétés diélectriques, par exemple la vitesse des ondes radar, est particulièrement intéressante car elle fournit une image à haute résolution pour décrire l'hétérogénéité. Néanmoins, l'estimation de la vitesse à haute résolution par une mesure de réflexion GPR à décalage commun n'est actuellement pas possible sans informations complémentaires sur le forage. En considérant la distribution de la vitesse sous la surface comme la superposition d'un champ de vitesse de fond à variation lisse et d'un champ de fluctuation de la vitesse à petite échelle, la première étude de cette thèse développe une approche utilisant les composantes de diffraction pour déduire le champ de vitesse de fond et les composantes de réflexion pour inverser le champ de perturbation de la vitesse. Les résultats des tests de données synthétiques et de terrain montrent l'efficacité de cette méthode. Cette méthode se distingue par le fait qu'elle ne nécessite aucune information sur le calibration ou le conditionnement du trou de forage, ce qui est plutôt intéressant pour les levés de réflexion GPR à décalage commun. L'étape d'estimation du modèle de vitesse de fond est importante car un biais local dans le modèle déduit peut largement affecter le résultat. Pour améliorer les performances de l'analyse de la vitesse basée sur la diffraction, une deuxième étude présente une stratégie de pondération basée sur une mesure de focalisation de la semblance locale. La fonction de pondération conçue varie en fonction de la sensibilité d'une courbe de diffraction aux changements de la vitesse de migration. Des tests de données synthétiques et de terrain montrent que la méthode peut augmenter la résolution des spectres de semblance, ce qui réduit de manière correspondante les incertitudes dans l'analyse de la vitesse de diffraction. Une autre façon de caractériser les sous-sols peu profonds complexes avec des données GPR est d'estimer les propriétés géostatistiques liées à l'hétérogénéité. Pour estimer efficacement les paramètres géostatistiques, une troisième étude propose un schéma d'apprentissage supervisé. Le réseau neuronal convolutif est entraîné sur une vaste base de données d'images d'autocorrélation 2D obtenues de la modélisation basée sur la convolution de données GPR synthétiques pour une gamme compréhensive de modèles stochastiques de vitesse subsurface. Les résultats des tests de données synthétiques et de terrain confirment la viabilité de l'utilisation d'un réseau entraîné pour estimer le rapport d'aspect structurel de l'hétérogénéité de la subsurface. L'estimation est encore robuste en présence de niveaux de bruit élevés. Dans la dernière étude de cette thèse, les données sismiques de la zone d'Ivrea-Verbano sont prises comme exemple pour tester la capacité d'utiliser les champs d'ondes diffractées et une vue géostatistique pour interpréter les structures à fort pendage dans un environnement cristallin. Des données de réflexion sismique synthétique pour des modèles canoniques d'hétérogénéité crustale structurés de

manière sub-verticale sont employées pour corroborer la viabilité des techniques proposées. À partir des données sismiques de terrain, la continuité et l'angle de pendage dominant des structures sub-verticales dans la région sondée sont évalués. Ce travail fournit des informations intéressantes et potentiellement de nouvelles perspectives concernant les levés de sismique réflexion dans les terraines cristallins. Finalement, les avancées techniques de cette thèse ont le potentiel d'améliorer notre compréhension des environnements complexes de subsurface. Bien que les applications actuelles soient limitées aux cas 2D, les extensions des quatre méthodes proposées à la 3D sont conceptuellement directes.

Abstract

The Earth's subsurface exhibits heterogeneity at multiple scales. Constraining the heterogeneity is critical to enhance our understanding of the geological conditions. Compared to borehole-based approaches that are limited to the vertical dimension, surface-based geophysical measurements, especially ground penetrating radar (GPR) and seismics, allow to effectively reveal the laterally correlated features of subsurface. GPR and seismics are based on the propagation of electromagnetic and seismic waves, respectively, which have strong mathematical analogies. The two methods both have good potential to constrain heterogeneity. In this thesis, four novel quantitative interpretation techniques are proposed, with three for exploring the radar wave velocity or correlation structure of the shallow subsurface using GPR, and one for characterizing sub-vertical structures in crystalline environment using seismics. For the shallow subsurface, detailed knowledge of dielectric properties, e.g., radar wave velocity, is of particular interest because it provides a fine-scale image to describe the heterogeneity. Nonetheless, high-resolution velocity estimation through common-offset GPR reflection measurement is currently not possible without complementary borehole information. Regarding the subsurface velocity distribution as the superimposition of a smoothly varying background velocity field and a small-scale velocity fluctuation field, the first study of this thesis develops an approach utilizing diffraction components to infer the background velocity field and reflection components to invert the velocity perturbation field. The results from synthetic and field data tests show the effectiveness of this method. A distinguishing feature of this method is that it does not require any borehole calibration or conditioning information, which is rather attractive for common-offset GPR reflection surveys. The step of background velocity model estimation is important because a local bias in the inferred model can largely affect the result. To improve the performance of diffraction-based velocity analysis, a second study presents a weighting strategy based on a local semblance focusing measure. The designed weight function varies in accordance with the sensitivity of a diffraction curve to changes in migration velocity. Synthetic and field data tests show that the method can increase the resolution of the semblance spectra, which correspondingly reduces the uncertainties in diffraction velocity analysis. An alternative way to characterize complex shallow subsurface with GPR data is to estimate the geostatistical properties related to the heterogeneity. To estimate the geostatistical parameters efficiently, a third study proposes a supervised-learning scheme. The convolutional neural network is trained on a vast database of 2D autocorrelation images obtained from convolution-based modeling of synthetic GPR data for a comprehensive range of stochastic subsurface velocity models. Results from synthetic and field data tests confirm the viability of using a trained network to estimate the structural aspect ratio of the subsurface heterogeneity. The estimation is still robust in the presence of high noise levels. In the last study of this thesis, the seismic data from Ivrea-Verbano Zone are taken as an example to test the ability of using diffracted wavefields and a geostatistical view to interpret the steeply dipping structures in crystalline environment. Synthetic seismic reflection data for sub-vertically structured canonical models of crustal heterogeneity are employed to corroborate the viability of the proposed techniques. From the field seismic data, the continuity and dominant dip angle of sub-vertical structures in the probed region are assessed. This work provides interesting insights and potentially new perspectives regarding the seismic reflection surveys in crystalline terranes. Ultimately, the technical advances in this thesis have the potential to improve our understanding of complex subsurface environments. Though the current applications are limited to 2D cases, the extensions of the four proposed methods to 3D

are conceptually straightforward.

Chapter 1

Introduction

1.1 Motivation: heterogeneous subsurface environments

The Earth's subsurface has heterogeneity on many scales. There is a wide variation of rock types, and the variations in rock composition can range on scales of a few mm to many km in the subsurface. Properties of minerals that make up the bulk of rocks in the subsurface vary significantly (Simmons and Wang, 1971). Except for mineralogy, fractures that range in size from submicroscopic to many tens of meters also influence the properties of a rock (Simmons and Nur, 1968). The mass of rocks erupted from volcanoes provides geochemical and geological evidence of heterogeneity within the earth. Tectonic processes such as folding, faulting, and large-scale crustal movements associated with plate tectonics contribute to making the lithosphere heterogeneous. The heterogeneity is also documented by borehole measurements of sonic velocities, neutron porosity, resistivity and many other physical properties (e.g., Holliger, 1996; Barrash and Clemo, 2002; Jiang et al., 2020).

Characterization of the underground heterogeneity can largely enhance our understanding of the geological conditions. For example, knowledge of the spatial correlation of hydraulic properties in the shallow subsurface allows for realistic simulation of groundwater flow and contaminant transport, and thus for sustainable management and effective remediation of groundwater resources. In petroleum exploration, the sediment cover is probed, and potential good reservoirs can be identified if the heterogeneity of subsurface structures is constrained. From a larger point of view, investigating the spatial heterogeneity of the lithosphere helps to reveal the mechanism by which the earth’s crust is formed, volcanic processes, and the nature of active seismic zones.

The most direct means to observe and characterize the subsurface heterogeneity is borehole drilling. Rocks sampled directly from boreholes show variation in mineral composition and changes in physical properties with depth, which illustrates the heterogeneity in the vertical direction. However, boreholes provide only sparse and 1D measurements. Lateral variability of lithologic, sedimentary, and hydrologic units, a common and important characteristic of the subsurface, cannot be sufficiently mapped with direct sampling approaches. To study the lateral subterranean features, geophysical techniques which can remotely sense changes or contrasts in the structure and physical properties play a key role. Typical geophysical methods include seismics, electrical resistivity tomography, electromagnetics, ground penetrating radar (GPR) and potential methods. Among them, GPR and seismics exploit the propagation of electromagnetic waves and elastic waves in the subsurface, respectively, and are particularly useful to constrain heterogeneity because of the high resolution

of their results. Except for imaging the geometry of underground structures, surface-based GPR and seismic reflection surveys also allow predicting the velocity field distribution (e.g., Virieux and Operto, 2009; Schmelzbach et al., 2012) and recovering the correlation structure of the heterogeneity (e.g., Irving and Holliger, 2010; Scholer et al., 2010; Xu et al., 2020). Besides, some small discontinuities will generate scattered and/or diffracted wavefield components that can be used separately with some special data processing techniques to better characterize the heterogeneous subsurface model (e.g., Fomel et al., 2007; Schwarz and Gajewski, 2017; Yuan et al., 2019). Nonetheless, the quantitative interpretation for some complex environments is still challengeable. One example is the shallow subsurface, where the reliable estimation of physical properties through common-offset GPR reflection measurement is currently not possible without complementary borehole information. An alternative way to characterize the shallow subsurface heterogeneity is to estimate the correlation structure, yet the computational cost of current geostatistical inversion technique is inherently high. Another example of challenges is the crystalline environment, where the continuity and dominant dip angle of geological structures are all difficult to quantify thorough seismic reflection surveys.

The general aim of this thesis is therefore to develop novel quantitative interpretation approaches for overcoming the aforementioned challenges with GPR and seismic reflection data. For that, the methodological background of GPR and seismics, as well as the analogy and differences of two wave-based methods, will first be introduced. Then, the current ways to constrain heterogeneity with GPR and seismic reflection data will be reviewed. Finally, the specific objectives and outline of this thesis will be presented.

1.2 Methodological background of GPR and seismics

1.2.1 Basic principles of GPR

The GPR technique employs electromagnetic (EM) waves, usually in the 10 MHz - 1 GHz range to probe the subsurface (Annan, 2005). GPR works by emitting a pulse into the ground and recording the responses that result from subsurface objects. The non-invasive way of data collection together with the broad EM wave frequency spectrum leads to a wide range of GPR applications. The use of GPR method requires understanding of the EM wave propagation through the subsurface, as summarized in the following.

The physics of EM fields is mathematically described by the well-known Maxwell's equations, while the material properties are quantified by constitutive equations. Combining the Maxwell's equations with constitutive equations provides the foundation for quantitatively describing GPR signals. The basic wave equations describing electric field vector $\mathbf{E}$ can be written as:

$$\nabla^2\mathbf{E} = \mu\varepsilon\frac{\partial^2\mathbf{E}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{E}}{\partial t}, \tag{1.1}$$

where μ denotes magnetic permeability, σ denotes the electrical conductivity and ε denotes the dielectric permittivity. GPR is effective in low-loss materials where energy dissipation (second term in above equation) is small compared to energy storage (third term in above equation). The solutions for a monochromatic, linearly polarized wave that propagates through a homogeneous media can be written as:

$$f(\beta, t) \approx f(\beta \pm vt)e^{\mp\alpha\beta}, \tag{1.2}$$

where β denotes the distance in the propagation direction, t denotes propagation time, v and α are EM wave velocity and attenuation, respectively. Except for velocity v and attenuation α, EM impedance Z is

another key wave field property. For high enough frequencies and low enough conductivities of the medium, the low-loss condition can be assumed. In this case, the three wave field properties can be expressed as follows:

$$v \approx \frac{1}{\sqrt{\varepsilon\mu}} \approx \frac{c}{\sqrt{\varepsilon_r}}, \tag{1.4}$$

$$\alpha \approx \frac{\sigma}{2}\sqrt{\frac{\mu}{\varepsilon}} \approx \frac{\sigma}{2\sqrt{\varepsilon_r}} Z_0, \tag{1.5}$$

$$Z \approx \sqrt{\frac{\mu}{\varepsilon}} \approx \frac{Z_0}{\sqrt{\varepsilon_r}}, \tag{1.6}$$

where ε_r denotes the relative permittivity, $Z_0 = 377$ Ω denotes the impedance of free space, and the right most expression in each equation is valid when magnetic property variations are assumed negligible, making $\mu = \mu_0$, where $\mu_0 = 1.25 \times 10^{-6}$ H/m is the free-space magnetic permeability. The above equations show that ε_r is the most important factor that will determine the velocity and EM impedance, and σ mainly contributes to the EM wave attenuation. To put the wave properties in perspective, typical values of velocity v are in the range between 0.07 and 0.15 m/ns. Typical values of attenuation α are 1 dB/m with high losses of 10-100 dB/m and very low losses being 0.01-0.1 dB/m. Typical EM impedance values are 100-150 Ω.

GPR reflection surveys using a single transmitter and a single receiver with a fixed spacing at each measurement location are the most common (Figure 1.1). EM wave reflection occurs when the wave encounters an impedance boundary. The ratio between the amplitudes of reflected and incident waves are given by the reflection coefficient R. For the special case of a normal incident plane wave in Transverse Electric (TE) mode (electric field vector in the interface plane) on a planar boundary between two layers n and $n+1$, the corresponding reflection coefficient R_n can be written as:

$$\mathrm{R}_n = \frac{\mathrm{Z}_{n+1} - \mathrm{Z}_n}{\mathrm{Z}_{n+1} + \mathrm{Z}_n}. \tag{1.7}$$

Note that reflected amplitudes can be positive or negative, depending on whether EM impedance decreases or increases across an interface. The amplitudes recorded in GPR reflection surveys can be utilized to infer the information of subsurface wave field properties. Variations in reflection amplitude and time delay indicate variations in velocity, attenuation and EM impedance.

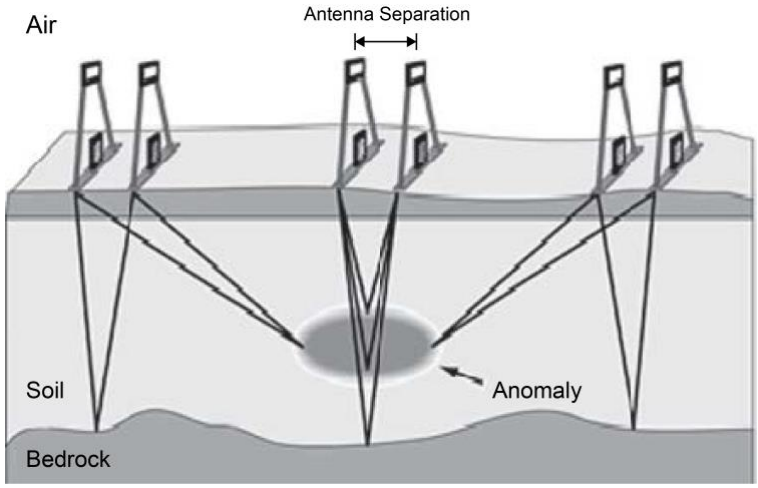


Figure 1.1: Schematic illustration of common-offset, single-fold GPR profiling along a line. Adapted from Jol (2008).

The parameters defining a common-offset GPR reflection survey involve dominant frequency of the GPR antenna, recording time window length, time sampling interval, spacing along the line, antenna separation

distance, and antenna orientation with respect to the profile direction. The resolution is determined by the dominant wavelength of the GPR pulse, which is a function of the dominant frequency and EM wave velocity. Usually, the resolution of GPR data can be increased by using higher frequency antenna. At the same time, the penetration depth reduces because the EM waves with higher frequency attenuate more strongly.

1.2.2 Basic principles of seismics

Seismic exploration is the use of seismic energy to probe the structures and physical properties of the Earth's subsurface. The corresponding surveys are conducted by deploying an array of energy sources and an array of receivers. Seismic waves include body waves which travel through the interior of the Earth, and surface waves which travel along the Earth's surface. Body waves further include P-waves and S-waves. The P-wave is a longitudinal wave, the force applied in the direction that the P-wave is propagating. For the S-wave, the medium is displaced in a transverse way, and the medium must move away from the material right next to it to generate the shear and transmit the wave.

A complete solution of the seismic wave problems requires the use of the wave equation. The equation of motion for a continuum can be written as

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = \nabla \mathbf{T} + \mathbf{F}, \tag{1.8}$$

where ρ is the density of medium, $\boldsymbol{u}$ is the displacement, $\mathbf{T}$ is the stress tensor and $\mathbf{F}$ is the body force density. In isotropic homogeneous linearly elastic medium, the equation of motion for seismic waves outside the source region takes the following form

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = (\lambda + 2\mu_L)\nabla(\nabla \cdot \boldsymbol{u}) - \mu_L \nabla \times (\nabla \times \boldsymbol{u}), \tag{1.9}$$

where λ and μ_L are the Lamé's constants of medium. The above equation provides the basis for most synthetic seismogram calculations of body waves. The displacement field $\boldsymbol{u}$ can be decomposed into a rotation-free part $\boldsymbol{u}_P = \nabla \cdot \boldsymbol{u}$, and a divergence-free part $\boldsymbol{u}_S = \nabla \times \boldsymbol{u}$. Since the divergence of a curl and the rotation of a divergence are zero, we can get two independent solutions for Equation (1.9) when forming its scalar product and vector product, respectively:

$$\frac{\partial^2 (\nabla \cdot \boldsymbol{u})}{\partial t^2} = \frac{(\lambda + 2\mu_L)}{\rho} \nabla^2 (\nabla \cdot \boldsymbol{u}), \tag{1.10}$$

and

$$\frac{\partial^2 (\nabla \times \boldsymbol{u})}{\partial t^2} = \frac{\mu_L}{\rho} \nabla^2 (\nabla \times \boldsymbol{u}). \tag{1.11}$$

The above two equations are solutions of the wave equation for the propagation of P-waves and S-waves, respectively. Their corresponding wave velocities are:

$$V_P = \sqrt{\frac{(\lambda + 2\mu_L)}{\rho}} \tag{1.12}$$

and

$$V_S = \sqrt{\frac{\mu_L}{\rho}}. \tag{1.13}$$

The above equations mathematically illustrate that P-waves travel always faster than S-waves for the same material. Fluids have no shear strength ($\mu_L = 0$) and thus do not propagate S-waves. Velocity is one of the

most important seismic wave properties. Typical P-wave velocities are between 500 m/s and 2500 m/s in unconsolidated sediments, and between 3000 m/s and 6500 m/s in solid crustal rocks. Of the common rocks of the crust, velocities are greatest in basalt and granite. Typical S-wave velocities are between 100 m/s and 800 m/s in soft sediments, and between 1500 m/s and 3800 m/s in solid rocks (Earle, 2015).

Seismic reflection surveys using multiple receivers at different locations for sensing returning P-waves generated by each source points are the most common. P-wave reflection occurs when the wave encounters an impedance boundary. It is noted that, the acoustic impedance here is defined as the product of density ρ and P-wave velocity V_P. For normal and low angles of incidence, sthe reflection coefficient of a boundary has a similar relationship with acoustic impedances of two adjacent layers with that depicted in Equation (1.7).

If a seismic wave impinges upon a surface with an edge to it, such as a faulted layer, the wavefront bends around the end of the feature and gives rise to a diffracted wave (Figure 1.2). Similarly, boulders whose scales are of the same order as the wavelength of the incident signal can generate diffractions. The curvature of the diffraction event is a function of the subsurface velocity. While diffractions are usually considered as noise and attempts are made at resolving them through data processing, they can be used as an interpretational aid.

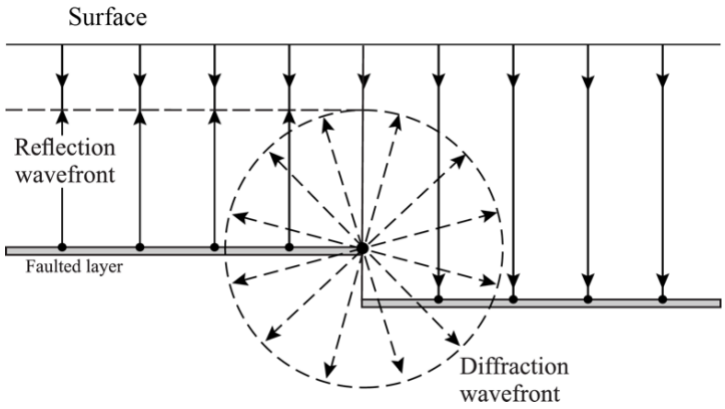


Figure 1.2: Ray schemes for reflected and diffracted waves. While reflected wavefronts obey Snell's Law, diffracted wavefronts are scattered radially when encountering the truncated end of the faulted layer. Adapted from Preine et al. (2020).

1.2.3 Analogy between GPR and seismics

In the past few decades, GPR has become one of the major tools to probe the shallow subsurface. The acquisition principle is very similar to reflection seismics, and therefore, a lot of seismic data processing and interpretation methods are reusable for GPR data, and vice versa. GPR and seismics are both wave-based methods. Comparing EM wave described by Equation (1.1) with elastic wave described by Equations (1.10) and (1.11), the forms are similar assuming the equivalence between parameters described in Table (1.1).

Table 1.1: Equivalence between EM and elastic wave parameters.

EM parameter	Elastic parameter
$1/\varepsilon$	$\begin{cases} \lambda + 2\mu_L & \text{(P wave)} \\ \mu_L & \text{(S wave)} \end{cases}$
μ	ρ

It has been demonstrated that the 2D Maxwell's equations describing the propagation of the transverse magnetic (TM) mode in anisotropic media are mathematically equivalent to the shear-horizontal (SH) wave equation in an anisotropic-viscoelastic solid (Carcione and Cavallini, 1995). Later, Carcione and Robinson (2002) investigated the corresponding mathematical analogy between elastic waves and electromagnetic waves. They obtained a complete parallelism for the reflection and refraction problem, considering the most general situation, i.e., a situation with the presence of anisotropy and attenuation. In the ideal isotropic and lossless case, the reflection coefficients of the TE wave and acoustic wave (or TM wave and SH wave) are identical (or opposite) along with the incidence angle (Laurain and Lecomte, 2001).

The mathematical equivalences allow us, for example, to use a finite-difference modelling code developed for seismic wave propagation to model EM wave propagation (e.g., Ernst et al., 2007; Belina et al., 2012). The analogies between EM and seismic wave propagation also implicate that GPR and seismic reflection surveys have plenty of similarities. Both are based upon the reflected waves from buried objects and have a good ability to depict the lateral continuity of subsurface structures. They both need to choose a suitable frequency range before data acquisition to balance the resolution and penetration depth. Standard seismic data processing steps such as amplitude scaling, bandpass frequency filter, deconvolution, velocity analysis and migration are also applicable to GPR data. The same conceptual model can be used for the two types of data. For example, both the migrated GPR and seismic data sections can be considered as the convolution product of the subsurface reflectivity field and source wavelet.

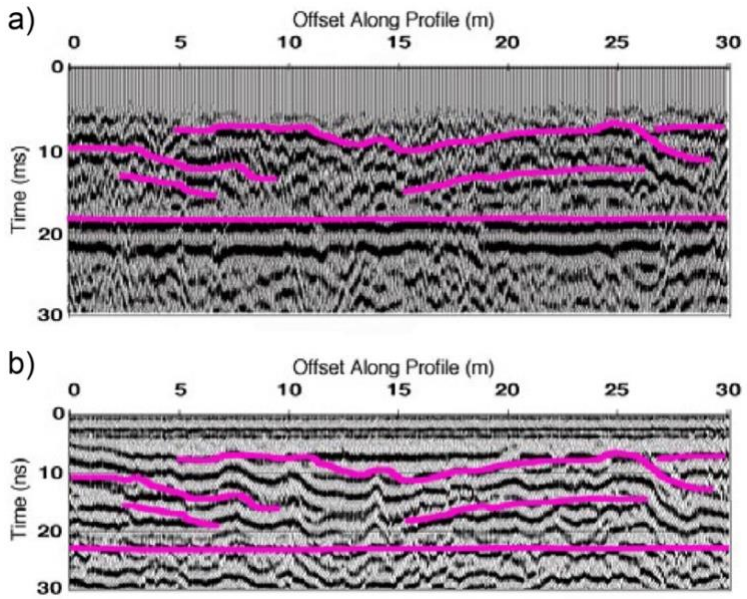


Figure 1.3: Example of using seismic and GPR reflection surveys to complement each other. a) Interpreted seismic reflection section and b) interpreted common-offset GPR reflection section. Both methods recorded strong reflections from the same interpreted interfaces, but the GPR record is more detailed. Adapted from Baker et al. (2001).

GPR and seismic reflection surveys also have differences in many aspects. As discussed before, several modes are present in seismic wave propagation: P-wave and S-wave, and surface wave. Each of these modes will have a distinct propagation velocity. While EM waves in TM and TE mode share the same velocity. Since the wavelength of EM wave is smaller than that of seismic wave, GPR usually offers greater resolution and shallower penetration depth than seismics. The two methods are sensitive to different physical properties. GPR uses radio frequency electromagnetic waves as an energy source, which are sensitive to the dielectric properties

and reflective objects within it. While seismic waves are sensitive to the elastic properties like seismic velocity and bulk density. They can be used complementarily to identify an interface across which both EM and elastic parameters vary. An example of combination is shown in Figure 1.3, where GPR and seismic surveys are used to image several reflectors within the same volume of alluvium at depths within 3 m. Characterization of the vadose zone stratigraphy by merging qualitatively similar results from GPR and seismic surveys allowed the two techniques to reinforce each other.

1.3 Methods to constrain heterogeneity with GPR and seismics

1.3.1 Primary reflection imaging and inversion

Imaging the reflectors in the subsurface can provide a basis for locating the discontinuities and characterizing the heterogeneity. Migration is the process that moves dipping reflections to their true subsurface positions and collapses diffractions to their scattering points based on the recorded reflection wavefields. This technique has been studied for a long time in seismic community. The one-way-in-depth scalar wave equation is the basis for common migration algorithms. These algorithms do not explicitly model multiple reflections, converted waves, surface waves, or noise. They require primary reflection data as input, i.e., wavefields with a single bounce in the subsurface and thus linear in reflectivity. This indicates that the multiples need to be removed or the data is dominated by single scattering. The corresponding migration algorithms can be classified under three main categories: those that are based on integral solution to the scalar wave equation, those that are based on the finite-difference solution, and those that are based on frequency-wavenumber implementations. Except for the migration algorithms, choice of an appropriate migration strategy is also necessary. The strategies include post-stack versus pre-stack migration, and time versus depth migration. Post-stack migration assumes that the stacked section is equivalent to a zero-offset section. This assumption is not valid to handle the case of conflicting dips, which needs pre-stack time migration. When the lateral variation of velocity is modest, the time migration methods perform adequately. When the lateral velocity variation is severe, more computationally intensive depth migration is required. Regardless of the migration algorithms and strategies, the key prerequisite is velocity. For time migration, velocities can be estimated by semblance velocity analysis of common-mid-point gathers. For depth migration, a more accurate velocity model is necessary and can be obtained by tomographic methods. The common-reflection-surface stack method is an alternative which does not depend on the velocity model. It provides a zero-offset simulation of primary reflections from seismic multi-coverage reflection data (e.g., Jäger et al., 2001).

The vast majority of surface-based GPR reflection surveys are performed using the common-offset approach, where a fixed transmitter–receiver antenna pair is incrementally moved along the measurement profile (e.g., Annan 2005). It has been shown that the difference between common-offset and zero-offset profiles beneath air and ground waves is often small enough to be ignored (Neal, 2004). Therefore, the common-offset GPR data are often considered analogous to the zero-offset or post-stack seismic section, and common post-stack time migration methods are applicable to them (e.g., Fisher et al., 1996). The velocity used for migration is usually obtained by diffraction hyperbola fitting. For complex and heterogeneous environments, multi-offset GPR surveys are more useful to build a high-resolution velocity model and perform accurate imaging (e.g., Bradford, 2006), though the labor cost is much higher. Specific GPR-based methods have also been developed

to overcome some of the limitations in the seismic routines. Examples include matched filter migration (Leuschen and Plumb, 2000), topographic migration (Lehman and Green, 2000; Allroggen et al., 2015), vector migration (Streich et al., 2007), frequency domain migration for lossy soils (Di and Wang, 2004; Oden et al., 2007), and polarimetric migration (Fent et al., 2015).

The primary reflection data can be used not only for obtaining a geometric description of the main structures but also for estimating the physical properties of subsurface. In seismic community, impedance inversion is a popular tool to estimate the acoustic impedance values, lithologies or fluids for purposes like reservoir characterization. Considering the seismic data as the convolution product of the subsurface reflectivity field and source wavelet, the goal of inversion is to find an impedance model that results in minimization of the misfit between the predicted traces and observed traces. There are several problems with inverting seismic reflection data to impedance, such as limited data bandwidth, noise in the data, and imperfect model parameterization. Additional information, usually from the borehole logs, are necessary to provide the low-frequency components or starting model for inversion. The impedance inversion can be posed as a deterministic problem or as a stochastic problem. Regarding deterministic inversion, there are mainly two categories: sparse-spike methods and model-based methods (Russel and Hampson, 1991). Sparse-spike techniques deconvolve the seismic trace under sparseness assumptions of the reflectivity series, and compute the impedance based on a recursive relationship, after which the missing low frequencies are integrated. Model-based techniques require a good initial impedance model and perturb the model until some minimization criteria are satisfied. For deterministic inversion, efforts have been made to exploit the prior information by using different regularization strategies (e.g., Gholami, 2015; Hamid and Pidlisecky, 2015). Regarding stochastic inversion, the seismic data constrain the inversion within the seismic bandwidth, while the higher spatial frequencies are stochastically constrained by the variograms obtained from well logs and the hard data at the wells. A review of seismic stochastic inversion methods can be found in Bosch et al. (2010).

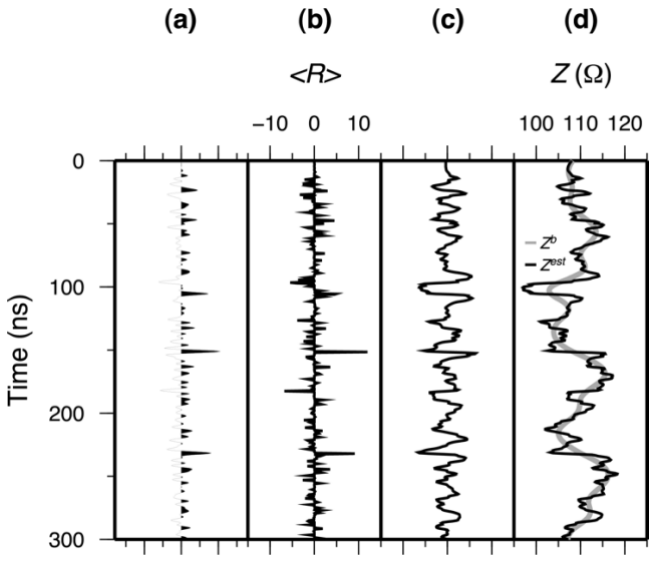


Figure 1.4: Trace example of GPR impedance inversion. a) Data trace extracted from a processed synthetic GPR reflection image; b) Bandlimited reflectivity computed from the trace; c) Band-limited impedance computed from the reflectivity; d) Low-frequency impedance background (gray line) and estimated full-bandwidth impedance (black line). Adapted from Schmelzbach et al. (2012).

In GPR community, the EM impedance inversion and its adapted velocity inversion has been studied in the past decade. Schmelzbach et al. (2012) present an impedance inversion approach for common-offset GPR data

that is based on a convolution model for the GPR traces, where borehole dielectric permittivity or porosity logs are used to recover the low-frequency part of the velocity structure that cannot be obtained from the reflection data (Figure 1.4). Zeng et al. (2015) and Liu et al. (2018) adopt similar approaches to estimate the distribution of soil water content and to characterize buried archaeological remains, respectively. Xu et al. (2021) also assume a convolution model for the GPR traces but combine stochastic simulation with simulated annealing optimization to generate velocity realizations that honor the GPR measurements and borehole porosity log data along the profile. Forte et al. (2013; 2014) assume a locally 1D layered subsurface structure and use picked reflection amplitudes to recursively estimate the GPR velocity in a series of identified subsurface layers, in which the velocity is assumed constant. Overall, the aforementioned methods mostly rely on complementary borehole information, and it is currently not possible to estimate high-resolution velocity structure only through common-offset GPR reflection measurement.

1.3.2 Full-wavefield imaging and inversion

Unlike common migration algorithms that are based on the one-way scalar wave equation, reverse time migration (RTM) is a technique that uses two-way wave equation and has the potential to achieve the best accuracy and imposes no limit on the variations in velocity, reflector dip and wavefield type. The strategy of RTM is to seek the reflection image as the best match based on an imaging condition between two fields mapped from two spaces. One mapped field is the extrapolation of time-reversed data waveforms from the data space, and the other mapped field is the prediction from the model space based on the reference velocity model and source parameters. Historically, RTM methods also need to go through pre-processing to attenuate those parts of the data that are not primary reflections (e.g., McMechan, 1982). Then the benefits of using full wavefields were realized (e.g., Levin, 1984). Now the RTM algorithms are able to use all computable wavefield types including primary reflections, converted waves, head waves, turning waves, multiple reflections and surface waves (Zhou et al., 2018). Considering multiple reflections as signal rather than noise provides two main benefits: First, this helps reduce the artifacts from mistaking multiples as primary reflections; and second, this expands the solution space because multiple reflections cover a wider range of space than that of primary reflections. At the same time, the RTM methods using full wavefields have a high computational cost and the successful imaging relies on an accurate velocity model.

Actually, migration is also a complement to velocity model building because the latter can use reflectivity images to establish the velocity interfaces. Lailly (1983) points out that computing the gradient in full-waveform inversion (FWI) is equivalent to applying the pre-stack RTM operator on data residuals. FWI, originally introduced to seismic exploration community by Lailly (1983) and Tarantola (1984), aims at the least-squares minimization of the misfit between observed and synthetic data based on full-wavefield forward modeling. It is shown that the gradient of the misfit function along which the perturbation model is searched can be built by cross-correlating the incident wavefield emitted from the source and the back-propagated residual wavefields. The perturbation model obtained after the first iteration of the local optimization looks like a migrated image obtained by RTM. One difference is that the seismic wavefields recorded at the receivers are back propagated in RTM, whereas the data misfit is back propagated in FWI (Virieux and Operto, 2009). When the velocity perturbations are added to the initial velocities, an updated velocity model is created as a starting model for the next iteration of minimizing the misfit function. Early applications of FWI suffered from high computational cost and the difficulties associated with short-offset data. In the 1990s, FWI was pushed forward with developments in frequency-domain formulation, which allowed to acknowledge the importance of long-offsets and transmission data to reconstruct large-scale structures and alleviated the numerical cost

(e.g., Pratt, 1999). Now the research interest in FWI continues, and efforts are put into aspects such as honoring as much as possible the physics of wave propagation (e.g., Trinh et al., 2019), alleviating the inherent ill-posed problem of FWI by using alternative misfit-functions (e.g., Bozdağ et al., 2011), and uncertainty quantification (e.g., Zhu et al., 2016).

The RTM and FWI algorithms have also been extended to GPR community. Applying post-stack RTM methods on common-offset GPR data is conceptually straightforward. At first, post-stack RTM based on the scalar wave equation was utilized (e.g., Fisher et al., 1992). Now most RTM applications solve the Maxwell's equations, which can account for complex velocity and conductivity distributions, as well as topography variations (e.g., Bradford et al., 2018). Figure 1.5 shows an example of using RTM to characterize geological structures based on GPR records. By contrast, it is more challenging to apply FWI on GPR reflection data. Though the FWI of cross-hole GPR data is quite successful (e.g., Ernst et al., 2007; Belina et al., 2012), the application of FWI is rather difficult for surface-based surveys where multi-offset GPR acquisition is usually needed (e.g., Busch et al. 2012; Lavoué et al. 2014). The success so far has been limited due to the rather narrow range of reflection angles and antenna radiation patterns that are highly complex, largely unknown, and site dependent.

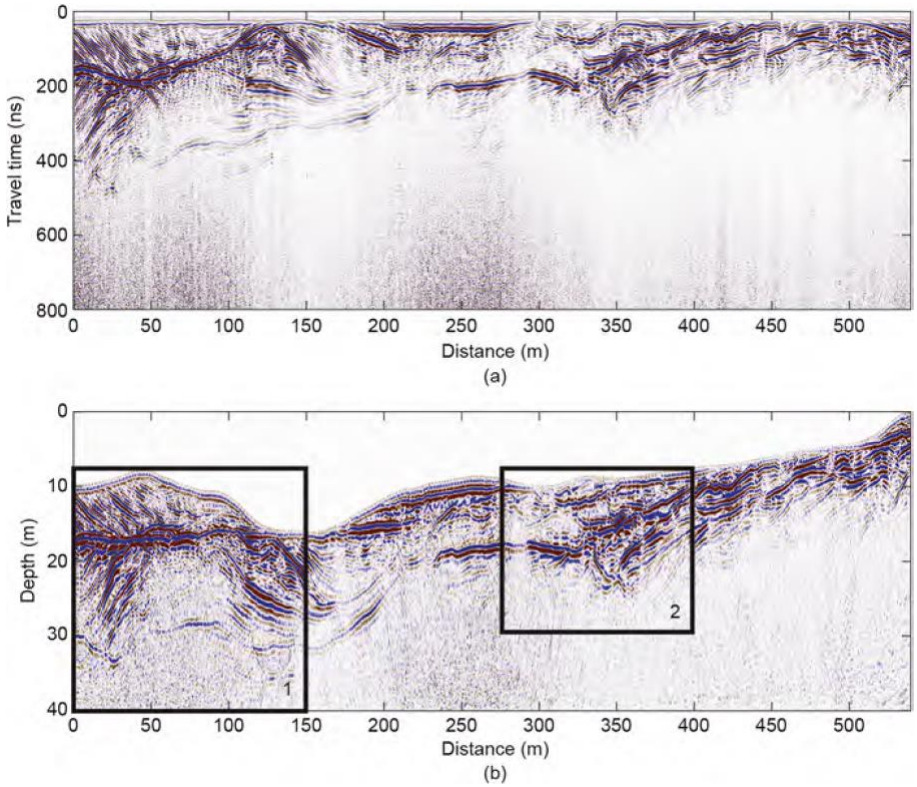


Figure 1.5: An example GPR profile showing a) the preprocessed data prior to migration, and b) the data after RTM from topography. Region 1 is a zone of steeply dipping bedforms within the bedrock stratigraphy. Region 2 contains a feature that is interpreted as a normal fault within the bedrock (Bradford et al., 2018).

1.3.3 Diffraction imaging

When characterizing the subsurface heterogeneities with seismic reflection survey, it is usually important to identify small features, e.g., faults, fractures, channels, and rough edges of crystalline rocks or small changes in seismic reflectivity such as those caused by fluid presence or fluid flow during reservoir production. Those small features can be considered as diffracting structures. The corresponding diffracted waves have the unique

property that they do not honor Snell's law and radiate uniformly, resulting in far superior illumination compared with reflections and favorable potential for velocity analysis (Klem-Musatov, 1994). Due to their generally low amplitudes, diffracted events are often barely recognizable in seismic data and mostly masked by dominant specular reflections. Therefore, diffracted wavefields need to be isolated from the full wavefields for imaging and characterization, and the process is called diffraction imaging (Figure 1.6). The study of diffraction imaging involves three aspects: 1) diffraction separation; 2) diffraction velocity analysis; 3) imaging of diffractions.

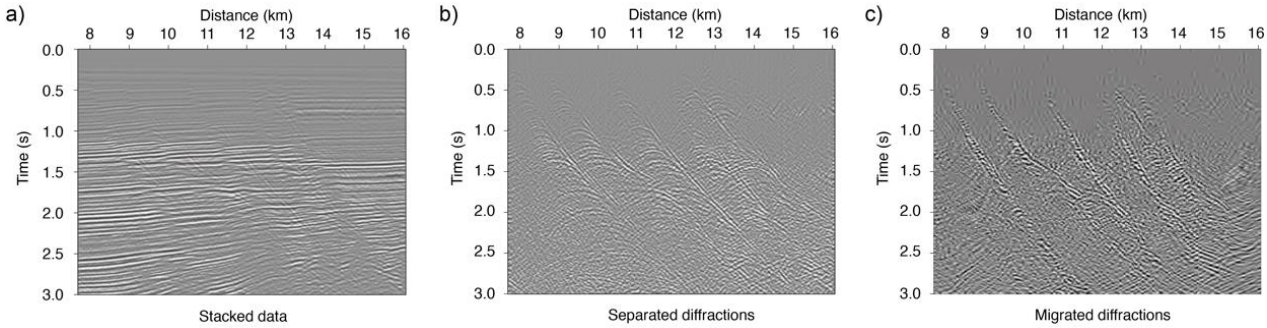


Figure 1.6: An example of seismic diffraction imaging. a) Stacked section from a Gulf of Mexico data set and its corresponding b) separated diffractions and c) migrated diffractions. Adapted from Fomel et al. (2007).

Regarding diffraction separation, the methods mainly contain plane-wave destruction, filters based on wave-front attributes, and identification in post-migration dip-angle domain. Plane-wave destruction removes the locally planar reflection events corresponding to slowly varying slopes, thereby isolating the diffracted energy (Fomel, 2002). It is a popular tool for post-stack and common-offset seismic reflection data (e.g., Fomel et al., 2007). Filters based on wave-front attributes also aim to remove the dominant reflection events, but in a kinematic way. For pre-stack seismic data, the filters are usually based on reflection focusing attributes (e.g., Khaidukov et al., 2004; Moser and Howard, 2008) and common-reflection-surface attributes (Dell and Gajewski, 2011). For post-stack or common-offset data, coherent subtraction filter is used (Schwarz et al., 2017). Identification of diffractions from reflections in post-migration dip-angle domain is easier than that in the time-space domain. In dip-angle common image gathers, specular reflections appear as hyperbolic events centered at the reflector dip and curving upward, even when over- or under-migrated. Conversely, diffractions will be flat in this domain when using the correct migration velocity, or curve upward or downward in the case of over- and under-migration, respectively (Reshef and Landa, 2009). Based on this feature, Klokov and Fomel (2012) proposed to use hybrid Radon transform to extract diffractions in this domain.

Regarding diffraction velocity analysis, there are generally three approaches available. The first one involves migration focusing analysis of diffraction events. Since a diffraction migrated with the correct velocity will collapse to a point at its apex, Harlan et al. (1984) proposed to remove reflections from the data and to estimate the velocity structure through an evaluation of diffraction focusing as a function of different migration velocities. In this regard, Fomel et al. (2007) proposed a constructive workflow which evaluates diffraction focusing based on velocity continuation and local kurtosis techniques. The second approach is to examine diffraction velocity in the post-migration dip-angle domain. As introduced, diffractions will have different shapes in this domain after using different migration velocities, based on which the correct velocity can be estimated (e.g., Reshef and Landa, 2009). Finally, the third approach for diffraction-based velocity analysis is to perform diffraction wavefront tomography. Bauer et al. (2017) performed ray-based tomographic inversion for the diffraction section.

Regarding imaging of diffractions, it is usually a simple migration operation once the diffractions have been isolated and velocity information is available (e.g., Fomel et al., 2007; Schwarz et al., 2017). This process focuses the diffracted energy at the subsurface locations from which the diffractions originated. However, there are also some methods that can perform the imaging of diffractions directly without the separation or velocity analysis, including diffraction semblance attribute (e.g., Schmelzbach et al., 2008), adapted imaging condition in migration (e.g., Yin and Nakata, 2019), and path summation of diffractions (e.g., Merzlikin and Fomel, 2017).

The shallow subsurface environments are rather complex and heterogeneous, and it is common to see diffraction components in GPR data (e.g., Grasmueck et al., 2005). Most seismic diffraction imaging techniques can be directly applied on GPR reflection data, but the velocity information carried by diffractions is particularly useful when no offset-dependent information is available, as is notably the case for typical common-offset GPR reflection surveys. Figure 1.7 shows an example of diffraction-based velocity analysis of GPR data. Arguably, migration focusing analysis is the most commonly used method of diffraction-based velocity analysis for common-offset GPR data (e.g., Novais et al., 2008; Yuan et al., 2019). However, this type of method has inherent uncertainties in the velocity picking process (Fomel et al., 2007). Also, the resolution of diffraction-based velocity model is usually not as good as its reflection-based counterpart for quantitative interpretation of the subsurface.

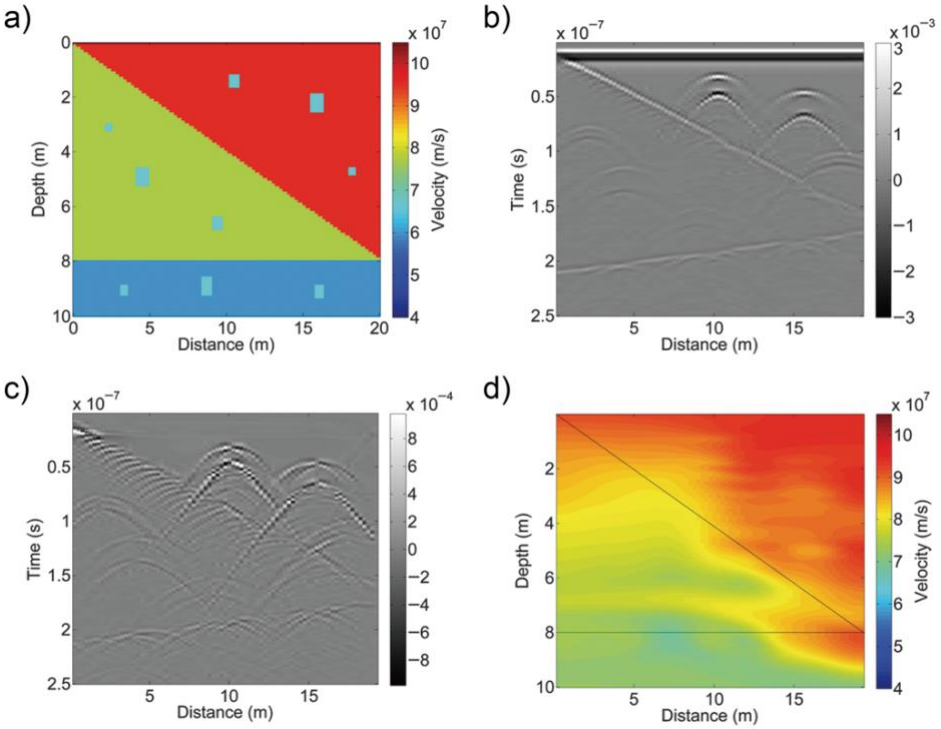


Figure 1.7: An example of diffraction-based velocity analysis of common-offset GPR data. a) The true velocity model and its corresponding b) synthetic common-offset GPR reflection image; c) The separated diffractions and the d) estimated velocity model. Adapted from Yuan et al. (2019).

1.3.4 Geostatistical inversion

The aforementioned methods using various wavefields all aim at probe the detailed structure or physical

properties of subsurface. To characterize strongly heterogeneous subsurface environments, geostatistical methods are required where we cannot possibly recover enough detail about the heterogeneity from the other methods. This means to estimate parameters describing the geostatistical nature of the heterogeneity. Specifically, the second-order statistical properties of seismic and/or GPR reflection data are expected to provide information about the correlation structure of the subsurface velocity field through which the waves have traveled (Figure 1.8). In that context, the lateral correlation properties are of particular interest because they cannot be obtained from borehole logs.

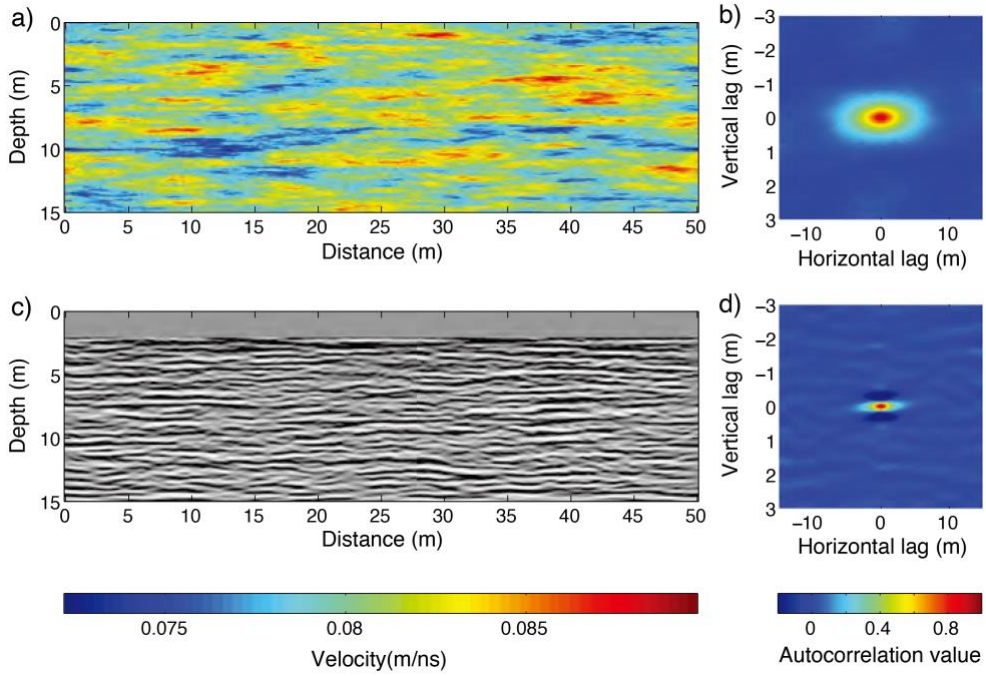


Figure 1.8: Autocorrelation structures of an example subsurface velocity field and its corresponding GPR reflection data. a) Synthetic EM velocity field generated using an exponential autocorrelation model and its corresponding b) 2D autocorrelation. c) Processed zero-offset GPR reflection image, obtained by finite-difference modeling on a), and its corresponding d) 2D autocorrelation. Adapted from Irving et al. (2010).

In seismic community, initial theoretical studies by Gibson (1991), Holliger et al. (1992, 1994), and Pullammanappallil et al. (1997) attempted to show that the lateral statistics of seismic reflection data and the underlying velocity fluctuations should be equivalent. However, empirical studies by Hurich (1996) and by Hurich and Kocurko (2000) show that although a strong degree of correlation does exist for weakly scattering media, those statistics are certainly not equivalent. In fact, Bean et al. (1999) and Carpentier and Roy-Chowdhury (2007) pointed out the fundamental dependence of the lateral correlation structure of a seismic section on bandwidth and on the vertical-derivative operator that acts to transform velocity fields to reflection coefficients. The two elements have not been considered properly in the previous theoretical work.

In GPR community, similar efforts led to essentially identical results and conclusions. Rea and Knight (1998) and Dafflon et al. (2006) observed good agreement between the lateral statistics of a cliff-face photograph and those of a GPR image collected along the top of the cliff, which led them to conclude that the lateral statistics of a GPR image and the underlying water-content distribution are likely equivalent. In both cases, the rather weak assumption is made that the grayscale tones in the photograph are related to sediment grain size and thus to water content. Knight et al. (2007) also found similarities between the lateral correlation statistics of GPR data and neutron-probe measurements. However, Knight et al. (2004) and Oldenborger et al. (2004) noticed

that the lateral correlation structure of a GPR image is significantly affected by its vertical resolution, which is controlled by the GPR antenna frequency. Again, this indicates that the assumption of equivalence between the lateral statistics of water content and GPR reflection data is not valid and that the physics of the underlying wave propagation phenomena should be accounted for more appropriately.

Irving et al. (2009) and Irving and Holliger (2010) followed up on the pioneering works, notably by clarifying the underlying theoretical and methodological foundations. They found that seismic and GPR reflection images allow to infer only the structural aspect ratio of the subsurface heterogeneity, whereas the horizontal and vertical correlation lengths cannot be independently resolved. The reflection data also seemed to exhibit only a very weak sensitivity to the decay of the associated power spectrum. The estimation of those geostatistical parameters involves an iterative Monte-Carlo-type approach (e.g., Irving et al., 2010; Xu et al., 2020), where sets of model parameters drawn from prescribed prior distributions are either accepted or rejected depending on how well they allow for prediction of the observed autocorrelation structure of the reflection image. The findings, as well as the fundamental validity of the estimation technique, have been corroborated through a range of follow-up studies (e.g., Scholer et al., 2010, Xu et al., 2020, 2021). While the Monte-Carlo inversion method can provide a comprehensive exploration of the parameter space and uncertainty analysis regarding the inferred parameters, the associated computational costs are considerably high when the method is applied to a large number of datasets.

1.4 Objectives and outline of the thesis

The primary objective of this thesis is to develop novel quantitative interpretation techniques to constrain heterogeneity using GPR and seismic reflection data. The thesis focuses on two challenging but important environments, i.e., the shallow subsurface and crystalline rocks. For the shallow subsurface, detailed knowledge of EM velocity distribution allows for the reliable prediction of soil water content, which is a key parameter for many hydrogeological, agricultural and engineering applications. Nonetheless, high-resolution velocity estimation through common-offset GPR reflection measurement is currently not possible without complementary borehole information. The use of diffracted wavefield components in GPR data to recover velocity information is therefore attractive. But the reliability and resolution of diffraction-based velocity model are usually not satisfying for quantitative interpretation of subsurface. An alternative way to characterize complex subsurface environment is to estimate the geostatistical parameters related to the subsurface heterogeneity, which is a prerequisite for realistic simulation of groundwater flow and contaminant transport. Yet the computational cost of current geostatistical inversion technique is inherently high. In crystalline rocks, the geological structures are rather complex. From seismic reflection survey, it is now difficult to extract quantitative information about the structures, especially those that are steeply dipping. Based on the aforementioned problems, the following specific objectives are to be addressed:

1) Use common-offset GPR reflection data to perform high-resolution velocity estimation.
2) Improve the performance of diffraction-based velocity analysis.
3) Alleviate the computational cost of geostatistical inversion.
4) Characterize steeply dipping structures in crystalline environment with seismic reflection data.

This thesis focuses on achieving these four objectives by providing novel solutions. Therefore, the following chapters are organized in the same sequence as the objectives.

Chapter 2 presents a novel methodology for estimating the detailed subsurface velocity structure from common-offset GPR reflection measurements, which does not require additional conditioning information. The proposed approach combines two key components: Diffraction analysis is used to infer the smooth, large-scale component of the velocity distribution, whereas the superimposed small-scale fluctuations are inferred via inversion of the reflected wavefield. The method is tested and validated with two synthetic datasets having increasing degrees of complexity and realism before applied to a field example from the BHRS, where independent control data in the form of neutron-neutron porosity logs are available for validation. The results obtained demonstrate the viability and robustness of the proposed approach.

Chapter 3 presents a weighting strategy whose aim is to improve the reliability of estimations of the root-mean-square velocity from common-offset GPR data. The resolution of the inferred semblance spectra is increased through a weighting function that varies in accordance with the sensitivity of a diffraction curve to changes in velocity. The weighting function is calculated from coherency and slope attributes of the diffracted wavefield components. To demonstrate the viability of the proposed method, it is applied in two synthetic test cases and one field GPR dataset. Compared with conventional unweighted local semblance spectra, their weighted counterparts allow for a significantly increased resolution and correspondingly reduced picking uncertainty.

Chapter 4 presents an alternative supervised-learning-based methodology for inferring subsurface geostatistical parameters from reflection GPR data in a highly efficient manner. This approach uses a convolutional neural network that is trained on a vast database of 2D autocorrelation images obtained from synthetic GPR images for a comprehensive range of stochastic subsurface models. Results from synthetic and field data tests confirm the viability of this method to estimate the structural aspect ratio of the subsurface heterogeneity. The estimation is still robust in the presence of high noise levels. For the estimation of the corresponding Hurst number, it is found that the method only performs well in a non-realistic synthetic scenario involving ideal GPR data generated using a convolution model.

Chapter 5 presents a high-resolution seismic reflection survey whose aim is to explore whether the sub-vertical structures prevailing at the surface can be expected to continue at depth. The acquisition and processing of the seismic reflection data were geared towards revealing weak backscattered events from local heterogeneities associated with the prevailing sub-vertical structural grain. The migrated sections show numerous short lineaments that seem to align sub-vertically. To substantiate the observation, synthetic seismic reflection data are generated for canonical models of sub-vertical structures associated with Gaussian- and binary-distributed heterogeneities. Both the observed and synthetic seismic data were then subjected to energy-based attribute analysis as well as geostatistical estimations of the structural aspect ratios and the associated dips. The results of these quantitative interpretation approaches are indicative of the overall consistency between the synthetic and the observed seismic data and, hence, support the original qualitative interpretation of the latter in that the sub-vertical structural grain evident at the surface seems to prevail throughout the imaged part of the upper crust.

Appendix A presents a collaborative project related to this thesis. A conditional stochastic inversion procedure is developed for common-offset GPR reflection measurements. Stochastic realizations of subsurface properties that offer an acceptable fit to GPR data are generated via simulated annealing optimization. The realizations are conditioned to borehole porosity measurements available along the GPR profile or equivalent measurements of another petrophysical property that can be related to the dielectric permittivity, as well as to geostatistical parameters derived from the borehole logs and the processed GPR image. Validation of the

inversion procedure is performed on a pertinent synthetic data set and indicates that the method is capable of reliably recovering strongly heterogeneous porosity structures associated with surficial alluvial aquifers. This finding is largely corroborated through application of the methodology to field measurements from the BHRS.

Chapter 2

High-resolution velocity estimation from surface-based common-offset GPR reflection data

---

Yu Liu, James Irving, Klaus Holliger

Published[1] in *Geophysical Journal International*

[1] Liu, Y., Irving, J. and Holliger, K. (2022). High-resolution velocity estimation from surface-based common-offset GPR reflection data. *Geophysical Journal International*, 230(1), 131-144. doi: 10.1093/gji/ggac058

2.1 Abstract

Surface-based common-offset ground-penetrating radar (GPR) reflection profiling is a popular geophysical exploration technique for obtaining high-resolution images of the shallow subsurface in a cost-effective manner. One drawback of this technique is that, without complementary borehole information in form of dielectric permittivity and/or porosity logs along the profile, it is currently not possible to obtain reliable estimates of the high-frequency electromagnetic velocity distribution of the probed subsurface region. This is problematic because adequate knowledge of the velocity is needed for accurate imaging and depth conversion of the data, as well as for quantifying the distribution of soil water content. To overcome this issue, we have developed a novel methodology for estimating the detailed subsurface velocity structure from common-offset GPR reflection measurements, which does not require additional conditioning information. The proposed approach combines two key components: Diffraction analysis is used to infer the smooth, large-scale component of the velocity distribution, whereas the superimposed small-scale fluctuations are inferred via inversion of the reflected wavefield. We test and validate our method on two synthetic datasets having increasing degrees of complexity and realism before applying it to a field example from the Boise Hydrogeophysical Research Site (BHRS), where independent control data in the form of neutron-neutron porosity logs are available for validation. The results obtained demonstrate the viability and robustness of the proposed approach. Further, due to its efficiency, both in terms of field effort and computational cost, the method can be readily extended to 3D, which further enhances its attractiveness compared to multi-offset-based GPR velocity estimation techniques.

2.2 Introduction

Ground-penetrating radar (GPR) is a high-resolution geophysical exploration technique that has the potential of providing images of shallow subsurface structure with a resolution in the meter to decimeter range (e.g., Knight 2001; Annan 2005; Klotzsche et al. 2018; Lai et al., 2018). Whereas borehole-based GPR transmission techniques have proven to be well-suited to full-waveform inversion approaches (e.g., Ernst et al. 2007; Klotzsche et al. 2019), estimating the detailed velocity structure of the subsurface from surface-based GPR reflection data is notoriously difficult. This is problematic because: (i) the overwhelming majority of GPR data are acquired in reflection mode along the Earth's surface; (ii) accurate velocity information is necessary for proper imaging of reflection data; and (iii) the high-frequency electromagnetic wave velocity in the GPR regime has a strong and direct sensitivity to soil water content, which is a key parameter for many hydrogeological, agricultural, and engineering applications (e.g., Huisman et al. 2003).

One common approach for subsurface velocity estimation from reflection GPR measurements is to collect data at multiple transmitter-receiver offsets and to perform either normal-moveout (NMO) velocity analysis (e.g., Greaves et al. 1996; Huisman et al. 2003; Perroud & Tygel 2005) or reflection tomography (e.g., Bradford 2008). With regard to NMO, the inherent assumption of 1D horizontal layering means that it cannot effectively deal with the highly heterogeneous velocity structures that are rather common in near-surface investigations. Although reflection tomography is able to overcome this issue, it comes at a rather high computational cost and requires inherently subjective horizon picking. Further, as pointed out by Bradford et al. (2009), reflection tomography only recovers the large-scale component of the subsurface velocity distribution that is needed to properly focus and image the GPR data, which is of substantially lower resolution than the reflection image itself. The latter problem can be potentially alleviated through waveform inversion approaches (e.g., Busch et

al. 2012; Lavoué et al. 2014); however their success so far has been limited due to the rather narrow range of reflection angles and antenna radiation patterns that are highly complex, largely unknown, and site-dependent (e.g., Lampe & Holliger 2003). Finally, a clear drawback of multi-offset GPR acquisitions is their high cost in terms of acquisition time, which increases approximately linearly with the considered number of transmitter-receiver offsets for the common case of GPR systems having a single transmitter and receiver antenna. Indeed, such surveys become largely impractical in the context of long 2D profiles and, particularly, 3D acquisitions.

For the above reasons, the vast majority of surface-based GPR reflection surveys are performed using the traditional bi-static, common-offset approach, where a single transmitter-receiver antenna pair, separated by a small fixed distance, is incrementally moved along the measurement profile (e.g., Annan 2005). While the estimation of the subsurface velocity distribution from such data is substantially more difficult than for multi-offset GPR surveys, significant efforts have been made during the past decade because of the high potential rewards. In this regard, Schmelzbach et al. (2012) present an impedance inversion approach for common-offset GPR data that is based on a convolution model for the GPR traces, where borehole dielectric permittivity or porosity logs are used to recover the low-frequency part of the velocity structure that cannot be obtained from the reflection data. Zeng et al. (2015) and Liu et al. (2018) adopt similar approaches to estimate the distribution of soil water content and to characterize buried archaeological remains, respectively. Xu et al. (2021) also assume a convolution model for the GPR traces, but combine stochastic simulation with simulated annealing optimization in order to generate velocity realizations that honor the GPR measurements and borehole porosity log data along the profile. Forte et al. (2013; 2014) assume a locally 1D layered subsurface structure and use picked reflection amplitudes to recursively estimate the GPR velocity in a series of identified subsurface layers, in which the velocity is assumed constant. Other authors estimate the spatial distribution of GPR velocity from common-offset data via the analysis of diffractions present in the recordings. Novais et al. (2008) use velocity continuation to generate several migrated sections and analyze the associated diffraction focusing to build a root-mean-square (RMS) velocity model. Clair & Holbrook (2017) apply the seismic diffraction imaging and velocity analysis workflow proposed by Fomel et al. (2007) to common-offset GPR data in order to estimate snow water equivalent. Yuan et al. (2019) employ a similar approach to characterize the velocity structure of surficial chalk deposits. Although all of the above methods have the ability to estimate subsurface properties from common-offset GPR measurements, they all suffer from inherent limitations. Notably, the reflection-based methods have the potential to provide high-resolution results, but they generally require complementary information such as borehole logs, which are usually not available. Conversely, diffraction-based methods require a suitably dense and even distribution of diffractions in the data and, even under ideal circumstances, can only resolve the large-scale velocity structure.

In this study, we present a novel velocity estimation method for surface-based common-offset GPR reflection data that combines the advantages of the reflection- and diffraction-based techniques described above. To estimate the low-frequency background velocity field, diffractions are separated from the unmigrated GPR data and subjected to migration velocity analysis based on a prescribed focusing measure. After migrating the GPR data using the derived velocity field, the reflected wavefield is isolated and used to deduce the small-scale velocity fluctuations. The latter is accomplished via sparse inversion based on an iteratively reweighted least-squares strategy assuming a convolutional model for each GPR trace. The final high-resolution velocity distribution is obtained by combining the large-scale diffraction-based and the fine-scale reflection-based estimates.

The paper proceeds as follows. We begin by describing the methodological background of the proposed velocity estimation method. Next, we show the application of our method to two synthetic datasets, which

differ in their degree of complexity and realism. Finally, we apply the proposed approach to common-offset 100-MHz GPR field data acquired at the Boise Hydrogeophysical Research Site (BHRS) near Boise, Idaho, USA.

2.3 Methodology

We assume in our work that the subsurface velocity distribution $v(x,z)$ can be regarded as the sum of a smoothly varying or constant background velocity field $v_0(x,z)$ and a small-scale velocity fluctuation field $\Delta v(x,z)$ (e.g., Pullammanappallil et al. 1997; Poppeliers 2007; Irving et al. 2009; Scholer et al. 2010)

$$v(x,z) = v_0(x,z) + \Delta v(x,z). \tag{2.1}$$

To determine $v(x,z)$ from a common-offset GPR reflection dataset, we separate the recorded wavefield into its diffracted and reflected components, which are used to estimate v_0 and Δv, respectively. This inherently assumes that the background velocity field is smooth at the scale of a dominant GPR wavelength and beyond, such that it does not contribute to the reflected wavefield. Figure 2.1 illustrates schematically the steps involved in our velocity estimation procedure. First, diffractions are separated and analyzed in order to infer the spatially variable RMS and interval velocity structures. The latter serves as v_0, whereas the former is used to migrate the common-offset GPR data, after which the dominant reflections are separated. Assuming a convolutional relationship between the velocity perturbation field and the reflection data based on an estimated mixed-phase wavelet, a L1-norm constrained inversion is then used to infer Δv. Below we describe in detail this inversion workflow in terms of the following four main components: (i) diffraction separation, (ii) background velocity estimation, (iii) reflected wavefield prediction, and (iv) velocity perturbation inversion.

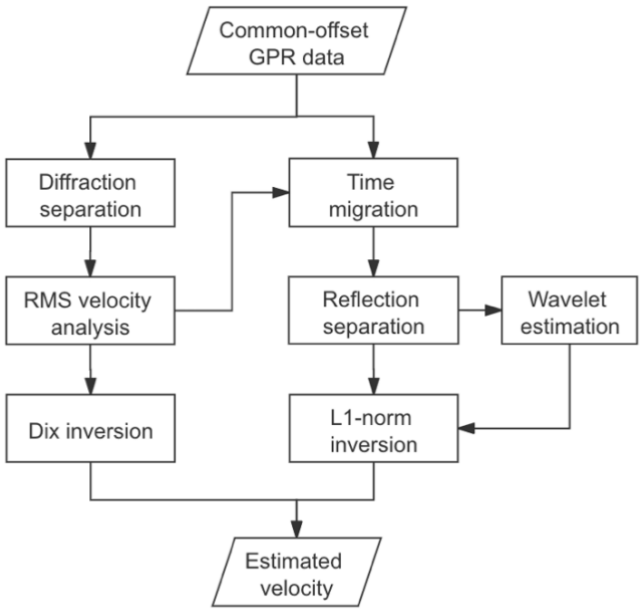


Figure 2.1: Flowchart illustrating the proposed method for estimating the detailed subsurface velocity structure from surface-based common-offset GPR reflection data.

2.3.1 Diffraction separation

The diffracted wavefield is obtained in our methodology via plane-wave destruction (PWD) filtering (Fomel, 2002). The underlying assumption when using this approach is that reflections correspond primarily to coherent events having slowly changing slopes in the *x*-*t* domain, whereas the slopes associated with diffractions are significantly more spatially variable. The goal of PWD filtering is to destroy locally planar events in the data corresponding to an estimated slope field $\sigma(x,t)$. By regularizing the estimation problem, it is possible to destroy only those events whose slopes change slowly in space, thereby isolating the diffracted energy.

A local plane wave in the *x*-*t* domain can be expressed by the following differential equation (e.g., Fomel, 2002):

$$\frac{\partial u}{\partial x} + \sigma \frac{\partial u}{\partial t} = 0, \tag{2.2}$$

where $u(x,t)$ is the wavefield and σ is the local slope. If the local slope in a seismic or GPR dataset is unchanging in time, the wavefields observed at two adjacent trace positions x_i and x_{i+1} are related by a time shift $\sigma\Delta x$, where Δx is the trace spacing. That is,

$$u(x_{i+1},t) = u(x_i, t+\sigma\Delta x), \tag{2.3}$$

which has the Fourier transform

$$U(x_{i+1},\omega) = U(x_i,\omega)e^{i\omega\sigma\Delta x}. \tag{2.4}$$

Eq. (2.4) shows that we can predict the trace at position x_{i+1} from the trace at position x_i by application of a linear phase shift. To apply this concept to data with temporally variable local slopes, Fomel (2002) used the fractional delay filter of Thiran (1971) to derive a localized, discrete, time-domain approximation to $e^{i\omega\sigma\Delta x}$ whose coefficients depend nonlinearly on the local slope values. Prediction of a trace using its neighbor can then be accomplished by matrix-vector multiplication

$$\mathbf{u}_{i+1} = \mathbf{P}_{i,i+1}\mathbf{u}_i, \tag{2.5}$$

where $\mathbf{P}_{i,i+1}$ is a time-variable convolution matrix linking trace vectors $\mathbf{u}_{i+1}$ and $\mathbf{u}_i$, whose entries are a nonlinear function of the local slope field $\sigma(x,t)$.

The PWD problem seeks to estimate $\sigma(x,t)$ by minimizing the prediction error for an entire seismic or GPR section, thereby destroying the local plane waves in the data. Considering the section $\mathbf{s}$ as a column vector containing all of the traces, i.e., $\mathbf{s} = [\mathbf{s}_1^{\mathrm{T}}\ \mathbf{s}_2^{\mathrm{T}}\ \ldots\ \mathbf{s}_{\mathrm{n}}^{\mathrm{T}}]^{\mathrm{T}}$, this is described by

$$\mathbf{r} = \mathbf{D}\mathbf{s}, \tag{2.6}$$

where $\mathbf{r}$ is the destruction residual, and $\mathbf{D}$ is the destructor matrix defined by:

$$\mathbf{D} = \begin{bmatrix} -\mathbf{P}_{1,2} & \mathbf{I} & 0 & \cdots & 0 \\ 0 & -\mathbf{P}_{2,3} & \mathbf{I} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & -\mathbf{P}_{\mathrm{n}-1,\mathrm{n}} & \mathbf{I} \end{bmatrix}, \tag{2.7}$$

with $\mathbf{I}$ representing identity operator. The estimation of $\sigma(x,t)$ is accomplished via regularized nonlinear least-squares minimization of Eq. (2.6), where shaping regularization (Fomel, 2007a) is used to control the smoothness of the estimated slope field. In our case, the considered lateral smoothing radius for the regularization must be large enough to estimate a slope field that well represents the reflections in the dataset, but not the diffractions. The prediction residual corresponding to the estimated slope field is simply the GPR section with the reflection events removed. All of the above steps are performed in our work using the programs 'sfdip' and 'sfpwd' in Madagascar (https://reproducibility.org/), an open-source data analysis package.

2.3.2 Background velocity estimation

Once the diffracted wavefield has been separated, the next step is to use it to estimate the low-frequency background velocity field v_0, which, due to its smoothness at the wavelength scale, does not to contribute to the reflected wavefield. To this end, we first estimate the RMS velocity distribution by examining the focusing of diffractions during migration using a suite of constant velocity values. Fomel et al. (2007) proposed a migration focusing metric based on local kurtosis, whereas Decker et al. (2017) considered the local semblance attribute. Here, we use the latter measure, which can be defined as

$$s(x,t,v) = \frac{\left(F_v\big(a(x,t)\big)\right)^2}{F_v\left(a^2(x,t)\right)}, \tag{2.8}$$

where $a(x,t)$ denotes the diffraction amplitude as a function of horizontal position x and time t, and operator F_v denotes time migration using constant velocity v. Migration is performed on both the diffracted wavefield and its square using the velocity continuation method of Fomel (2003), which results in two space-time-velocity cubes. In Madagascar, this step is accomplished using the program ‘sfvelcon’. The division in Eq. (2.8) is then performed in a regularized manner using the program 'sfdivn' in order to constrain the smoothness of the resulting local semblance cube (Fomel, 2007b). Using the automatic picking algorithm 'sfpick' (Fomel, 2009), the maxima on each time-velocity panel are next selected, which yields a 1D RMS velocity curve at each trace location. These curves are combined into a 2D RMS velocity model, which is finally provided as input to the constrained Dix inversion program 'sfdix' (Fomel and Guitton, 2006) to estimate $v_0(x,t)$.

2.3.3 Reflection separation

In order to obtain the reflected wavefield that is used in our inversion procedure to estimate the velocity perturbation Δv, two steps are performed. First, the GPR profile is time-migrated based on the inferred RMS velocity model from Section 2.2 using the velocity continuation method described in Fomel (2003). This has the effect of collapsing diffractions and moving dipping reflectors to their correct positions in terms of vertical traveltime, and is accomplished using the Madagascar programs 'sfvelcon' and 'sfslice'. Then, we apply the PWD method to the migrated reflection section in order to estimate the local slopes, which in this case are used to predict the time-migrated reflected wavefield void of migration artifacts and random noise (Fomel and Guitton, 2006). The latter step is accomplished using the Madagascar program ‘sfpwdsmooth2’. Note that our use of PWD here is different compared to what was presented in Section 2.1, where the method was used to suppress reflected energy in the data and isolate the diffracted wavefield. In this regard, it is important to note that a high-quality reflection section cannot be obtained by simply subtracting the diffracted wavefield from the GPR data. Indeed, predicting the reflected wavefield from the estimated slopes of the time-migrated image results in a cleaner section that is much more amenable to the velocity perturbation inversion described next.

2.3.4 Velocity perturbation inversion

To estimate the velocity perturbation field Δv, we perform sparse inversion of the time-migrated reflected GPR wavefield obtained in Section 2.3. To this end, we assume that the wavefield can be effectively described using the so-called primary reflectivity section (PRS) model (e.g., Gibson and Levander, 1990; Holliger et al.,

1994; Irving et al., 2009), whereby the time-migrated data, $d(x,t)$, are expressed as the convolution product of the GPR source wavelet, $\mathrm{w}(t)$, and the subsurface reflectivity distribution, $r(x,t)$

$$d(x,t) = \mathrm{w}(t) * r(x,t). \tag{2.9}$$

Eq. (2.9) is well known to provide an adequate model for zero-offset seismic or GPR reflection data when single scattering prevails and dispersion is absent (e.g., Yilmaz, 2001). Although the second assumption is only strictly valid for GPR data acquired in perfectly electrically resistive environments, experience has shown that this model is able to accommodate the limited dispersion effects linked to low-loss conditions for which the GPR method is suitable (e.g., Irving et al., 2009; Xu et al., 2020). Indeed, such effects in GPR data tend to be inherently rather minor, as it is simply impossible to acquire high-quality GPR reflection data in strongly dispersive environments.

As the subsurface reflectivity can be approximated using the temporal derivative of the velocity perturbation field (e.g., Pullammanappallil et al. 1997; Poppeliers 2007), and because the temporal derivative operator may be treated as a finite-difference filter whose position within a convolution equation can be shifted (Irving et al. 2009; Scholer et al. 2010), Eq. (2.9) leads to

$$\begin{aligned} d(x,t) &\approx \mathrm{w}(t) * \frac{\partial}{\partial t}\Delta v(x,t) \\ &= g(t) * \Delta v(x,t), \end{aligned} \tag{2.10}$$

where $g(t)$ represents the time-differentiated GPR wavelet. Expression (10) provides a linear relationship between the time-migrated reflected GPR wavefield and the velocity perturbation field, which forms the basis for our inversion procedure. Indeed, considering data vector $\mathbf{d}$ containing all of the GPR traces arranged into a single column, i.e., $\mathbf{d} = [\mathbf{d}_1^{\mathrm{T}}\ \mathbf{d}_2^{\mathrm{T}}\ ...\ \mathbf{d}_{\mathrm{n}}^{\mathrm{T}}]^{\mathrm{T}}$, and model vector $\mathbf{m}$ containing the corresponding velocity perturbations underlying each trace arranged into a single column, i.e., $\mathbf{m} = [\boldsymbol{\Delta}\mathbf{v}_1^{\mathrm{T}}\ \boldsymbol{\Delta}\mathbf{v}_2^{\mathrm{T}}\ \boldsymbol{\Delta}\mathbf{v}_3^{\mathrm{T}}\ ...\ \boldsymbol{\Delta}\mathbf{v}_{\mathrm{n}}^{\mathrm{T}}]^{\mathrm{T}}$, we have

$$\mathbf{d} = \mathbf{G}\mathbf{m}, \tag{2.11}$$

where $\mathbf{G}$ is a block-diagonal matrix containing n replicates of the convolution matrix associated with the time-differentiated wavelet $g(t)$.

To define the kernel matrix $\mathbf{G}$, information on the GPR source wavelet is required. In this work, we estimate $w(t)$ from the reflected wavefield using the method of Schmelzbach & Huber (2015), which assumes that a typical mixed-phase GPR source wavelet can be considered as a minimum-phase wavelet that has been shifted by a constant phase angle. To first estimate the corresponding minimum-phase wavelet, we perform standard least-squares spiking deconvolution on the reflected wavefield and take the inverse of the deconvolution operator (e.g., Buttkus, 2000). A search of the phase rotation angle that maximizes the kurtosis when applied to this minimum-phase wavelet is then used to obtain the final mixed-phase GPR source wavelet. The practical validity of this source wavelet estimation procedure was recently demonstrated by Xu et al. (2021). Note that the effects of minor dispersion in the GPR data are, at least in part, accounted for in the sense that an effective wavelet that best fits the considered dataset in its entirety, rather than the true emitted GPR source signal, is estimated.

To invert for the velocity perturbation $\mathbf{m}$ given the reflection data $\mathbf{d}$, we minimize the following objective function

$$\theta(\mathbf{m}) = ||\mathbf{G}\mathbf{m} - \mathbf{d}||_{\mathbf{2}}^{\mathbf{2}} + \lambda\, \|\mathbf{D}\mathbf{m}\|_1, \tag{2.12}$$

where $\|\cdot\|_p$ denotes the L-p norm, λ is a trade-off parameter that controls the desired balance between fitting the data and honoring the prescribed prior information about the model, and matrix $\mathbf{D}$ is given by

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_x \\ \eta\, \mathbf{D}_t \end{bmatrix}. \tag{2.13}$$

Here, $\mathbf{D}_x$ and $\mathbf{D}_t$ are finite-difference matrices that approximate the first derivatives of the velocity perturbation model in the horizontal and temporal directions, respectively, and η is an anisotropy parameter that controls the degree of desired smoothing between the temporal and horizontal directions. The choice of α should reflect the expected aspect ratio of the underlying GPR velocity heterogeneity.

Eq. (2.12) corresponds to a regularized least-squares inversion with blocky model prior constraints. That is, in seeking to minimize the L1-norm of the first derivatives of the velocity perturbation field, we tend to recover models that have a sparse first derivative structure, meaning a piecewise-constant or blocky appearance. Note that this approach has similarities to sparse spike deconvolution in seismic data processing, which uses sparsity constraints to recover the underlying reflectivity series from a seismic trace (e.g., Claerbout and Miur, 1973; Oldenburg et al., 1983; Velis, 2008). Our method differs, however, in the sense that (i) we use sparsity applied to the first derivative of the velocity perturbation field and invert for the latter directly, rather than inverting for a sparse reflectivity series; and (ii) we invert all traces at once with both vertical and lateral regularization constraints in order to estimate the full 2D velocity perturbation field.

Due to the presence of the L1-norm, the minimization of Eq. (2.12) is nonlinear. To address this, we use an iteratively reweighted least squares (IRLS) approach based on the following approximation of the Lp-norm proposed by Ekblom (1973):

$$\|\mathbf{x}\|_p \approx \sum_{i=1}^{n} \left(x_i^2 + \epsilon^2\right)^{p/2}, \tag{2.14}$$

where ϵ is a small user-defined value (e.g., Farquharson and Oldenburg, 1998). Taking the derivative of Eq. (2.12) with respect to $\mathbf{m}$ and setting it to zero, and considering approximation (2.14), we arrive at

$$\left(2\mathbf{G^T G} + \lambda \mathbf{D}^T \mathbf{R D}\right) \mathbf{m} = 2\mathbf{G^T d}, \tag{2.15}$$

where

$$\mathbf{R}_{ii} = \frac{1}{|(\mathbf{Dm})_i| + \epsilon} \tag{2.16}$$

is a diagonal reweighting matrix. We solve for $\mathbf{m}$ iteratively as follows:

(1) Set $\mathbf{R} = 2\mathbf{I}$.
(2) Solve Eq. (2.15) for $\mathbf{m}$ using the conjugate gradient method.
(3) Update $\mathbf{R}$ using Eq. (2.16) and the result for $\mathbf{m}$ obtained in Step 2.
(4) Return to Step (2) and iteratively update $\mathbf{m}$ until a defined maximum number of iterations or desired data fit is reached.

In carrying out the above steps, the first iteration of our inversion procedure solves for the velocity perturbation field corresponding to an L2-norm constraint on the model derivative term in Eq. (2.12). This and subsequent solutions are then used within the IRLS reweighting scheme in order to gradually converge to the L1-norm solution, typically within a few iterations. Once $\Delta v(x,t)$ has been obtained, it is added to the estimated background velocity model $v_0(x,t)$ from Section 2.2. As a final step, the resulting subsurface velocity field in terms of vertical two-way traveltime, $v(x,t)$, is converted to depth to obtain the desired $v(x,z)$.

2.4 Results

2.4.1 Application to synthetic data

In the following, we test and validate the velocity estimation technique outlined in Section 2.3 and illustrated in Figure 2.1 by applying it to synthetic common-offset GPR reflection data. We first consider a layered subsurface velocity model containing a small number of well-defined point-type diffractors. We then move to an arguably more realistic scenario involving a stochastic velocity distribution characterized by the explicit absence of idealized diffracting structures.

2.4.1.1 Layered model

Our layered velocity model, which is shown in Figure 2.2a, is similar to that recently considered by Yuan et al. (2019) in a diffraction imaging study. The model contains two main velocity units separated by a dipping interface. A thin horizontal bed, with a thickness of 0.5 m, is present in the underlying unit. Both the upper and lower units contain three circular diffractors with diameters ranging from 0.4 to 0.6 m. The corresponding relative dielectric permittivities of the upper and lower layers, the thin bed, and the diffractors are 9, 16, 25, and 4, respectively. The electrical conductivity of all materials is fixed at a constant value of 1 mS/m, and the magnetic permeability is assumed to be equal to its value in free space.

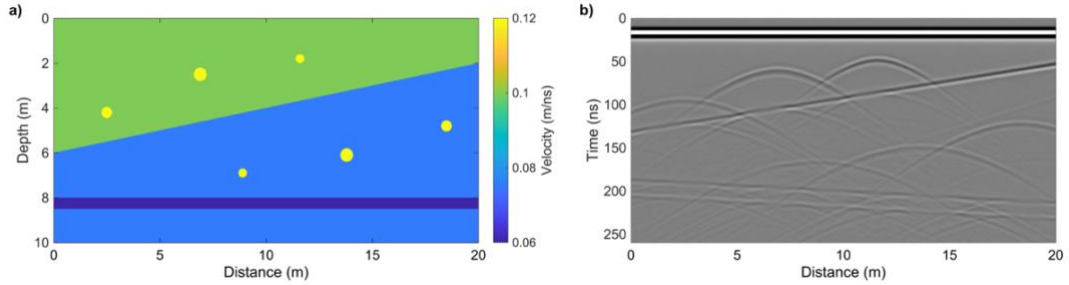


Figure 2.2: a) Layered velocity model with discrete diffractors and b) corresponding synthetic common-offset GPR reflection data with 2% Gaussian random noise added.

Synthetic common-offset GPR reflection data were simulated over the layered velocity model using the gprMax software (Warren et al. 2016), which solves Maxwell's equations using the finite-difference time-domain (FDTD) method. The transmitter and receiver antennas, which are approximated by point electric dipoles, were spaced 0.5 m apart and moved at 0.1 m increments along the survey profile. The source antenna current function was specified as a Ricker wavelet having a dominant frequency of 100 MHz, which resulted in a propagating electromagnetic pulse corresponding to the first derivative of this function. The resulting synthetic GPR data were then contaminated with 2% Gaussian random noise (Figure 2.2b) prior to being subjected to a standard processing flow involving (i) elimination of the direct air and ground arrivals from the data by subtracting the average trace calculated over a time window from 0 to 36 ns using a moving spatial window of 50 traces; (ii) amplitude scaling to compensate for energy spreading, absorption, and scattering using a gain function of the form $g(t) = (1 + at)e^{bt}$; and (iii) 5-300 Hz bandpass filtering. With time measured in nanoseconds, the parameters a and b were chosen empirically to be 0.3 ns^{-1} and 0.2 ns^{-1},

respectively, such that the gain function brought all amplitudes along a given trace to the same average level. Figure 2.3a shows the resulting unmigrated processed data section.

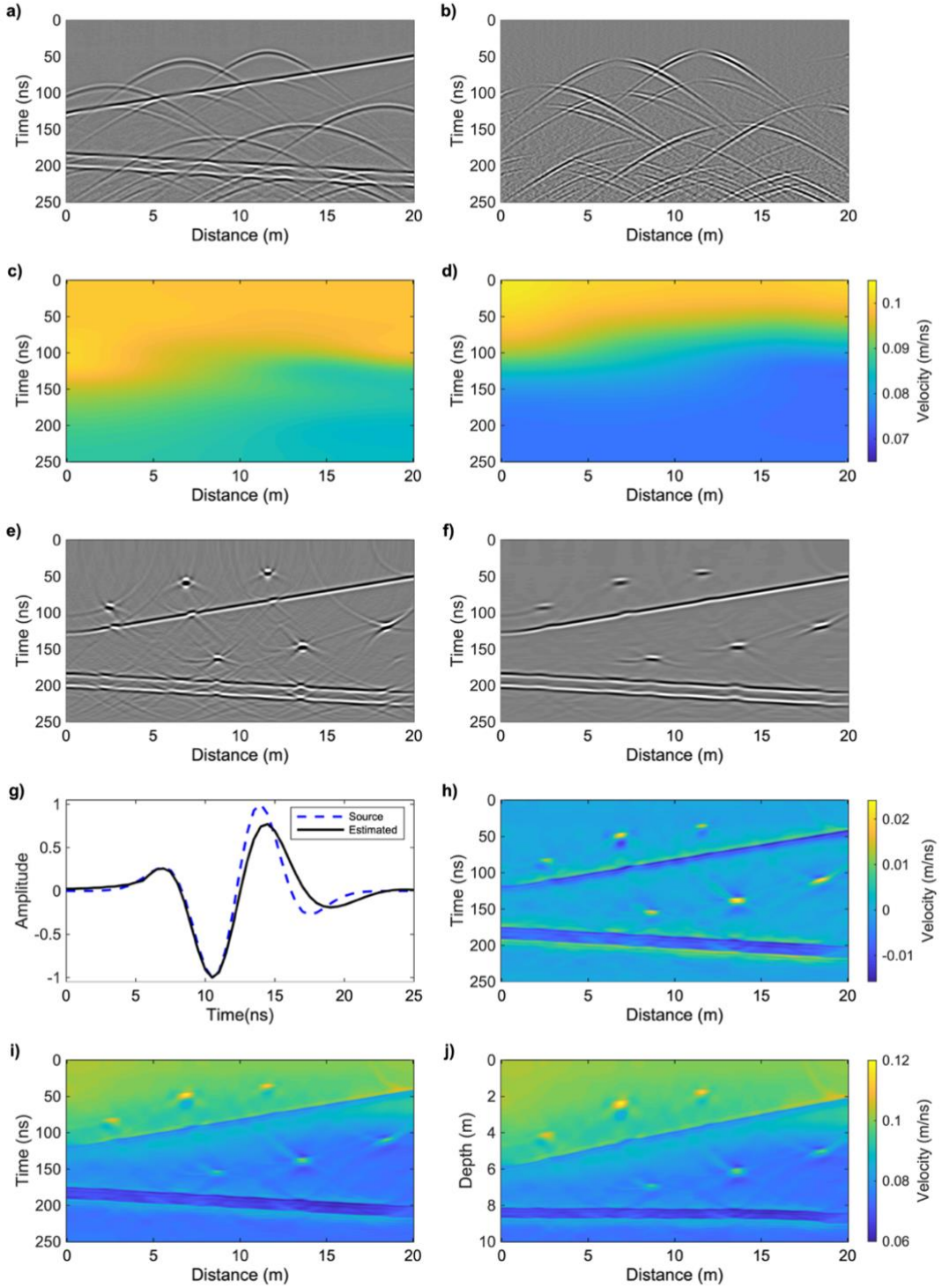


Figure 2.3: Velocity estimation process for the layered synthetic velocity model presented in Figure 2.2a. From the a) processed GPR section, the b) diffracted wavefield is separated and used to estimate the c) RMS velocity structure. Using Dix inversion, the d) low-frequency background velocity field v_0 (x,t) is obtained. e) Time-migrated GPR section based on the RMS velocity structure, from which the f) reflected wavefield is obtained. g) Comparison of estimated GPR wavelet with the true source wavelet. h) Velocity perturbation field Δv(x,t) obtained by inverting the reflected wavefield. The final estimated velocity structure (background+perturbation) is shown in terms of i) traveltime and j) depth.

Following the velocity estimation workflow outlined in Section 2.3 and illustrated in Figure 2.1, diffractions were first separated from the processed data using PWD filtering (Figure 2.3b). The diffracted wavefield was then subjected to velocity continuation and local kurtosis analysis in order to estimate the RMS velocity structure (Figure 2.3c), which was used in a Dix inversion procedure to obtain the low-frequency background velocity field diplayed in Figure 2.3d. Note that this result shows some resemblance to the underlying model in Figure 2.2a. However, it fails to adequately represent the dipping interface as a sharp discontinuity having a constant slope, and it entirely misses the thin bed. In Figure 2.3e, we show the GPR reflection data after time migration based on the estimated RMS velocity structure in Figure 2.3c. We see that the data have been adequately imaged apart from some residual "smiles", which are attenuated through the application of PWD to isolate the specular reflections (Figure 2.3f). From the separated reflection image, a mixed-phase GPR wavelet was estimated (Schmelzbach & Huber 2015), which is compared with the first derivative of the Ricker source current function in Figure 2.3g. Figure 2.3h shows the velocity perturbation field inferred through our L1-norm inversion approach using a value of $\eta = 10$ and after 5 IRLS iterations. We observe that the high-frequency elements present in Figure 2.2a have now been estimated, but not the low-frequency velocity trend. Finally, Figures 2.3i and 2.3j show the complete estimated velocity model, equal to the sum of the background and perturbation fields, in terms of two-way traveltime and after conversion to depth, respectively. The comparison with the reference velocity model (Figure 2.2a) is quite favorable, which clearly illustrates the potential benefits of the proposed diffraction- and reflection-based velocity estimation approach. In this context, is important to emphasize that the former can only resolve the smooth large-scale velocity structure and, hence, entirely misses the presence of the thin bed (e.g., Yuan et al. 2019) whereas, on its own, the latter requires coincident borehole information for calibration and recovery of the large-scale component of the velocity structure (e.g., Schmelzbach et al. 2012; Xu et al. 2021).

2.4.1.2 Heterogeneous model

We now test our proposed methodology on an arguably more realistic model of the shallow subsurface. In this regard, we consider the stochastic velocity distribution shown in Figure 2.4a, which is meant to emulate a heterogeneous surficial alluvial environment. The model was geostatistically generated based on the von Kármán autocorrelation function, which describes a band-limited fractal medium (e.g., Tronicke & Holliger 2005) and is given by

$$C(r) = \frac{r^{\nu} K_{\nu}(r)}{2^{\nu-1}\Gamma(\nu)}, \tag{2.17}$$

where $K_{\nu}(r)$ is the modified Bessel function of the second kind of order $0 \leq \nu \leq 1$, Γ is the gamma function and

$$r = \sqrt{(x/a_x)^2 + (z/a_z)^2} \tag{2.18}$$

is the weighted radial autocorrelation lag with a_x and a_z denoting the correlation lengths along horizontal and vertical directions x and z, respectively. Values of $\nu = 0.5$, $a_x = 2.0$ m, and $a_z = 0.2$ m were considered, along with a mean velocity of 0.1 m/ns and a standard deviation equal to 0.01 m/ns. The generated multi-Gaussian velocity realization was then transformed into a facies-type distribution through thresholding, whereby six units having constant velocities equal to 0.079, 0.092, 0.100, 0.105, 0108, and 0.116 m/ns were specified.

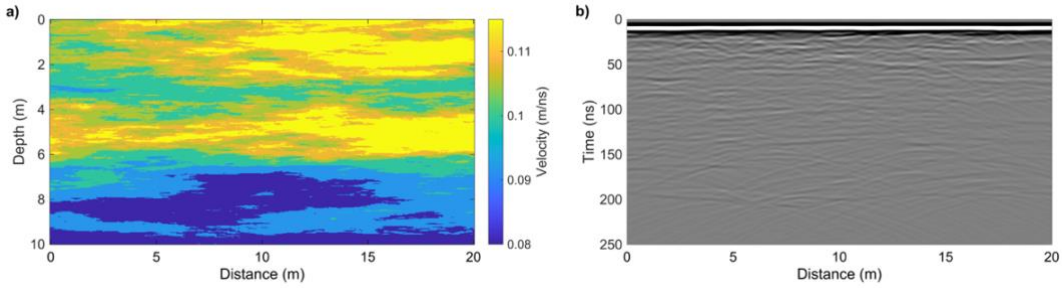


Figure 2.4: a) Stochastic velocity model and b) corresponding synthetic common-offset GPR reflection data with 2% Gaussian random noise added.

To generate synthetic common-offset GPR reflection data over the velocity model in Figure 2.4a, we again used the gprMax software (Warren et al. 2016). Velocity values v were converted to relative dielectric permittivity ε for the FDTD modeling using the low-loss approximation $v \approx 1/\sqrt{\varepsilon\mu}$, where the magnetic permeability μ was assumed equal to its value in free space. As was done previously, the electrical conductivity was fixed at a constant value of 1 mS/m. For the source antenna current function, we considered for this simulation the derivative of a Blackman-Harris window having a dominant frequency of 100 MHz (Irving and Knight, 2006). The spacing between the transmitter receiver antennas was again set to 0.5 m, and traces were simulated every 0.1 m along the profile. Figure 2.4b shows the resulting synthetic GPR data with the addition of 2% Gaussian noise. Processing of these data was essentially identical to that for the layered synthetic velocity model except that the averaging window used for the first-arrival removal was set from 0 to 25 ns, and the gain parameters a and b were set to 0.2 ns^{-1} and 0.2 ns^{-1}, respectively. The processed GPR section is shown in Figure 2.5a.

Figure 2.5b shows the diffracted wavefield estimated from the processed data in Figure 2.5a, which was used to infer the RMS velocity structure (Figure 2.5c) and, subsequently, the background velocity field through Dix inversion (Figure 2.5d). In Figure 2.5e, we show the time-migrated GPR section based on the RMS velocity field, from which the reflected wavefield was obtained (Figure 2.5f). The latter was used to estimate an effective mixed-phase source wavelet, which is shown in Figure 2.5g and seen to compare favorably to the true source wavelet corresponding to the derivative of the considered input current function. Finally, in Figures 2.5h, 2.5i, and 2.5j, we show the inverted velocity perturbation field obtained after 5 IRLS iterations using a value of $\eta = 10$, along with the final estimated velocity model in terms of two-way traveltime and depth, respectively. Comparison of Figure 2.5j with the underlying velocity model in Figure 2.4a demonstrates remarkably good agreement, but also points to two interesting aspects of the proposed velocity estimation method that did not become evident in its application to the more idealized layered model (Figures 2.2 and 2.3). The first concerns the influence of the direct wave and its muting, which, in the presence of small-scale heterogeneity, inherently affects the viability and accuracy of the velocity estimation over a depth range corresponding to approximately one dominant wavelength, that is, the first ~1 m depth. The second observation concerns the importance of the background velocity model and its impact on the final result. This is illustrated by the fact that our final velocity model (Figure 2.5j) misses the pervasive low-velocity zone between ~2 and ~3.5 m depth in the central and right-hand side of the model from ~7 m to ~20 m lateral distance, which can be directly related to the limited resolution of the estimated background velocity model.

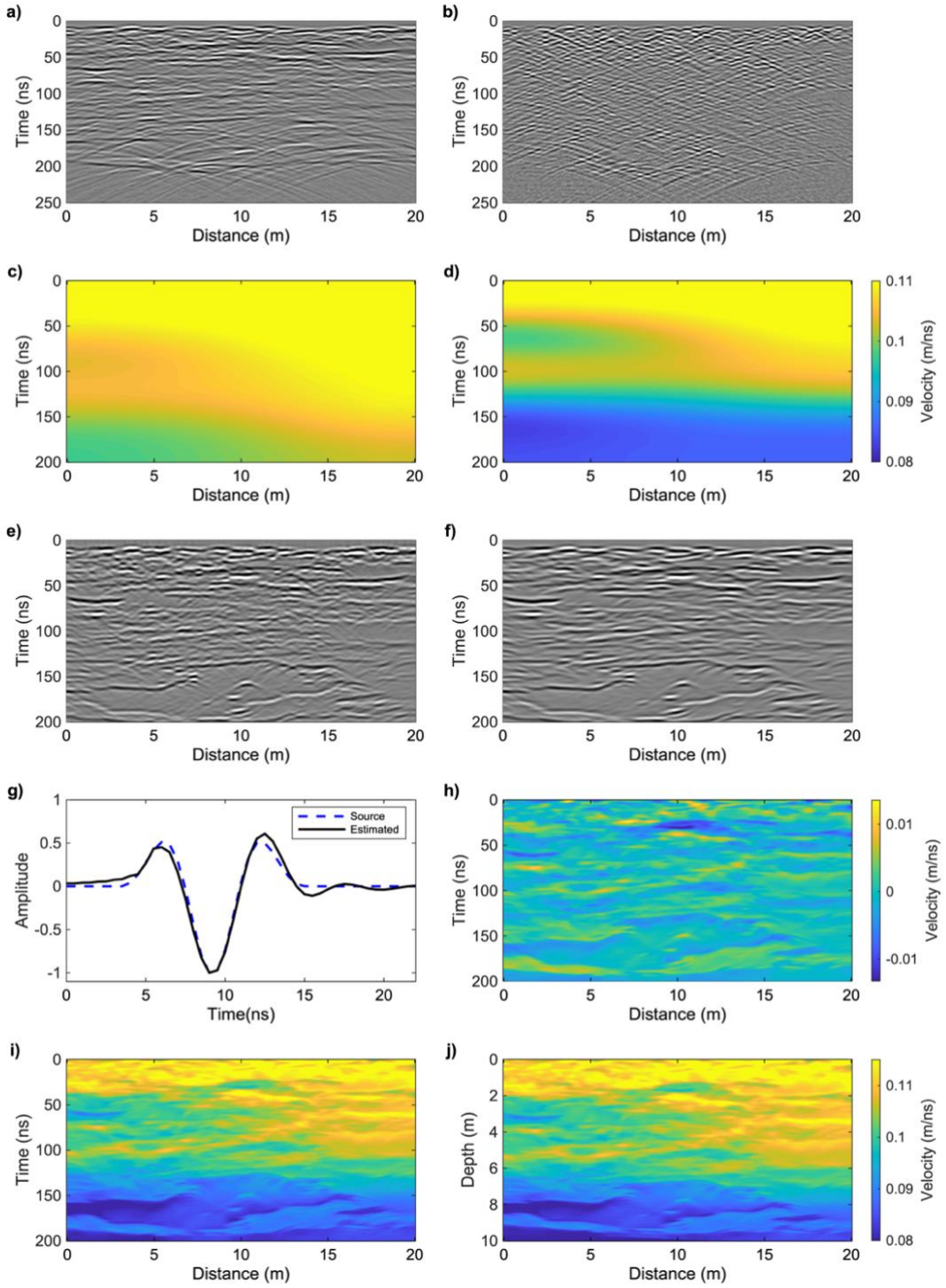


Figure 2.5: Velocity estimation process for the stochastic synthetic velocity model presented in Figure 2.4a. From the a) processed GPR section, the b) diffracted wavefield is separated and used to estimate the c) RMS velocity structure. Using Dix inversion, the d) low-frequency background velocity field v_0 (x,t) is obtained. e) Time-migrated GPR section based on the RMS velocity structure, from which the f) reflected wavefield is obtained. g) Comparison of estimated GPR wavelet with the true source wavelet. h) Velocity perturbation field Δv(x,t) obtained by inverting the reflected wavefield. The final estimated velocity structure (background+perturbation) is shown in terms of i) traveltime and j) depth.

2.4.2 Application to field data

2.4.2.1 Database

We now apply the proposed velocity estimation method to a field GPR dataset acquired at the Boise Hydrogeophysical Research Site (BHRS). The BHRS is a research wellfield located on a gravel bar adjacent to the Boise River near Boise, Idaho, USA (Figure 2.6). The surficial aquifer consists of late Quaternary fluvial deposits dominated by gravel and sand, and is underlain by a layer of red clay at ~20 m depth (Barrash & Clemo 2002). The depth of the groundwater table at the site varies seasonally between ~2 m and ~4 m. Over the past two decades, the BHRS has been extensively utilized for the testing, validation, and improvement of a wide variety of geophysical and hydrogeological characterization methods (e.g., Tronicke et al. 2004; Bradford et al. 2009; Dafflon et al. 2009, 2011; Hochstetler et al. 2016; Xu et al. 2020, 2021).

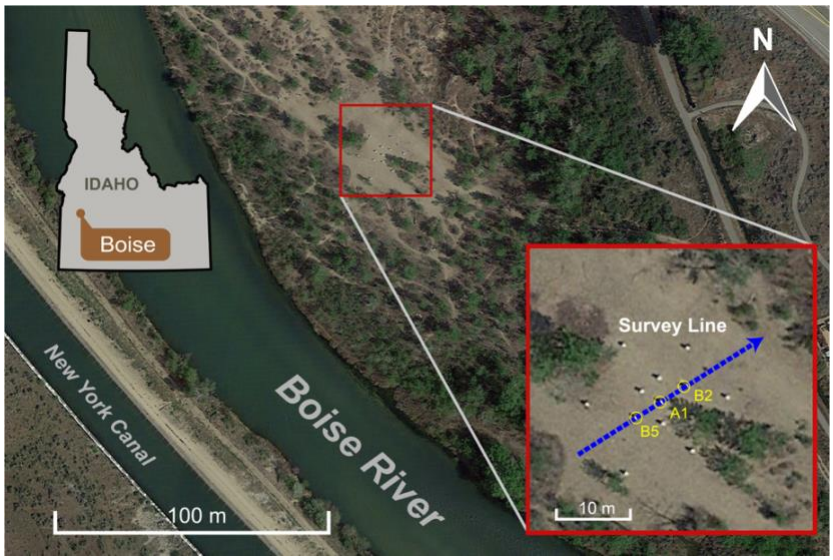


Figure 2.6: Map of the BHRS (43°32'32" N, 116°05'52" W) showing the location of considered common-offset GPR reflection profile (blue dashed line). The profile is aligned with boreholes B5, A1, and B2 (yellow circles).

The considered common-offset, bi-static GPR reflection prosfile is a part of 3D survey performed at the BHRS in 1998 using a Pulse Ekko Pro 100 system (Sensors & Software Inc.) with 100 MHz nominal center frequency antennas, and can be considered as a reference for surface-based GPR reflection data collected in alluvial environments (e.g., Xu et al., 2020, 2021). The profile has a length of 30 m and crosses three boreholes, B5, A1, and B2, for which neutron-neutron porosity logs are available below the groundwater table (Figure 2.6). While the exact values have not been reported, the depth of the latter at the time of acquisition of the GPR data and neutron-neutron logs was approximately 2 m. The GPR data were collected using a constant antenna spacing of 1 m, a lateral trace increment of 0.1 m, and a time sampling interval of 0.8 ns. For each recorded trace, 32 stacks were performed to improve the signal-to-noise ratio. Antenna positioning errors and differences in antenna coupling across the profile were estimated to be negligible. Figure 2.7 shows the GPR reflection section after minor pre-processing.

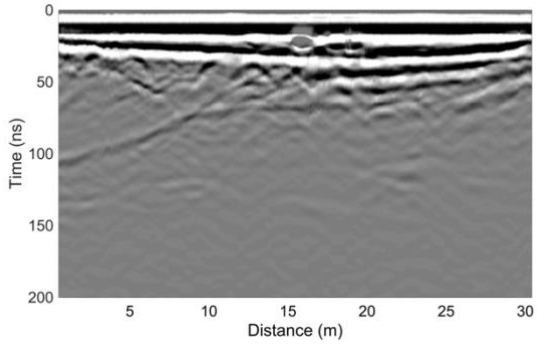


Figure 2.7: Common-offset GPR reflection section from the BHRS after minor preprocessing consisting of time-zero correction and "de-wow" filtering.

2.4.2.2 Velocity estimation

The considered field GPR data were subjected to a processing flow consisting of, in order, time-zero correction, DC shift removal, "de-wow" filtering, removal of the direct air and ground arrivals, correction for the antenna offset, and amplitude scaling. Time-zero was determined based on the first deflection of the data above the ambient noise level. While slightly different approaches are possible, we estimate that the corresponding uncertainties do not exceed ~2 ns. To correct for antenna offset, we used the average velocity of the vadose zone of 0.14 m/ns inferred from previous work (e.g., Bradford, 2008; Bradford et al., 2009). Contrary to our synthetic examples which involved an antenna spacing of 0.5 m, correction for the larger offset between the antennas in the case of the BHRS data was deemed necessary and should lead to negligible differences in traveltime beneath the direct air and ground arrivals compared to the corresponding zero-offset acquisition. Due to the proximity of the direct arrivals to the reflection from the groundwater table, we used a manual surgical mute to remove them as opposed to the average trace subtraction technique considered previously. As was done for the synthetic data, amplitude scaling was performed using a gain function of the form $g(t) = (1 + at)e^{bt}$, where the parameters a and b that best balanced the amplitudes along any given trace were found to be 0.5 ns^{-1} and 0.8 ns^{-1}, respectively. It is important to emphasize that, with this choice of gain function that smoothly varies in time, the relative reflection amplitudes along the GPR traces are importantly well preserved, which would not be the case with the use of an AGC-type amplitude scaling.

Figure 2.8a shows the processed GPR section, to which we then applied the proposed velocity estimation methodology. Following the workflow described in Section 2.2 and illustrated in Figure 2.1, we began with the separation of the diffractions (Figure 2.8b) which, although not evident in the original processed data, turn out to be quite abundant, particularly in the central part of the profile. This was followed by the estimation of the RMS velocity structure (Figure 2.8c) and subsequent Dix inversion to infer the background velocity field (Figure 2.8d). The latter points to the presence of a rather prominent low-velocity zone at intermediate depths in the left-hand side of the profile. The inferred RMS velocity structure was then used to perform time-migration of the GPR section (Figure 2.8e) which, overall, appears to result in an adequate focusing and imaging of the data. An exception are the artefacts introduced into the uppermost part of the section, which are likely related to the harsh surgical muting of the direct wave as well as its potential interference with the neighboring reflection from the groundwater table. Figure 2.8f shows the reflected wavefield that was extracted from the migrated section, which we see to be largely devoid of these artefacts. After estimating the effective mixed-phase source wavelet (Figure 2.8g) using the method of Schmelzbach and Huber (2015), we proceeded to invert the imaged reflected wavefield for the underlying velocity perturbations using a value of $\eta = 10$,

whose choice was based on the typical aspect ratio encountered in similar heterogeneous environments as well as previous BHRS studies (Gelhar, 1993; Xu et al., 2020, 2021). The results, which are shown in Figure 2.8h, clearly depict the dramatic velocity discontinuity associated with the groundwater table. Finally, Figures 2.8i and 2.8j show the superposition of the large-scale background velocity structure (Figure 2.8d) and the inverted small-scale velocity perturbation field (Figure 2.8h) in terms of two-way traveltime and depth, respectively. Note that in deriving the latter, we also accounted for some mild topographic variations that were present along the profile. Note that, in the case of significant topographic variations, such variations would need to be corrected for earlier in our analysis procedure.

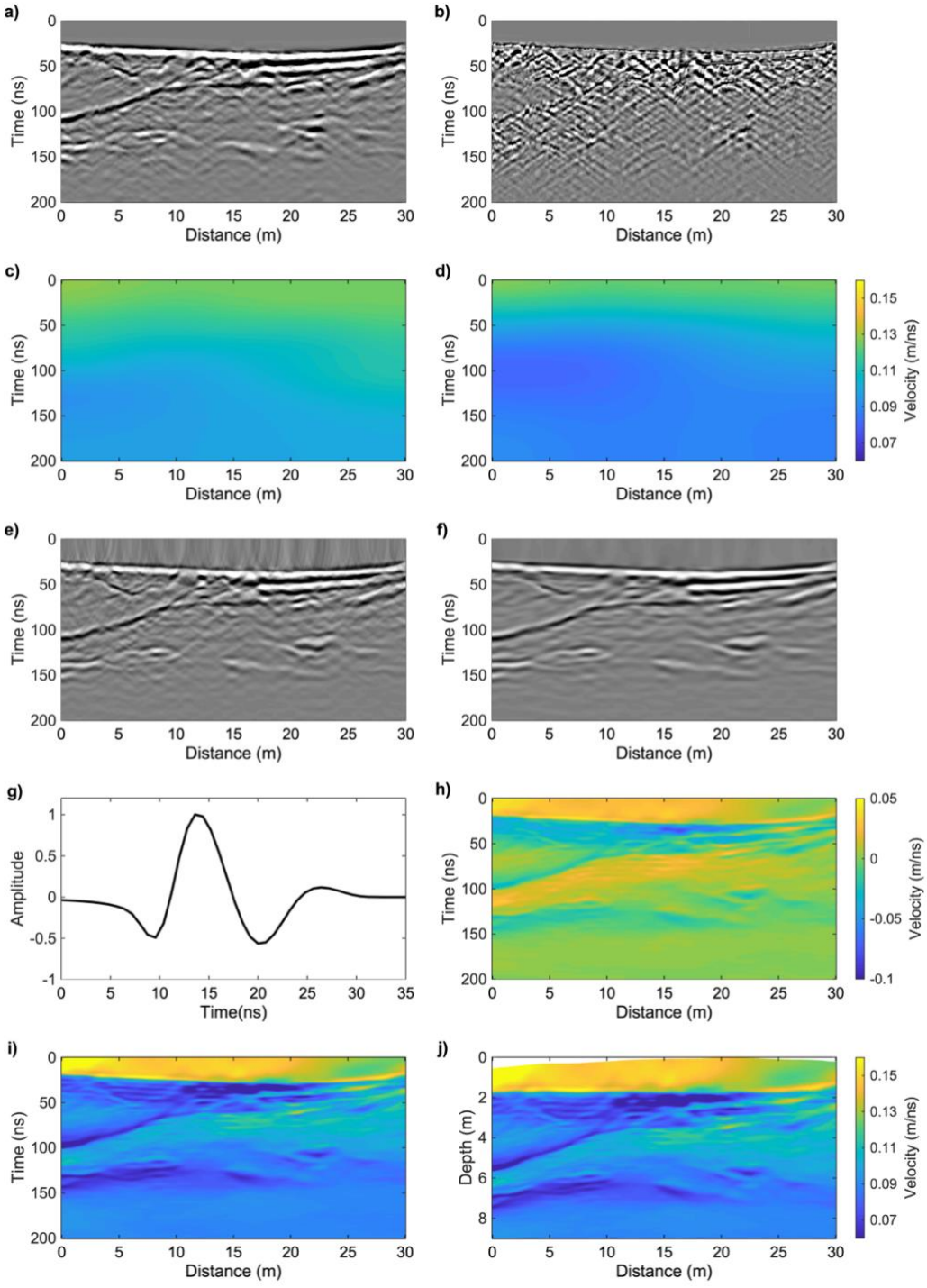


Figure 2.8: Velocity estimation process for the BHRS field data presented in Figure 2.7. From the a) processed GPR section, the b) diffracted wavefield is separated and used to estimate the c) RMS velocity

structure. Using Dix inversion, the d) low-frequency background velocity field v_0 (x,t) is obtained. e) Time-migrated GPR section based on the RMS velocity structure, from which the f) reflected wavefield is obtained. g) Estimated GPR source wavelet. h) Velocity perturbation field Δv(x,t) obtained by inverting the reflected wavefield. The final estimated velocity structure (background+perturbation) is shown in terms of i) traveltime and j) depth.

Although the inferred velocity distribution presented in Figure 2.8j is clearly dominated by the sharp transition from high to low velocities across the water table, the underlying saturated zone shows a significant degree of velocity heterogeneity. This heterogeneity largely emulates the structure depicted by the reflected wavefield in Figure 2.8d and, as such, is geologically plausible. To further assess the realism of our results, we compare them with the neutron-neutron porosity logs available in the saturated zone for boreholes B5, A1, and B2 (Figure 2.6). To this end, we transform the porosity logs to GPR velocity v using a standard petrophysical mixing model (e.g., Huisman et al. 2003)

$$v = \frac{c}{\sqrt{\varepsilon_r^s}(1-\phi) + \sqrt{\varepsilon_r^w}\phi}, \tag{19}$$

where $c =$ 0.3 m/ns is the speed of light in free space, ϕ is the porosity, and $\varepsilon_r^s = 4.6$ and $\varepsilon_r^w = 80$ are the relative dielectric permittivities of the dry solid matrix and water, respectively. A relative dielectric permittivity of 4.6 for the dry matrix corresponds to the average value for quartz (e.g., Schön, 2015) and, as such, is widely regarded as being suitable for alluvial environments in general and the BHRS in particular. Indeed, Dafflon et al. (2009) demonstrated the overall suitability of a relative dielectric permittivity of 4.6 for the solid matrix at the BHRS. While variations in this parameter on the order of 10 to 15% are conceivable, the associated uncertainties are minor and, hence, largely irrelevant compared to other sources of uncertainty in our inversion results and the neutron-neutron logs.

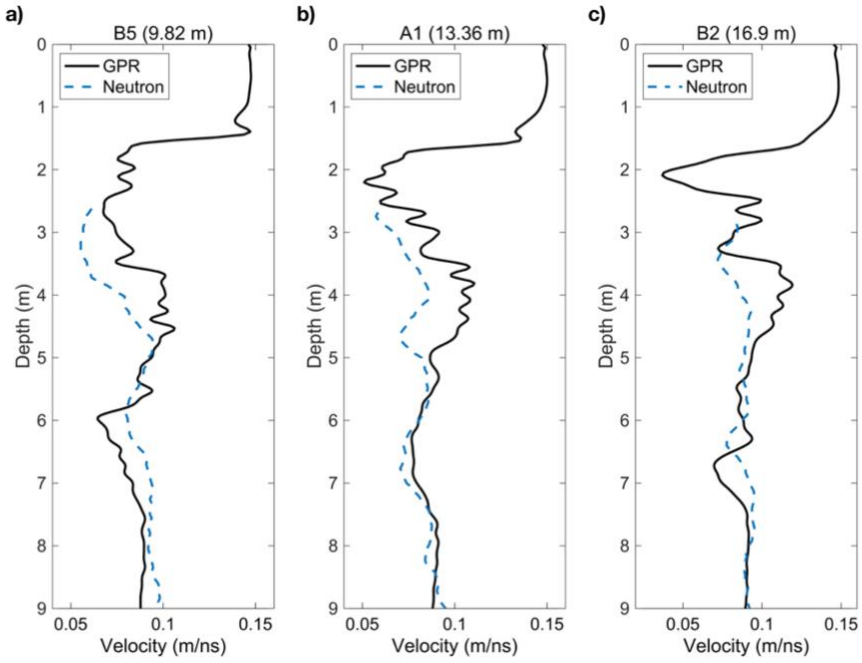


Figure 2.9: Comparison of the velocity estimated from the common-offset GPR reflection data from the BHRS along boreholes a) B5, b) A1 and c) B2 (black solid lines) with the corresponding converted neutron-neutron porosity logs (blue dashed lines).

Figure 2.9 shows the comparison between the GPR-derived velocity and the converted porosity logs at the borehole locations. Overall, we see that the curves are in good agreement, not only in terms of the trend, but also with regard to the absolute values. In this context, it is important to note that the inferred velocity estimates are at least as accurate as those previously inferred from multi-offset GPR reflection tomography (Bradford et al., 2009), while at the same time exhibiting a significantly higher resolution. Arguably, the most conspicuous mismatch between the GPR- and neutron-neutron-based velocity profiles is a seemingly systematic depth shift of ~+0.5 m of the former with regard to the latter, which was recently discussed by Xu et al. (2021). This could be related to the depth calibration of the neutron-neutron logs and/or to a systematic overestimation of the GPR velocity in the vadose zone, the latter of which is the most poorly constrained part of our inferred velocity model due to partial interference between the direct arrivals and the reflection from the water table. Conversely, this apparent mismatch is unlikely to be related to the time-zero determination, whose uncertainty is estimated to be on the order of 2 ns.

2.5 Discussion and Conclusions

We have presented in this paper a novel method for estimating the detailed high-frequency electromagnetic velocity distribution in the shallow subsurface from surface-based common-offset GPR reflection data. The smooth background component of the velocity structure is estimated from the diffracted part of the recorded wavefield, whereas the superimposed small-scale fluctuations are inferred from the associated reflected component. An important and distinguishing feature of our methodology is that, in contrast to previous related approaches (e.g., Schmelzbach et al. 2012; Liu et al. 2018; Xu et al. 2021), it does not require any borehole calibration and/or conditioning information. It does, however, inherently rely upon the presence of diffractions in the GPR data. In this regard, it is important to note that, although diffractions are often not immediately obvious in a GPR profile, they can become much more evident after wavefield separation. This is clearly illustrated in Figure 2.8.

The proposed technique was tested and validated on synthetic data corresponding to two velocity models of differing complexity and realism: one an idealized layered model containing a small number of discrete diffractors, and the other a stochastic facies-based model emulating the typical heterogeneity observed in surficial alluvial environments (e.g., Gelhar 1993; Tronicke et al. 2004; Tronicke and Holliger 2005). These synthetic tests not only illustrate the fundamental validity and robustness of our method, but also allow us to identify a number of features that merit attention during its application. Successful validation of our approach on the BHRS field data further illustrates its capacity for estimating complex velocity structures.

Results for our synthetic test involving the stochastic subsurface model (Figures 2.4 and 2.5) showed a loss of accuracy in the shallowest part of the inferred velocity distribution due to the elimination of the direct air and ground arrivals, which removed important reflections from this zone. Processing techniques used for this purpose, such as subtracting the average trace in a corresponding time window, will thus affect the inferred velocity model over an initial depth range of roughly 1-2 dominant wavelengths (Figures 2.4 and 2.5). Another interesting feature emerging from the stochastic synthetic example is the fact that we fail to fully resolve the thin low-velocity zone between ~2 m and ~3.5 m depth, notably in the central and right-hand side of the profile. A bias in the estimated background velocity distribution over these depths (Figure 2.5d) is likely the cause of this result. Given that this region has a density of diffractions that is comparable to the rest of the model, this may point to the inherently limited resolution of the inferred background velocity field. Under ideal circumstances, diffraction-based velocity analysis can be expected to achieve a resolution on the order of one

dominant wavelength, which for the considered synthetic data is of the order of ~1 m. In practice, however, the achievable resolution critically depends on the so-called smoothing radius parameter, which controls the regularization of a number of steps in the diffraction velocity analysis procedure (Fomel et al., 2007). As recently illustrated by Yuan et al. (2019), a smoothing radius that is too small leads to unstable estimates of the diffraction-derived velocity model, whereas one that is too large will lower unnecessarily its resolution. While we made every effort to determine an optimal value of the smoothing radius for all data considered in this study, there may be regions in the final velocity model where the large- and small-scale components inadequately complement each other due to the limited resolution of the former.

Arguably, the most important criterion that must be fulfilled for our method to perform satisfactorily is the presence of an ample amount evenly distributed diffractions throughout the recorded constant-offset GPR section. Given the inherent heterogeneity of the shallow subsurface (e.g., Gelhar, 1993; Rubin and Hubbard, 2006; Dafflon et al., 2009; Xu et al., 2021), this condition is likely to be fulfilled even if the diffracted energy is not directly obvious in the original recorded data (Figures 2.5 and 2.8). Nonetheless, in the case where an sufficient amount of diffracted energy cannot be retrieved through wavefield separation and/or where the diffraction events are highly unevenly distributed throughout the probed subsurface region, standard common-midpoint-type analyses may still be used to estimate the large-scale velocity structure with our methodology. Under these circumstances, only the estimation of the background field would change, and the inversion for the small-scale velocity fluctuations would remain the same.

Two final assumptions upon which our method relies are that the recorded GPR wavefield is largely non-dispersive and is dominated by single scattering. This allows us to use a convolutional model to describe the reflection data, which in turn permits us to pose the velocity perturbation estimation procedure as a highly efficient linear inverse problem. Limited signal attenuation, and thus limited dispersion, is a prerequisite for acquiring surface-based common-offset GPR reflection data of adequate quality and depth of penetration. The practical validity of this assumption is notably underscored by the success of numerous studies explicitly relying upon an adequate estimation of the GPR source wavelet (e.g., Schmelzbach et al., 2012; Schmelzbach and Huber, 2015; Liu et al., 2018; Xu et al., 2021). While it is theoretically conceivable that there exist environments where multiple scattering becomes sufficiently important in GPR studies, the results of extensive testing of the convolutional model on synthetic and field data suggest that the effects of multiples are largely negligible in near-surface environments (e.g., Irving et al., 2009; Schmelzbach et al., 2012; Xu et al., 2020,2021). Indeed, in such environments, the combination of small reflection coefficients and signal attenuation due to conductivity-related losses means that multiply reflected energy is not strong. The latter is consistent with a methodological study involving acoustic waves in strongly heterogeneous environments, where it was found that multiple scattering only becomes important for strong local velocity fluctuations, corresponding to standard deviations of the order of 10% and more, in combination with long propagation paths in excess of ~20 dominant wavelengths (e.g., Holliger, 1997). These conditions are generally not fulfilled for surface-based constant-offset GPR reflection data.

An important characteristic of this work is that the proposed methodology is rather straightforward. After basic processing of the GPR data in MATLAB, wavefield separation, diffraction velocity analysis, and Dix inversion are carried out using the Madagascar software package, which is well-established for this purpose. The subsequent wavelet estimation and L1-norm inversion are then again performed in MATLAB. For all of the datasets considered in this study, the total time required to complete all of the steps in our workflow is on the order of one day. The IRLS inversion procedure itself proved to be stable and to converge to consistent estimates of the velocity perturbation field after approximately five iterations. In practical terms, the latter

amounted to less than one minute of CPU time on a modest laptop computer. Finally, the inherent computational efficiency of the convolutional model used in our approach implies that the extension of the proposed method to 3D is conceptually straightforward. The only challenge that we anticipate in this regard is the still somewhat limited practical maturity of 3D diffraction velocity analysis techniques (e.g., Merzlikin et al. 2017; Bauer et al., 2020). A direct benefit of 3D analysis is that errors introduced into the background velocity estimation procedure by out-of-plane diffractions can be avoided.

Chapter 3

Weighted diffraction-based migration velocity analysis of common-offset GPR reflection data

---

Yu Liu, James Irving, Klaus Holliger

Resubmitted after revision to *IEEE Transactions on Geoscience and Remote Sensing*

3.1 Abstract

Migration focusing analysis of diffractions is an increasingly important tool for estimating the large-scale subsurface velocity structure from surface-based common-offset ground-penetrating radar (GPR) reflection data. We present a weighting strategy whose aim is to improve the reliability of estimations of the root-mean-square (RMS) velocity obtained using a local semblance focusing measure. In this regard, we increase the resolution of the inferred semblance spectra through a weighting function that varies in accordance with the sensitivity of a diffraction curve to changes in velocity. The weighting function is derived from coherency and slope attributes of the diffracted wavefield components. To demonstrate the viability of our proposed method, we consider its application in two synthetic test cases and one field GPR dataset. Compared with conventional unweighted local semblance spectra, their weighted counterparts allow for a significantly increased resolution and correspondingly reduced picking uncertainty.

3.2 Introduction

Estimating a reliable model of the velocity distribution in the subsurface is a critical step of the seismic and ground-penetrating radar (GPR) reflection processing and imaging workflows. Among the various methods available, diffraction-based velocity analysis is particularly useful when no offset-dependent information is available, as is notably the case for typical surface-based GPR reflection measurements acquired using a single bistatic, common-offset transmitter-receiver antenna configuration (e.g., Grasmueck et al., 2005; Yuan et al., 2019).

There are generally three approaches available to infer the subsurface velocity structure based on diffractions, all of which originated in seismic data processing. The first one involves migration focusing analysis of diffraction events. Since a diffraction migrated with the correct velocity will collapse to a point at its apex, Harlan et al. (1984) proposed to remove reflections from the data and to estimate the velocity structure through an evaluation of diffraction focusing as a function of different migration velocities. In this regard, Fomel et al. (2007) performed diffraction separation using the so-called plane-wave destruction method and evaluated diffraction focusing based on velocity continuation and local kurtosis techniques. Burnett and Fomel (2011) extended this procedure for 3D azimuthally anisotropic velocity analysis, whereas Decker et al. (2017) further decomposed the diffractions into slope components and estimated the velocity based on the local semblance of diffractions. The second approach for diffraction-based velocity analysis is to examine diffraction events in the post-migration dip-angle domain. In dip-angle common image gathers, specular reflections appear as hyperbolic events centered at the reflector dip and curving upward, even when over- or under-migrated. Conversely, diffractions will be flat in this domain when using the correct migration velocity, or curve upward or downward in the case of over- and under-migration, respectively. Based on this feature, Reshef and Landa (2009) performed migration velocity analysis in the dip-angle domain by measuring diffraction flatness. Klokov and Fomel (2012) further proposed the use of the hybrid Radon transform to improve diffraction separation and velocity analysis in the dip-angle domain. Finally, the third approach for diffraction-based velocity analysis is to perform diffraction wavefront tomography. Based on the idea that the wavefront attributes together with traveltimes can be used to locally approximate the kinematic response of reflection points (Duveneck, 2004), Bauer et al. (2017) performed ray-based tomographic inversion for zero-offset seismic data, where wavefront attributes were extracted directly from the diffraction section. Recently, Preine et al. (2020) applied diffraction wavefront tomography to single-channel marine seismic data acquired in a

volcanic environment.

Arguably, migration focusing analysis is the most commonly used method of diffraction-based velocity analysis for common-offset GPR data (e.g., Novais et al., 2008; Clair and Holbrook, 2017; Yuan et al., 2019, 2020; Dong et al., 2020; Li and Zhang, 2021). In this regard, diffraction semblance has proven to be an effective focusing indicator (e.g., Liu et al., 2022). Diffraction semblance can be regarded as the normalized squared correlation of the diffracted event with a constant (Fomel, 2009). This inherently assumes that there is no amplitude or phase variation along the hyperbolic trajectory. When this assumption is violated, for example due to the effects of attenuation or the presence of noise, diffraction semblance may no longer provide robust velocity estimates. A similar problem exists in classic common mid-point (CMP) velocity analysis, where semblance is computed along flattened reflection hyperbolae (e.g., Taner and Koehler, 1969) and becomes less robust in the presence of strong variations in reflection amplitude. In the latter case, the problem was addressed by introducing semblance weighting terms (Luo and Hale, 2012; Chen et al., 2015), which allow for an improvement in the resolution of CMP-based velocity spectra by placing more weight on data that are more sensitive to changes in velocity in the semblance calculation.

In this paper, we develop a diffraction semblance weighting function in order to enhance diffraction-based velocity analysis for common-offset GPR reflection data. Linear coherency analysis (e.g., Schwarz and Gajewski, 2017; Schwarz, 2019) is used to estimate the local slopes and coherencies of the diffracted wavefield components, which in turn are used to derive the weights. We begin by presenting the methodological background of our approach. We then assess the viability of the proposed technique by testing it on two synthetic common-offset GPR data examples. Finally, the method is applied to a field GPR dataset from a typical surficial alluvial environment, where borehole logs are available to assess the results obtained.

3.3 Methodology

Our diffraction velocity analysis procedure begins with the separation of the diffracted wavefield from the common-offset GPR data. To this end, we employ plane-wave destruction (PWD) filtering (Fomel, 2002; Chen et al., 2013) to eliminate the specular reflections. This is accomplished using the programs ‘sfdip’ and ‘sfpwd’ in Madagascar (https://reproducibility.org/), an open-source data analysis package. Once the diffracted wavefield has been isolated, the data are ready for migration focusing analysis to estimate the corresponding root-mean-square (RMS) velocity structure.

Migration focusing of diffractions can be measured by the diffraction semblance, which is defined as

$$s(t,x,v) = \frac{\{F_v[a(t,x)]\}^2}{F_v[a^2(t,x)]}, \tag{3.1}$$

where $a(t,x)$ is the amplitude of the diffracted wavefield with t and x denoting the time and space coordinates, respectively, and F_v represents the time migration operator for a constant velocity v. Diffraction semblance is a normalized coherency measure performed along a hyperbolic trajectory. More generally, we can consider the weighted diffraction semblance, which is defined as

$$s_w(t,x,v) = \frac{\{F_v[a(t,x)w(t,x)]\}^2}{F_v[a^2(t,x)]F_v[w^2(t,x)]}. \tag{3.2}$$

The weighted semblance can be regarded as the squared correlation between the diffracted wavefield and the weighting function $w(t,x)$. When $w(t,x)$ is equal to a constant, Eq. (3.2) is equivalent to Eq. (3.1).

For CMP-type seismic velocity analysis, Luo and Hale (2012) showed that the use of a weighted semblance function, where greater emphasis is placed on parts of the normal moveout (NMO) curve that are particularly sensitive to velocity changes, can significantly increase the resolution of the velocity spectra obtained. To improve the resolution of diffraction velocity analysis for common-offset GPR data, we similarly aim to develop a weighting function, to be used in Eq. (3.2), that places more emphasis on parts of the diffraction curve that have greater sensitivity to changes in the migration velocity. To this end, we consider a point diffractor whose apex appears on a GPR section at the two-way traveltime τ (Figure 3.1). Assuming that the diffractor is embedded in a homogeneous medium having a value of the squared slowness of $\gamma = 1/v^2$, the traveltime of the diffraction curve in terms of the zero-offset traveltime is approximately given by

$$t = \sqrt{\tau^2 + 4\gamma {x_d}^2}, \tag{3.3}$$

where x_d denotes the horizontal distance between the diffraction apex and the observation location at the surface. Taking the derivative of Eq. (3.3) with respect to the squared slowness, we arrive at

$$\frac{\partial t}{\partial \gamma} = \frac{2{x_d}^2}{\sqrt{\tau^2 + 4\gamma {x_d}^2}} = \frac{2{x_d}^2}{t}. \tag{3.4}$$

We see from Eq. (3.4) that the change in traveltime along the diffraction hyperbola that results from a change in the squared slowness is proportional to the horizontal distance from the apex squared and inversely proportional to the travel time. The distance x_d can be expressed in terms of the local slope p of the diffraction curve, which is equal to the derivative of Eq. (3.3) with respect to x_d. Taking $\partial t/\partial x_d$, solving for x_d, and substituting the result into Eq. (3.4), yields

$$\frac{\partial t}{\partial \gamma} = \frac{tp^2}{8\gamma^2}. \tag{3.5}$$

Eq. (3.5) suggests that our weighting function for diffraction semblance should vary proportionally to the product of the traveltime and the squared local slope of the diffracted wavefield. To estimate the latter, we perform so-called C2 coherency analysis, where we seek to find the value of p that maximizes a linear semblance measure over a particular time and trace window (e.g., Schwarz, 2019)

$$\max_p \frac{1}{n} \frac{\sum_{\delta t} \{\sum_{i=1}^{n} a[t_0 + p(\Delta x_i), x_0 + \Delta x_i]\}^2}{\sum_{\delta t} \sum_{i=1}^{n} a^2 [t_0 + p(\Delta x_i), x_0 + \Delta x_i]}, \tag{3.6}$$

where δt represents the interval over which vertical summation is performed, n is the number of considered neighboring traces, position (t_0, x_0) is where the slope p is estimated, and Δx_i denotes the horizontal distance from position (t_0, x_0). The C2 coherency analysis procedure based on Eq. (3.6) yields not only a slope attribute $p(t, x)$, but also a maximum-semblance-based coherency attribute $c(t, x)$. For parasitic slopes related to random noise, this coherency value will be small because of the inherent lack of continuity. Conversely, for the diffracted parts of the wavefield, the coherency attribute will take on a high value, even in the case of weak amplitudes, because the hyperbolae can be locally approximated by continuous linear events.

To formulate our weighting function for diffraction semblance velocity analysis, which must be defined everywhere across the diffraction section, we first evaluate Eq. (3.5) using the local slope estimate derived from the C2 coherency analysis procedure. Next, to avoid the consideration of regions that do not correspond to coherent diffraction events, we set to zero all parts of the section where the estimated coherency value is below a chosen threshold c_0, which is determined empirically. This masking procedure can be expressed mathematically as

$$u(t, x) = f(t, x) \cdot H[c(t, x) - c_0], \tag{3.7}$$

where $u(t, x)$ denotes the masked data, $f(t, x)$ the results of Eq. (3.5), and H the Heaviside step function.

Finally, we tailor our weight function $w(t,x)$ such that we can easily control the degree of weighting of diffraction events based on Eq. (3.5). This is done by normalizing $u(t,x)$ by its maximum value u_{max} and introducing the balance parameter λ, which can take on values between 0 and 1

$$w(t,x) = 1 - \lambda + \lambda \left[\frac{u(t,x)}{u_{max}}\right]. \tag{3.8}$$

When $\lambda = 0$, we see that $w(t,x) = 1$, meaning that the weighted semblance in Eq. (3.2) is equivalent to the normal semblance described by Eq. (3.1). Conversely, when $\lambda = 1$, the weighting is based entirely on the sensitivity relationship expressed by Eq. (3.5).

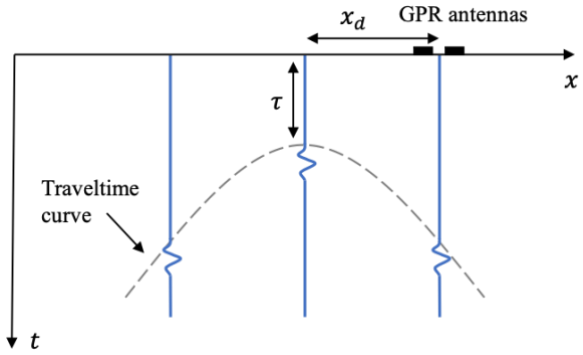


Figure 3.1: Schematic illustration of a diffraction whose apex appears at two-way traveltime τ in a common-offset GPR section. Parameter x_d denotes the horizontal distance between the considered observation point at the surface and the diffraction apex.

To perform the diffraction semblance velocity analysis, Eq. (3.2) is evaluated over a suite of velocity values. This yields a *t*-*x*-*v* cube of weighted diffraction semblance values. Using the automatic picking program 'sfpick' in Madagascar [18], the maxima on each *t*-*v* panel are selected and a 2D RMS velocity model is generated. This result, if desired, can then be transformed to a corresponding interval velocity model using the constrained Dix inversion program 'sfdix' (Fomel and Guitton, 2006). The flowchart in Figure 3.2 illustrates the essential elements of our proposed approach.

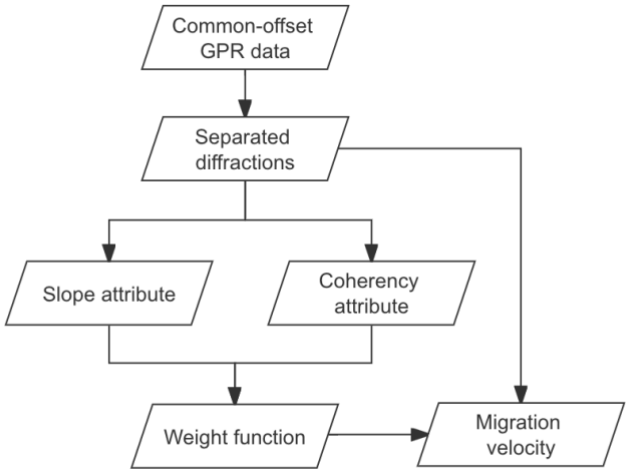


Figure 3.2: Flowchart illustrating the essential elements of the proposed diffraction semblance weighting strategy for estimating the subsurface velocity structure from surface-based common-offset GPR reflection data.

3.4 Synthetic data examples

3.4.1 Vertical gradient model

We first show the application of our diffraction velocity analysis strategy to a synthetic example characterized by a negative velocity gradient with depth, which is commonly encountered in surficial environments in response to a gradual increase of the soil water content (Figure 3.3a). Ten point-type diffractors are distributed randomly throughout this model, for which we generate synthetic common-offset GPR reflection data using the gprMax software (Warren et al., 2016), which solves Maxwell's equations using the finite-difference time-domain (FDTD) method. The subsurface dielectric permittivity ε is assumed to be frequency-independent and obtained from the GPR velocity v using the low-loss approximation $v \approx 1/\sqrt{\varepsilon\mu}$ (Annan, 2005). The magnetic permeability μ is assumed to be equal to its value in free space, and the electrical conductivity is fixed at a constant value of 1 mS/m. The source current is defined as Ricker wavelet having a dominant frequency of 200 MHz which, in turn, results in a propagating electromagnetic pulse corresponding to the first derivative. The distance between the transmitter and receiver antennas is set to 0.5 m and traces are simulated every 0.1 m along the survey profile. The fact that the antenna offset is not strictly zero leads to an inherent underestimation of the velocity that decreases with increasing depth. For the standard antenna offsets used in surface-based bistatic GPR reflection surveys, the resulting errors can be generally regarded as negligible from a practical point of view.

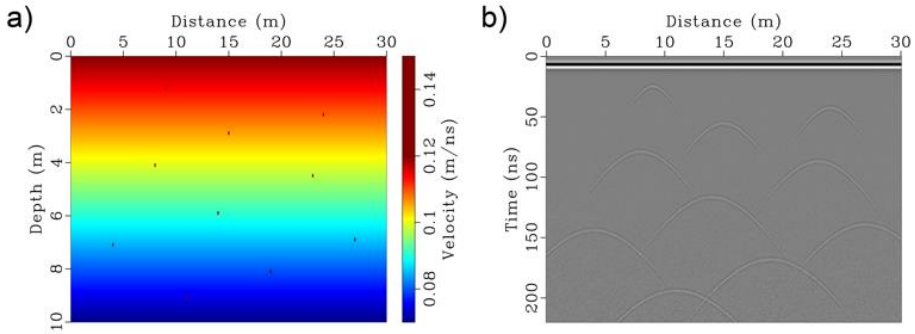


Figure 3.3: a) Velocity model characterized by a negative velocity-depth gradient. Ten point-type diffractors are distributed randomly throughout the model. b) Corresponding synthetic common-offset GPR reflection data with 5% Gaussian random noise added.

The resulting synthetic GPR data are contaminated with 5% Gaussian random noise with regard to the maximum absolute amplitude below the direct wave, and are subjected to a standard signal processing sequence consisting of (i) amplitude scaling to compensate for energy spreading, absorption, and scattering using a gain function consisting of a linear and an exponential part, (ii) elimination of the direct air and ground arrivals by subtracting the average trace calculated over a time window from 0 to 25 ns, and (iii) 20-600 MHz bandpass filtering. Figure 3.3b shows the raw noise-contaminated GPR data after amplitude scaling only. Figure 3.4a then illustrates the final processed GPR data, which, due to the absence of reflecting interfaces in the underlying velocity model, only consist of diffractions and noise.

We next performed C2 coherency analysis described by Eq. (3.6) on the processed GPR data, which leads to the local coherency and slope attribute sections shown in Figures 3.4b and 3.4c, respectively. We see that the estimated coherencies are strong for the diffractions and rather weak and disorganized for the noise, which

allows us to easily identify the diffractions and estimate their slopes. Based on this information, Eq. (3.5), (3.7) and (3.8) were used to calculate the semblance weighting function to be used in diffraction velocity analysis, where balance parameter λ was set to a value of 0.5 (Figure 3.4d). Through empirical testing, we have found that intermediate values of λ lead to the best results in the velocity spectra, and, for consistency, we therefore consider $\lambda = 0.5$ for all of the examples presented in the manuscript. The weights are seen to increase with distance from the apex of each diffraction hyperbola, as the diffraction curves have a greater sensitivity to changes in velocity in these locations.

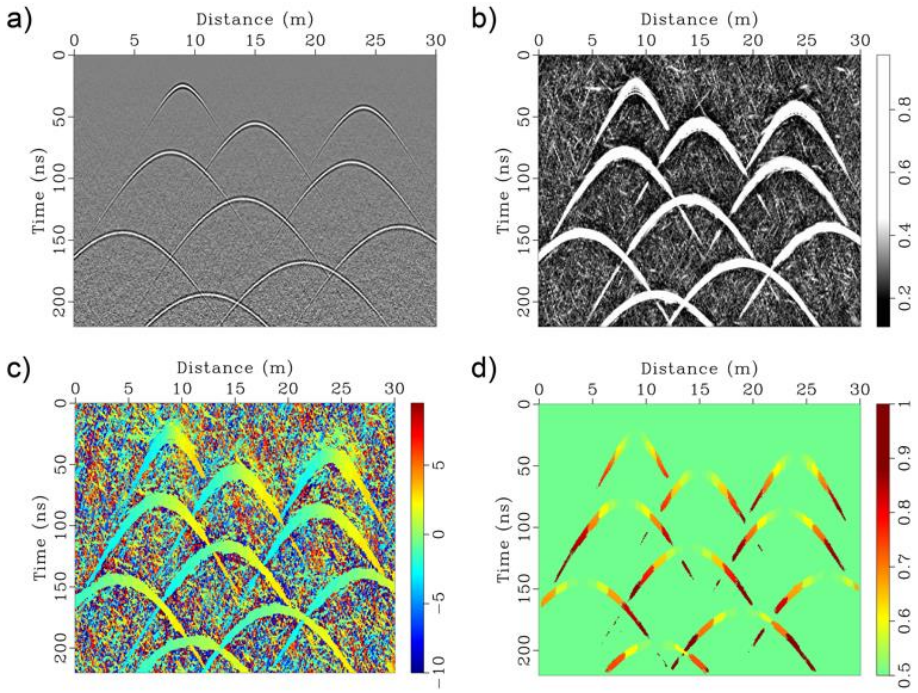


Figure 3.4: Diffraction semblance weight function calculation for the vertical gradient velocity model example presented in Figure 3.3. a) Synthetic common-offset GPR data from Figure 3.3a after processing. b) Corresponding coherency attribute. c) Corresponding slope attribute. d) Inferred weights for diffraction velocity analysis.

We computed both the unweighted and weighted local semblance based on Eq. (3.1) and (3.2) for a suite of velocity values ranging from 0.07 m/ns to 0.13 m/ns. The corresponding velocity spectra for lateral positions of 3, 9, 15, 21 and 27 m are shown in Figure 3.5. Note that the velocity resolution for the weighted panels is significantly higher than that for the unweighted panels, thanks to a heavier emphasis in the semblance calculation on parts of the diffraction hyperbolae that are most sensitive to velocity changes. The weighted panels are clearly more amenable to velocity picking, particularly if done manually. In Figure 3.6a, we show the derived 2D RMS velocity structure for the entire GPR section, which was obtained by automatically picking the maxima on the time-velocity panels for all trace locations considering an average velocity at the surface of 0.12 m/ns. The impact of using different velocities in the immediate subsurface on automatic picking has been investigated by Yuan et al. (2019). In practice, the surficial velocity can be rather easily and reliably estimated based on the linear moveout of the direct ground wave as a function of antenna offset. The result in Figure 3.6a agrees well with the true RMS velocity structure of the vertical gradient model, which is presented in Figure 3.6b. This indicates that the inherent velocity bias related to the non-zero antenna offset of surface-based bistatic GPR reflection surveys is, as mentioned above, unlikely to be of any practical significance. Figure 3.6c shows the corresponding depth-converted interval velocity model, which was obtained via Dix inversion. We see that the derived velocity field is largely consistent with the underlying gradient model

(Figure 3.3a). The discrepancy in the lower part of the profile is related to the relative scarcity of complete diffraction hyperbolae in this part of the synthetic data. Finally, Figure 3.6d shows the migrated GPR image in the depth domain, which was obtained by first performing Kirchhoff time migration using the estimated RMS velocity structure, and then converting from time to depth based on the corresponding interval velocities. We see in the depth image that the diffraction hyperbolae have collapsed effectively into their apexes whose positions are consistent with those in the underlying model (Figure 3.3a).

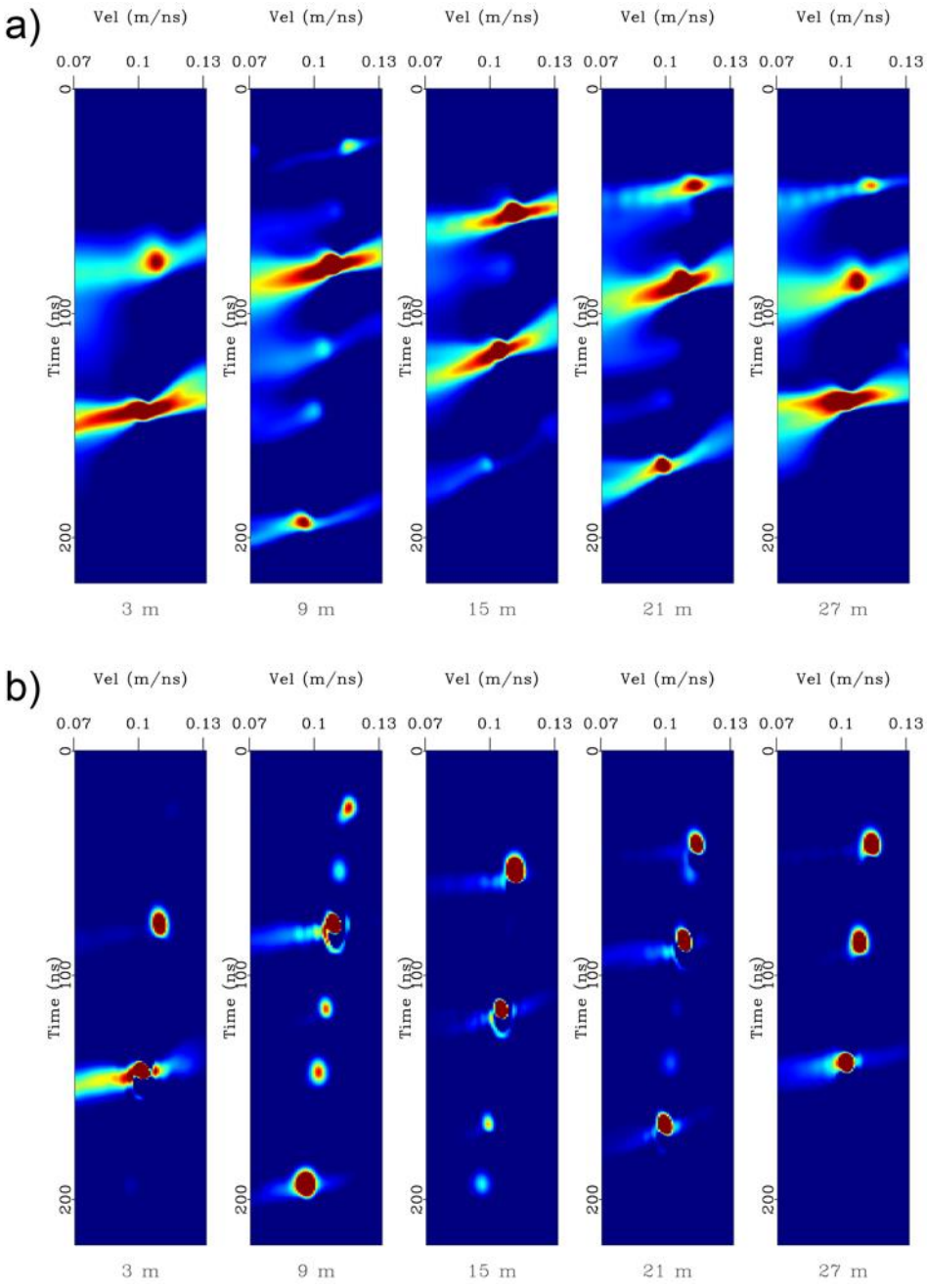


Figure 3.5: a) Unweighted and b) weighted diffraction semblance spectra inferred for the vertical gradient velocity model example at lateral distances of 3, 9, 15, 21, and 27 m.

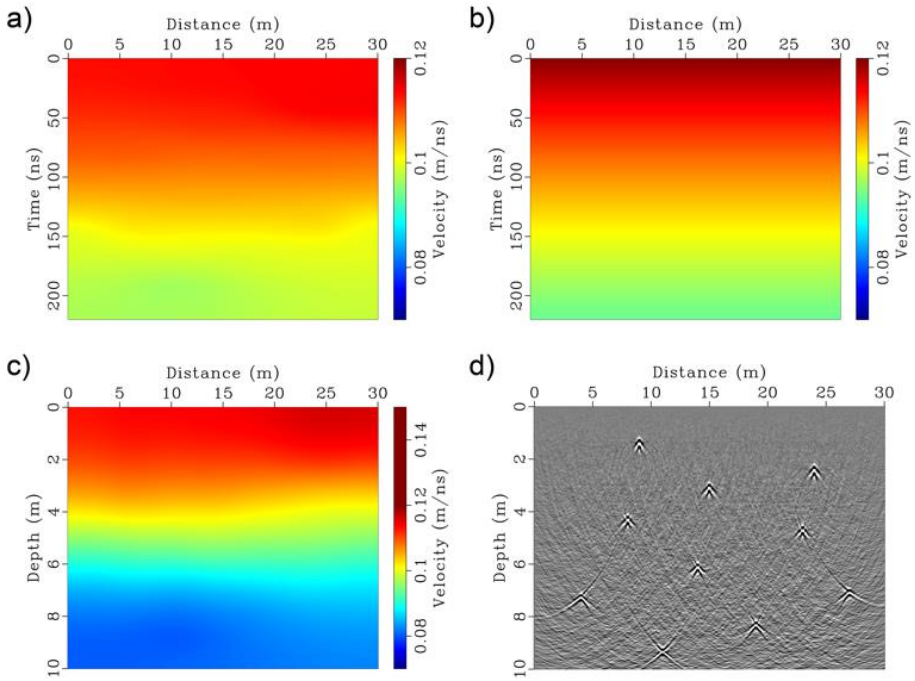


Figure 3.6: a) Estimated and b) true RMS velocity distribution for the vertical gradient velocity model example (Figure 3.3a). c) Estimated interval velocity model obtained by Dix inversion. d) Corresponding migrated depth image of the common-offset GPR data (Figure 3.3b).

3.4.2 Layered model

Our second synthetic example, presented in Figure 3.7a, is based upon a layered velocity model and is very similar to the example considered by Yuan et al. (2019) in a recent GPR diffraction velocity analysis study. The model has two major units of constant velocity, which are separated by a dipping interface. Each of the two constant velocity units contains three diffractors. A thin horizontal layer is also present in the lower part of the second unit. As before, we simulated a common-offset GPR reflection survey, assuming low-loss conditions for the conversion of velocity to dielectric permittivity and using a constant electrical conductivity of 1 mS/m and a magnetic permeability equal to its value in free space. The gprMax FDTD software was again used to perform the GPR simulations with a Ricker wavelet having a dominant frequency of 200 MHz as the source current function, a transmitter-receiver spacing of 0.5 m, and a lateral trace increment of 0.1 m. After contaminating the resulting data with 5% Gaussian random noise with regard to the maximum absolute amplitude below the direct wave, they were subjected to the same signal processing sequence as described previously. Figure 3.7b shows the raw noise-contaminated GPR data after amplitude scaling only, where we observe that, in addition to the diffraction hyperbolae arising from the point-type scatterers, specular reflections corresponding to the layer boundaries are present.

As the data in Figure 3.7b contain both reflections and diffractions, we applied PWD filtering after processing in order to eliminate the specular reflections. Figure 3.8a shows the result, where we see that the diffractions have been preserved and that the reflections have been largely eliminated. In Figures 3.8b and 3.8c, we show the local coherency and slope attributes corresponding to the diffracted wavefield in Figure 3.8a, respectively. In Figure 3.8d, we show the weight function calculated from these attributes based on Eq. (3.5), (3.7) and (3.8), where the balance parameter λ was again set to a value of 0.5. As before, we computed both the unweighted and weighted local semblance for a suite of velocity values ranging from 0.07 m/ns to 0.13 m/ns. The

corresponding velocity spectra for lateral positions of 2, 6, 10, 14 and 18 m are shown in Figure 3.9. Consistent with the results presented previously, the weighted semblance panels are seen to have a considerably higher resolution along the velocity axis and contain fewer artifacts than their unweighted counterparts.

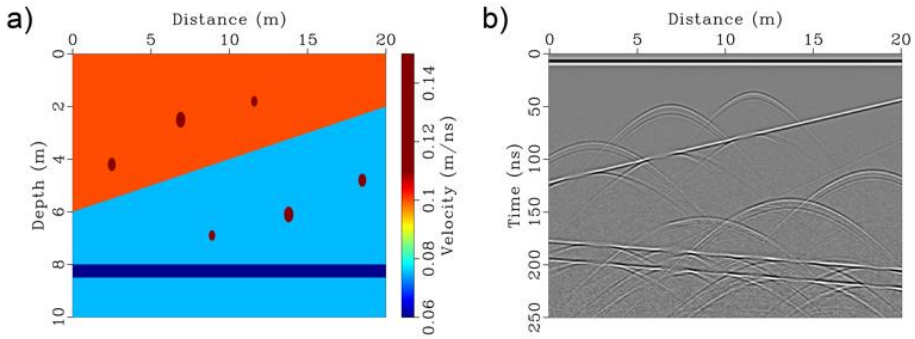


Figure 3.7: a) Layered velocity model containing six point-type diffractors. b) Corresponding synthetic common-offset GPR reflection data with 5% Gaussian random noise added.

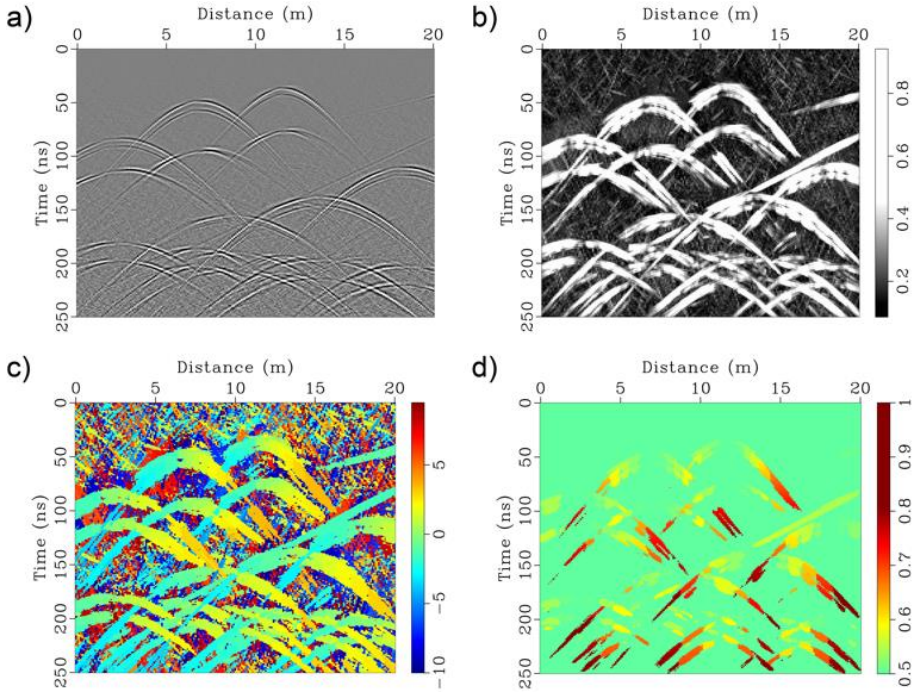


Figure 3.8: Diffraction semblance weight function calculation for the layered velocity model example presented in Figure 3.7. a) Synthetic common-offset GPR data from Figure 3.7b after processing and diffraction separation. b) Corresponding coherency attribute. c) Corresponding slope attribute. d) Inferred weights for diffraction velocity analysis.

Using automatic velocity picking on all of the weighted semblance panels assuming an average velocity at the surface of 0.1 m/ns, the 2D RMS velocity model presented in Figure 3.10a is generated. This result compares reasonably well with true RMS velocity structure of the layered model, which is presented in Figure 3.10b. The estimated RMS velocity structure was then used to estimate interval velocities in the time domain through Dix inversion (Figure 3.10c). Finally, Figure 3.10d shows the corresponding depth image of the GPR data. From this depth image, we see that the position of dipping interface is imaged satisfactorily, that the diffracted parts of the wavefield have been well focused, and that the thin layer is well resolved with its geometric characteristics closely emulating those of the underlying model with the exception of a mild pull-down towards

the right-hand side. The latter results from errors in the estimation of the RMS and interval velocity distributions in the overlying part of the model due the relative scarcity of diffractors. Compared to the migrated images of Yuan et al. (2019), the positioning of the prevailing structures has been significantly improved.

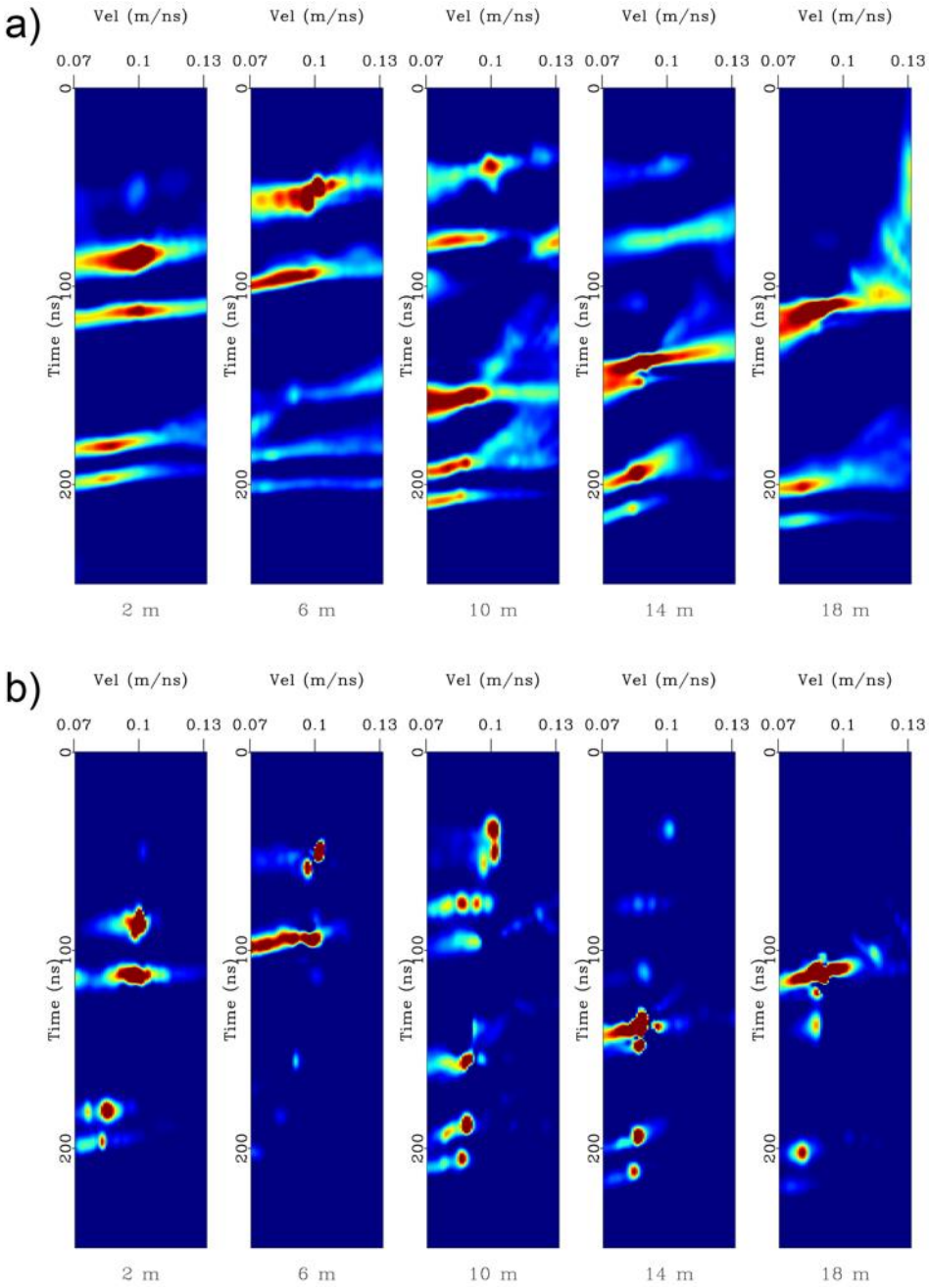


Figure 3.9: a) Unweighted and b) weighted diffraction semblance spectra inferred for the layered velocity model example at lateral distances of 2, 6, 10, 14, and 18 m.

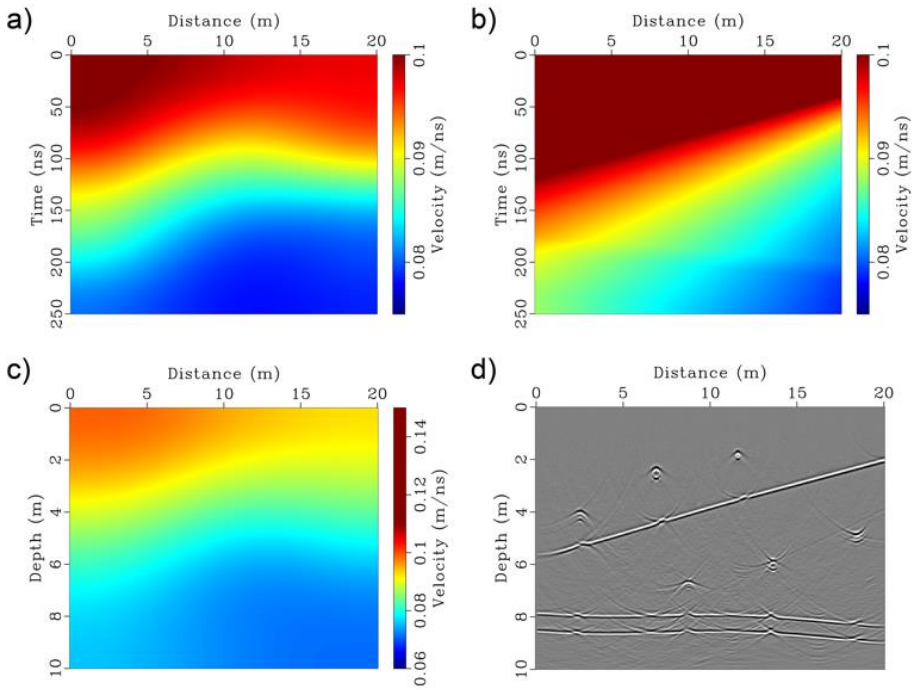


Figure 3.10: a) Estimated and b) true RMS velocity distribution for the layered velocity model example (Figure 3.7a). c) Estimated interval velocity model obtained by Dix inversion. d) Corresponding migrated depth image of the common-offset GPR data (Figure 3.7b).

3.5 Field data examples

We now consider the application of our proposed methodology to constant-offset GPR reflection data acquired at the Boise Hydrogeophysical Research Site (BHRS), which is located on a gravel bar adjacent to the Boise River near Boise, Idaho, USA (Figure 3.11). The corresponding aquifer is unconfined and consists predominantly of late Quaternary fluvial deposits dominated by gravel and sand. The groundwater table is, with some seasonal variations, located around 2 m depth. A layer of red clay at approximately 18 m depth acts as an aquitard and forms the base of the aquifer (Barrash & Clemo 2002). Over the past two decades, a wide range of geophysical and hydrogeological studies have been conducted at the site (e.g., Bradford, 2009; Xu et al., 2021).

The considered GPR data are from an inline profile acquired in 1998 during a 3D survey at the BHRS using a PulseEkko IV system (Sensors & Software Inc.) with a nominal antenna frequency of 200 MHz. The 30-m-long survey line crosses three boreholes, B5, A1, and B2, where neutron-neutron porosity logs are available below the groundwater table, which was located at a depth of ~2.8 m at the time of the survey. The spacing between the transmitter and receiver antennas was set to 0.5 m and traces were acquired every 0.1 m along the profile line. The corresponding time sampling interval was 0.8 ns, and 32 stacks were performed at each trace location in order to improve the signal-to-noise ratio. In the following, we exclusively focus our analysis to the saturated part imaged by the GPR data.

Figure 3.12a shows the raw BHRS GPR reflection data. These data were subjected to a typical signal processing flow consisting of time-zero correction, “de-wow” filtering, surgical mute of the direct air and ground arrivals, and amplitude scaling using a gain function consisting of a linear and an exponential part (Figure 3.12b). Next, we applied PWD filtering to isolate the diffracted part of the recorded wavefield (Figure

3.13a). Though not evident in the original data after standard processing, the diffractions do indeed turn out to be quite abundant after wavefield separation, particularly in the left part of the profile.

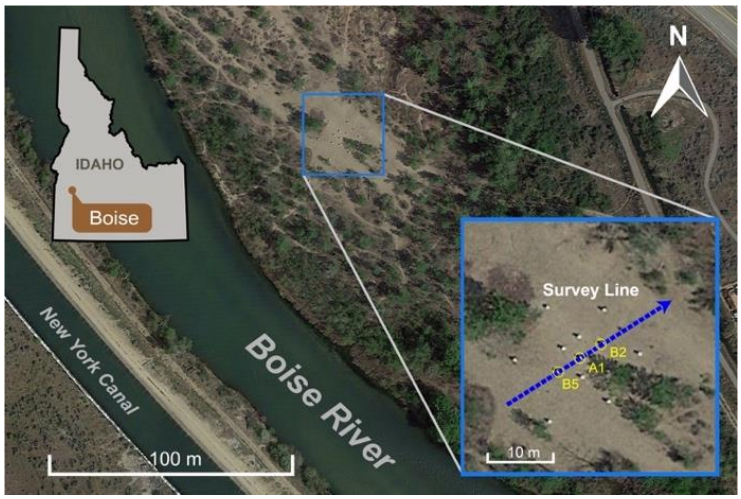


Figure 3.11: Map of the BHRS (43°32'32″ N, 116°05'52" W) showing the location of the considered common-offset GPR reflection profile (blue dashed line). The profile is aligned with boreholes B5, A1, and B2 (yellow circles).

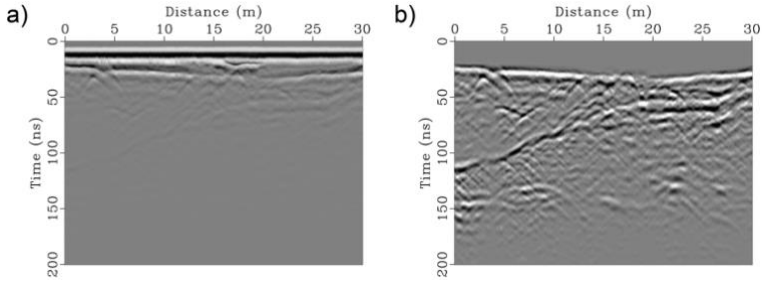


Figure 3.12: a) Raw and b) processed common-offset GPR reflection data from the BHRS.

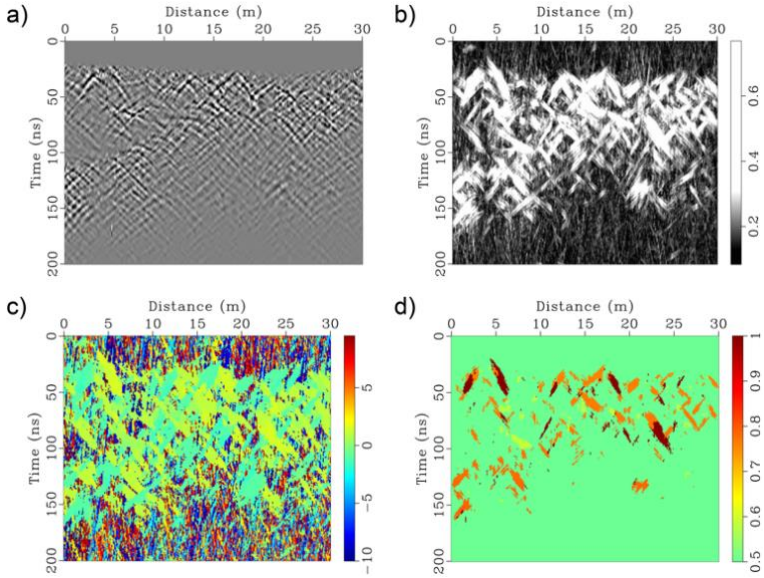


Figure 3.13: Diffraction semblance weight function calculation for the common-offset field GPR reflection data from the BHRS. a) Diffracted wavefield obtained from the processed data shown in Figure 3.12b. b) Corresponding coherency attribute. c) Corresponding slope attribute. d) Inferred weights for diffraction velocity analysis.

In Figures 3.13b and 3.13c, we show the local coherency and slope attributes corresponding to the diffracted wavefield shown in Figure 3.13a. Figure 3.13d shows the weight function calculated for diffraction velocity analysis, where the balance parameter λ was again set to 0.5. We computed both the unweighted and weighted local semblance for a suite of velocity values ranging from 0.07 m/ns to 0.13 m/ns. The corresponding velocity spectra for lateral positions of 3, 9, 15, 21 and 27 m are shown in Figure 3.14. As for the synthetic test cases, we see that the velocity resolution in the weighted panels shows distinct improvements compared to the unweighted panels and that the uncertainties in velocity picking are decreased by the proposed approach. The corresponding estimated 2D RMS velocity structure is shown in Figure 3.15a. Through Dix inversion and time-to-depth conversion, we then obtained the interval velocity model shown in Figure 3.15b. Quite interestingly, this smooth, large-scale diffraction-based velocity model clearly senses the change in water content associated with transition from the vadose zone to the saturated zone at the BHRS. Figure 3.15c shows the corresponding depth image of migrated GPR data.

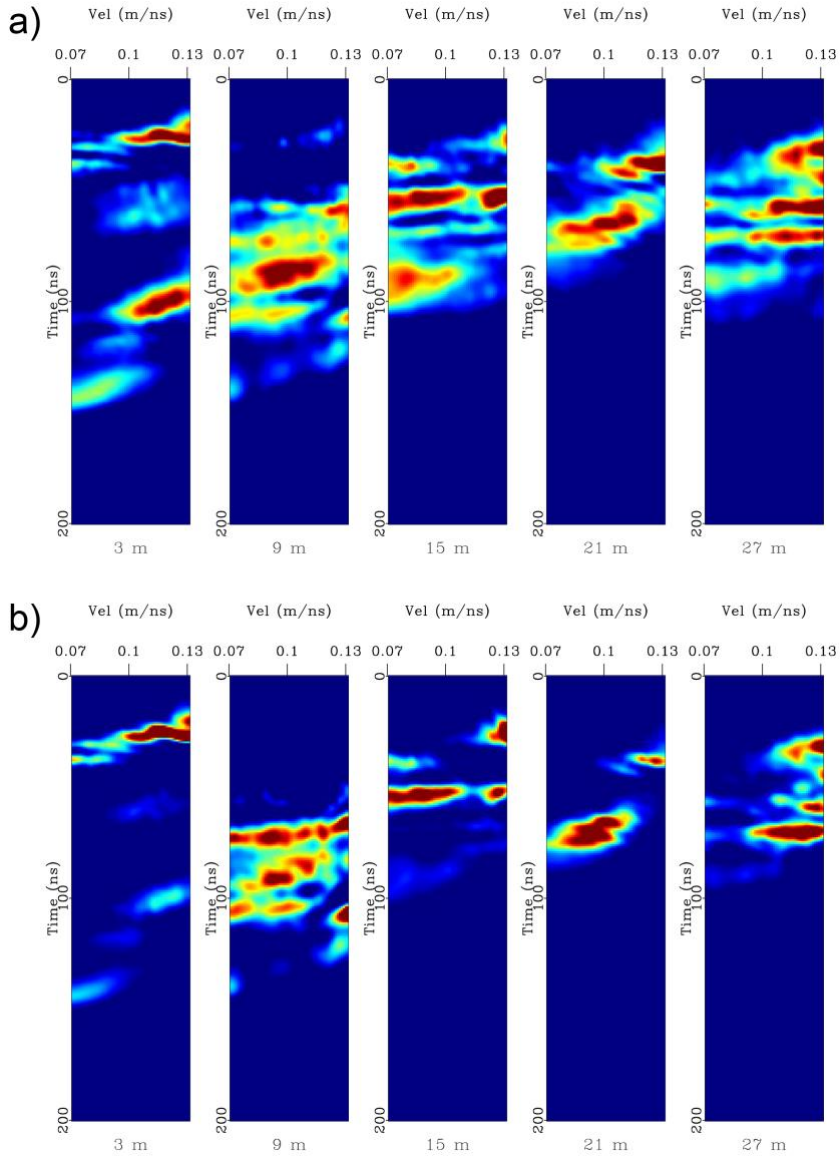


Figure 3.14: a) Unweighted and b) weighted diffraction semblance spectra obtained from the BHRS field data at lateral distances of 3, 9, 15, 21, and 27 m.

To further assess these results of our methodology, we compare the inferred interval velocity structure with the

neutron-neutron porosity logs, which were acquired only within the water-saturated zone, at the borehole locations B5, A1, and B2 (Figure 3.16). To this end, we convert the logged porosities to GPR velocity using a common petrophysical mixing model for the water-saturated part of subsurface (e.g., Huisman et al. 2003)

$$v = \frac{c}{\sqrt{\varepsilon_r^s}(1-\phi) + \sqrt{\varepsilon_r^w}\phi}, \tag{3.9}$$

where $c = 0.3$ m/ns is the speed of light in free space, ϕ is the porosity, and $\varepsilon_r^s = 4.6$ and $\varepsilon_r^w = 80$ are the relative dielectric permittivities of the dry solid matrix and water, respectively. Figure 3.16 shows the comparison between the estimated and logged velocities at the three borehole locations, which is overall quite favorable.

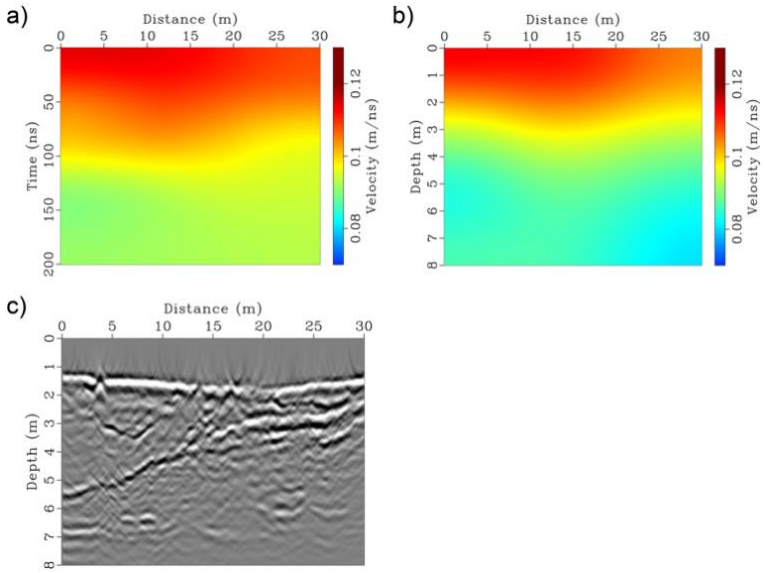


Figure 3.15: a) RMS velocity structure estimated from the BHRS field data. b) Estimated interval velocity model obtained by Dix inversion. c) Corresponding migrated depth image of the BHRS field data (Figure 3.12b).

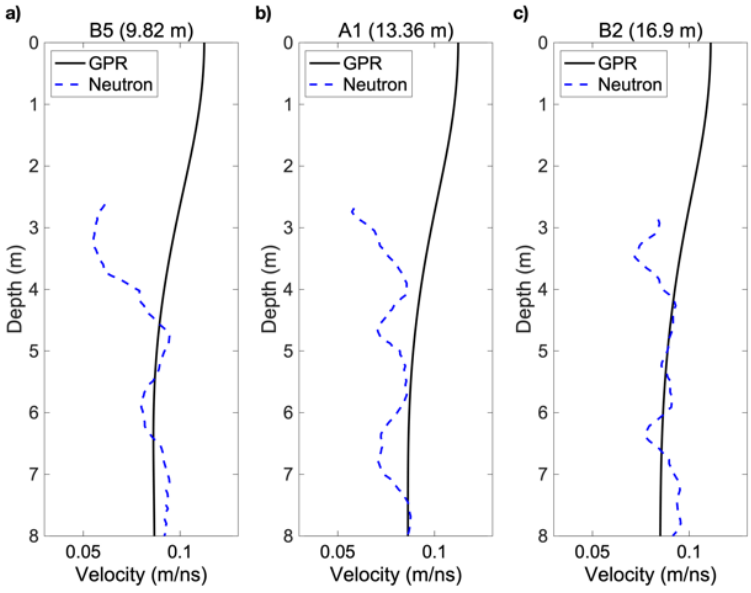


Figure 3.16: Comparison of the velocity profiles estimated from GPR diffraction analysis (black solid lines) at the borehole locations a) B5, b) A1 and c) B2 (Figure 3.11c) with the corresponding velocity profiles

inferred from the neutron-neutron porosity logs acquired within the water-saturated zone (blue dashed lines).

3.6 Discussion and conclusions

We have presented in this paper a weighting strategy to improve the performance of diffraction-based velocity analysis of surface-based common-offset GPR reflection data using local semblance as the focusing measure. Our weighting function is calculated throughout the GPR section from coherency and slope attributes that are derived from linear coherency analysis of the diffracted wavefield. Through its use, far-distance components of diffraction hyperbolae, whose positions are particularly sensitive to the prevailing RMS velocity, are favoured. This leads to increased focusing and resolution of the resulting semblance spectra, which correspondingly reduces the uncertainties for velocity picking. Tests on synthetic common-offset GPR reflection data for two canonical models and subsequent application to field data corroborate the viability of the proposed weighting method.

Our weight function is designed in an analogous manner to Luo and Hale (2012), who estimated the NMO velocity from CMP gathers, where far-offsets tend to be more sensitive to velocity changes and the offset information is available. In our case, however, the information about the diffractor positions and the associated lateral distances of the interesting parts of the diffraction curves is not explicitly available, and we use attribute analysis to retrieve it. Diffraction-based migration velocity analyses provide the smooth large-scale velocity structure of the probed subsurface region, which is needed for subsequent imaging and depth-conversion. Liu et al. (2022) recently illustrated how to combine this information with small-scale velocity fluctuations inferred from the reflected part of the GPR wavefield in order to obtain comprehensive estimates of the detailed velocity distribution. The diffraction velocity analysis component of the latter approach is likely to benefit from the corresponding improvements provided by the methods proposed in the current study. This is expected to be particularly pertinent in the presence of sub-optimal signal-to-noise ratios.

Arguably, the most important condition for a successful application of the proposed methodology is the presence of an ample amount of diffraction hyperbolae that are distributed reasonably evenly throughout the recorded common-offset GPR profile. Due to the inherent heterogeneity of the shallow subsurface, much of the incident GPR energy is indeed scattered/diffracted back to the surface, rather than being reflected from specular interfaces (e.g., Yuan et al., 2019, 2020), meaning that this criterion is often satisfied. If such diffracted energy is missing along parts of the GPR profile, it was observed in our synthetic tests that the corresponding parts of the estimated velocity field will lose resolution. Another critical part of the proposed methodology is the effective isolation of the diffracted wavefield. In this study, standard PWD filtering was considered, but we expect that comparable results would be obtained through alternative approaches, such as, for example, coherent wavefield subtraction (Schwarz et al., 2017). In this regard, we have observed that much of the diffracted energy on a GPR section only becomes clearly visible after this step.

Although our weighting strategy contributes to increase the resolution of velocity spectra, there are still inherent uncertainties in the velocity picking process. For the automatic picking code ‘sfpick’, for example, two key input parameters required are the velocity at the surface and the smoothing radii in the horizontal and vertical directions, all of which have an effect on the results obtained (Yuan et al., 2019). Further, even with an optimal choice of these parameters, uncertainties remain in the obtained RMS velocity fields due to the sparse and/or uneven distribution of diffractors (Yuan et al., 2019). These uncertainties are, however, not specific to migration focusing analysis, and similarly prevail in all other types of diffraction-based velocity

analysis techniques (e.g., Reshef and Landa, 2009; Bauer et al., 2017). The exploration, characterization, and quantification of these uncertainties is an important topic for future research. Another focus of future research will be the extension of the proposed technique from 2D to 3D, which is conceptually straightforward, but algorithmically cumbersome. This, in turn, is expected to significantly enhance the 3D migration velocity analysis of common-offset GPR data as, due to the inherent heterogeneity of the shallow subsurface, the diffracted energy present in corresponding 2D sections is likely to come from out-of-plane sources.

Chapter 4

Estimating subsurface geostatistical properties from GPR reflection data using a supervised learning approach

---

Yu Liu, James Irving, Klaus Holliger

In preparation for *Computers & Geosciences*

4.1 Abstract

The quantitative characterization of near-surface heterogeneity using ground-penetrating radar (GPR) is an important but highly challenging task. To date, the estimation of subsurface geostatistical parameters from common-offset GPR reflection data has relied upon a Monte-Carlo-type inversion approach. While such an approach allows for a comprehensive exploration of the parameter space, the associated computational costs are inherently high. Here, we present an alternative supervised-learning-based methodology for inferring subsurface geostatistical parameters from reflection GPR data in a highly efficient manner. This novel approach uses a convolutional neural network (CNN) that is trained on a vast database of 2D autocorrelation images obtained from synthetic GPR images for a comprehensive range of stochastic subsurface models. Results from synthetic and field data tests confirm the viability of using a trained CNN to estimate the structural aspect ratio of the subsurface heterogeneity. These results also demonstrate a remarkable robustness of this approach even in the presence of high noise levels.

4.2 Introduction

The shallow subsurface consists of soil, glacial and alluvial deposits as well as weathered bedrock. It is from this thin surficial layer that we draw our drinking water and produce our food, and it is also where we ultimately discharge much of our waste. In densely populated and highly industrialized regions, the sustainable use of this increasingly fragile surficial environment has become a societal and political issue of critical importance. Near-surface geophysics in general, and ground-penetrating radar (GPR) in particular, allow for the exploration and characterization of the surficial heterogeneity and thus have an essential role to play in this endeavor.

Surface-based GPR reflection profiling has the potential to provide high-resolution images of the shallow subsurface. Due to the inherently complex nature of the near-surface environment, these images are, however, often better suited to stochastic than to deterministic interpretation approaches. This means that we need to estimate parameters describing the geostatistical nature of the heterogeneity. Specifically, the second-order statistical properties of GPR reflection data are expected to provide information about the correlation structure of the subsurface velocity field through which the electromagnetic waves have traveled. Rea and Knight (1998) and Dafflon et al. (2006) observed good agreement between the spatial correlation structure of cliff-face photographs and those of GPR images collected along the top of the cliff. This led them to conclude that the lateral statistics of a GPR image and the underlying water-content distribution are likely equivalent. Knight et al. (2007) also found similarities between the lateral correlation statistics of GPR data and neutron-probe measurements. However, Knight et al. (2004) and Oldenborger et al. (2004) noticed that the lateral correlation structure of a GPR image is significantly affected by its vertical resolution, which is controlled by the frequency content of the emitted source signal. These findings indicate that the assumption of a direct equivalence between the lateral statistics of water content and GPR reflection data is not valid and that the physics of the underlying wave propagation phenomena should be accounted for more appropriately. Irving et al. (2009) and Irving and Holliger (2010) followed up on these pioneering studies, notably by clarifying the underlying theoretical and methodological foundations. They found that GPR reflection images allow us to infer only the structural aspect ratio a_x/a_z of the subsurface heterogeneity, whereas the associated horizontal and vertical correlation lengths, a_x and a_z, cannot be independently resolved. GPR data also seem to exhibit only a very weak sensitivity to the decay of the associated power spectrum. All of these findings, as well as the fundamental validity of the technique, have been corroborated through a range of follow-up studies (e.g.,

Scholer et al., 2010, Xu et al., 2020, 2021; Liu et al., 2021).

To date, research on the estimation of subsurface geostatistical parameters from reflection GPR data has involved an iterative Monte-Carlo-type inversion approach (e.g., Irving et al., 2009; Xu et al., 2020), whereby sets of model parameters drawn from prescribed prior distributions are either accepted or rejected depending on how well they allow for prediction of the observed autocorrelation structure of the GPR image. While this approach has the advantage of allowing for a comprehensive exploration of the parameter space and providing a measure of uncertainty with regard to the inferred parameters, the associated computational costs are inherently high, in particular for 3D datasets (Xu et al., 2020). In this work, we seek to alleviate this problem through a fundamentally different approach based on supervised deep learning.

In recent years, deep learning has attracted increasing attention in the geophysical community as a novel data-driven technique compared to conventional model-driven approaches. The latest review of deep learning applications in geophysics can be found in Yu and Ma (2021). For wave-based geophysical exploration techniques, deep learning demonstrates remarkable utility in many areas, including data processing (e.g., Kaur et al., 2020), interpretation (e.g., Wu et al., 2019), and inversion (e.g., Yang et al., 2019; Di and Abubakar, 2022; Geng et al., 2022). Deep learning techniques employ artificial neural networks having multiple layers, such as convolutional neural networks (CNN) and recurrent neural networks, to learn complex relationships between the input data and the targeted output (LeCun et al., 2015; Goodfellow et al., 2016). In our case, a CNN architecture is designed to build a functional relationship between the spatial autocorrelation of a GPR image and the underlying subsurface geostatistical parameters, which serve as the input data and targeted output, respectively. To this end, our workflow involves the efficient generation of vast synthetic GPR reflection datasets, which are used for the supervised learning.

In the following, we first present the methodological foundations of the proposed approach, followed by its application to realistic synthetic GPR data generated using a finite-difference time-domain (FDTD) solution of Maxwell's equations. Here, we also examine the impact of noise and variations in the frequency spectrum of the data on the results obtained. Finally, the method is applied to a field GPR dataset from a typical surficial alluvial environment.

4.3 Methodology

The overall idea behind our proposed approach is (i) to generate a vast synthetic GPR database corresponding to a comprehensive range of stochastic subsurface models, (ii) to train a CNN in a supervised setting in order to build a link between the 2D autocorrelation images of these synthetic data and the underlying geostatistical parameters, and (iii) to use the trained network to estimate the geostatistical parameters corresponding to a particular GPR dataset from its 2D autocorrelation image (Figure 4.1).

A key assumption underlying our work is that the subsurface heterogeneity can be adequately described by its second-order statistics. Our experience from previous Monte-Carlo-based inversions for subsurface geostatistical parameters indicates that, to a first approximation, this is generally the case as long as the analysis focuses on a given lithological and/or hydrogeological unit (e.g., Xu et al., 2021). To relate the second-order statistics of the subsurface velocity field with the corresponding GPR reflection section, we briefly summarize the derivation presented in Irving et al. (2009) and refer the reader to that paper for full details. We begin with the assumption that the subsurface GPR velocity distribution, $v(x,z)$, can be characterized as the

superposition of a slowly varying or constant background velocity model, $v_0(x,z)$, and a stochastic velocity perturbation field, $\Delta v(x,z)$:

$$v(x,z) = v_0(x,z) + \Delta v(x,z). \tag{4.1}$$

It is the stochastic component of the velocity structure, $\Delta v(x,z)$, that gives rise to the reflected GPR wavefield recorded at the surface and whose second-order statistics we wish to estimate. To this end, the underlying subsurface GPR reflectivity distribution, $r(x,z)$, can be approximated as the vertical derivative of the velocity-perturbation field (e.g., Pullammanappallil et al. 1997; Poppeliers, 2007; Liu et al., 2022)

$$r(x,z) \approx \frac{\partial}{\partial z}\Delta v(x,z). \tag{4.2}$$

Assuming that the recorded GPR wavefield is largely non-dispersive and dominated by single scattering, a zero-offset GPR reflection section, after proper processing and migration, can be approximated as the convolution product of the subsurface reflectivity, the input GPR wavelet, and a horizontal resolution filter that accounts for the limited lateral resolution of a migrated GPR image (e.g., Irving et al., 2009; Scholer et al., 2010):

$$\begin{aligned} p(x,z) &= w(z) * r(x,z) * h(x) \\ &= w(z) * \frac{\partial}{\partial z}\Delta v(x,z) * h(x) \\ &= \Delta v(x,z) * f(x,z), \end{aligned} \tag{4.3}$$

where $p(x,z)$ is the so-called primary reflection section, $w(z)$ is the GPR wavelet expressed in depth, $h(x)$ is the horizontal resolution filter, and we define $f(x,z) = \frac{\partial}{\partial z} w(z) * h(x)$.

Eq. (4.3) provides a simple convolutional relationship between the velocity perturbation field $\Delta v(x,z)$ and the GPR reflection image $p(x,z)$. Making use of the Fourier transform and the Wiener-Khintchine theorem linking the power spectrum and autocorrelation functions, a similar convolutional relationship is implied for the spatial autocorrelations of all quantities:

$$R_{pp}(\delta x, \delta z) = R_{vv}(\delta x, \delta z) * R_{ff}(\delta x, \delta z). \tag{4.4}$$

Eq. (4.4) states that the 2D spatial autocorrelation of the GPR image, $R_{pp}(\delta x, \delta z)$, where δx and δz refer to the horizontal and vertical lags, respectively, is related to the 2D spatial autocorrelation of the velocity perturbation field, $R_{vv}(\delta x, \delta z)$, through 2D convolution with the filter autocorrelation, $R_{ff}(\delta x, \delta z)$. This equation provides an effective link between the second-order statistics of the radar wave velocity field with those of the corresponding GPR reflection section. Based on this link, we consider using 2D autocorrelation images of GPR reflection data R_{pp} as the input of our CNN which learns to predict geostatistical properties of R_{vv}.

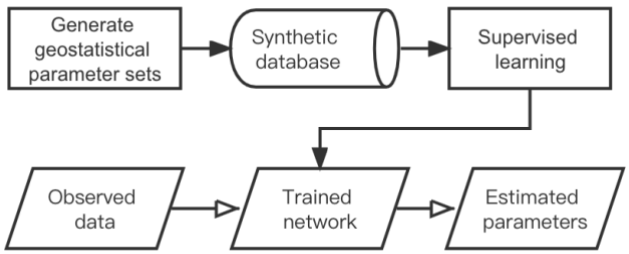


Figure 4.1: Flowchart illustrating the proposed supervised-learning-based approach for the estimation of subsurface geostatistical parameters from common-offset GPR reflection images.

To describe subsurface spatial variability with a versatile parametric autocorrelation model that is capable of describing the expected range of subsurface heterogeneity, we consider the so-called von Karman autocorrelation function (e.g., Goff and Jordan, 1988) for the velocity perturbation field

$$R_{vv}(\delta x, \delta z) = \frac{r^{\nu} K_{\nu}(r)}{2^{\nu-1}\Gamma(\nu)}, \tag{4.5}$$

where $K_{\nu}(r)$ is the modified Bessel function of the second kind of order $0 \leq \nu \leq 1$, Γ is the gamma function, and

$$r = \sqrt{(x/a_x)^2 + (z/a_z)^2} \tag{4.6}$$

is the weighted radial autocorrelation lag with a_x and a_z denoting the horizontal and vertical correlation lengths, respectively. In addition to a_x and a_z, the von Karman autocorrelation function is characterized by the ν-value, also known as the Hurst number, which quantifies the decay of the corresponding power spectrum. Values of ν close to zero characterize the ubiquitous, but enigmatic, pink or flicker noise. ν=0.5 corresponds to the well-known exponential correlation function describing brown noise. Finally, for ν values close to 1, the von Karman autocorrelation function resembles its Gaussian equivalent.

Based on the von Karman autocorrelation function described in Eq. (4.5), we generate 3000 stochastic subsurface GPR velocity models representing a wide range of combinations of a_x, a_z, and ν. For the current study, all models have a fixed size of 20 m wide by 10 m deep. We consider values of a_z ranging from 0.1-1 m. After a_z is determined, the corresponding value of a_x is generated based on the aspect ratio a_x/a_z ranging from 1-20. The ν-value is varied over the range from 0.1 to 0.9. We fix the mean and standard deviation of the realizations at 0.1 m/ns and 0.01 m/ns, respectively. From these subsurface velocity models, we generate the corresponding synthetic GPR reflection images by convolving the vertical reflectivity series at each transmitter-receiver location calculated by Eq. (4.2) with a Ricker source wavelet having a central frequency of 100 MHz and, following Eq. (4.3), applying a Gaussian filter $h(x)$ with a width of one dominant wavelength in the horizontal direction (Irving et al., 2010). It is important to note that we use the convolution method to generate our synthetic database because it has a much lower computational cost compared to the FDTD method. Although this convolution model does not capture the full physics of the GPR experiment, previous work has shown that it is capable of emulating a remarkable degree of realism present in field data, even in complex scenarios (e.g., Xu et al., 2021; Liu et al., 2022). The thus resulting synthetic GPR data are contaminated with zero-mean Gaussian random noise having a standard deviation of 2% with regard to the maximum absolute amplitude of the synthetic data.

From the 3000 synthetic GPR reflection images, we compute their 2D autocorrelations, which we use together with the corresponding values of a_x/a_z and ν to perform the supervised training of the CNN. Figure 4.2 shows an example of a stochastic subsurface velocity model characterized by $a_x = 2$ m, $a_z = 0.2$ m and $\nu = 0.5$ (Figure 4.2a), the corresponding synthetic convolution-based GPR image (Figure 4.2b), and its 2D autocorrelation function (Figure 4.2c). For the design and training of the CNN, we employ the MATLAB Deep Learning Toolbox. We use a basic CNN architecture consisting of six layers: an input layer, a convolution layer, a rectified linear unit layer, a max pooling layer, a fully connected layer, and a regression layer. The 3000 autocorrelation images are labeled with the values of a_x/a_z and ν corresponding to their underlying velocity models. We train two CNNs with the same hyperparameters to predict the values of a_x/a_z and ν, respectively. The stochastic gradient descent with momentum algorithm is used as the optimizer during training. The mini-batch size and epoch number are set to 128 and 30, respectively. We employ an initial learning rate of 0.001 and lower it after 20 epochs. The training process for one network can be completed within ~50 minutes using a standard laptop computer with 16 GB of memory and a 2.2 GHz Intel i7-8750 processor.

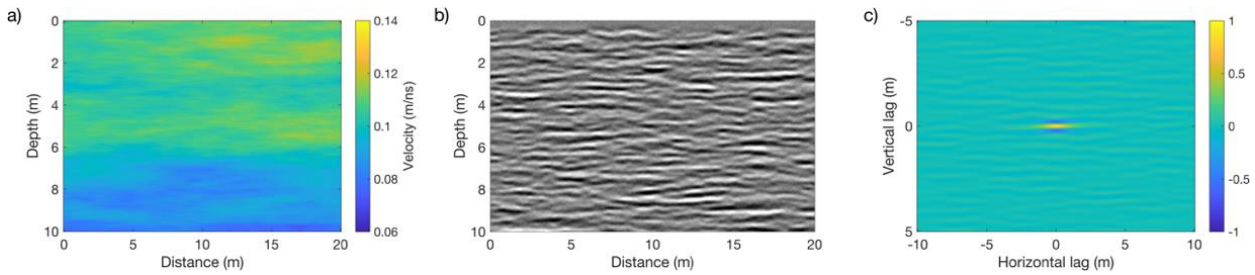


Figure 4.2: Example of a) a stochastic subsurface GPR velocity model, b) the corresponding convolution-based synthetic GPR image, and c) the 2D autocorrelation of the GPR image.

4.4 Results

4.4.1 FDTD-based synthetic data

To assess whether and to what extent the trained network has learned to estimate the values of a_x/a_z and ν from the 2D GPR autocorrelation image, we first consider tests on realistic FDTD-based synthetic data. To this end, we generated 20 stochastic subsurface models (Figure 4.3) using the same manner of creating synthetic models for training database and randomly chosen values of a_x, a_z, and ν. Then we simulated the corresponding GPR data using the MATLAB-based FDTD code of Irving and Knight (2006). For each simulation, the transmitter and receiver antennas were spaced 1 m apart and moved at 0.1 m increments along the survey profile. The source antenna current function was specified to be the first derivative of a Blackman–Harris window having a dominant frequency of 100 MHz.

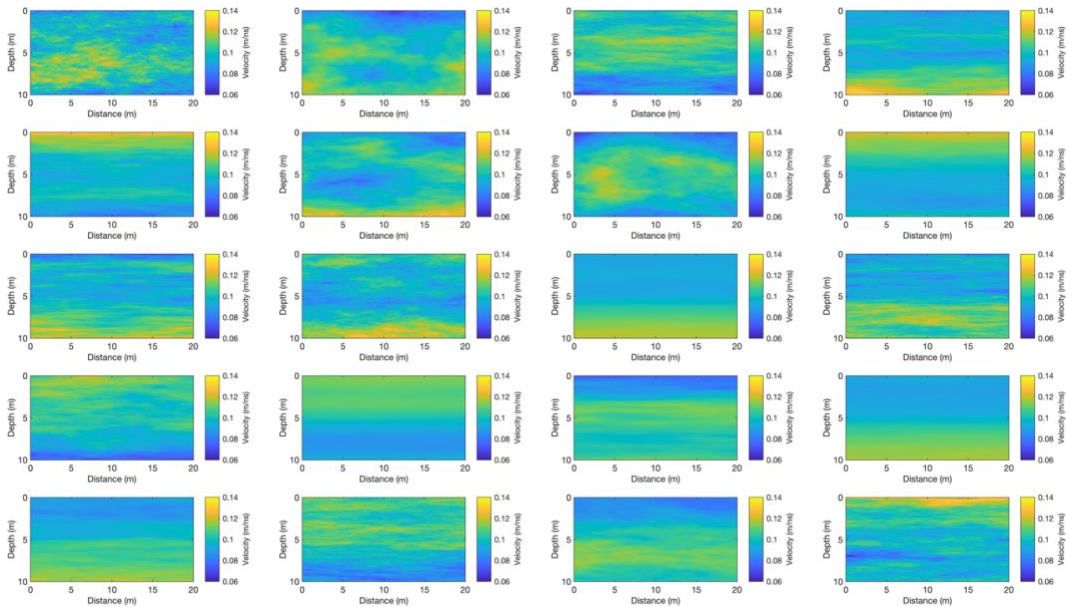


Figure 4.3: Velocity models used to generate the FDTD-based synthetic GPR datasets.

The resulting synthetic GPR data were then contaminated zero-mean Gaussian random noise having a standard deviation of 2% of the maximum absolute signal amplitude below direct waves, and were subjected to a standard processing sequence involving: (i) elimination of the direct air and ground arrivals by removing mean

trace computed over a sliding window; (ii) 30-300 MHz bandpass filtering; (iii) amplitude scaling to compensate for the geometric spreading of energy along with scattering and transmission losses using a smooth gain function based on the mean absolute amplitude curve for the whole section; and (iv) 2D F-K migration and time-to-depth conversion using a constant velocity of 0.08 m/ns, which was found to optimally focus diffraction events in the un-migrated section. Figure 4.4 shows the final processed GPR data sections.

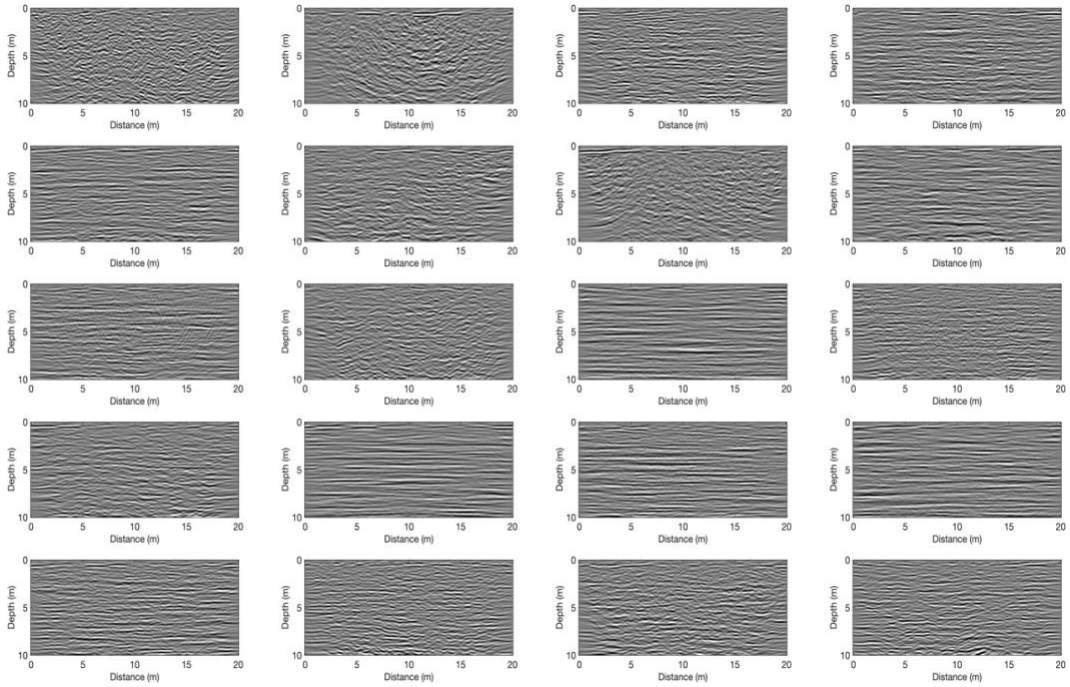


Figure 4.4: Processed 100-MHz FDTD-based GPR datasets corresponding to the velocity models shown in Figure 4.3.

We calculated the autocorrelation images of the FDTD-generated GPR datasets, which were then passed through the trained CNN to estimate the geostatistical parameters of the corresponding subsurface models. Figure 4.5a compares the estimated versus true values of the structural aspect ratio a_x/a_z. The correlation coefficient is 0.93. Overall, the estimated results are remarkably accurate. This corroborates that the autocorrelation of GPR reflection data are inherently sensitive to the aspect ratio of the underlying heterogeneity, as suggested by Irving and Holliger (2010), and that our trained CNN is seemingly an effective tool to estimate it. Figures 4.5b shows the corresponding comparison for the ν-value. The correlation coefficient between the predicted and true values is 0.36. These results, again, confirm previous studies that GPR reflection data have very little sensitivity to this parameter (Irving and Holliger, 2010).

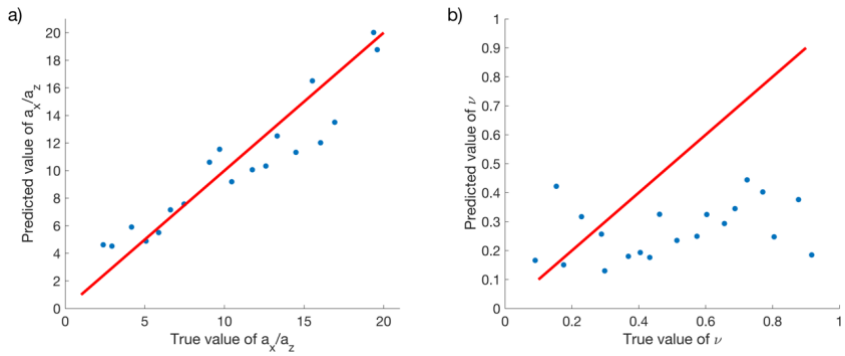


Figure 4.5: Scatter plot of the estimated versus true a) aspect ratios a_x/a_z and b) ν-values for the 100-

MHz FDTD-based synthetic GPR datasets shown in Figure 4.4. The red line denotes perfect agreement.

4.4.2 Sensitivity to data noise

To test the sensitivity of the proposed method to the level of noise in the GPR measurements, we use the same stochastic model as shown in Figure 4.2a. The synthetic data are then contaminated with gradually increasing amounts of zero-mean Gaussian random noise before being subjected to the same signal processing flow used in the previous test. In each case, the standard deviation of the added noise is set relative to the maximum absolute signal amplitude below the direct arrivals.

Figures 4.6a and 4.6b show the resulting GPR images for noise levels of 2% and 50%, respectively. In this context, it is interesting and important to note that the FDTD-based GPR image shown in Figure 4.6a bares a reasonably strong resemblance to its convolution-based counterpart shown in Figure 4.2b, which corroborates that it is reasonable to use convolution-based synthetics for training database of CNN. Although the GPR image shown in Figure 4.6b is clearly similar, it has a mass of "smiles" generated from the strong random noise in the migration operation. Figure 4.7 shows the inferred values for a_x/a_z as a function of the noise level. Quite interestingly, the results suggest that the estimated parameters remain stable even in the presence of very high noise levels. We observe that the values of a_x/a_z decrease with an increase in noise level, which we expect results from the degradation of the lateral autocorrelation of GPR image.

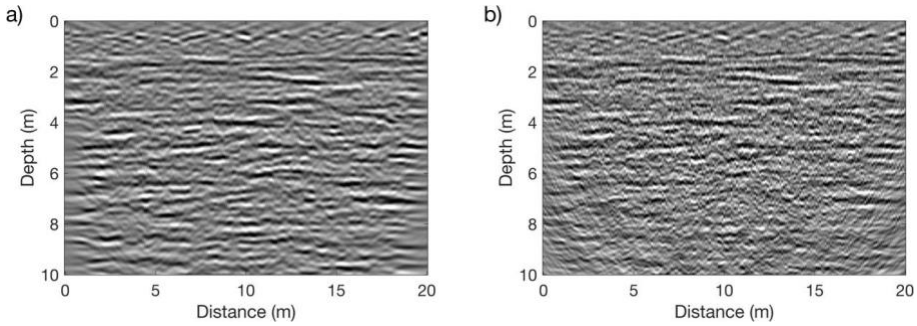


Figure 4.6: Processed 100-MHz FDTD-based synthetic GPR data corresponding to the velocity model shown in Figure 4.2a with noise levels of a) 2% and b) 50%.

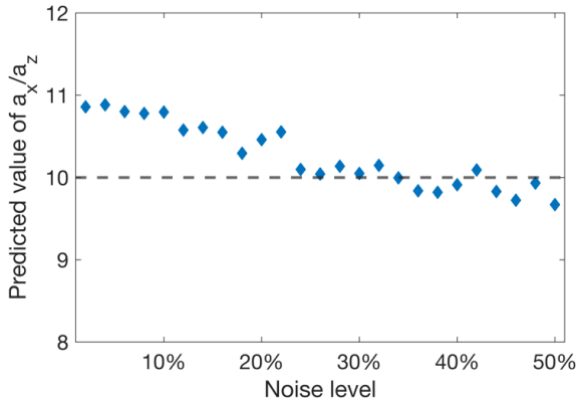


Figure 4.7: Aspect ratio a_x/a_z inferred from the 100-MHz FDTD-based synthetic GPR data as a function of the level of added noise. The dashed line denotes the true value.

4.4.3 Differences in signal frequency content between training and observed data

Since our training is based on datasets having a central frequency of 100 MHz, the trained CNN is intended for use on GPR data having similar frequency characteristics. In field surveys, however, GPR data collected with nominal 100-MHz frequency antennas may have a significantly lower dominant frequency as a results of antenna coupling effects. To explore the impact of such discrepancies between the frequency content of the training database and that of the considered GPR dataset, we again use the 20 stochastic subsurface models shown in Figure 4.3 and simulate new FDTD-based GPR datasets using a source antenna current function having a dominant frequency of 70 MHz. The other settings in the simulation were kept the same. The resulting 70 MHz GPR datasets were again contaminated with 2% Gaussian random noise and subjected to the same signal processing flow as in the 100-MHz data test. Figure 4.8 shows the processed GPR data sections.

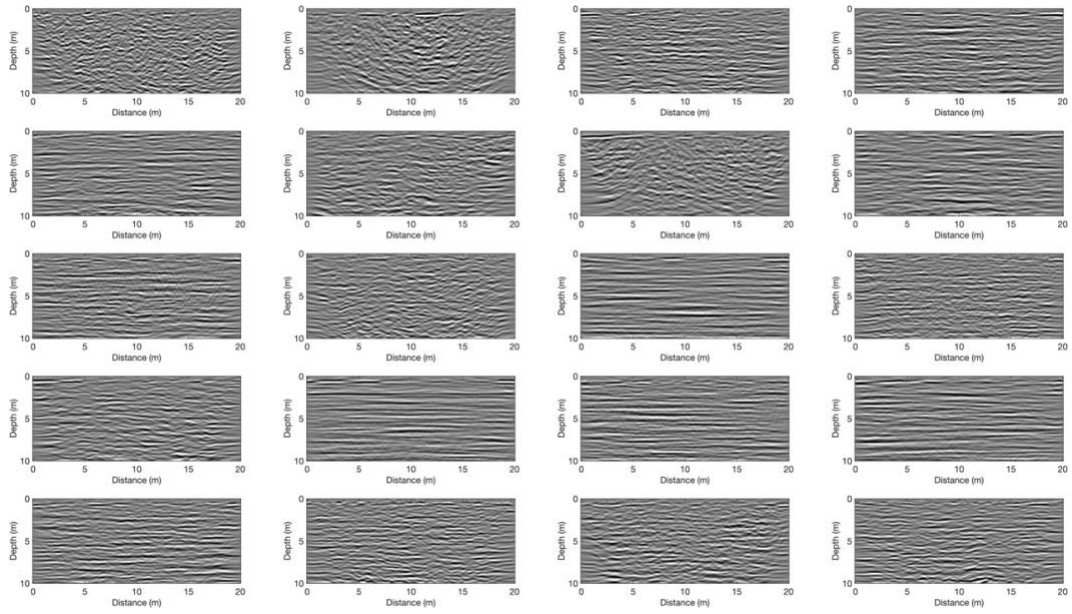


Figure 4.8: Processed 70-MHz FDTD-based GPR datasets corresponding to the velocity models shown in Figure 4.3.

To deal with the frequency differences between training and observed data, we consider a rather simple adjustment of the frequency content of the GPR measurements prior to estimating the geostatistical parameters using our trained CNN. Specifically, we: (i) calculate the Fourier transform of each 70-MHz trace, (ii) shift its spectrum 30 MHz higher by padding with zeros, and (iii) transform the shifted spectrum back to the time domain. The original and shifted Fourier amplitude spectra, averaged over all traces for each GPR dataset in Figure 4.8, are shown in Figure 4.9. After the spectral adjustment, the dominant frequency of observed GPR data is close to 100 MHz, the dominant frequency of training data.

The autocorrelation images of the GPR datasets, before and after spectral shifting, were input to the trained CNN. Figure 4.10 summarizes the estimation results of structural aspect ratio a_x/a_z. The correlation coefficients between predicted and true values are 0.93 and 0.96, respectively. The predicted values from the 70-MHz data show an overall deviation compared to the true values (Figure 4.10a), which is attributed to the frequency bias. After the adjustment of spectra, the CNN-predicted values become closer to the true values

(Figure 4.10b), indicating that spectral shifting is a useful tool to match the frequency content of input data with that of the training database and the proposed method offers a reliable means of inferring the subsurface geostatistical parameters.

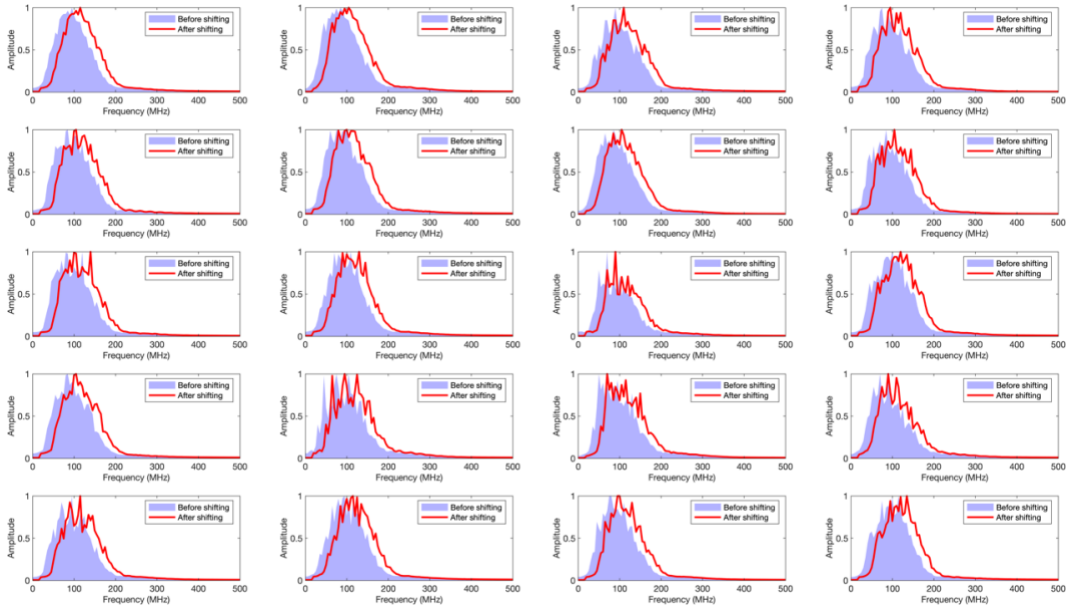


Figure 4.9: Original and shifted Fourier amplitude spectra corresponding to the 70-MHz FDTD-based synthetic GPR datasets shown in Figure 4.8.

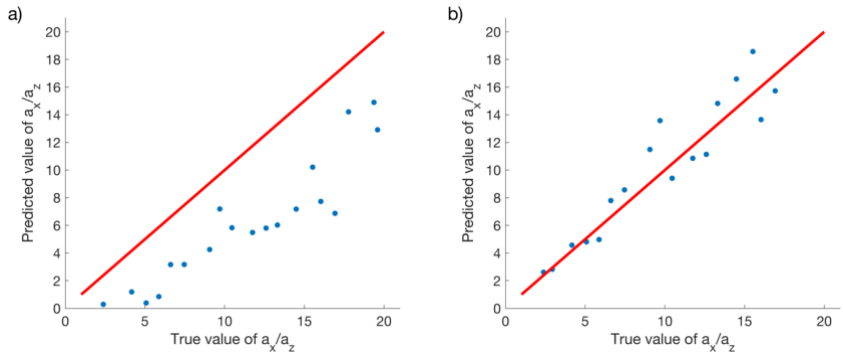


Figure 4.10: Scatter plot of the estimated versus true aspect ratios a_x/a_z for the 70-MHz FDTD-based synthetic GPR datasets shown in Figure 4.8 a) before and b) after spectral shifting. The red line denotes perfect agreement.

4.4.4 Application to field data

Following the preceding tests of our proposed methodology on FDTD-based synthetic data, we now consider its application to common-offset GPR reflection data acquired at the Boise Hydrogeophysical Research Site (BHRS), which is located on a gravel bar adjacent to the Boise River, ~15 km from downtown Boise, Idaho, USA (Barrash and Clemo, 2002). The shallow, unconfined alluvial aquifer at the BHRS is bounded by a clay layer at approximately 20-m depth. The depth to the groundwater table varies seasonally between 0 and 3 m. Over the past two decades, a wide variety of geophysical and hydrogeological studies have been conducted at this site for characterizing heterogeneous aquifers (e.g., Tronicke et al., 2004; Bradford et al., 2009; Dafflon and Barrash, 2012; Pirot et al., 2016; Liu et al., 2022). The considered GPR reflection profile is a part of 3D

survey, which is performed in the summer of 1998 using a PulseEkko Pro 100 system (Sensors & Software Inc.) having a nominal antenna frequency of 100 MHz. The data were collected in common-offset mode using a transmitter-receiver antenna spacing of 1 m. Traces were recorded every 0.2 m along the 18-m-long profile. The time sampling interval was 0.8 ns and 32 vertical stacks were performed for each recorded GPR trace to improve the signal-to-noise ratio.

The considered GPR data were subjected to a processing flow similar to their synthetic counterparts, which included time-zero correction, dewow filtering, elimination of direct waves, 30-300 MHz bandpass filtering, smooth time-varying amplitude gain, and 2D F-K migration using a constant velocity of 0.08 m/ns. The velocity value was found to optimally focus diffraction events in the data below the water table, and it is consistent with the average velocity of saturated sediments at the BHRS determined by Bradford et al. (2009). Figure 4.11a shows the processed GPR image. From the spectral analysis of the processed data, we found the central frequency of the data is ~80 MHz, and thus corrected the frequency content to 100 MHz for consistency with the training database (Figure 4.11b). The average Fourier amplitude spectra of the original and shifted are shown in Figure 4.11c. For characterizing the heterogeneity of the water saturated zone, the top 40 ns in the spectral-shifted data, which corresponds to the unsaturated zone, was not considered in the estimation.

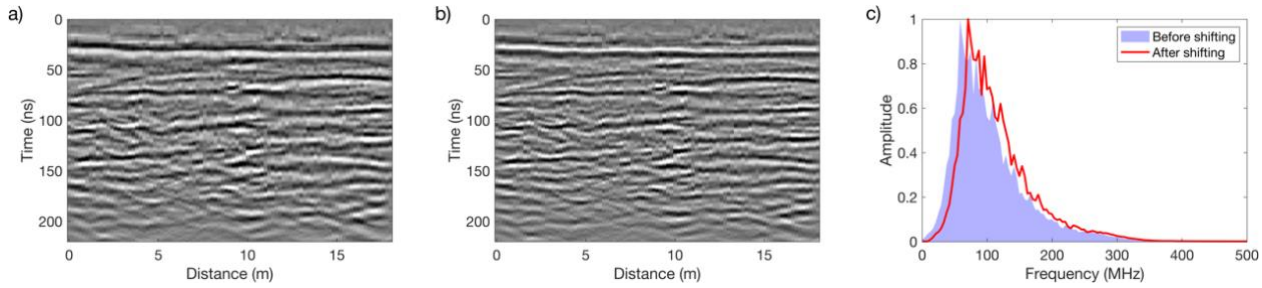


Figure 4.11: The a) processed and b) spectral-shifted field GPR data from the BHRS, along with c) the corresponding Fourier amplitude spectra.

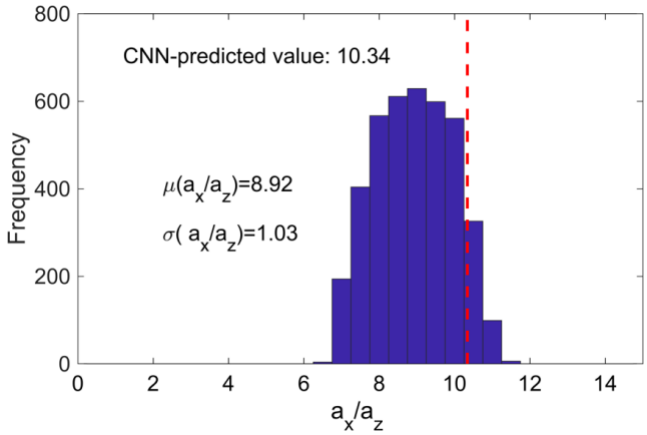


Figure 4.12: Comparison of the results obtained from a Monte-Carlo-type inversion for the subsurface geostatistical parameters with those obtained from the proposed supervised-learning-based algorithm. The blue histogram denotes the Monte-Carlo results for the aspect ratio, whereas the red dashed line denotes the CNN-estimated aspect ratio.

We calculated the autocorrelation image of the spectral-shifted data section and passed it through the trained CNN, which yields an aspect ratio a_x/a_z of 10.34. For comparison, we also applied the Monte Carlo inversion approach to estimate the aspect ratio (e.g., Irving et al., 2009; Irving and Holliger, 2010) for the

original processed GPR data. The prior range of acceptable values for the horizontal and vertical correlation lengths was set between 0.1 and 3 m. For the ν-value, the prior range was set between 0.1 and 0.9. Similarly, the top 40 ns was not considered in the inversion procedure, which was run until 4000 accepted sets of von Karman autocorrelation model parameters were obtained. The resulting histogram for the horizontal-to-vertical aspect ratio of the underlying heterogeneity is shown in Figure 4.12, where the mean value and standard deviation are 8.92 and 1.03, respectively. The CNN-predicted value falls well within the range of high probability given by the histogram, indicating that the results from two types of methods are consistent.

4.5 Discussion and conclusions

We have explored a novel deep-learning-based method for inferring subsurface geostatistical parameters from surface-based GPR reflection images. The proposed approach uses a CNN that is trained on a database of 2D autocorrelation images obtained from convolution-based modeling of synthetic GPR data for a comprehensive range of stochastic subsurface velocity models. Tests on FDTD-based synthetic GPR reflection data and subsequent application to field data corroborate the viability of using the proposed method to estimate the structural aspect ratio a_x/a_z of the subsurface heterogeneity. For the estimation of the ν-value, it is demonstrated that our method has very little sensitivity to this parameter. Compared to a Monte-Carlo-type inversion, our CNN-based approach can perform the estimation with almost no computational cost once the training of the network has been completed. The workflow we use to generate vast synthetic GPR reflection datasets is quite efficient. Thus, an eventual extension to 3D is methodologically straightforward and practically feasible at a reasonable computational cost.

Testing reveals that the CNN-based method is robust in the presence of high noise levels in the GPR data. The inferred a_x/a_z from noisy data, however, is typically smaller than its noise-free counterpart because the lateral autocorrelation of GPR image is degraded. The frequency spectrum of the GPR data also needs to be approximately consistent with that of the training datasets for reliable results. We have demonstrated that a simple shifting of the spectrum of the observed data can effectively address this problem. We fully acknowledge that other operations, e.g., application of a shaping filter, might be equally if not more effective.

Like most deep learning applications in geophysics, one remaining question is whether the estimation results, not having a solid theoretical foundation, can be trusted. Current results of the proposed method cannot show the uncertainties as the histogram results obtained from a Monte-Carlo inversion. Future work may need to consider CNN-based prediction with uncertainty assessment, such as using the probability distribution as output target to provide confidence in the estimation (e.g., Cao et al., 2020; Grana et al., 2020). The GPR datasets used for supervised training in our study share the same central frequency, which leads to that the trained CNN only works for one type of GPR data. Further training can use data in different central frequencies and adapt the network design to make the trained model applicable to all types of GPR reflection data. In this study, we use the 2D autocorrelation of the GPR image as input, which significantly simplifies the relationship between input and output of the neural network. However, this operation causes the loss of some valuable information, e.g., the phase of the GPR data. In the following study, multiple inputs, e.g., both GPR image and its 2D autocorrelation, could be considered to improve the performance of the trained network (e.g., Lowney et al., 2021). In some cases, the dominant dip angle of subsurface heterogeneity is an important parameter (e.g., Xu et al., 2020; Liu et al., 2021), the automatic estimation of which is an issue that needs to be considered in the future.

The deep-learning-based approach described in this study provides a simple and efficient alternative to estimating the subsurface geostatistical properties from GPR reflection images. Results from synthetic and field data show the capacity and robustness of our trained neural network. Although this study focuses on GPR reflection data, the proposed methodology is readily applicable to seismic reflection data.

Chapter 5

High-resolution seismic reflection surveys crossing the Insubric Line into the Ivrea-Verbano Zone: Novel approaches for interpreting the seismic response of steeply dipping structures

---

Yu Liu, Andrew Greenwood, György Hetényi, Ludovic Baron, Klaus Holliger

Published[1] in *Tectonophysics*

[1] Liu, Y., Greenwood, A., Hetényi, G., Baron, L., Holliger, K. (2021). High-resolution seismic reflection survey crossing the Insubric Line into the Ivrea-Verbano Zone: Novel approaches for interpreting the seismic response of steeply dipping structures. *Tectonophysics*, 816, 229035. doi: 10.1016/j.tecto.2021.229035

5.1 Abstract

A high-resolution seismic reflection survey has been conducted across the Insubric Line from the Sesia Zone into the Ivrea-Verbano Zone (IVZ), where a remarkably complete cross-section of lower continental crust is exposed. The survey was carried out in preparation for the DIVE (Drilling the Ivrea-Verbano zonE) project, which was recently approved by the International Continental Scientific Drilling Program (ICDP). DIVE aims to gain new insights into the characteristics of the lower continental crust through targeted drilling, sampling, and borehole logging. A key borehole is planned near the Insubric Line at Balmuccia, where the deepest parts of the lower continental crust are exposed. As such, the primary objective of this seismic survey was to explore whether the sub-vertical structures prevailing at the surface can be expected to continue at depth or whether there are any indications for major flattening or fault-related offsets. Correspondingly, the acquisition and processing of the seismic reflection data were geared towards revealing weak backscattered events from local heterogeneities associated with the prevailing sub-vertical structural grain. The migrated sections, contain coherent backscattered events to a depth of ~1 km, which form numerous short lineaments that seem to align sub-vertically. To substantiate this observation, we have generated synthetic seismic reflection surveys for canonical models of sub-vertical structures associated with Gaussian- and binary-distributed heterogeneities. Both the observed and synthetic seismic data were then subjected to energy-based attribute analysis as well as geostatistical estimations of the structural aspect ratios and the associated dips. The results of these quantitative interpretation approaches are indicative of the overall consistency between the synthetic and the observed seismic data and, hence, support the original qualitative interpretation of the latter in that the sub-vertical structural grain evident at the surface seems to prevail throughout the imaged part of the upper crust.

5.2 Introduction

The Insubric Line corresponds to the western end of the Periadriatic fault system delineating the boundary between the European and Adriatic plates, which can be followed over more than 700 km from Slovenia to the southwest of Torino (e.g., Schmid et al., 1989; Handy et al., 2015). Amongst the exposed paleo-plate boundaries, the Insubric Line is unique as it represents a particularly well-preserved suture zone documenting the late continental collision stage of the Alpine orogeny (e.g., Schmid et al., 1987). As such, the Insubric Line separates the Austroalpine gneiss units of the Sesia Zone to the west from Adriatic lower crustal rocks exposed in the Ivrea-Verbano Zone (IVZ) to the east. The IVZ is associated with a very prominent positive gravity anomaly (e.g., Berckhemer, 1968; Kissling et al., 1984; Scarponi et al., 2020) and an associated seismic high-velocity anomaly (e.g., Diehl et al., 2009; Lu et al., 2018), whose origins extend to shallow crustal levels. This so-called Ivrea Geophysical Body (IGB) is assumed to represent a sliver of Adriatic lower crust and upper mantle, which was “upwarped” in the course of the continental collision process (e.g., Schmid et al., 2017). The IVZ thus corresponds to a partial exposure of the IGB.

The geology of the IVZ has been extensively studied, as it is widely regarded to represent arguably the most complete exhumed cross-section of the lower continental crust and its transition into the uppermost mantle (e.g., Fountain, 1976; Schmid et al., 1996; Quick et al., 2003; Brack et al., 2010) (Figure 5.1). As a consequence, the IVZ is a prime locality to gain insights into the composition, structure, and evolution of this critically important, but still largely enigmatic part of the continental lithosphere. These questions will be addressed in a unified and focused manner in the framework of an international research project called DIVE (Drilling the Ivrea-Verbano zonE) (Pistone et al., 2017), which recently has been approved by the International Continental

Scientific Drilling Program (ICDP). A ~4-km-deep borehole is planned just east of the Insubric Line near Balmuccia, where the deepest parts of the lower continental crust have been exhumed. To locate and orient this borehole, and to develop an effective drilling, sampling, and logging strategy, the targeted subsurface region needs to be geophysically characterized. This is generally best achieved through seismic reflection surveys, which are indeed widely regarded as a *conditio sine qua non* for drilling campaigns in general and ICDP projects in particular (e.g., Demirel-Schlueter et al., 2005; Juhlin et al., 2010; Simon and Buske, 2017). To this end, we have carried out high-resolution seismic reflection measurements across the Insubric Line from the Sesia Zone into the IVZ near the planned DIVE drilling location at Balmuccia. The primary objective of this seismic survey is to clarify whether the sub-vertical structures prevailing at the surface can be expected to be continuous throughout the uppermost crust, that is, to a depth of ~1 km, or whether there is evidence of major structural flattening or fault-related offsets.

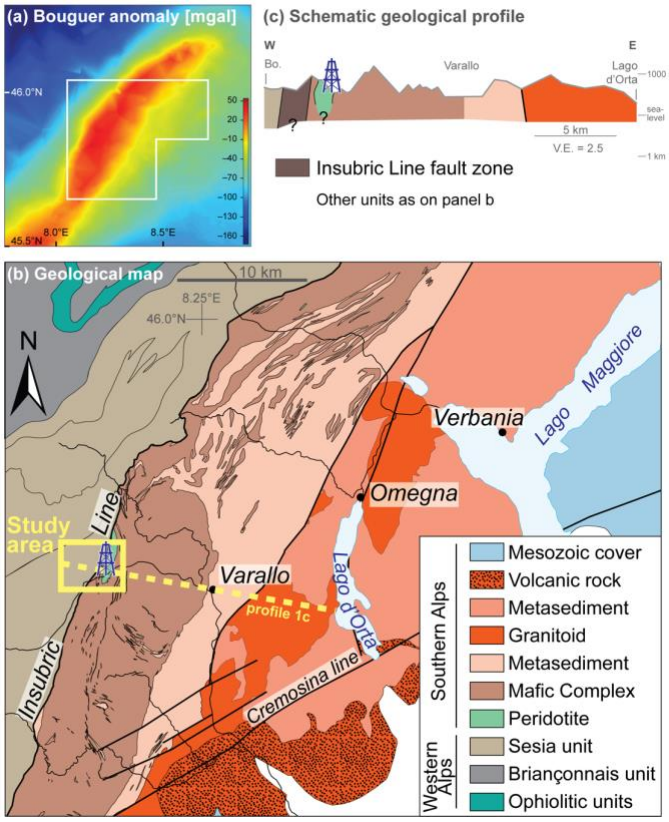


Figure 5.1: Geophysical and geological setting of the study area. a) Bouguer anomaly map of the IVZ highlighting the large positive gravity anomaly due to dense rocks at and near the surface. Adapted from Scarponi et al. (2020). b) Lithotectonic map of the central IVZ, framed in white in a). Adapted from Petri et al. (2019). The drilling rig denotes the location of a planned borehole. c) Schematic geological cross-section along the yellow dashed line in b) based on maps and profiles of Hunziker and Zingg (1980), Quick et al. (2003), Berger et al. (2012), and Petri et al. (2019).

The sub-vertical surficial structures and the high seismic velocities associated with the crystalline rocks present in DIVE's target zone differ strikingly from the layered sedimentary environments, for which the seismic reflection method was originally designed. While seismic reflection profiling has demonstrated significant potential in complex crystalline environments, the acquisition and processing need to be tuned to the corresponding targets, which tend to be non-specular, small-scale, and complex (e.g., Eaton et al., 2003; Malehmir et al., 2012). Correspondingly, the interpretation of the resulting unconventional seismic images is notoriously challenging and tends to lend itself to modelling-based approaches (e.g., Bongajum et al., 2012).

In view of this, we have made efforts to acquire seismic reflection data with high spatial and temporal resolution as well as high redundancy, and we geared the processing flow towards enhancing weak seismic events backscattered from complex, non-specular heterogeneities associated with the prevailing sub-vertical structural grain. For the interpretation, we complement the conventional qualitative visual assessment of crustal seismic reflection images with numerical simulations as well as with attribute and geostatistical analyses.

5.3 Database

Steeply dipping structures are inherently difficult to image with surface-based seismic reflection surveys, as specular interfaces with dips in excess of ~70 degrees generally do not return any seismic energy back to the surface (e.g., Juhlin et al., 2000). A possibility to address this problem is to focus on the energy backscattered from small-scale, non-specular heterogeneities associated with larger-scale structures (e.g., Schmelzbach, et al., 2008; Tertyshnikov et al., 2015; Khoshnavaz et al., 2016; Schwarz and Krawczyk, 2020). Correspondingly, our acquisition and processing strategies are geared towards collecting high-fold seismic profiles with high spatial and temporal resolution, such that weak backscattered seismic energy can be effectively recorded and enhanced through targeted signal processing and wavefield separation techniques (e.g., Salisbury et al., 2000; Eaton et al., 2003; Schijns et al., 2009; Malehmir and Juhlin, 2010; Juhlin et al., 2010).

The three seismic lines presented in this study, denoted as L1, L2, and L3, were collected along the Boccioleto-Balmuccia and Balmuccia-Isola roads (Figure 5.2a). The acquisition and processing methodologies for all three lines are largely identical. In the following, we focus on the data acquired along L1, which crosses from the Sesia Zone in the west into the Insubric Line in the east, because it exemplarily illustrates our methodological approach, quasi-perpendicularly crosses some of the most pertinent structures, and has a significantly higher signal-to-noise ratio (S/N) than L2 and L3.

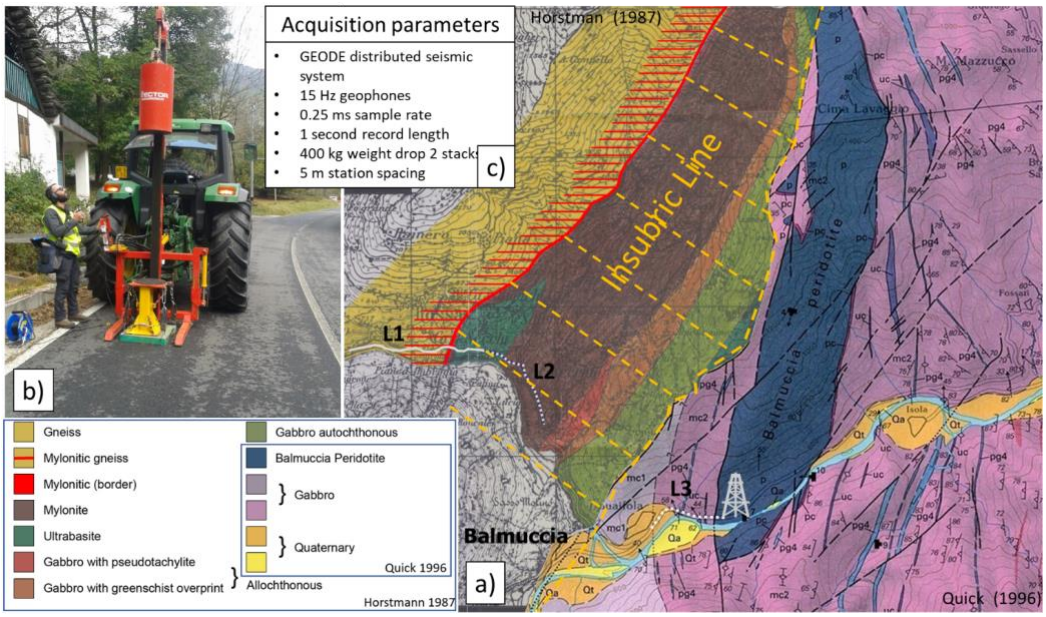


Figure 5.2: a) Location of the seismic profiles L1, L2, and L3 and local geological setting based on Horstmann (1981) and Quick (2003). The drilling rig denotes the location of a planned borehole. b) 400 kg weight-drop source used to acquire the seismic data. c) List of key acquisition parameters.

5.3.1 Acquisition

The data were acquired using a 400 kg weight-drop source (Figure 5.2b) in conjunction with 15 Hz vertical-component geophones and a GEODE distributed seismograph system. The acquisition parameters are listed in Figure 5.2c. The receiver geometry consisted of a geophone spread with 144 live channels with 5 m station spacing. The receiver spread was rolled along every 24 stations, except for L1, which consisted of a single 48-station roll-on that resulted in a two-spread geometry as shown in Figure 5.3a. The source stations were collocated with receiver stations, that is, every 5 m for L1 and L2, which resulted in a dense array of common mid-points or CMPs (Figure 5.3b) with a nominal spacing of 2.5 m and a fold of 77 for the 144 live receivers. This fold was increased in the centre of the line by having overlapping receiver spreads (Figure 5.3c). For L3, the source increment had to be increased to 10 m due to time limitations, which corresponds to a nominal fold of 36. The Boccioleto-Balmuccia road section of L1 is only mildly crooked, such that all CMP locations fall within a 30-m-wide swath of the central trend line. Geophones were planted alongside the road and over embankments. Where outcropping solid rock prevented this, 8-mm-diameter holes were drilled into the concrete verge of the road to plant the geophones. Consequently, some receiver station elevations were not vertically collocated with the sources (Figure 5.3d), which resulted in erratic first-arrival times (Figure 5.4a).

The 400 kg weight-drop source provided signals with a dominant frequency of ~80 Hz (Figure 5.4b) and had sufficient energy to reach far-offset receivers, except for the westernmost (~1-30) source positions. In these field records, the first-arriving seismic energy was weak for the easternmost receivers. Lateral variations in signal quality and strength can be seen from west to east in the example field records shown in Figure 5.4a, for example, between the dashed vertical white lines. These variations are likely caused by surficial geological variations across the profile and are assumed to be the primary reason for the source energy barely reaching the most distant receivers. This resulted in data with relatively low S/N recorded at these distant channels until the source progressed passed position 30. Additionally, the operation of the weight-drop source was associated with an uncontrollable "bounce", which resulted in ghost shots with onsets at ~300-400 ms. However, due to the high seismic velocities of the study area, these ghost shots, for the most part, do not affect the travel time interval of interest ranging from 0 to ~300 ms. A prominent example of such a ghost shot is shown in Figure 5.4a.

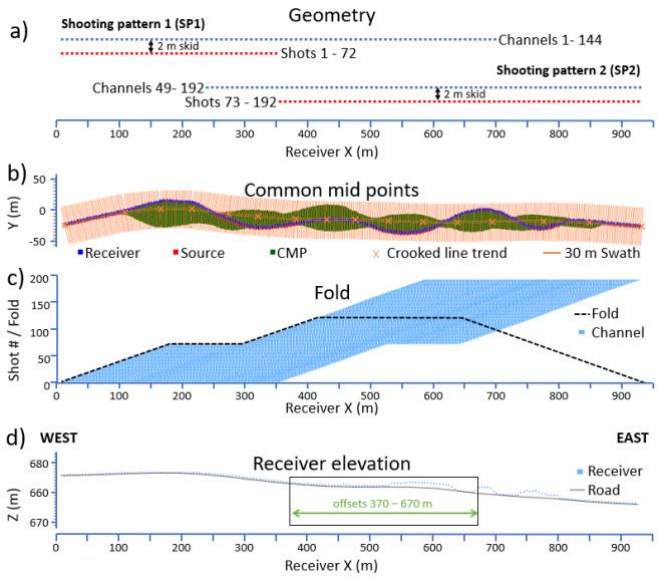


Figure 5.3: a) The data along L1 were collected using two overlapping static receiver spreads, SP1 and SP2.

Source and receiver stations were offset laterally by 2 m. b) CMP and c) fold distributions. d) Elevation differences between the receiver positions (blue dots) and the source points located on road (grey line).

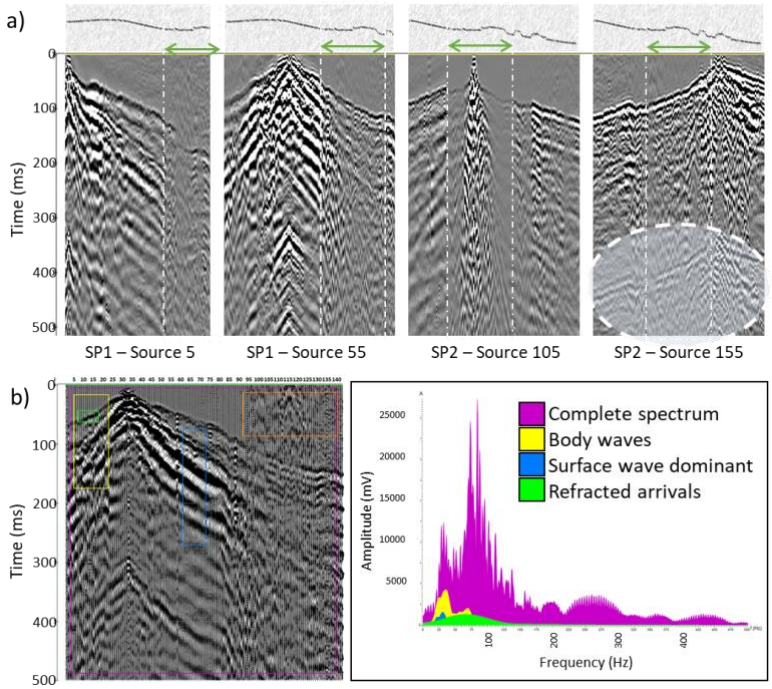


Figure 5.4: a) Typical examples of field records along L1 (Figure 5.3a) after band-pass filtering with their corresponding receiver elevation profiles above. The mylonitized units associated with the Insubric Line are indicated with green arrows and associated with a change in signal strength and quality. The shaded elliptic area in SP2-Source 155 denotes a particularly prominent example of a ghost shot. b) Amplitude frequency spectrum of shot record 32 (left) and selected parts thereof.

5.3.2 Processing

Topographic elevation variations and low-velocity alluvial and/or weathering layers close to the surface are known to significantly distort reflections in hardrock seismic surveys (e.g., Palmer and Jones, 2005; Urosevic and Juhlin, 2007). To alleviate this problem, static corrections were performed prior to pre-stack processing. Firstly, receiver elevation statics were applied to reduce the receivers to a common source-receiver datum corresponding to the Bocioletto-Balmuccia road. Following this, sub-surface velocity variations were evaluated using refracted ray travel-time tomography (Figure 5.5). The tomographic approach was preferred over standard refraction seismic methods, such as the plus-minus method (Hagedoorn, 1959) or the generalized reciprocal method (Palmer and Jones, 2005), due to the pronounced lateral velocity variations along the profile, which are clearly observable in the first-arrival travel-time curves (Figure 5.5a). Static correction times at each receiver location were computed using the inferred surficial velocity structure (Figure 5.5b) down to a depth of 20 m (Figures 5.5c and 5.5d), and then applied to the data (Figures 5.5e and 5.5f).

Following these static corrections, a targeted seismic reflection processing flow was applied, which is summarized in Figure 5.6 and the most essential parts of which are outlined in the following. Conditioning of the pre-stack data involved band-pass filtering (10-20-85-100 Hz) to subdue the airwave, which has a dominant frequency of ~100 Hz, and the subsequent application of spiking deconvolution to balance and enhance the spectral characteristics of data (Figure 5.6a). Spiking deconvolution was successful in boosting the central

frequency of the data to ~80 Hz, which corresponds to the upper limit of the previously applied band-pass filter. A non-standard technique was used to remove the direct arrivals and their associated reverberations (Figure 5.6b). This was achieved using a 2D alpha-trimmed mean filter, which required the flattening of the first-arrivals to an indefinite datum and the reversal of this procedure after filtering (e.g., Greenwood et al., 2019; Caspari et al., 2020). Please note that this filter also naturally removes the P-wave component of the ghost shots. S-waves, including those associated with the ghost shots, and groundroll were removed using a polygon-type *f-k* filter. Examples of the pre-stack processing of shot records are shown in Figure 5.6c. Here, we see little to no evidence of continuous hyperbolic reflection events that might be associated with shallow gently to moderately dipping structures. Conversely, we see many features that are cross-cutting each other in a chevron-type manner with similar move-outs as the first-arrivals. These opposing dips are analogous to those observed for reflections in VSP data with a vertical receiver line crossing horizontal structures, which represents the transpose to our survey geometry and the prevailing geological structures (Figure 5.6a through 5.6c).

After static corrections, removal of unwanted wavefield components, and CMP-sorting, a corresponding stacked section can be produced. This requires a velocity model to correct for the normal move-out (NMO). NMO velocity analysis is typically based on the assumption of quasi-hyperbolic move-out associated with a layered subsurface structure. However, in complex, steeply dipping environments, this kind of analysis is not meaningful. Therefore, we visually analysed constant-velocity CMP stacks to determine the stacking velocity, which optimally enhanced the overall strength and coherence of the backscattered seismic energy. While this process is inherently subjective, our key observation was that, while the apexes of the backscattered events changed in time as a function of the stacking velocity, their strength and coherency remained largely constant. Ultimately, we decided on a single, constant stacking velocity of 7000 m/s, which is broadly consistent with evidence from laboratory measurements (e.g., Fountain, 1976). In this context, it is important to note that the use of constant or quasi-constant stacking velocities is quite common in hardrock seismic exploration in response to the small move-outs prevailing in high-velocity crystalline environments (e.g., Eaton et al., 2003).

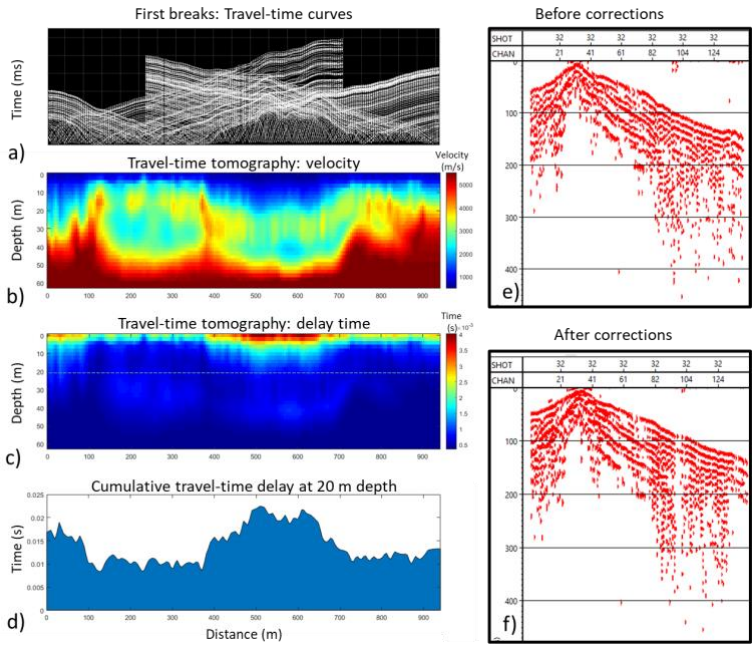
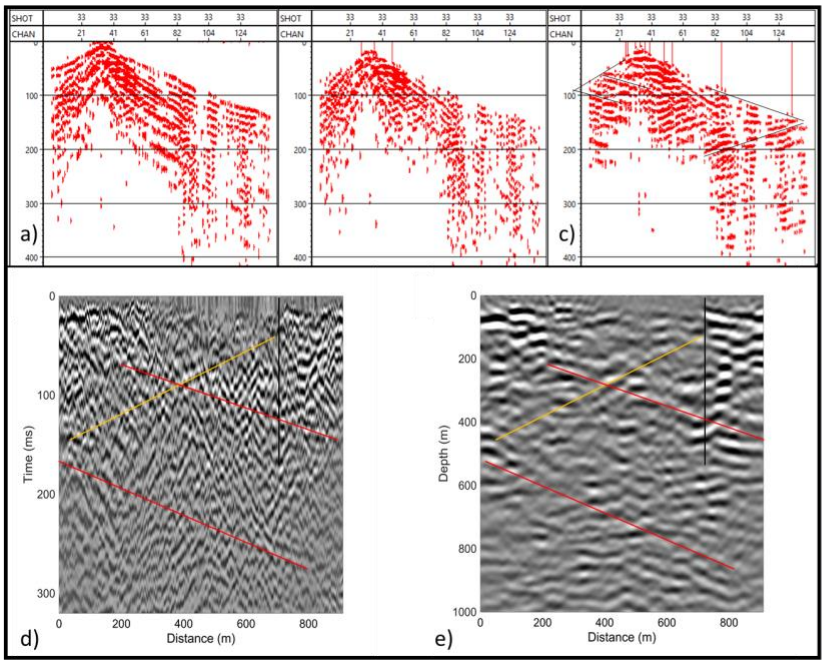


Figure 5.5: Refracted ray travel-time tomography and static corrections. a) Travel-time curves from first-arrival analysis; b) velocity tomogram; c) velocity tomogram converted to delay-time; d) cumulative travel-

time delay correction derived at each receiver location down to a depth of 20 m; and shot record 32 e) before and f) after application of static corrections.



	Processing step	Description
1	Read SEG 2 data	
2	Trace edit	Automatic and manual
3	Common source stack	Diversity stacking
4	Amplitude recovery	Spherical divergence and mean trace balance
	Static corrections:	Elevation and tomographic refraction statics
5	Precondition data and suppress airwave (100 Hz)	Bandpass filter (10-15-85-100)
6	Recover high frequencies	Spiking deconvolution: 0.25 ms (1 sample) gap
7	Removal of first arrival and ghost-shot direct arrivals	Application of linear move-out to flatten first arrivals; 2D alpha-trimmed median filter (33% rejection; 13 traces; 1 sample) to remove first arrival; undo linear moveout corrections
8	S-wave and ground roll suppression	f-k polygon reject filter
9	Velocity analysis	Constant velocity stack panels: 4,500 – 14,000 m/s with increments of 100 m/s
10	NMO correction and CMP stack	NMO stacking with constant velocity of 7000 m/s. CMP binning at 2.5 m
11	Backscattered S-wave suppression	f-k fan reject filter
12	Migration	Kirchhoff post-stack time migration with a constant velocity of 7000 m/s and aperture of 100 m
13	Time-to-depth conversion	Velocity: 7000 m/s

Figure 5.6: L1 processing flow (right) and selected results. a) Statically corrected shot record 32 from Figure 5.4 after preconditioning and recovery of high frequencies (processing steps 5 and 6); b) after removal of direct wave (processing step 7); c) after S-wave and groundroll removal (processing step 8); d) CMP stack, and e) depth-converted post-stack Kirchhoff migration thereof. Moderately dipping VSP-type reflection events (black lines) from vertical boundaries are present in shot profiles before and after wavefield separation. They remain weakly visible in the CMP-stacked and in the migrated sections (red and yellow lines), where the presumed location of the reflecting sub-vertical boundary is denoted by a corresponding black line.

The resulting stacked section of L1 unsurprisingly does not exhibit any prominent events (Figure 5.6d) and, hence, there are also no laterally continuous reflections that could be associated with gently to moderately dipping structures. Conversely, there are many cross-cutting, moderately dipping events, which show coherence over a few traces. These short and discontinuous seismic events, which are quite typical of hardrock environments, are likely to represent seismic energy backscattered from small-scale heterogeneities (e.g., Holliger, 1997; Eaton et al., 2003; Bongajum et al., 2012). There are also weak side reflections coming from ~100 m east of the profile that are back-reflected within the profile at ~750 m (Figure 5.6d). These events are interpreted as VSP-type reflections associated with direct waves being reflected from sub-vertical structures. Finally, imaging and subsequent time-to-depth conversion of the data was performed using Kirchhoff post-stack time migration with a constant velocity of 7000 m/s. In accordance with the presumed backscattered energy present in the stacked section, a small migration aperture of 100 m was used. While the use of a small migration aperture limits the proper imaging of steeply dipping specular reflections, this approach was deemed appropriate due to the absence of such signal characteristics in the stacked data. In turn, this allows to effectively focus the backscattered seismic events and, at the same time, to reduce artefacts, that is, “smiling” effects, in the imaged data. The VSP-type reflections observed in the CMP stack are also visible in the migrated section (Figure 5.6e).

Prior to migration there are pervasive, steeply dipping coherent noise events observable throughout the stacked data. The apparent velocity of this noise is ~2000 m/s, thus eliminating the possibility of it being P-waves reflected from vertical boundaries, which would manifest themselves with an apparent velocity of ~3500 m/s. Conversely, the linear move-out of this coherent noise is consistent with an S-wave velocity of ~4000 m/s, which is realistic in the given context. Interpreting this steeply dipping coherent noise as backscattered S-waves and/or groundroll is also consistent with the work of Chopra and Marfurt (2014). Due to their steep dips, these events remain unaffected by migration, which was confirmed by testing migration velocities between 2800 and 7000 m/s specifically for this purpose. To obtain a cleaner final image, we therefore applied a targeted *f-k* fan filter to the stacked section prior to migration. Extensive tests with varying filter designs demonstrated that this process is robust and largely devoid of artefacts. Weak, yet pervasive, VSP-type side reflections are still present after migration and cut through the section as shown by the red and yellow lines in Figure 5.6e.

The processing flow of L1 outlined above, was also used for seismic lines L2 and L3, and the corresponding stacked and imaged sections are shown in Figures 5.7a and 5.7b, and Figures 5.7c and 5.7d, respectively. L2 runs oblique to the geological strike (Figure 5.2) and had an active river gravel quarry operating at its eastern bound. Consequently, the data contained strong and abundant cultural noise, which required very rigorous trace editing. For this reason, the effective average fold of L2 was reduced to ~50 with regard to its original nominal value of 77. The course of L3 is strongly crooked, running sub-parallel to strike (Figure 5.2) and on top of Quaternary sediments for the first ~200 m (Figure 5.7b). As mentioned above, the shot spacing for L3 had to be increased to 10 m, which reduced its nominal fold to 36. In spite of their lower fold and lower S/N, the overall character of the stacked (Figures 5.7a and 5.7b) and migrated (Figures 5.7c and 5.7d) sections of L2 and L3 rather closely resembles that of L1 (Figures 5.6d and 5.6e).

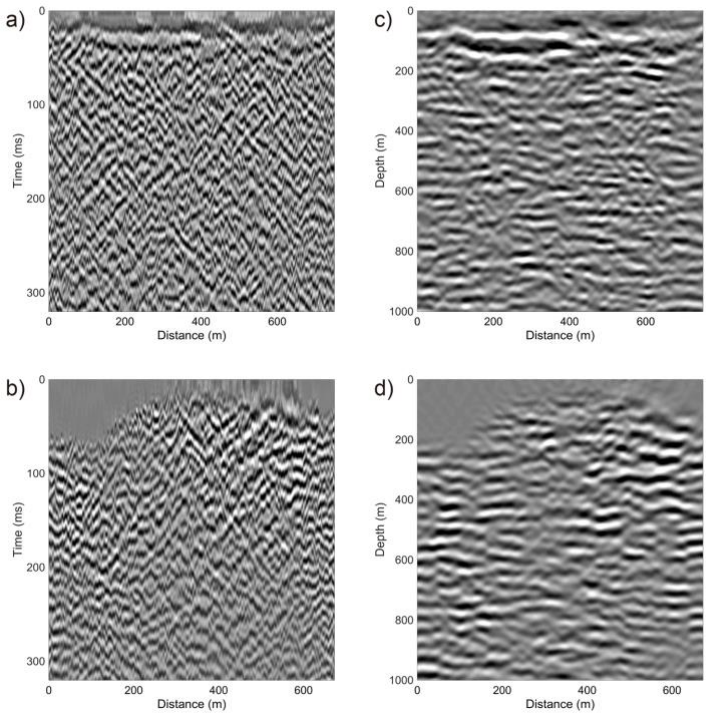


Figure 5.7: CMP stacks of a) L2 and b) L3. Kirchhoff migration images after time-to-depth conversion of c) L2 and d) L3. Please note that the course L3 is strongly crooked with the first ~200 m running sub-parallel to strike.

5.4 Interpretation

The resulting stacks and migrated images of all three seismic lines are devoid of laterally coherent reflectors (Figures 5.6 and 5.7), which is consistent with little to no energy being reflected to the surface from larger-scale, specular-type structures. This, in turn, indicates that the Insubric Line, as well as its associated secondary structures, are likely to remain sub-vertical within the subsurface region imaged by our seismic survey. This first-order assessment is consistent with the fact that the migrated and depth-converted images of L1, L2 and L3, are characterized by local bright lineaments, which subtly change in dip and seem to align sub-vertically. As such, these events could represent the apexes of seismic energy backscattered from local heterogeneities associated with the prevailing larger-scale sub-vertical structural grain (Horstmann, 1981).

In the following, we assess the above hypothesis through the application of attribute and geostatistical analyses of the seismic images. The validity of these approaches, which, as of yet, have not been used in a related context, is tested on synthetic seismic reflection surveys for canonical models of crustal heterogeneity with a sub-vertical structural grain.

5.4.1 Synthetic seismic data

To explore the seismic expressions of a sub-vertical structural grain associated with local heterogeneities, we evaluate and analyse synthetic seismic data for canonical crustal models based on Holliger et al.'s (1993) geostatistical conceptualization of Ivrea-type lower crust (Figure 5.8). We consider both a standard Gaussian-distributed stochastic model as well as a binarized version thereof, which allows us to explore the end-member-type seismic responses of smoothly varying and abruptly changing material properties along sub-vertical structures (Figure 5.8). The density fluctuations in both models are characterized by a so-called von Karman autocovariance function (e.g., Tronicke and Holliger, 2005) with vertical and horizontal correlation lengths a_z and a_x of 800 m and 200 m, respectively, a Hurst number ν of 0.3, and a constant mean density of 2800 kg/m^3. Both the P- and S-wave velocities are kept constant at 7000 m/s and 4200 m/s, respectively, to facilitate the subsequent imaging and to reduce associated biases. The Gaussian-distributed model is characterized by standard deviation of the density fluctuations of ~100 kg/m^3, while the density values in the binary-distributed model are either 2600 kg/m^3 or 2800 kg/m^3. As such, the average impedance contrasts in the binary-distributed model are significantly higher than those of its Gaussian-distributed counterpart. The considered autocovariance model as well as its correlation lengths and ν-value are consistent with the results of Holliger et al. (1993) inferred from the combined stochastic analysis of geological maps and rock physical properties for some key locations in the IVZ. The binary-distributed model can be regarded as a first-order approximation of a complex interlayering of mafic and felsic rocks locally prevailing in the IVZ (Holliger et al., 1993), while the Gaussian-distributed model emulates compositionally less heterogeneous upper crustal granitic and gneissic rocks (e.g., Holliger, 1996, 1997) and, hence, can be expected to rather representative of the gneisses and mylonites prevailing in the Sesia Zone and the Insubric Line, respectively.

For these models, we then generated synthetic seismic reflection surveys using a staggered-grid finite-difference solution of the elastic wave equation, which is fourth-order accurate in space and second-order accurate in time (e.g., Levander, 1988). All key acquisition and processing parameters were chosen to emulate those of the field data. The source is a Ricker wavelet with a centre frequency of 100 Hz. Figures 5.9 and 5.10 show the resulting stacked and migrated synthetic seismic sections. The data for the binary-distributed model

exemplarily illustrate the potential of the chosen acquisition and processing strategy for detecting and imaging seismic energy backscattered from local heterogeneities associated with larger-scale sub-vertical structures characterized by strong impedance contrasts. This backscattered energy is also present, albeit in a much less prominent manner, in the synthetic data for the Gaussian-distributed stochastic model. In this context, it is important to note that, in terms of the overall characteristics, both the stacked and migrated field data of L1 bear a conceptual similarity with their synthetic counterparts for the Gaussian-distributed model. This observation is consistent with the fact, that L1 is primarily located in gneissic and mylonitic rocks, for which the fluctuations in the material properties are indeed expected to be continuous and quasi-Gaussian (e.g., Holliger, 1996, 1997).

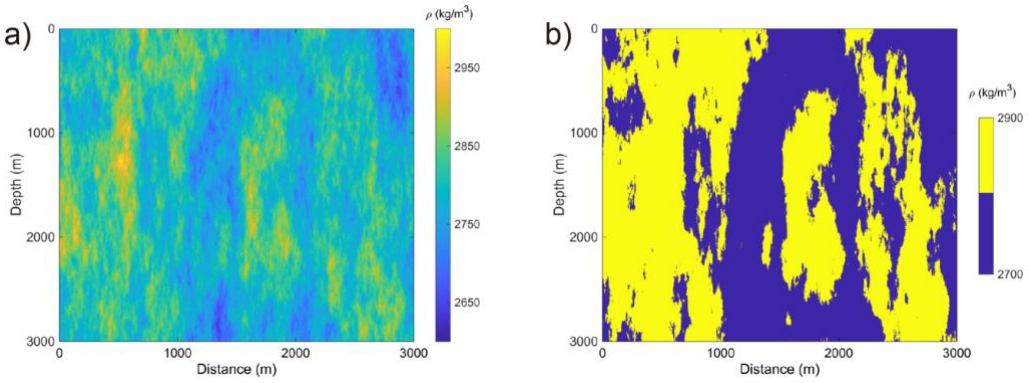


Figure 5.8: a) Gaussian- and b) binary-distributed density models of crustal heterogeneity (Holliger et al., 1993; Holliger, 1996, 1997). The density fluctuations of these models are characterized by a von Karman autocovariance function with vertical and horizontal correlation lengths, a_z and a_x, of 800 m and 200 m, respectively, and a ν-value of 0.3. The P- and S-wave velocities are constant at 7000 m/s and 4200 m/s, respectively.

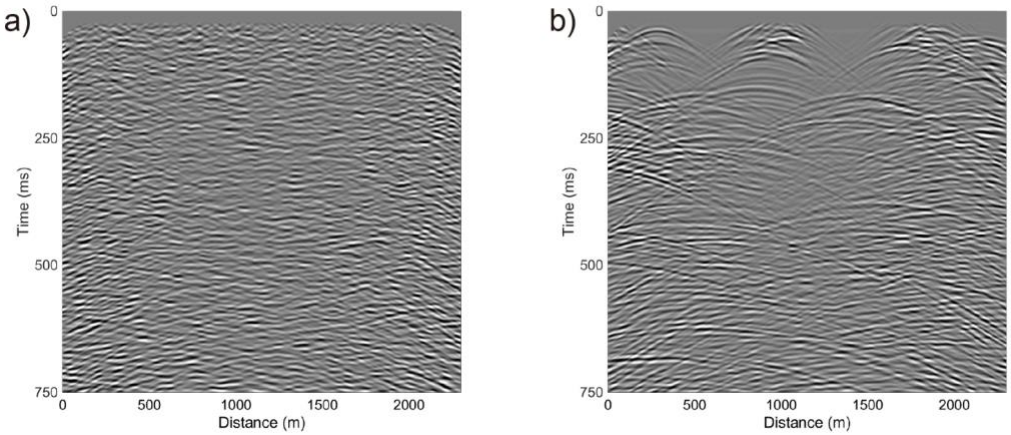


Figure 5.9: CMP-stacked synthetic seismic reflection data obtained for the a) Gaussian- and b) binary-distributed heterogeneous crustal models shown in Figure 5.8.

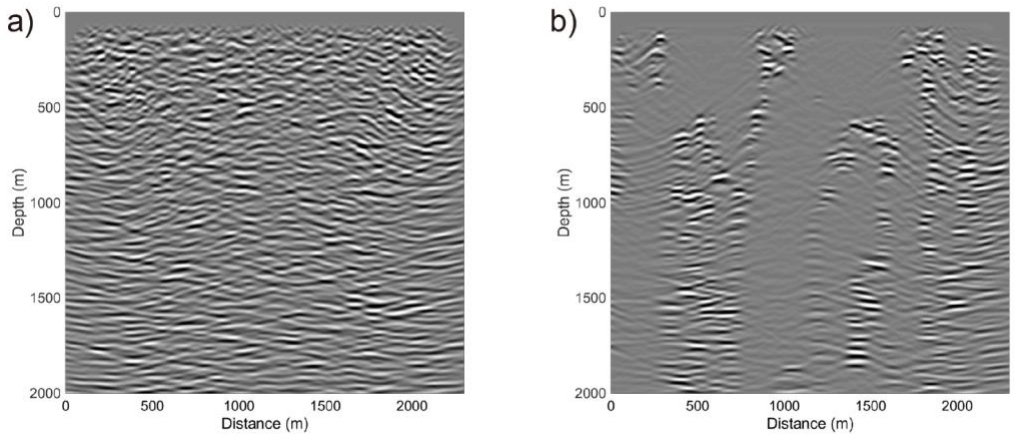


Figure 5.10: Migrated and depth-converted images of the stacked synthetic seismic reflection data obtained from the a) Gaussian- and b) binary-distributed heterogeneous crustal models shown in Figure 5.8.

5.4.2 Attribute analysis and geostatistical inversion

To enhance the seismic energy backscattered from heterogeneities associated with sub-vertical structures, we subjected the migrated versions of both synthetic and field data to energy-based attributed analysis (Figure 5.11). This attribute is commonly used in sedimentary environments to assist the detection and characterization of lateral structural discontinuities associated with, for example, faults and chimneys (e.g., Avseth et al., 2010). Here, we explored the potential utility of this attribute in complex hardrock environments for revealing the coherence between individual "bright lineaments" associated with seismic energy backscattered seismic from local heterogeneities associated with the overall structure grain. The backscattered energy was calculated using a moving window with a length of 60 ms along each trace and defined as the sum of the squared sample values within the window normalized by the number of samples. Figure 5.11 shows a comparison of the energy images of synthetic seismic data for Gaussian- and binary-distributed models with that for the field data recorded along L1, L2, and L3. This illustrates the consistency amongst the observed data as well as their similarity in character with the synthetic data in general and those for the Gaussian-distributed model in particular, which, in turn, qualitatively supports the hypothesis outlined above and, thus, the prevalence of the sub-vertical structures throughout the imaged part of the upper crust.

To further corroborate and quantify this interpretation, we proceeded to analyse the migrated synthetic and field data using the method of Irving et al. (2009), which relates the geostatistical properties of images of the backscattered wavefields to those of the scattering media. Specifically, this method allows to estimate the structural aspect ratio of the underlying heterogeneity and, to a lesser extent, also its complexity, which for the von Karman autocovariance model is quantified by the Hurst number ν. While this technique has been applied successfully to heterogeneous sub-horizontally structured environments (e.g., Irving et al., 2009; Irving and Holliger, 2010; Scholer et al., 2010; Xu et al., 2020), its applicability in the presence of sub-vertical structures was as of yet unproven. In the following, we therefore first assess the viability of this approach for our synthetic data and subsequently apply it to the observed data.

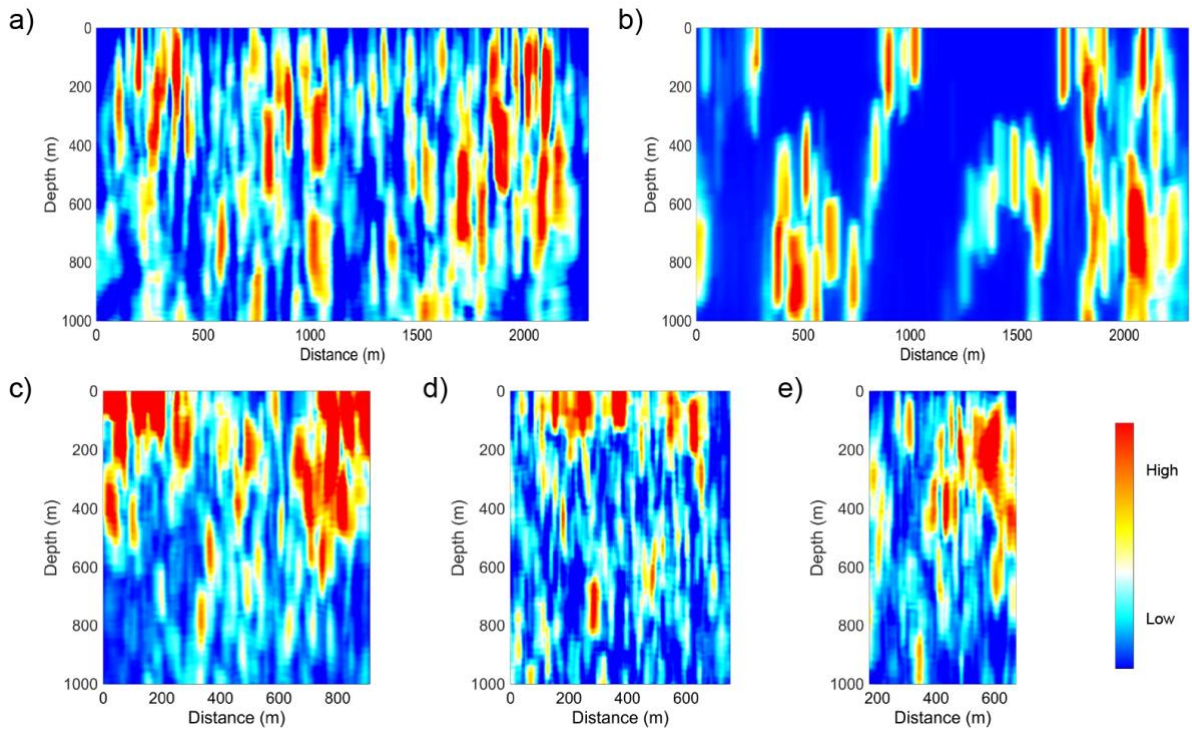


Figure 5.11: Energy-based attribute analysis of the migrated depth images of the synthetic data for a) Gaussian- and b) binary-distributed heterogeneous crustal models as well as for the field data recorded along c) L1, d) L2, and e) L3.

Figure 5.12 shows the corresponding validation for the migrated synthetic seismic images for the Gaussian- and binary-distributed models as well as the results obtained for the field data recorded along L1, L2, and L3 based on a Monte-Carlo-type inversion approach (e.g., Xu et al., 2020). Please note that the first ~200 m of L3, which run sub-parallel to strike, have been excluded from this analysis. We see that the resulting statistics for the synthetic data (Figures 5.12a and 5.12b) are largely consistent with the ratio of the vertical to horizontal correlation lengths a_z/a_x of 4 for the underlying models. The results for the Hurst number ν are, as expected from previous studies (e.g., Xu et al., 2020), less well constrained, but nevertheless in the correct range. These results thus seem to confirm the validity of the method of Irving et al. (2009) even in the presence sub-vertical structures. The corresponding analyses of the field data also point to a structural aspect ratio a_z/a_x that is significantly larger than unity as well as to a rather small ν-value. While the former agrees with the presumed dominance of sub-vertically aligned heterogeneites , the latter is consistent with the observation that the stochastic distribution of elastic properties in crystalline rocks seems to be universally characterized by low ν-values (e.g., Holliger, 1996, 1997). An interesting observation is that the character of the histogram of the a_z/a_x-values for L2 differs from that for the synthetic data as well as that for L1 and L3, which may be due to the fact the L2 runs oblique to strike and has a lower S/N than the two other lines.

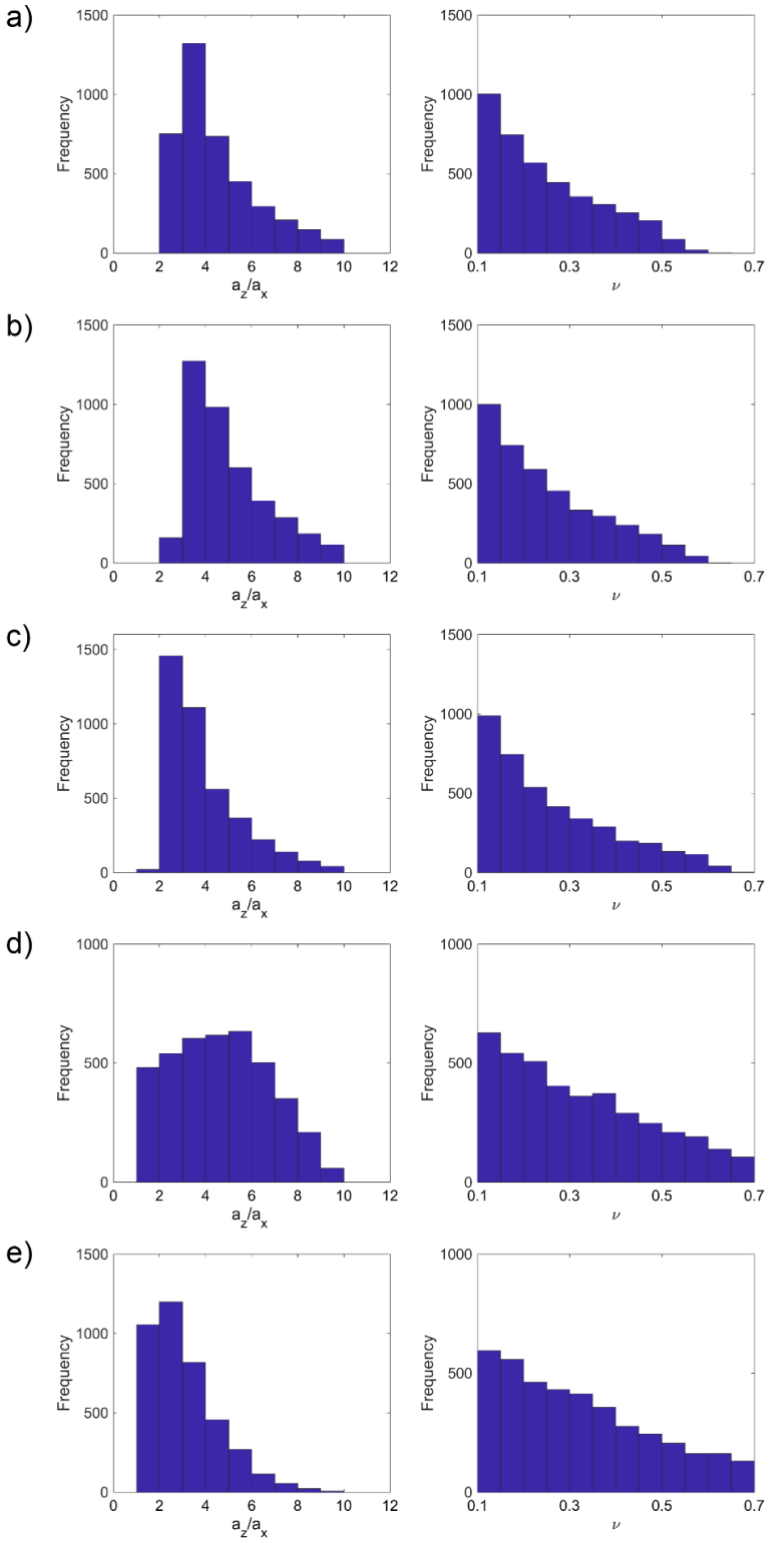


Figure 5.12: Results of the Monte-Carlo-type geostatistical inversion based on the method of Irving et al. (2009) of the migrated depth images of for vertical-to-horizontal aspect ratio and the ν-value for the synthetic seismic data based on a) Gaussian- and b) binary-distributed heterogeneous crustal models and the field data recorded along c) L1, d) L2, and e) L3. The mean values for a_z/a_x in a), b), c), d, and e) are 4.5, 4.9, and 3.9, 4.7, and 3.1 respectively. The mean values for ν in a), b), c), d), and e) are 0.25, 0.25, 0.26, 0.32, and 0.32, respectively.

5.4.3 Dip angle analysis

The above qualitative and quantitative analyses indicate that the outcropping sub-vertical structures are likely to persist throughout the upper crustal section imaged by our seismic data. The question that still needs to be addressed concerns the average dip of the structural grain, which, in turn, provides information as to whether the dominant dips observed at the surface remain more or less constant or undergo significant changes at depth. We found that this problem can be approached by analysing the secondary lobes of the autocorrelations of the energy images (Figure 5.11), which reveal the larger-scale alignment of the backscattered energy across the entire section. To assess and validate this novel approach, we again generated synthetic seismic reflection data for our stochastic models with the same structures as before (Figure 5.8), but inclined by 10, 20, and 30 degrees with regard to the vertical. Figures 5.13 and 5.14 show the corresponding Gaussian- and binary-distributed seismic models together with the 2D autocorrelations of the corresponding energy images of the synthetic seismic data. While there is some interpretational leeway, the trend of secondary lobes of the autocorrelations is clearly indicative of the overall structural dip. Figure 5.15 shows the autocorrelation of the energy image of field data collected along L1, L2, and L3. Note that the first ~200 m of L3, which run sub-parallel to strike (Figure 5.3), have again not been considered in this analysis. The autocorrelations of the energy images of L1 and L3 exhibit prominent side lobes whose trends point to steep dips of the order of 10 degrees in the east-west direction. This is consistent with the predominant dips prevailing at the surface (Horstmann, 1981) and seems to confirm that the outcropping structural grain seems to persist throughout the seismically imaged part of the upper crust. Conversely, the result for L2, which runs oblique to strike over its entire length and has a lower S/N than L1 and L3, is more ambiguous due to the absence of well-defined side lobes.

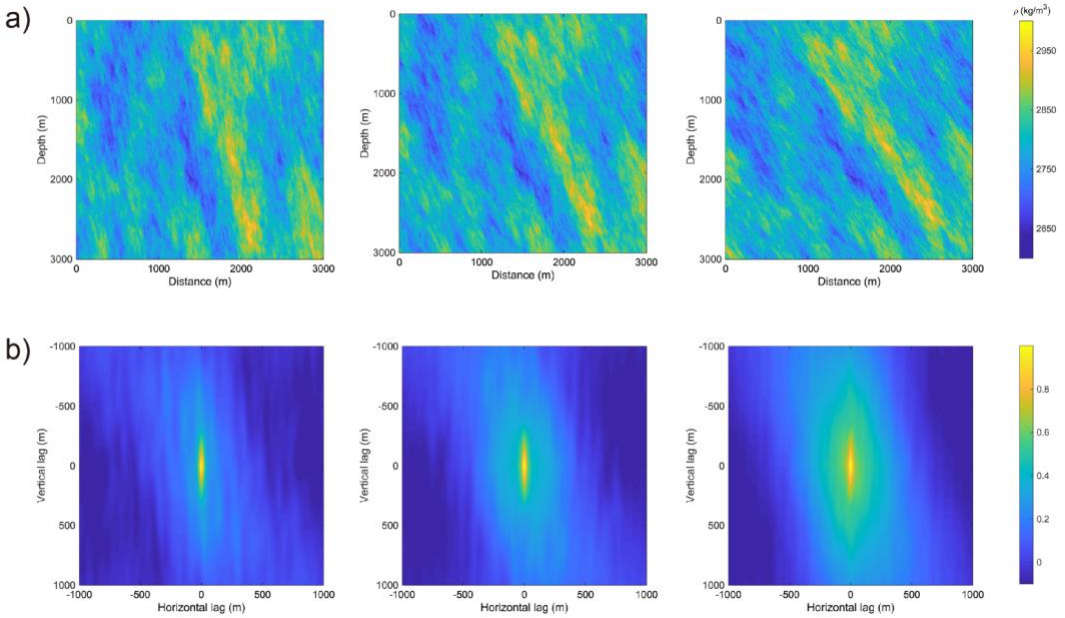


Figure 5.13: a) Gaussian-distributed crustal models with the same structures as in Figure 5.7 but dipping with 10, 20, and 30 degrees with regard to the vertical. b) 2D autocorrelations of the energy images of corresponding synthetic seismic data.

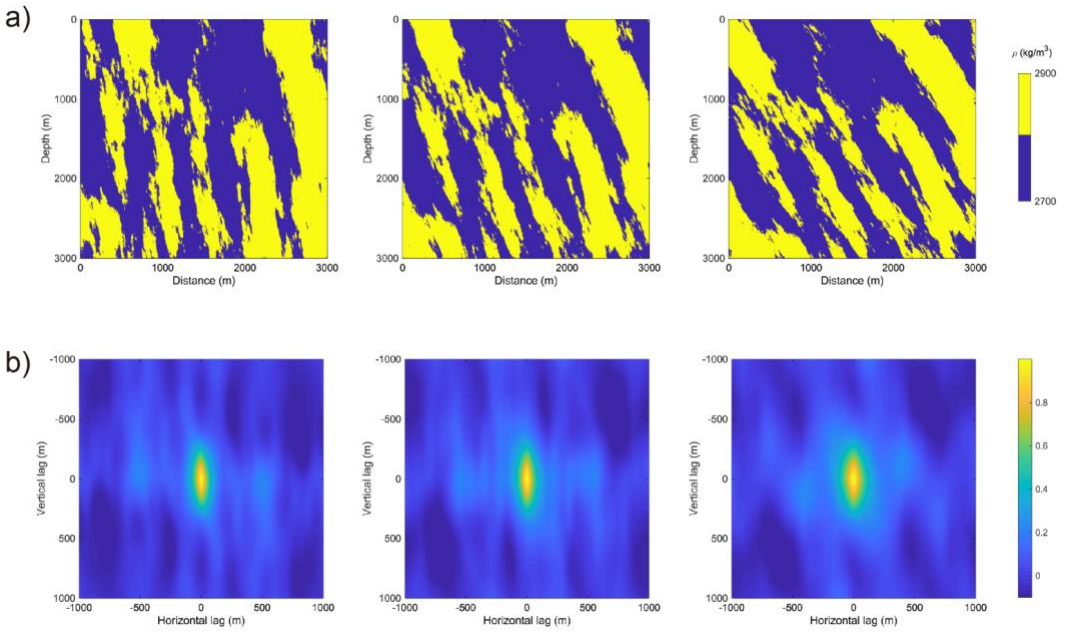


Figure 5.14: a) Binary-distributed crustal models with the same structures as in Figure 5.7 but dipping with 10, 20, and 30 degrees with regard to the vertical. b) 2D autocorrelations of the energy images of corresponding synthetic seismic data.

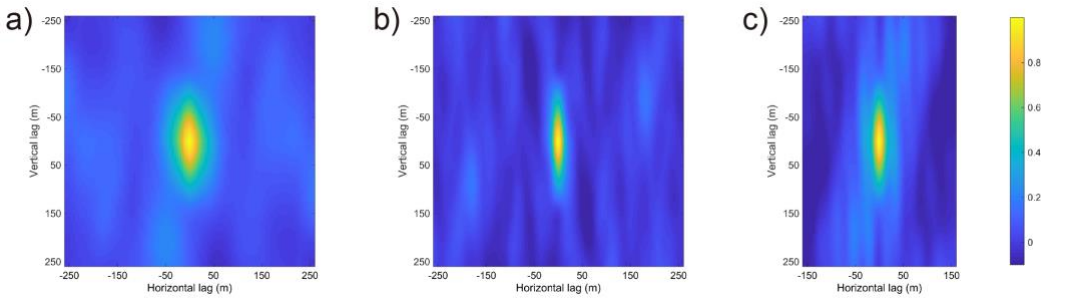


Figure 5.15: 2D autocorrelation of the energy images of the field data recorded along a) L1, b) L2, and c) L3. Please note that the westernmost ~200 m of L3, which run sub-parallel to strike, have not been included in the analysis.

Finally, Figure 5.16 provides a synoptic consolidation of the results of this study in form of an overlay-type representation of the seismic depth images and the associated energy-based attributed of L1, L2, and L3, juxtaposed with the local surface geology mapped by Horstmann (1981) and Quick (2003). Also shown are the dips of the predominant structures, namely the mylonitic border zone of Insubric Line crossed by L1 and the gabbro/peridotite contact near Balmuccia crossed by L3. For both of these sub-vertical structures, there is a clear correlation with changes in seismic character whose vertically continuous and laterally discontinuous nature is highlighted by the energy-based attributes. This representation of the seismic data is consistent with our previous qualitative and quantitative analyses, which indicate that the sub-vertical structural grain prevailing at the surface is likely to continue at least throughout the first one to two kilometres of subsurface imaged in our survey.

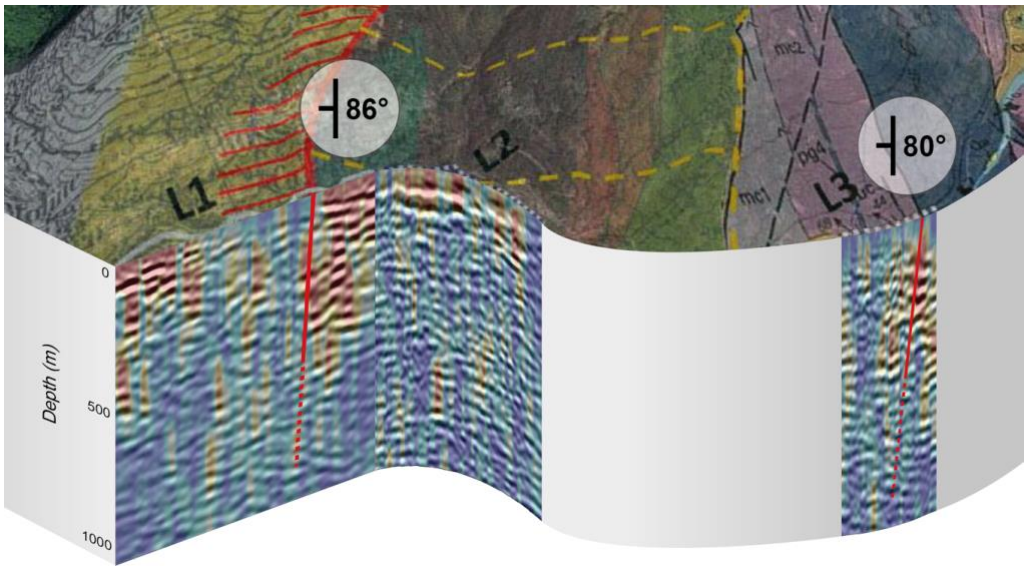


Figure 5.16: Presentation of the seismic data of L1, L2, and L3 relative to the geologic mapping of Horstmann (1981) and Quick (2003). The seismic data are displayed in greyscale overlain with the colorscale energy-based attributes from Figure 5.11. Please note that the first ~200 m of L3, which run sub-parallel to strike are not shown. The major structural dips mapped at surface (grey circles) are denoted within the sections by solid red lines, while their presumed continuation at depth is denoted by dashed red lines.

5.5 Discussion and Conclusions

We have presented the acquisition, processing, and interpretation of a high-resolution seismic reflection survey crossing the Insubric Line from the Sesia Zone into the IVZ. The objective of this survey, which consists of three ~1-km-long lines of variable S/N, was to characterize the first ~1 km of the subsurface in preparation for an ICDP drilling project aiming to explore lower continental crust exposed in the IVZ. Given that the structures exposed at the surface are sub-vertical, the key challenge for this project was to reliably assess whether the outcropping structural grain is continuous at depth or whether there are any indications for major flattening or shallow-angle cut-offs. Given this scenario, specular-type reflections were unlikely and the backscattered energy, which seemingly manifests itself as local bright lineaments in the seismic images, was tentatively interpreted as originating from small-scale heterogeneities associated with the sub-vertical large-scale structural grain.

To test this hypothesis and to corroborate the associated qualitative interpretation of the seismic images, we have generated synthetic seismic reflection surveys for sub-vertically structured canonical models of crustal heterogeneity. The acquisition and processing parameters of these synthetic seismic surveys emulate those of the field data. We consider both Gaussian- and binary-distributed heterogeneities for our models. While the stacked and imaged synthetic data for the Gaussian-distributed models bear a clear resemblance with the field data, the synthetic data for binary-distributed models exemplarily illustrate that the recorded backscattered energy originates from local heterogeneities associated with the overall sub-vertical large-scale structural grain. This interpretation was corroborated by performing energy-based attribute analysis as well as by a geostatistical inversion of the depth images of the observed and synthetic seismic data with regard to the structural aspect ratios of the backscattering structures. For the latter, we first needed to demonstrate the validity of the underlying method, which was originally conceived for sub-horizontal structured environments, in the presence of sub-vertical structures. Although, this geostatistical analysis is less sensitive to the

complexity of the underlying medium than to its structural aspect ratio, the corresponding results are consistent with the ubiquitous and seemingly universal characteristics of crystalline rocks. Finally, we illustrate that the alignment of the side lobes of the 2D autocorrelation of the seismic energy images allows to assess the overall dip of the prevailing larger-scale structural grain, which for the observed data are consistent with the sub-vertical east-west dip of the predominant structures mapped at the surface.

The qualitative and quantitative analyses of our seismic reflection surveys thus indicate that the sub-vertical structural grain exposed at the surface is likely to prevail throughout the imaged part of the crust. Beyond addressing this specific objective, the multi-faceted and unconventional approach pursued in this study provides interesting insights and potentially new perspectives with regard to the acquisition, processing, and interpretation of high-resolution seismic reflection data in crystalline terranes and their capacity for providing information on complex structures whose dips are too steep for deterministic imaging. Conversely, it is important to bear in mind that this seismic reflection survey violated the method's founding assumptions and, while the results of our unconventional interpretation approaches seem interesting and promising, we are unable to authoritatively validate and generalize them at this point. A critical question in this context, and indeed for most, if not all, 2D seismic reflection profiles, concerns the importance of 3D effects. This is particularly pertinent in our case, as, although the strike and dip of the larger-scale geological structures are quite stable along the survey, this quasi-2D assumption cannot be extended to the smaller-scale heterogeneities, which are superimposed on these larger structures and which represent the very target of our interpretation efforts. Rigorously addressing the associated question regarding the importance of 3D scattering effects on our results would require extensive 3D finite-difference-type elastic modelling for a wide range of scenarios. While this is clearly beyond the scope of the current work, it represents an interesting and pertinent research problem in its own right, which ultimately will need to be addressed in order to assess the reliability and robustness of the proposed interpretation approaches. At present, our admittedly quite speculative hypothesis in this regard is that, due to their inherent 3D nature, out-of-plane scattering effects would have a tendency of reducing, rather than enhancing, the sub-vertical alignment of backscattered seismic events with regard to the considered 2D reference case. This, in turn, would imply that we tend to underestimate, rather than overestimate, the ratio of the vertical to horizontal correlation lengths a_z/a_x.

Chapter 6

Conclusions and outlook

As stated in the introduction, characterization of heterogeneous subsurface environments is a challenging, yet important task. This PhD thesis contributes to the improving the ability of using GPR and seismic reflection surveys to constrain subsurface heterogeneous. Two different environments, i.e., the shallow subsurface and crystalline rocks, are explored with four novel quantitative interpretation techniques, based on which the four specific objectives outlined in the introduction has been largely met:

1) The approach presented in Chapter 2 makes it possible to conduct high-resolution velocity estimation using common-offset GPR reflection image only.
2) The weighting strategy presented in Chapter 3 improves the performance of diffraction-based velocity analysis.
3) The supervised-deep-learning scheme presented in Chapter 4 alleviates the computational cost of estimating geostatistical parameters.
4) The quantitative interpretation methods presented in Chapter 5 provide a novel view to characterize steeply dipping structures in crystalline environment.

In the following, the main conclusions and contributions of this PhD thesis are summarized.

6.1 Conclusions

In this thesis, four novel quantitative interpretation techniques for GPR and seismic reflection data are presented. The first two of them, described in Chapter 2 and 3, aim at constraining the shallow subsurface heterogeneity by inferring the spatially varying velocity structure from common-offset GPR reflection data. I assume that subsurface velocity distribution can be regarded as the sum of a smoothly varying background velocity field and a small-scale velocity fluctuation field. This assumption is common in geostatistical studies, yet not often used in deterministic type of inversion. Chapter 2 presented a novel idea, i.e., to separate the data into diffracted and reflected wavefields, use them to estimate the background velocity model and velocity fluctuation model, respectively, and combine them. An important and distinguishing feature of this methodology is that it does not require any borehole calibration or conditioning information (e.g., Schmelzbach et al., 2012; Xu et al., 2021), which is rather attractive for common-offset GPR reflection surveys. The step of background velocity model estimation with diffraction components is important because a local bias in the inferred model can largely affect the result. To improve the performance of diffraction-based velocity analysis, the weighting strategy based on a local semblance focusing measure presented in Chapter 3 is proposed and

tested. The use of weight function, which varies in accordance with the sensitivity of a diffraction curve to changes in migration velocity, can increase the resolution of the semblance spectra and reduce the uncertainties in velocity analysis. Generally, the methods described in Chapter 2 and 3 rely on the presence of an ample amount of evenly distributed diffractions throughout the recorded GPR section, as well as an effective isolation of the diffracted wavefields. However, when there is a lack of diffractions, complementary data such as common-midpoint-type analyses can be used to provide the large-scale velocity structure for the method described in Chapter 2.

The last two methods, presented in Chapter 4 and 5, adopt a geostatistical view to deal with the heterogeneity. More specifically, the main objective is to estimate quantitative parameters, such as aspect ratio of heterogeneity and dominant dip angle, to describe the geostatistical nature of subsurface. The estimated geostatistical properties can be further utilized for stochastic-type inversion, as described in Appendix A. Currently, the estimation of geostatistical parameters involves a Monte-Carlo-type inversion method. In Chapter 5, I applied the method to synthetic and field seismic reflection data for characterizing the heterogeneity that is sub-vertically structured. Yet this method has the limitation in computational speed. When there are a large number of datasets to be analyzed, it would be quite time-consuming to perform the estimation. To alleviate the computation cost, I presented a supervised-deep-learning scheme in Chapter 4, which provides a simple and efficient alternative to infer the aspect ratio of heterogeneity from surface-based GPR reflection images. But a disadvantage of the proposed method is that the results cannot show the uncertainties as the histogram results obtained from Monte Carlo inversion. Also, only horizontally layered structures were considered for simplicity, which is different from the case in Chapter 5. In Chapter 5, I presented an approach based on the autocorrelation of energy attribute of the reflection image to assess the dominant dip angle of steeply dipping structures. This work provides interesting insights and potentially new perspectives regarding the high-resolution seismic reflection surveys in crystalline terranes and their capacity for providing information on complex structures whose dips are too steep for deterministic imaging.

This thesis explored the capacities of recorded diffracted wavefields to constrain subsurface heterogeneity in Chapter 2, 3 and 5. For the shallow subsurface, the diffraction components of GPR data are used only for estimation of a smooth-varying background velocity model, which is the basis of inferring the fine-scale velocity distribution as described in Chapter 2. To improve the performance of diffraction-based velocity analysis, the weight function involving coherency and slope attributes of diffracted wavefields is designed and applied. The use of weight function for diffraction-related study is still rare and the methodology presented in Chapter 3 provides new perspectives. In crystalline rocks, seismic diffractions generated from the local heterogeneities associated with the prevailing sub-vertical structural grain are shown to be rather valuable to image complex structures in the subsurface.

Ultimately, the technical advances in this thesis have the potential to improve our understanding of the complex subsurface environments with GPR and seismic surveys. Though the current applications are limited to 2D cases, the extensions of the four proposed methods to 3D are conceptually straightforward.

6.2 Outlook

Though this thesis provides several valuable tools to constrain subsurface heterogeneity, there are still some limitations. The methodologies presented in Chapter 2, 3 and 5 utilize diffracted energy in 2D GPR or seismic reflection data, which is likely to come from out-of-plane sources. For diffraction-based velocity analysis, this

may lead to errors in the resulting velocity model. Though the computational cost of 3D diffraction velocity analysis is considerably high, it is worthwhile to explore the influence of such errors. In the seismic interpretation described in Chapter 5, rigorously addressing the associated question regarding the importance of 3D scattering effects is absent because it requires extensive 3D finite-difference-type elastic modelling for a wide range of scenarios, which is still a challenge. In Chapter 4 and 5, the geostatistical inversion of structural aspect ratio of the heterogeneity employs a standard Cartesian coordinate system, which means that the dip angle is assumed to be zero. However, this assumption is often not realistic enough for field surveys (e.g., Xu et al., 2020). Geostatistical inversion involving more parameters to describe the heterogeneity can largely increase our knowledge of subsurface.

In the following, an overview of some specific perspectives that the author considers important for the continuing advancement of subsurface heterogeneity characterization with GPR and seismic reflection data is presented.

1) The velocity estimation method introduced in Chapter 2 can be applied to marine seismic or deep land seismic reflection surveys. Most marine seismic reflection data are collected with single-channel systems or multichannel seismic cables with an offset-depth ratio, which is too small for velocity analysis based on common-midpoint-type processing. The case is the same for deep seismic profiling, where the target depth is considerably larger than the offsets. By using the proposed velocity estimation approach, a high-resolution velocity model can be provided for further data processing or quantitative interpretation of geological structures.

2) The exploration, characterization, and quantification of these uncertainties in diffraction-based velocity analysis process is necessary. As mentioned in Chapter 3, uncertainties in the resulting velocity model are not specific to migration focusing analysis, but similarly prevail in all other types of diffraction-based velocity analysis techniques. Though the proposed weighting strategy contributes to increasing the resolution of velocity spectra, the degree of improvement is not explored. Since the velocity corresponding to an out-of-plane diffraction has inherent deviation from the true subsurface velocity, it is also valuable to extend current diffraction-based velocity analysis techniques from 2D to 3D and quantify the errors caused by out-of-plane sources.

3) The supervised-deep-learning approach for geostatistical parameter estimation presented in Chapter 4 can be improved by considering multiple inputs to the neural network and/or training database consisting of GPR sections in different central frequencies. Also, the dominant dip angle of subsurface heterogeneity is an important parameter, of which the automatic estimation is an issue that needs to be considered. A complex but more realistic circumstance is that the subsurface is composed by several parts/facies holding different geostatistical properties. In this case, the automatic estimation could consider a two-step solution, i.e., a classification of facies first and parameter estimation for each classified part in the following.

4) In the canonical models of subsurface heterogeneities presented in Chapter 5, the sub-vertical structures can generate large amounts of scattered wavefields that are valuable for imaging and interpretation. This is an informative numerical study because the relationship between geological structure and heterogeneity is characterized in a novel way. Currently, most geophysical synthetic studies assume smooth geological interfaces in model building, which is obviously violated in the crystalline rock environment. Stochastic function seems to have the ability to shape realistic boundaries of hard rocks, and a focus of future research

could be the combination between stochastic function and geological map. To this end, the technique of multi-point geostatistics might be a powerful tool.

Bibliography

Allroggen, N., Tronicke, J., Delock, M., Böniger, U. (2015). Topographic migration of 2D and 3D ground-penetrating radar data considering variable velocities. *Near Surface Geophysics*, 13(3), 253–259.

Annan, A.P. (2005). Ground-Penetrating Radar. In Butler, D.K., editor, *Near-Surface Geophysics*, pp. 357-438. Society of Exploration Geophysicists.

Avseth, P., Mukerji, T., Mavko, G. (2010). *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press.

Babcock, E., and Bradford, J. H. (2015). Reflection waveform inversion of ground-penetrating radar data for characterizing thin and ultrathin layers of nonaqueous phase liquid contaminants in stratified media. *Geophysics*, 80(2), H1-H11.

Baker, G.S., Steeples, D.W., Schmeissner, C., Pavlovic, M., Plumb, R. (2001). Near-surface imaging using coincident seismic and GPR data. *Geophysical Research Letters*, 28(4), 627–630.

Barrash, W., Clemo, T. (2002). Hierarchical geostatistics and multifacies systems: Boise Hydrogeophysical Research Site, Boise, Idaho. *Water Resource Research*, 38(10), 14-1-14–18.

Bauer, A., Schwarz, B., Gajewski, D. (2020). Velocity inversion and scatterer detection with 3D P-cable data. In *SEG technical program expanded abstracts 2020*, pp. 3644–3648. Society of Exploration Geophysicists.

Bauer, A., Schwarz, B., Gajewski, D. (2017). Utilizing diffractions in wavefront tomography. *Geophysics*, 82(2), R65–R73.

Belina, F.A., Irving, J., Ernst, J.R., 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.

Berckhemer, H. (1968). Topographie des “Ivrea-Körpers” abgeleitet aus seismischen und gravimetrischen Daten. *Schweizerische Mineralogische und Petrographische Mitteilungen*, 48, 235–246.

Berger, A., Mercolli, I., Kapferer, N., Fügenschuh, B. (2012). Single and double exhumation of fault blocks in the internal Sesia-Lanzo Zone and the Ivrea-Verbano Zone (Biella, Italy). *International Journal of Earth Sciences*, 101(7), 1877–1894.

Binley, A., Hubbard, S.S., Huisman, J.A., Revil, A., Robinson, D.A., Singha, K. and Slater, L.D. (2015). The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. *Water Resources Research*, 51(6), 3837-3866.

Blindow, N. (2006). Ground penetrating radar in groundwater geophysics. In Kirsch Reinhard, editor, *Groundwater Geophysics*, pp. 226-252. Springer.

Bosch, M., Mukerji, T., Gonzalez, E.F. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics*, 75(5), 75A165-75A176.

Bozdağ, E., Trampert, J., Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2), 845–870.

Brack, P., Ulmer, P., Schmid, S.M. (2010). A crustal-scale magmatic system from the Earth's mantle to the Permian surface: Field trip to the area of lower Valsesia and Val d'Ossola (Massiccio dei Laghi, Southern Alps, Northern Italy). *Swiss Bulletin für angewandte Geologie*, 15, 3–21.

Bradford, J.H. (2008). Measuring Water Content Heterogeneity Using Multifold GPR with Reflection Tomography. *Vadose Zone Journal*, 7(1), 184–193.

Bradford, J.H., Clement, W.P., Barrash, W. (2009). Estimating porosity with ground-penetrating radar reflection tomography: A controlled 3-D experiment at the Boise Hydrogeophysical Research Site. *Water Resource Research*, 45(4), 1–11.

Bradford, J.H. and Deeds, J.C. (2006). Ground-penetrating radar theory and application of thin-bed offset-dependent reflectivity. *Geophysics*, 71(3), K47-K57.

Bradford, J.H., Privette, J., Wilkins, D., Ford, R. (2018). Reverse-Time Migration from Rugged Topography to Image Ground-Penetrating Radar Data in Complex Environments. *Engineering*, 4(5), 661–666.

Bueno, A., Benitez, C., De Angelis, S., Diaz Moreno, A., Ibanez, J.M. (2020). Volcano-Seismic Transfer Learning and Uncertainty Quantification with Bayesian Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2), 892–902.

Burnett, W., Fomel, S. (2011). Azimuthally anisotropic 3D velocity continuation. *International Journal of Geophysics*, 2011.

Busch, S., Van Der Kruk, J., Bikowski, J., Vereecken, H. (2012). Quantitative conductivity and permittivity estimation using full-waveform inversion of on-ground GPR data. *Geophysics*, 77(6), H79-H91.

Buttkus, B. (2012). *Spectral analysis and filter theory in applied geophysics*. Springer Science & Business Media.

Caers, J. (2007). Comparing the gradual deformation with the probability perturbation method for solving inverse problems. *Mathematical Geology*, 39(1), 27-52.

Carcione, J.M., Cavallini, F. (1995). On the acoustic-electromagnetic analogy. *Wave Motion*, 21(2), 149–162.

Carcione, J.M., Robinson, E.A. (2002). On the acoustic-electromagnetic analogy for the reflection-refraction problem. *Studia Geophysica et Geodaetica*, 46(2), 321–346.

Cardiff, M., Bakhos T., Kitanidis P. K., Barrash W. (2013), Aquifer heterogeneity characterization with oscillatory pumping: Sensitivity analysis and imaging potential. *Water Resources Research*, 49(9), 5395-5410.

Caspari, E., Greenwood, A., Baron, L., Egli, D., Toschini, E., Hu, K., Holliger, K. (2020). Characteristics of a fracture network surrounding a hydrothermally altered shear zone from geophysical borehole logs. *Solid Earth*, 11(3), 829–854.

Chan, C. Y., and Knight, R J. (2001). Laboratory measurements of electromagnetic wave velocity in layered sands. *Water Resources Research*, 37(4), 1099-1105.

Chen, Y., Liu, T., Chen, X. (2015). Velocity analysis using similarity-weighted semblance. *Geophysics*, 80(4), A75–A82.

Chen, Z., Fomel, S., Lu, W. (2013). Accelerated plane-wave destruction. *Geophysics*, 78(1), 1–9.

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

Claerbout, J.F. (1985). *Imaging the Earth's Interior*. Oxford: Blackwell Scientific Publications.

Clair, J.S., Steven Holbrook, W. (2017). Measuring snow water equivalent from common-offset GPR records through migration velocity analysis. *The Cryosphere*, 11(6), 2997–3009.

Cui, T., and Margrave, G. F. (2014). Seismic wavelet estimation. *University of Calgary CREWES Project Research Report*, 26, 1-16.

Dafflon, B., Barrash, W. (2012). Three-dimensional stochastic estimation of porosity distribution: Benefits of using ground-penetrating radar velocity tomograms in simulated-annealing- based or Bayesian sequential simulation approaches. *Water Resource Research*, 48(5), 1–13.

Dafflon, B., Irving, J., Barrash, W. (2011). Inversion of multiple intersecting high-resolution crosshole GPR profiles for hydrological characterization at the Boise Hydrogeophysical Research Site. *Journal of Applied Geophysics*, 73(4), 305–314.

Dafflon, B., Irving, J., Holliger, K. (2009). Use of high-resolution geophysical data to characterize heterogeneous aquifers: Influence of data integration method on hydrological predictions. *Water Resource Research*, 45(9), 1–12.

Dafflon, B., Tronicke, J., Holliger, K. (2006). Inferring the lateral subsurface correlation structure from georadar data: Methodological background and experimental evidence, In *Geostatistics for Environmental Applications*, pp. 467–478. Springer.

Decker, L., Merzlikin, D., Fomel, S. (2017). Diffraction imaging and time-migration velocity analysis using oriented velocity continuation. *Geophysics*, 82(2), U25–U35.

Dell, S., Gajewski, D. (2011). Common-reflection-surface-based workflow for diffraction imaging. *Geophysics*, 76(5), S187-S195.

Demirel-Schlueter, F., Krastel, S., Niessen, F., Demirbag, E., Imren, C., Toker, M., Litt, T., Sturm, M. (2005). Seismic pre-site survey for a potential new ICDP site—PaleoVan—at Lake Van, Turkey. In *Geophysical Research Abstracts*, pp. J-07997. European Geosciences Union.

Deparis, J., and Garambois, S. (2009). On the use of dispersive APVO GPR curves for thin-bed properties estimation: Theory and application to fracture characterization. *Geophysics*, 74(1), J1-J12.

Di, H., Abubakar, A. (2022). Estimating subsurface properties using a semi-supervised neural networks approach. *Geophysics*, 87(1), IM1-IM10.

Di, Q., Wang, M. (2004). Migration of ground-penetrating radar data with a finite-element method that considers attenuation and dispersion. *Geophysics*, 69(2), 472–477.

Diehl, T., Husen, S., Kissling, E., Deichmann, N. (2009). High-resolution 3-DP-wave model of the Alpine crust. *Geophysical Journal International*, 179(2), 1133–1147.

Dolan, S., and Bean, C. J. (1997). Some remarks on the estimation of fractal scaling parameters from borehole wire-line logs. *Geophysical Research Letters*, 24(10), 1271-1274.

Dong, Z., Feng, X., Zhou, H., Liu, C., Zeng, Z., Li, J., Liang, W. (2020). Properties analysis of lunar regolith at chang'E-4 landing site based on 3D velocity spectrum of lunar penetrating radar. *Remote Sensing*, 12(4),

629.

Duveneck, E. (2004). Velocity model estimation with data-derived wavefront attributes. *Geophysics*, 69(1), 265–274.

Earle, S. (2015). *Physical geology*. BCcampus.

Eaton, D.W., Milkereit, B., Salisbury, M.H. (2003). *Hardrock seismic exploration*. Society of Exploration Geophysicists.

Ekblom, H. (1973). Calculation of linear best Lp-approximations. *BIT Numerical Mathematics*, 13(3), 292–300.

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

Farquharson, C.G., Oldenburg, D.W. (1998). Non-linear inversion using general measures of data misfit and model structure. *Geophysical Journal International*, 134(1), 213–227.

Feng, X., Yu, Y., Liu, C., Fehler, M. (2015). Subsurface polarimetric migration imaging for full polarimetric ground-penetrating radar. *Geophysical Journal International*, 202(2), 1324–1338.

de Figueiredo, L.P., Grana, D., Bordignon, F.L., Santos, M., Roisenberg, M., Rodrigues, B.B. (2018). Joint Bayesian inversion based on rock-physics prior modeling for the estimation of spatially correlated reservoir properties. *Geophysics*, 83(5), M49-M61.

Fisher, E., McMechan, G.A., Annan, A.P., Cosway, S.W. (1992). Examples of reverse-time migration of single-channel, ground-penetrating radar profiles. *Geophysics*, 57(4), 577–586.

Fisher, S.C., Stewart, R.R., Jol, H.M. (1996). Ground Penetrating Radar (GPR) Data Enhancement Using Seismic Techniques. *Journal of Environmental and Engineering Geophysics*, 1(2), 89–96.

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

Fomel, S. (2003). Time-migration velocity analysis by velocity continuation. *Geophysics*, 68(5), 1662–1672.

Fomel, S. (2007a). Local seismic attributes. *Geophysics*, 72(3), 29–33.

Fomel, S. (2007b). Shaping regularization in geophysical-estimation problems. *Geophysics*, 72(2), R29–R36.

Fomel, S. (2009). Velocity analysis using AB semblance. *Geophysical Prospecting*, 57(3), 311–321.

Fomel, S., Guitton, A. (2006). Regularizing seismic inverse problems by model reparameterization using plane-wave construction. *Geophysics*, 71(5), 43–47.

Fomel, S., Landa, E., Taner, M.T. (2007). Poststack velocity analysis by separation and imaging of seismic diffractions. *Geophysics*, 72(6), 2559–2563.

Forte, E., Dossi, M., Colucci, R.R., Pipan, M. (2013). A new fast methodology to estimate the density of frozen materials by means of common offset GPR data. *Journal of Applied Geophysics*, 99, 135–145.

Forte, E., Dossi, M., Pipan, M., Colucci, R.R. (2014). Velocity analysis from common offset GPR data inversion: Theory and application to synthetic and real data. *Geophysical Journal International*, 197(3), 1471–1483.

Forte, E. and Pipan, M. (2017). Review of multi-offset GPR applications: Data acquisition, processing and

analysis. *Signal Processing*, 132, 210-220.

Fountain, D.M. (1976). The Ivrea—Verbano and Strona-Ceneri Zones, Northern Italy: A cross-section of the continental crust—New evidence from seismic velocities of rock samples. *Tectonophysics*, 33(1-2), 145–165.

Frankel, A., and Clayton, R. W. (1986). Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity. *Journal of Geophysical Research: Solid Earth*, 91(B6), 6465-6489.

Gelhar, L.W. (1993). *Stochastic subsurface hydrology*. Prentice-Hall.

Geng, Z., Zhao, Z., Shi, Y., Wu, X., Fomel, S., Sen, M. (2022). Deep learning for velocity model building with common-image gather volumes. *Geophysical Journal International*, 228(2), 1054–1070.

Gholami, A. (2015). Nonlinear multichannel impedance inversion by total-variation regularization. *Geophysics*, 80(5), R217–R224.

Giannopoulos, A. (2005). Modelling ground penetrating radar by gprMax. *Construction and Building Materials*, 19(10), 755-762.

Gibson, B.S., Levander, A.R. (1990). Apparent layering in common-midpoint stacked images of two-dimensionally heterogeneous targets. *Geophysics*, 55(11), 1466–1477.

Goff, J.A., Jordan, T.H. (1988). Stochastic modeling of seafloor morphology: Inversion of sea beam data for second-order statistics. *Journal of Geophysical Research: Solid Earth*, 93(B11), 13589–13608.

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

Grana, D., Azevedo, L., Liu, M. (2020). A comparison of deep machine learning and Monte Carlo methods for facies classification from seismic data. *Geophysics*, 85(4), WA41–WA52.

Grasmueck, M., Weger, R., Horstmeyer, H. (2005). Full-resolution 3D GPR imaging. *Geophysics*, 70(1), 12–19.

Greaves, R.J., Lesmes, D.P., Lee, J.M., Toksöz, M.N. (1996). Velocity variations and water content estimated from multi-offset, ground-penetrating radar. *Geophysics*, 61(3), 683–695.

Greenwood, A., Caspari, E., Egli, D., Baron, L., Zahner, T., Hunziker, J., Holliger, K. (2019). Characterization and imaging of a hydrothermally active near-vertical fault zone in crystalline rocks based on hydrophone VSP data. *Tectonophysics*, 750, 153–176.

Gueting, N., Klotzsche, A., van der Kruk, J., Vanderborght, J., Vereecken, H., Englert, A. (2015). Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests. *Journal of Hydrology*, 524, 680-695.

Hamid, H., Pidlisecky, A. (2015). Multitrace impedance inversion with lateral constraints. *Geophysics*, 80(6), M101–M111.

Handy, M.R., Ustaszewski, K., Kissling, E. (2015). Reconstructing the Alps–Carpathians–Dinarides as a key to understanding switches in subduction polarity, slab gaps and surface motion. *International Journal of Earth Sciences*, 104(1), 1–26.

Harlan, W.S., Claerbout, J.F., Rocca, F. (1984). Signal/noise separation and velocity estimation. *Geophysics*, 49(11), 1869–1880.

Hartzell, S., Harmsen, S., Frankel, A. (2010). Effects of 3D random correlated velocity perturbations on predicted ground motions. *Bulletin of the Seismological Society of America*, 100, 1415-1426.

Hochstetler, D.L., Barrash, W., Leven, C., Cardiff, M., Chidichimo, F., Kitanidis, P.K. (2016). Hydraulic Tomography: Continuity and Discontinuity of High-K and Low-K Zones. *Groundwater*, 54(2), 171–185.

Holliger, K. (1996). Upper-crustal seismic velocity heterogeneity as derived from a variety of P-wave sonic logs. *Geophysical Journal International*, 125(3), 813–829.

Holliger, K. (1997). Seismic scattering in the upper crystalline crust based on evidence from sonic logs. *Geophysical Journal International*, 128(1), 65–72.

Holliger, K., Levander, A., Carbonell, R., Hobbs, R. (1994). Some attributes of wavefields scattered from Ivrea-type lower crust. *Tectonophysics*, 232(1-4), 267–279.

Holliger, K., Levander, A.R., Goff, J.A. (1993). Stochastic modeling of the reflective lower crust: Petrophysical and geological evidence from the Ivera Zone (northern Italy). *Journal of Geophysical Research: Solid Earth*, 98(B7), 11967–11980.

Horstmann, U. (1981). Petrographische, tektonische und mikrothermometrische Untersuchungen an der Insubrischen Line im Val Sermenza (Provinz Vercelli, Norditalien). Diploma thesis, University of Göttingen.

Hubbard, S., and Linde, N. (2010). Hydrogeophysics. In Wilderer P.A., editor, *Treatise on Water Science*, pp. 401-434. Elsevier.

Hu, B.X., Meerschaert, M.M., Barrash, W., Hyndman, D.W., He, C., Li, X., Guo, L. (2009). Examining the influence of heterogeneous porosity fields on conservative solute transport. *Journal of Contaminant Hydrology*, 108(3-4), 77-88.

Hu, H., Zheng, Y. (2020). Stochastic Inversion of Gaussian Random Media Using Transverse Coherence Functions for Reflected Waves: Theory and Method. *Journal of Geophysical Research: Solid Earth*, 125(12), 1–13.

Huisman, J.A., Hubbard, S.S., Redman, J.D., Annan, A.P. (2003). Measuring Soil Water Content with Ground Penetrating Radar: A Review. *Vadose Zone Journal*, 2(4), 476–491.

Hunziker, J.C. (1980). Lower Palaeozoic amphibolite to granulite facies metamorphism in the Ivrea Zone (Southern Alps, northern Italy). *Schweizerische Mineralogische und Petrographische Mitteilungen*, 60 (2–3), 181–213.

Hurich, C.A., Kocurko, A. (2000). Statistical approaches to interpretation of seismic reflection data. *Tectonophysics*, 329(1-4), 251–267.

Ikelle, L.T., Yung, S.K., Daube, F. (1993). 2-D random media with ellipsoidal autocorrelation functions. *Geophysics*, 58(9), 1359-1372.

Irving, J., Holliger, K. (2010). Geostatistical inversion of seismic and ground-penetrating radar reflection images: What can we actually resolve?. *Geophysical Research Letters*, 37(21), 1–5.

Irving, J., Knight, R. (2006). Numerical modeling of ground-penetrating radar in 2-D using MATLAB. Computers & Geosciences, 32(9), 1247–1258.

Irving, J., Knight, R., Holliger, K. (2009). Estimation of the lateral correlation structure of subsurface water content from surface-based ground-penetrating radar reflection images. *Water Resources Research.* 45(12),

1–14.

Irving, J., Scholer, M., Holliger, K. (2010). Inversion for the stochastic structure of subsurface velocity heterogeneity from surface-based geophysical reflection images. In *Advances in near-surface seismology and ground-penetrating radar*, pp. 77–96. Society of Exploration Geophysicists.

Jäger, R., Mann, J., Höcht, G., Hubral, P. (2001). Common-reflection-surface stack: Image and attributes. *Geophysics*, 66(1), 97–109.

Jiang, C., Igel, J., Dlugosch, R., Müller-Petke, M., Günther, T., Helms, J., Lang, J., Winsemann, J. (2020). Magnetic resonance tomography constrained by ground-penetrating radar for improved hydrogeophysical characterization. *Geophysics*, 85(6), JM13–JM26.

Johnson, D.S., Aragon, C.R., McGeoch, L.A., Schevon, C. (1991). Optimization by simulated annealing: an experimental evaluation; part II, graph coloring and number partitioning. *Operations Research,* 39(3), 378-406.

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

Jones, A.G. and Holliger, K. (1997). Spectral analyses of the KTB sonic and density logs using robust nonparametric methods. *Journal of Geophysical Research: Solid Earth*, 102(B8), 18391-18403.

Juhlin, C., Dehghannejad, M., Lund, B., Malehmir, A., Pratt, G. (2010). Reflection seismic imaging of the end-glacial Pärvie Fault system, northern Sweden. *Journal of Applied Geophysics*, 70(4), 307–316.

Kaur, H., Pham, N., Fomel, S. (2020). Improving the resolution of migrated images by approximating the inverse Hessian using deep learning. *Geophysics*, 85(4), WA173–WA183.

Kelkar, M., Perez, G., Chopra, A. (2002). *Applied geostatistics for reservoir characterization*. Society of Petroleum Engineers.

Khaidukov, V., Landa, E., Moser, T.J. (2004). Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution. *Geophysics*, 69(6), 1478–1490.

Khoshnavaz, M.J., Bóna, A., Hossain, M.S., Urosevic, M., Chambers, K. (2016). Diffractivity—Another attribute for the interpretation of seismic data in hard rock environment, a case study. *Interpretation*, 4(4), B23–B32.

Kissling, E., Wagner, J.J., Mueller, S. (1984). Three-dimensional gravity model of the northern Ivrea-Verbano Zone. In Wagner, J.-J., St. Müller, Editors, *Geomagnetic and Gravimetric Studies of the Ivrea Zone: Matér. Géol. Suisse. Géophys*, pp. 55–61. Kümmerly & Frey.

Klem-Musatov, K. (1994). *Theory of seismic diffractions*. Society of Exploration Geophysicists.

Klokov, A., Fomel, S. (2012). Separation and imaging of seismic diffractions using migrated dip-angle gathers. *Geophysics*, 77(6), S131-S143.

Klotzsche, A., Jonard, F., Looms, M.C., van der Kruk, J., Huisman, J.A. (2018). Measuring soil water content with ground penetrating radar: A decade of progress. *Vadose Zone Journal*, 17(1), 1–9.

Klotzsche, A., Vereecken, H., 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.

Knight, R. (2001). Ground penetrating radar for environmental applications. *Annual Review of Earth and*

Planetary Sciences, 29(1), 229–255.

Knight, R., Tercier, P., Irving, J. (2004). The effect of vertical measurement resolution on the correlation structure of a ground penetrating radar reflection image. *Geophysical Research Letters*, 31(21), 2–5.

Knight, R.J., Irving, J.D., Tercier, P., Freeman, G.J., Murray, C.J., Rockhold, M.L. (2007). A comparison of the use of radar images and neutron probe data to determine the horizontal correlation length of water content. In *Subsurface hydrology: Data integration for properties and processes*. American Geophysical Union.

Kobr, M. (2021). Geophysical well logging. In *Encyclopedia of solid earth geophysics*, pp. 527-537. Springer International Publishing.

Lai, W.W.-L., Derobert, X., Annan, P. (2018). A review of Ground Penetrating Radar application in civil engineering: A 30-year journey from Locating and Testing to Imaging and Diagnosis. *NDT & E International*, 96, 58–78.

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Conference on inverse scattering, theory and application. Society of Industrial and Applied Mathematics.

Lampe, B., Holliger, K. (2003). Effects of fractal fluctuations in topographic relief, permittivity and conductivity on ground-penetrating radar antenna radiation. *Geophysics*, 68(6), 1934–1944.

Lauzon, D. and Marcotte, D. (2019). Calibration of random fields by FFTMA-SA. *Computers & Geosciences*, 127, 99-110.

Lavoué, F., Brossier, R., Métivier, L., Garambois, S., 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.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Lehmann, F., Green, A.G. (2000). Topographic migration of georadar data: Implications for acquisition and processing. *Geophysics*, 65(3), 836–848.

Leuschen, C., Plumb, R. (2000). A matched-filter approach to wave migration. *Journal of Applied Geophysics*, 43(2-4), 271–280.

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

Leven, C. and Dietrich, P. (2006). What information can we get from pumping tests?-comparing pumping test configurations using sensitivity coefficients. *Journal of Hydrology*, 319(1-4), 199-215.

Levin, S.A. (1984). Principle of reverse-time migration. *Geophysics*, 49(5), 581–583.

Le Ravalec, M., Hu, L.Y., Roggero, F. (2008). Reconstruction of existing reservoir model for its calibration to dynamic data. *Earth Science Frontiers*, 15(1), 176-186.

Le Ravalec, M. and Mouche, E. (2012). Calibrating transmissivities from piezometric heads with the gradual deformation method: An application to the Culebra Dolomite unit at the Waste Isolation Pilot Plant (WIPP), New Mexico, USA. *Journal of Hydrology*, 472, 1-13.

Le Ravalec, M., Noetinger, B.. Hu, L.Y. (2000). The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning Gaussian simulations. *Mathematical Geology*, 32(6), 701-723.

Le Ravalec, M., Coureaud, B., Nicolas, L., Roggero, F. (2004). Conditioning an underground gas storage site to well pressures. *Oil & Gas Science and Technology*, 59(6), 611-624.

Li, C., Zhang, J. (2021). Velocity analysis using separated diffractions for lunar penetrating radar obtained by yutu-2 rover. *Remote Sensing*, 13(7), 1387.

Liang, M., Marcotte, D., Shamsipour, P. (2016). Simulation of non-linear coregionalization models by FFTMA. *Computers & Geosciences*, 89, 220-231.

Liu, Y., Greenwood, A., Hetényi, G., Baron, L., Holliger, K. (2021). High-resolution seismic reflection survey crossing the Insubric Line into the Ivrea-Verbano Zone: Novel approaches for interpreting the seismic response of steeply dipping structures. *Tectonophysics*, 816, 229035.

Liu, Y., Irving, J., Holliger, K. (2022). High-resolution velocity estimation from surface-based common-offset GPR reflection data. *Geophysical Journal International*, 230(1), 131–144.

Liu, Y., Shi, Z., Wang, B., Yu, T. (2018). GPR impedance inversion for imaging and characterization of buried archaeological remains: A case study at Mudu city cite in Suzhou, China. *Journal of Applied Geophysics*, 148, 226–233.

Lowney, B., Lokmer, I., O'Brien, G.S. (2021). Multi-domain diffraction identification: A supervised deep learning technique for seismic diffraction classification. *Computers & Geosciences*, 155, 104845.

Lu, Y., Stehly, L., Paul, A., Group, A.W. (2018). High-resolution surface wave tomography of the European crust and uppermost mantle from ambient seismic noise. *Geophysical Journal International*, 214(2), 1136–1150.

Luo, S., Hale, D. (2012). Velocity analysis using weighted semblance. *Geophysics*, 77(2), U15-U22.

Malehmir, A., Juhlin, C. (2010). An investigation of the effects of the choice of stacking velocities on residual statics for hardrock reflection seismic processing. *Journal of Applied Geophysics*, 72(1), 28–38.

Maliva, R.G. (2016). *Aquifer Characterization Techniques*. Springer.

Mangel, A.R., Moysey, S.M., Bradford, J. (2020). Reflection tomography of time-lapse GPR data for studying dynamic unsaturated flow phenomena. *Hydrology and Earth System Sciences*, 24(1), 159-167.

McMechan, G.A. (1982). Determination of source parameters by wavefield extrapolation. *Geophysical Journal International*, 71(3), 613–628.

Merzlikin, D., Meckel, T.A., Fomel, S., Sripanich, Y. (2017). Diffraction imaging of high-resolution 3D P-cable data from the Gulf of Mexico using azimuthal plane-wave destruction. *First Break*, 35(2).

Moser, T.J., Howard, C.B. (2008). Diffraction imaging in depth. *Geophysical Prospecting*, 56(5), 627–641.

Mount, G.J., Comas, X., Cunningham, K.J. (2014). Characterization of the porosity distribution in the upper part of the karst Biscayne aquifer using common offset ground penetrating radar, Everglades National Park, Florida. *Journal of Hydrology*, 515, 223-236.

Neal, A. (2004). Ground-penetrating radar and its use in sedimentology: principles, problems and progress. *Earth-science Reviews*, 66(3-4), 261–330.

Nichols, J., Mikesell, D., Van Wijk, K. (2010). Application of the virtual refraction to near-surface characterization at the Boise Hydrogeophysical Research Site. *Geophysical Prospecting*, 58(6), 1011-1022.

Novais, A., Costa, J., Schleicher, J. (2008). GPR velocity determination by image-wave remigration. *Journal of Applied Geophysics*, 65(2), 65–72.

Nussbaumer, R., Linde, N., Mariethoz, G., Holliger, K. (2019). Simulation of fine-scale electrical conductivity fields using resolution-limited tomograms and area-to-point kriging. *Geophysical Journal International,* 218(2), 1322-1335.

Oden, C.P., Powers, M.H., Wright, D.L., Olhoeft, G.R. (2007). Improving GPR image resolution in lossy ground using dispersive migration. *IEEE Transactions on Geoscience and Remote Sensing*, 45(8), 2492–2499.

Oldenborger, G.A., Knoll, M.D., Barrash, W. (2004). Effects of signal processing and antenna frequency on the geostatistical structure of ground-penetrating radar data. *Journal of Environmental and Engineering Geophysics*, 9(4), 201–212.

Oliver, D.S. (1995). Moving averages for Gaussian simulation in two and three dimensions. *Mathematical Geology*, 27(8), 939-960.

Palmer, D., Jones, L. (2005). Computing detailed refraction statics in a hard rock terrain with the GRM and the RCS, in *SEG technical program expanded abstracts 2005*, pp. 1037–1040. Society of Exploration Geophysicists.

Perroud, H., Tygel, M. (2005). Velocity estimation by the common-reflection-surface (CRS) method: Using ground-penetrating radar data. *Geophysics*, 70(6), 43–52.

Petri, B., Duretz, T., Mohn, G., Schmalholz, S.M., Karner, G.D., Müntener, O. (2019). Thinning mechanisms of heterogeneous continental lithosphere. *Earth and Planetary Science Letters*, 512, 147–162.

Pirot, G., Linde, N., Mariethoz, G., Bradford, J.H. (2017). Probabilistic inversion with graph cuts: Application to the Boise Hydrogeophysical Research Site. *Water Resources Research*, 53(2), 1231–1250.

Pistone, M., Müntener, O., Ziberna, L., Hetényi, G., Zanetti, A. (2017). Report on the ICDP workshop DIVE (Drilling the Ivrea–Verbano zonE). *Scientific Drilling*, 23, 47–56.

Poppeliers, C. (2007). Estimating vertical stochastic scale parameters from seismic reflection data: Deconvolution with non-white reflectivity. *Geophysical Journal International*, 168(2), 769–778.

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.

Preine, J., Schwarz, B., Bauer, A., Hübscher, C. (2020). When There Is No Offset: A Demonstration of Seismic Diffraction Imaging and Depth-Velocity Model Building in the Southern Aegean Sea. *Journal of Geophysical Research: Solid Earth*, 125(9), 1–19.

Pullammanappallil, S., Levander, A., Larkin, S.P. (1997). Estimation of crustal stochastic parameters from seismic exploration data. *Journal of Geophysical Research: Solid Earth*, 102(B7), 15269–15286.

Quick, J.E., Sinigoi, S., Snoke, A.W., Kalakay, T.J., Mayer, A., Peressini, G. (2003). *Geologic map of the southern Ivrea-Verbano Zone, northwestern Italy*. US Geological Survey.

Rea, J., Knight, R. (1998). Geostatistical analysis of ground-penetrating radar data: A means of describing spatial variation in the subsurface. *Water Resources Research*, 34(3), 329–339.

Reshef, M., Landa, E. (2009). Post-stack velocity analysis in the dip-angle domain using diffractions.

Geophysical Prospecting, 57(5), 811–821.

Romero-Ruiz, A., Linde, N., Keller, T., Or, D. (2018). A review of geophysical methods for soil structure characterization. *Reviews of Geophysics*, 56(4), 672-697.

Rubin, Y., and Hubbard, S. (2006). *Hydrogeophysics*. Springer.

Russell, B., Hampson, D. (1991). Comparison of poststack seismic inversion methods. In *SEG technical program expanded abstracts 1991*, pp. 876–878. Society of Exploration Geophysicists.

Salamon, P., Fernandez-Garcia, D., Gómez-Hernández, J.J. (2007). Modeling tracer transport at the MADE site: The importance of heterogeneity. *Water Resources Research*, 43(8).

Salisbury, M.H., Milkereit, B., Ascough, G., Adair, R., Matthews, L., Schmitt, D.R., Mwenifumbo, J., Eaton, D.W., Wu, J. (2000). Physical properties and seismic imaging of massive sulfides. *Geophysics*, 65(6), 1882–1889.

Savran, W.H., Olsen, K.B. (2016). Model for small-scale crustal heterogeneity in Los Angeles basin based on inversion of sonic log data. *Geophysical Journal International*, 205(2), 856–863.

Scarponi, M., Hetényi, G., Berthet, T., Baron, L., Manzotti, P., Petri, B., Pistone, M., Müntener, O. (2020). New gravity data and 3D density model constraints on the Ivrea Geophysical Body (Western Alps). *Geophysical Journal International*, 222(3), 1977–1991.

Schijns, H., Heinonen, S., Schmitt, D.R., Heikkinen, P., Kukkonen, I.T. (2009). Seismic refraction traveltime inversion for static corrections in a glaciated shield rock environment: a case study. *Geophysical Prospecting*, 57(6), 997–1008.

Schmelzbach, C., Huber, E. (2015). Efficient deconvolution of ground-penetrating radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 53(9), 5209–5217.

Schmelzbach, C., Simancas, J.F., Juhlin, C., Carbonell, R. (2008). Seismic reflection imaging over the South Portuguese Zone fold-and-thrust belt, SW Iberia. *Journal of Geophysical Research: Solid Earth*, 113(B8).

Schmelzbach, C., Tronicke, J., Dietrich, P. (2012). High-resolution water content estimation from surface-based ground-penetrating radar reflection data by impedance inversion. *Water Resources Research*, 48(8), 1–16.

Schmid, S.M., Aebli, H.R., Heller, F., Zingg, A. (1989). The role of the Periadriatic Line in the tectonic evolution of the Alps. *Geological Society, London, Special Publications*, 45(1), 153–171.

Schmid, S.M., Kissling, E., Diehl, T., van Hinsbergen, D.J.J., Molli, G. (2017). Ivrea mantle wedge, arc of the Western Alps, and kinematic evolution of the Alps–Apennines orogenic system. *Swiss Journal of Geosciences*, 110(2), 581–612.

Schmid, S.M., Pfiffner, O.A., Froitzheim, N., Schönborn, G., Kissling, E. (1996). Geophysical-geological transect and tectonic evolution of the Swiss-Italian Alps. *Tectonics*, 15(5), 1036–1064.

Schmid, S.M., Zingg, A., Handy, M. (1987). The kinematics of movements along the Insubric Line and the emplacement of the Ivrea Zone. *Tectonophysics*, 135(1-3), 47–66.

Schon, J.H. (1998). *Physical properties of rocks: Fundamentals and principles of petrophysics*. Pergamon Press.

Scholer, M., Irving, J., Holliger, K. (2010). Estimation of the correlation structure of crustal velocity

heterogeneity from seismic reflection data. *Geophysical Journal International*, 183(3), 1408–1428.

Schwarz, B. (2019). Coherent wavefield subtraction for diffraction separation. *Geophysics*, 84(3), V157–V168.

Schwarz, B., Gajewski, D. (2017). Accessing the diffracted wavefield by coherent subtraction. *Geophysical Journal International*, 211(1), 45–49.

Schwarz, B., Krawczyk, C.M. (2020). Coherent diffraction imaging for enhanced fault and fracture network characterization. *Solid Earth*, 11(5), 1891-1907.

Simmons, G., Nur, A. (1968). Granites: Relation of properties in situ to laboratory measurements. *Science*, 162(3855), 789–791.

Simmons, G., Wang, H. (1971). *Single crystal elastic constants and calculated aggregate properties*. MIT Press.

Simon, H., Buske, S. (2017). High-resolution seismic survey for the characterization of planned PIER-ICDP fluid-monitoring sites in the Eger Rift zone. In *AGU fall meeting abstracts 2017*, pp. S33A-0835. American Geophysical Union.

Stolt, R.H. (1978). Migration by Fourier transform. *Geophysics*, 43(1), 23-48.

Streich, R., van der Kruk, J., Green, A.G. (2007). Vector-migration of standard copolarized 3D GPR data. *Geophysics*, 72(5), J65-J75.

Taner, M.T., Koehler, F. (1969). Classic paper. Velocity spectra-digital computer derivation and applications of velocity functions. *Geophysics*, 34(6), 859–881.

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

Tertyshnikov, K., Pevzner, R., Bóna, A., Alonaizi, F., Gurevich, B. (2015). Steered migration in hard rock environments. *Geophysical Prospecting*, 63(3), 525–533.

Thiran, J.P. (1971). Recursive digital filters with maximally flat group delay. *IEEE Transactions on Circuit Theory*, 18(6), 659–664.

Trinh, P.T., Brossier, R., Métivier, L., Tavard, L., Virieux, J. (2019). Efficient time-domain 3D elastic and viscoelastic full-waveform inversion using a spectral-element method on flexible Cartesian-based mesh. *Geophysics*, 84(1), R75–R97.

Tronicke, J., Holliger, K. (2005). Quantitative integration of hydrogeophysical data: Conditional geostatistical simulation for characterizing heterogeneous alluvial aquifers. *Geophysics*, 70(3), H1–H10.

Tronicke, J., Holliger, K., Barrash, W., Knoll, M.D. (2004). Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation. *Water Resources Research*, 40(1), 1–14.

Urosevic, M., Juhlin, C. (2007). An analysis of seismic information obtained from crooked line seismic surveys in crystalline rocks, Australia. In *69th EAGE conference and exhibition incorporating SPE EUROPEC 2007*, pp. cp-27. European Association of Geoscientists & Engineers.

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

Warren, C., Giannopoulos, A., Giannakis, I. (2016). gprMax: Open source software to simulate

electromagnetic wave propagation for Ground Penetrating Radar. *Computer Physics Communications*, 209, 163–170.

Wu, X., Liang, L., Shi, Y., Fomel, S. (2019). FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, 84(3), IM35–IM45.

Xu, Z., Irving, J., Lindsay, K., Bradford, J., Zhu, P., Holliger, K. (2020). Estimation of the 3D correlation structure of an alluvial aquifer from surface-based multi-frequency ground-penetrating radar reflection data. *Geophysical Prospecting*, 68(2), 678–689.

Xu, Z., Irving, J., Liu, Y., Peimin, Z., Holliger, K. (2021). Conditional stochastic inversion of common-offset GPR reflection data. *Geophysics*, 86(5), 1–49.

Yang, F., Ma, J. (2019). Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics*, 84(4), R583–R599.

Yang, X. and Zhu, P. (2017). Stochastic seismic inversion based on an improved local gradual deformation method. *Computers & Geosciences*, 109, 75-86.

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

Yin, J., Nakata, N. (2019). Diffraction imaging using geometric-mean reverse time migration and common-reflection surface. *Geophysics*, 84(4), S355–S364.

Yu, S., Ma, J. (2021). Deep Learning for Geophysics: Current and Future Trends. *Reviews of Geophysics*, 59(3), 1–36.

Yuan, H., Looms, M.C., Nielsen, L. (2020). On the usage of diffractions in ground-penetrating radar reflection data: Implications for time-lapse gas migration monitoring. *Geophysics*, 85(5), H83–H95.

Yuan, H., Montazeri, M., Looms, M.C., Nielsen, L. (2019). Diffraction imaging of ground-penetrating radar data. *Geophysics*, 84(3), H1–H12.

Zeng, Z., Chen, X., Li, J., Chen, L., Lu, Q., Liu, F. (2015). Recursive impedance inversion of ground-penetrating radar data in stochastic media. *Applied Geophysics*, 12(4), 615–625.

Zhou, H., Hu, H., Zou, Z., Wo, Y., Youn, O. (2018). Reverse time migration: A prospect of seismic imaging methodology. *Earth-science Reviews*, 179, 207–227.

Zhu, H., Li, S., Fomel, S., Stadler, G., Ghattas, O. (2016). A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration. *Geophysics*, 81(5), R307–R323.

Appendix A

Conditional stochastic inversion of common-offset ground-penetrating radar reflection data

---

Zhiwei Xu, James Irving, Yu Liu, Peimin Zhu, Klaus Holliger

Published[1] in *Geophysics*

[1] Xu, Z., Irving, J., Liu, Y., Zhu, P., Holliger, K. (2021). Conditional stochastic inversion of common-offset ground-penetrating radar reflection data. *Geophysics*, 86(5), WB89-WB99. doi: 10.1190/geo2020-0639.1

A.1 Abstract

We present a stochastic inversion procedure for common-offset ground-penetrating radar (GPR) reflection measurements. Stochastic realizations of subsurface properties that offer an acceptable fit to the GPR data are generated via simulated annealing optimization. The realizations are conditioned to borehole porosity measurements available along the GPR profile, or equivalent measurements of another petrophysical property that can be related to the dielectric permittivity, as well as to geostatistical parameters derived from the borehole logs and the processed GPR image. Validation of our inversion procedure is performed on a pertinent synthetic data set and indicates that the proposed method is capable of reliably recovering strongly heterogeneous porosity structures associated with surficial alluvial aquifers. This finding is largely corroborated through application of the methodology to field measurements from the Boise Hydrogeophysical Research Site near Boise, Idaho, USA.

A.2 Introduction

Adequate characterization of the spatial heterogeneity of the vadose and/or saturated zones is a prerequisite for the reliable prediction of groundwater flow and contaminant transport in the subsurface. As such, it represents a key objective of many hydrogeological studies (e.g., Salamon et al., 2007; Hu et al., 2009; Cardiff et al., 2013; Maliva, 2016). Traditionally, this objective is approached through local borehole-based studies and larger-scale hydraulic tests. However, the associated gap in terms of spatial resolution and coverage can render an integrated interpretation difficult (e.g., Kobr et al., 2005; Leven and Dietrich, 2006; Gueting et al, 2015). This problem can be alleviated through targeted geophysical measurements (e.g., Rubin and Hubbard, 2006; Hubbard and Linde, 2010; Binley et al., 2015; Romero-Ruiz et al., 2019).

One geophysical method that has attracted significant interest with regard to subsurface hydrogeological studies is surface-based ground-penetrating radar (GPR) reflection profiling. This method has the potential to provide images of shallow subsurface structure with extremely high spatial resolution in comparison with other applied geophysical techniques, and a number of previous studies have investigated how such reflection images might be used in the context of aquifer characterization (e.g., Annan, 2005; Blindow, 2006; and references therein). As the corresponding data tend to be acquired in bi-static mode with a small constant offset between the transmitting and receiving antennae, one important challenge associated with typical GPR reflection measurements is that they do not readily provide detailed information on the spatial distribution of petrophysical properties in the probed subsurface region. This is of key interest as the underlying high-frequency electromagnetic wave propagation phenomena are largely governed by the dielectric permittivity, which is highly sensitive to soil water content and thus to soil texture and porosity above and below the water table, respectively (e.g., Knight, 2001). Although the analysis of diffraction hyperbolas in common-offset GPR data may be used to obtain useful information on the permittivity distribution (e.g., Mount et al., 2014; Yuan et al., 2019), the results are strongly limited in terms of spatial resolution and the method inherently requires a high density of diffractions throughout the GPR profile to be reliable.

To address the above limitation and recover detailed information on the dielectric properties of the probed subsurface from reflection GPR measurements, a variety of approaches have been developed. The vast majority of these rely upon the acquisition of multi-offset data, whereby multiple receiver antenna positions are considered for each transmitter antenna position in an analogous manner to seismic reflection surveying

(Forte and Pipan, 2017). The corresponding measurements can then be examined using a variety of different techniques, which include quantification of reflector moveout and reflection tomographic approaches (e.g., Greaves et al., 1996; Bradford et al., 2009; Mangel et al., 2020), amplitude-versus-offset (AVO) analysis (e.g., Bradford and Deeds, 2006; Deparis and Garambois, 2009), and waveform inversion (e.g., Lavoué et al., 2014; Babock and Bradford, 2015). While all of these approaches can provide important information on the spatial distribution of the subsurface dielectric permittivity, and to a lesser extent the electrical conductivity, one important drawback is that the acquisition of multi-offset GPR data for lower-frequency geological applications is time-consuming and logistically cumbersome, and is thus not routinely done. Indeed, lower-frequency GPR surveys typically involve only a single transmitter and receiver antenna, meaning that multi-offset survey time will increase by a factor equal to the number of desired offsets when compared to a common-offset acquisition. Further, methods such as reflection tomography, which rely upon ray theory and the use of traveltimes to reconstruct the permittivity distribution, are well known to suffer from limited spatial resolution. Finally, large uncertainties associated with GPR antenna radiation patterns in complex near-surface media mean that methods like AVO analysis or full-waveform inversion, which require accurate signal amplitudes, may be adversely affected.

If suitable borehole measurements, such as porosity or dielectric permittivity logs, are available for calibration along the GPR profile, one promising alternative for the recovery of detailed electrical property information from common-offset reflection GPR measurements is impedance inversion. In this regard, Schmelzbach et al. (2012) present a workflow to recover spatially distributed electromagnetic impedance, which is closely related to soil dielectric permittivity and water content, from reflection GPR data. Their approach involves sparse-spike deconvolution of an amplitude-corrected and migrated GPR image, followed by band-limited integration and scaling, whereby the low-frequency component of the impedance variability, which is not possible to recover from the noisy GPR data alone, is derived from direct-push logs acquired along the profile. Zeng et al. (2015) and Liu et al. (2018) use similar approaches to recover water content from common-offset GPR profiles in complex environments and to characterize buried archaeological remains, respectively.

In this paper, we complement and extend previous work on the determination of high-resolution subsurface properties from common-offset reflection GPR measurements with the aim of addressing one important limitation. This is the fact that the approaches developed to date are deterministic in nature and therefore provide only a single solution to an inverse problem that is well known to be highly non-unique due to the heterogeneous subsurface environment and complex physics associated with high-frequency electromagnetic wave propagation. A single deterministic solution makes an assessment of the inferred subsurface models difficult, notably with regard to their uncertainties, and poses a strong limitation on their utility in hydrogeological investigations where groundwater flow and contaminant transport must be evaluated within a statistical, risk-based framework. To this end, we consider the impedance-type inversion of surface-based common-offset GPR reflection data from a stochastic perspective, whereby we seek to match observed radargrams to spatial distributions of subsurface properties that honor, a priori, pertinent in situ information derived from borehole-type measurements as well as prescribed geostatistical constraints. This is done via conditional geostatistical simulation within a stochastic optimization procedure, the repeated application of which allows for the generation of multiple acceptable models in order to explore and quantify uncertainty. Additional advantages of our approach are that realistic lateral continuity is guaranteed in the inversion results and the accumulation of errors, associated with the along-trace integrations performed in traditional impedance inversion approaches, is avoided.

The paper proceeds as follows. We begin by describing the methodological foundations of our proposed approach. Next, we proceed to assess its viability on a pertinent synthetic case study. Finally, we apply the approach to a field data set from the Boise Hydrogeophysical Research Site (BHRS), Idaho, USA.

A.3 Methodology

Our inversion method assumes the availability of at least one high-resolution borehole or direct-push log of the dielectric permittivity, or a closely related petrophysical property, along the GPR profile that can be used to condition the generation of stochastic subsurface property realizations. This is done within a global optimization loop in order to fit the recorded GPR reflection waveforms. In the current study, we assume full water saturation and perform all analyses in terms of subsurface porosity, meaning that borehole porosity logs are taken to be available and the relationship between GPR velocity and porosity is assumed known via a suitable petrophysical transform. Note, however, that the generalization of our methodology to work with GPR velocity and/or cases with partial water saturation is straightforward.

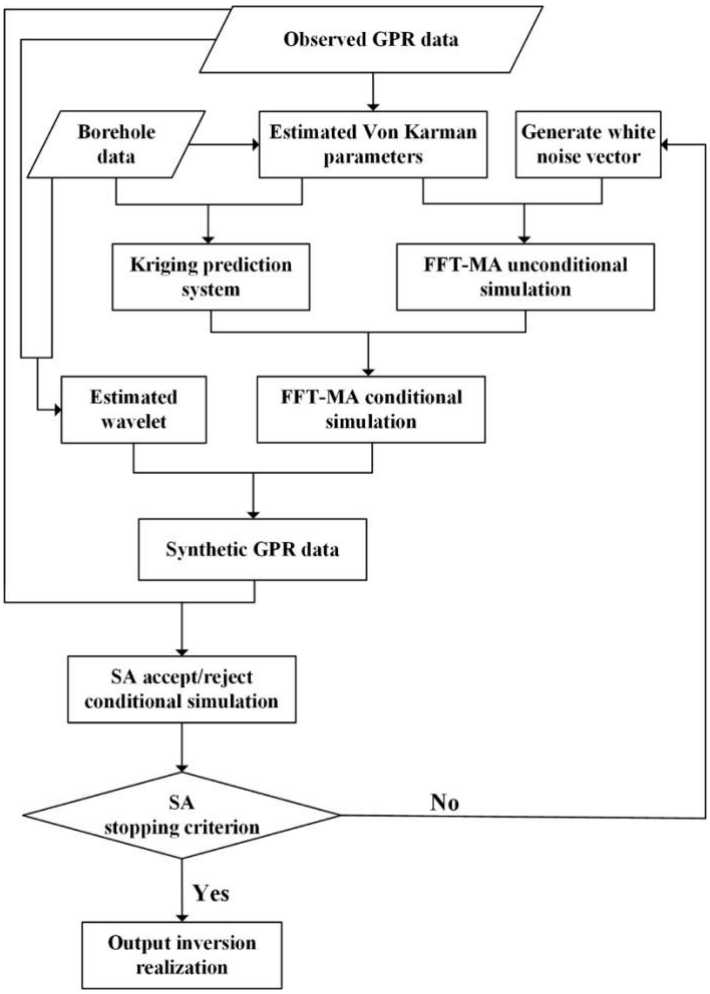


Figure A.1: Workflow summarizing the proposed conditional stochastic inversion procedure for generating a single output realization.

The steps involved in our inversion procedure to produce a single subsurface porosity realization are schematically illustrated in Figure A.1. Uncertainty can be assessed by running the procedure multiple times

and examining the corresponding collection of realizations. The overall inversion methodology can be broken down into four major components, which are:

(1) estimating a set of 2D autocovariance parameters that we assume are capable of adequately describing the stochastic variability of porosity in the probed subsurface region;
(2) generating a stochastic porosity realization honoring these parameters and conditioned to the borehole or direct-push porosity log measurements;
(3) calculating the corresponding synthetic reflection GPR profile and evaluating its misfit with the field GPR measurements; and
(4) applying simulated annealing (SA) optimization by iterating over steps (2) and (3) in order to find a porosity realization that honors (i) the estimated stochastic subsurface structure; (ii) the borehole porosity log data; and (iii) the common-offset GPR reflection measurements.

These components are described in further detail in the subsections below.

A.3.1 Estimation of subsurface stochastic parameters

We assume that the stochastic variability of subsurface porosity can be adequately captured by a Gaussian two-point geostatistical model. This assumption is generally considered to be valid for a given hydrogeological unit (e.g., Kelkar and Perez, 2002; Dafflon et al., 2009). For the parameterization of this model, we consider the so-called von Kármán autocorrelation function which, due to its immense versatility, has been used for a wide variety of research objectives, such as seafloor morphology quantification (e.g., Goff and Jordan, 1988), borehole data analysis (e.g., Dolan and Bean, 1997; Jones and Holliger, 1997), numerical simulations of wave propagation (e.g., Frankel and Clayton, 1986; Hartzell et al., 2010), and aquifer characterization (e.g., Tronicke and Holliger, 2005; Dafflon et al., 2009). In 2D and for anisotropic porosity heterogeneity aligned along coordinate axes x and z, the von Kármán autocorrelation equation takes the following form (e.g., Goff and Jordan, 1988):

$$R_{\theta\theta}(\delta x, \delta z) = \frac{r^{\nu} K_{\nu}(r)}{2^{\nu-1}\Gamma(\nu)} \tag{A.1}$$

where $R_{\theta\theta}$ is the porosity autocorrelation, δx and δz are the spatial autocorrelation lags in the x- and z-directions, respectively, $K_{\nu}(r)$ is the modified Bessel function of the second kind having order $0 \leq \nu \leq 1$, Γ is the gamma function, and

$$r = \sqrt{\left(\frac{\delta x}{a_x}\right)^2 + \left(\frac{\delta z}{a_z}\right)^2} \tag{A.2}$$

is a normalized lag parameter with a_x and a_z denoting the spatial correlation lengths along x and z, respectively. Eq. (A.1) defines an anisotropic heterogeneous medium which is self-similar, or fractal, at scales shorter than the correlation lengths. The decay of the autocorrelation function at small lags, and thus the local variability of the associated heterogeneity, is controlled by ν, which is generally referred to as the Hurst number. Values of ν close to 0 and 1 characterize locally highly complex and smooth fluctuations, respectively, whereas for $\nu = 0.5$, the von Kármán autocorrelation function reduces to its well-known, Brownian-noise-type, exponential equivalent (e.g., Goff and Jordan, 1988).

Variables a_x, a_z, and ν parameterize our geostatistical porosity model and must be determined from available data. Following Tronicke and Holliger (2005), we estimate a_z and ν, as well as the porosity mean and variance, from the high-resolution porosity logs available along the GPR profile. Parameter a_x can then be inferred from the GPR data using the inversion approach of Irving et al. (2009), which relates the

geostatistical properties of the backscattered wavefield to those of the underlying scattering medium. Specifically, this approach allows for inference of the structural aspect ratio a_x/a_z of the probed subsurface medium (Irving and Holliger, 2010) from which, knowing a_z, we can readily determine a_x. It is important to note that our use of this procedure inherently assumes the geostatistical structures of GPR velocity and porosity to be identical. This is reasonable in saturated soils given the approximately linear relationship between GPR velocity and water content over the range of porosities typically encountered in near-surface materials (e.g., Irving et al., 2009; Xu et al, 2020a).

A.3.2 Generation of conditional porosity realizations

The core of our inversion procedure involves the generation of stochastic realizations of the subsurface porosity field that (i) honor the geostatistical parameters inferred from the borehole porosity logs and common-offset GPR reflection data, and (ii) fit exactly the borehole porosity measurements, which are treated as hard data. These realizations are then tested with regard to how well they allow us to predict the observed GPR reflection data, and are iteratively perturbed within a SA optimization loop to generate a single inversion output realization (Figure A.1).

We generate unconditional stochastic realizations using the fast Fourier transform moving average (FFT-MA) technique (LeRavalec et al., 2000), which is a convenient and attractive implementation of the moving average (MA) geostatistical simulation method of Oliver (1995) in the discrete wavenumber domain. The FFT-MA method has been recognized for its efficiency and flexibility (e.g., Caers, 2007; Le Ravalec-Dupin et al. 2008; de Figueiredo et al. 2018). Any permissible autocovariance model may be considered and, given that the random and deterministic components of the algorithm are separated in the spatial, rather than in the wavenumber, domain, local re-simulations of specific areas of the model grid are possible. The latter cannot be done with standard power spectral simulation techniques (e.g., Ikelle, 1993) and has led to the common application of FFT-MA for stochastic modeling and inversion (e.g., Le Ravalec-Dupin et al., 2004; Le Ravalec and Mouche, 2012; Liang and Marcotte, 2016; Yang and Zhu, 2017; Lauzon and Marcotte, 2019).

From an unconditional FFT-MA-generated realization, which is not constrained to respect the porosity log data at the borehole locations, a corresponding conditional realization can be generated as follows (e.g., Chilès and Delfiner, 2012; Nussbaumer et al., 2019):

$$Z_c(x,z) = Z^*(x,z) + [Z_u(x,z) - Z_u^*(x,z)], \tag{A.3}$$

where Z_c is the output conditional porosity realization, Z_u is the unconditional realization generated using FFT-MA, and Z^* and Z_u^* the ordinary-kriging-based porosity estimates based on the log and unconditional simulation values at the borehole locations, respectively.

It is important to note that the conditional porosity simulations that are generated using the above procedure are effectively parameterized by a Gaussian white noise vector in the spatial domain, whose individual elements correspond to each location in the model grid. Re-simulating the values in this noise vector will produce different stochastic realizations, all of which honor the underlying von Kármán geostatistical model and the porosity values at the borehole locations. We exploit this in our SA optimization procedure, wherein conditional realizations are perturbed by re-simulating a certain percentage of elements in the Gaussian white noise vector. The number of re-simulated points and their location in the model grid govern the magnitude and the local-versus-global nature of the corresponding model perturbation.

A.3.3 GPR forward model

To compute the synthetic GPR reflection profile corresponding to a conditional stochastic porosity realization, we utilize the so-called primary reflectivity section (PRS) model (e.g., Gibson and Levander, 1990; Holliger et al., 1994; Irving et al., 2009), whereby an amplitude-corrected and time-migrated seismic or GPR reflection image $d(x,t)$ is expressed as the convolution of a source wavelet function $w(t)$ with the underlying subsurface reflectivity coefficient field $r(x,t)$:

$$d(x,t) = w(t) * r(x,t), \tag{A.4}$$

where t is the vertical two-way travel time and $*$ denotes convolution in time. Assuming that single scattering prevails and that dispersion is absent from the data, Eq. (A.4) is widely recognized to provide an adequate model for zero-offset geophysical reflection data (e.g., Yilmaz, 2001). Although the second assumption is only strictly valid for GPR data acquired under perfectly electrically resistive conditions, past experience has shown that this model is able to accommodate the limited dispersion effects associated with low-loss environments for which the GPR method has been conceived (e.g., Irving et al., 2009; Xu et al., 2020).

Reflection coefficients as a function of two-way travel time in Eq. (A.4) are obtained from the conditional porosity realization by first transforming it to dielectric permittivity. This is done using the following mixture model for water-saturated media (e.g., Schön, 1998):

$$\sqrt{\varepsilon_r} = \sqrt{\varepsilon_r^s}(1-\phi) + \sqrt{\varepsilon_r^w}\phi, \tag{A.5}$$

where ϕ is the porosity, ε_r is the relative dielectric permittivity, and ε_r^s and ε_r^w are the relative dielectric permittivities of the dry solid matrix and water, respectively, for which we assume values of 4.6 and 80 (e.g., Chan and Knight, 2001). For low-loss media amenable to GPR wave propagation, the high-frequency electromagnetic wave velocity v is related to the relative dielectric permittivity through (e.g., Annan, 2005)

$$v = \frac{c}{\sqrt{\varepsilon_r}} \tag{A.6}$$

where $c = 3 \times 10^8$ m/s is the speed of light in free space. This equation is used to transform depth into two-way vertical traveltime. Vertical incidence reflection coefficients in the traveltime domain are then obtained using

$$R = \frac{\sqrt{\varepsilon_{r_1}} - \sqrt{\varepsilon_{r_2}}}{\sqrt{\varepsilon_{r_1}} + \sqrt{\varepsilon_{r_2}}} \tag{A.7}$$

where indices 1 and 2 refer to the materials above and below an interface, respectively. Eq. (A.7) is applied iteratively to each column of the subsurface dielectric permittivity model to yield $r(x,t)$.

To estimate the GPR source wavelet $w(t)$ from field reflection measurements, we employ the constant phase method (Cui and Margrave, 2014). With this approach, the wavelet's Fourier domain amplitude spectrum is estimated from the GPR data assuming a statistically white series of reflection coefficients, whereas the phase spectrum is estimated based on borehole log measurements. In particular, a series of constant-phase rotations are applied to the amplitude spectrum derived from the GPR data, and each of the corresponding time-domain wavelets is then convolved with the reflectivity calculated from the borehole data. The phase rotation that provides the highest correlation between the resulting synthetic trace and the measured trace at the borehole location is chosen as the estimated wavelet phase. Based on numerous synthetic tests, we have found this approach to provide a reliable enough estimation of the GPR wavelet for use in our stochastic inversion methodology.

A.3.4 SA optimization

We wish to find conditional realizations of subsurface porosity, generated using the technique described above, whose corresponding synthetic GPR reflection data offer a good fit to the field GPR measurements. To this end, we build on previous work (e.g., Tronicke and Holliger, 2005; Dafflon et al., 2009, Lauzon and Marcotte, 2019) and use SA, a directional Monte-Carlo-type approach, to iteratively perform the optimization. The objective function to be minimized is the simple sum-of-squares error

$$O = \sum_k \sum_j \left(d^{syn}(x_j, t_k) - d^{obs}(x_j, t_k) \right)^2, \tag{A.8}$$

where d^{syn} and d^{obs} denote the synthetic and field reflection GPR profiles, respectively, and indices j and k sum over the number of points per trace and number of traces in the data, respectively.

We begin the SA procedure with a conditional porosity realization generated using a fully random vector of FFT-MA uncorrelated Gaussian noise elements, which is unconditionally accepted in the first iteration after the corresponding objective function value is evaluated. In subsequent iterations, a new conditional realization is created by re-simulating a randomly chosen subset of elements in the Gaussian noise vector, where the number of points in the subset N_i is given by

$$N_i = N_0 \cdot \gamma^i. \tag{A.9}$$

Here, i denotes the SA iteration number, N_0 is the total number of vector elements, and constant γ determines the rate at which the size of the model perturbation decreases as the iterations proceed (Lauzon and Marcotte, 2019). In this way, and similar to the effect of the temperature parameter discussed below, exploration of the model space is encouraged in the beginning of the SA algorithm, whereas exploitation is encouraged towards the end. Note that when the value of N_i in Eq. (A.9) becomes less than 1, only a single vector element is re-simulated. For both the synthetic and field examples presented in the next section, γ was set to an empirically determined value of 0.97.

After evaluating the objective function for the perturbed porosity realization using Eq. (A.8), the realization is either accepted or rejected according to a stochastic decision rule, whose probability of acceptance is given by

$$P_i^{acc} = \begin{cases} 1, & \text{if } O_i < O_{i-1} \\ \exp\left(-\left(\dfrac{O_i - O_{i-1}}{T_i} \right) \right), & \text{otherwise} \end{cases} \tag{A.10}$$

where T_i is a temperature parameter whose progressive decrease with increasing number of iterations defines the "cooling schedule" of the optimization process. Higher T_i values imply a greater probability of accepting random model perturbations that do not decrease the value of the objective function, which tends to encourage greater exploration of the model space. Here we use

$$T_i = T_0 \cdot \alpha^i, \tag{A.11}$$

where T_0 is the intial temperature and α is the temperature reduction factor, whose value also controls the balance between exploration and exploitation as the SA iterations proceed. Based on the arguments presented in Johnson et al. (1991), T_0 was set to a value of 1.5×10^{10} in our synthetic example and to a value of 1×10^8 for our field study. The parameter α was set equal to 0.95 in both cases.

For simplicity, SA iterations are continued in our inversion algorithm until a prescribed maximum number of iterations is reached. For our synthetic example, setting the maximum number of iterations to 350 was found to yield an acceptable fit to the observed GPR traces given the prescribed errors in the data. For the field study, a maximum number of 320 iterations was considered based on similar arguments.

A.4 Results

A.4.1 Synthetic study

To assess the viability of our proposed approach, we first apply it to a synthetic case study. The underlying "true" porosity model, which aims to emulate conditions in a heterogeneous alluvial aquifer, is shown in Figure A.2 and was created using the FFT-MA unconditional simulation method described previously. It is based upon the porosity distribution considered in Tronicke and Holliger (2005) and is characterized by a von Kármán autocovariance function having a ν-value of 0.3, a mean porosity value of 0.19, and a standard deviation of 0.026. The horizontal and vertical correlation lengths used in the model generation process are a_x = 133 m and a_z = 13 m, respectively. This implies that the resulting stochastic medium is pervasively self-similar and, hence, exhibits the typical combination of small- and large-scale heterogeneity observed in alluvial aquifers (Tronicke and Holliger, 2005). In this context, it is interesting to note that the larger-scale heterogeneities in Figure A.2, such as the high-porosity channel running across the model, assume a quasi-deterministic appearance. We consider the presence of three boreholes, located at lateral distances of 5, 15, and 27 m from the left model edge, along which high-resolution porosity logs are assumed to be available. The porosity data from the left- and right-hand boreholes are used as conditioning information in our inversions, whereas the data from the center borehole are used for validation purposes.

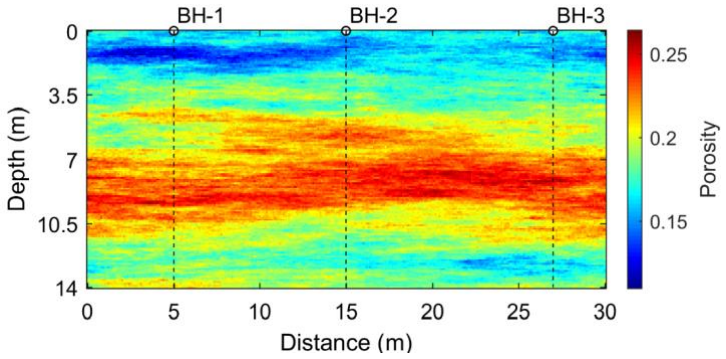


Figure A.2: Synthetic porosity model with dashed vertical lines denoting the prescribed borehole locations.

Using Eq. (5), the porosity realization in Figure A.2 was transformed into a distribution of subsurface relative dielectric permittivity and a synthetic common-offset GPR reflection survey was simulated using the gprMax software, which solves Maxwell's equations using the finite-difference time-domain (FDTD) method (Giannopoulos, 2005). The values of the electrical conductivity and relative magnetic permeability were set to 1 mS/m and 1, respectively. For the FDTD modeling, we used a transmitter-receiver antenna separation of 0.5 m and a Ricker source wavelet with a center frequency of 100 MHz. Synthetic GPR traces, sampled every 0.5 ns, were generated every 0.1 m along the profile.

After adding 2% uncorrelated Gaussian noise, the synthetic GPR data were subjected to a standard processing sequence consisting of (i) 10-300 MHz bandpass filtering to remove noise along with the low-frequency transient or "wow" upon which the GPR reflections are superimposed; (ii) application of a smooth time-varying gain based on the inferred average energy decay curve to compensate for the geometrical spreading of energy along with scattering and transmission losses; and (iii) 2D f-k time migration (Stolt, 1978) using a constant velocity of 0.09 m/ns. This velocity, which roughly corresponds to the average GPR velocity of the model, was found to optimally focus hyperbolic diffraction events. In Figure A.3, we show the resulting processed GPR image together with the source wavelet that was estimated from these data using the constant phase method and the true source wavelet for reference. We see that, in the GPR image, there are no apparent general trends in signal amplitude with time, which suggests that our choice of smooth gain function has effectively compensated for spreading and scattering/transmission losses in the data while leaving the relative reflection amplitudes intact. The latter is critical for an effective use of the PRS convolution model described previously. With regard to the estimated GPR source wavelet, we see that the constant phase method yields a result that is close in form to the true source wavelet (Figure A.3b), meaning that it should allow for reliable forward modeling of predicted data in our inversion procedure.

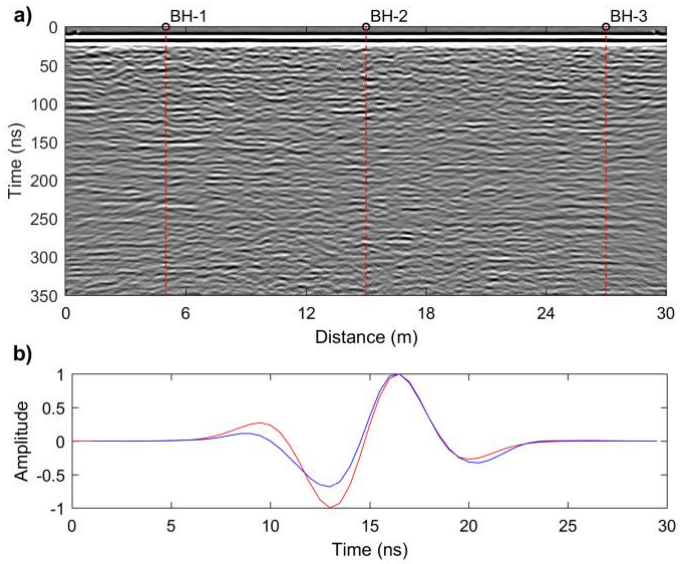


Figure A.3: a) Processed synthetic GPR section corresponding to the porosity model from Figure A.2 and b) the estimated and true source wavelets in blue and red, respectively. The dashed vertical red lines in a) show the considered borehole locations.

The processed GPR data shown in Figure A.3a, together with the high-resolution porosity information from the left- and right-hand boreholes, were subjected to the conditional stochastic inversion workflow outlined in Figure A.1. Based on geostatistical analysis of the porosity log data, we estimated a mean and standard deviation of 0.193 and 0.0259, respectively, and a Hurst number of ν = 0.30. These estimates are all close to the true values. Analysis of the GPR image using the stochastic inversion methodology of Irving et al. (2009) yielded a mean value for the aspect ratio of the subsurface porosity field of 10.5, which is again close to the true value and was used with the considered vertical correlation length of a_z = 13.0 m to obtain a value for the lateral correlation length of a_x = 136.5 m. These results, along with the measured porosity data at the borehole locations, were used to generate conditional porosity simulations (Figure A.4) that were optimized to fit the GPR measurements using SA (Figure A.1). Note that only the GPR data beyond 50 ns in time were considered in the inversion procedure in order to avoid the zone containing the direct air and ground arrivals (Figure A.3).

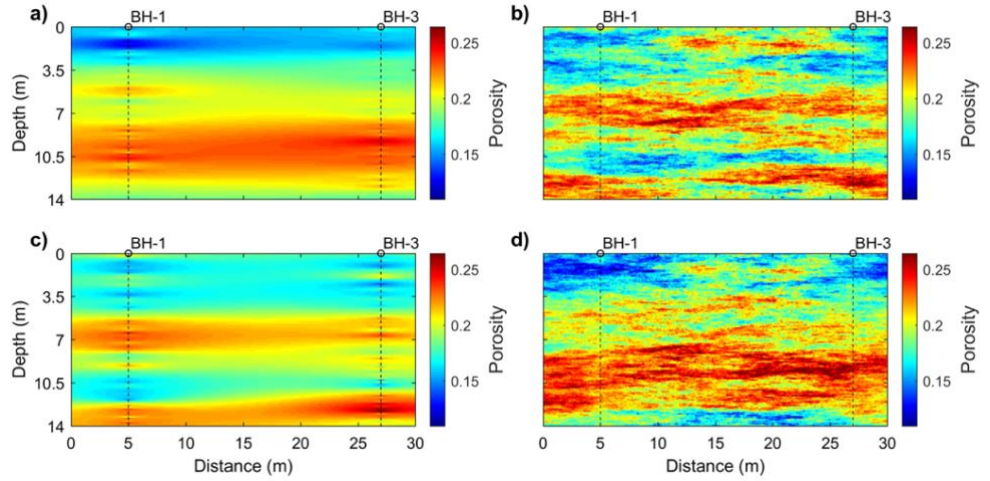


Figure A.4: Generation of a single conditional stochastic porosity realization for our synthetic example using the method summarized in equation (A.3). Shown are a) Z^*, the ordinary-kriging-based estimate of porosity based on the porosity-log values along boreholes BH-1 and BH-3; b) Z_u, an unconditional stochastic porosity realization generated using the FFT-MA method; c) Z_u^*, the ordinary-kriging-based estimate of porosity based on the unconditional porosity values from b) at the borehole locations; and d) Z_u, the final conditional realization.

In Figure A.5 we compare the underlying "true" porosity model from Figure A.2 with three output realizations obtained using our inversion methodology, along with maps of the mean and standard deviation of porosity inferred from an ensemble of 100 such realizations. Overall, the conditional stochastic inversion results are seen to faithfully reproduce both the smaller- and larger-scale features of the true porosity heterogeneity, and we observe a close match of the ensemble mean to the target model, even away from the left- and right-hand conditioning borehole locations, all of which suggests that the inversion procedure has successfully converged and that the reflection GPR measurements have greatly helped in characterizing the subsurface porosity distribution. Unsurprisingly, the standard deviation of the output ensemble is highest in the middle of the model domain where borehole conditioning data are not available. Note, however, that the corresponding values ($<$ 0.012) are still significantly lower than the global standard deviation of the porosity distribution (0.026), suggesting that the reflection GPR measurements have notably reduced our uncertainty in this region. A zone of higher uncertainty is present near the top of the model because the inversion results were not conditioned to the GPR data for times less than 50 ns.

We compare in Figure A.6 the true and inverted results for the central validation borehole location, in terms of both porosity logs (Figure A.6a) and the corresponding GPR traces (Figure A.6b). The curves for 100 inversion realizations along with the ensemble mean are shown. Also shown in Figure A.6c are the porosity curves at the central borehole location corresponding to 100 conditional stochastic realizations that were not constrained to fit the GPR data. The latter represents the prior in our inversion procedure. We observe in Figure A.6a that, overall, the larger-scale trends in the true porosity distribution are well captured by the inversion realizations, in the sense that the mean porosity curve follows reasonably closely the true one. The levels of small-scale variability in the inversion results and the true porosity distribution are also similar. With regard to fitting the GPR data, all of the modeled traces corresponding to the porosity inversion results offer a close match to the observed trace at the validation borehole location (Figure A.6b). Finally, in comparing the range of the prior stochastic realizations shown in Figure A.6c with that of the inverted realizations in Figure A.6a, we see that consideration of the GPR data has significantly reduced our uncertainty with regard to the porosity distribution

in the middle of the model domain. Regions where the true porosity distribution falls close to the limits of the range of the inverted realizations, for example between 10 m and 12 m depth, are also seen to be regions where the true porosity is less likely in the context of the considered prior.

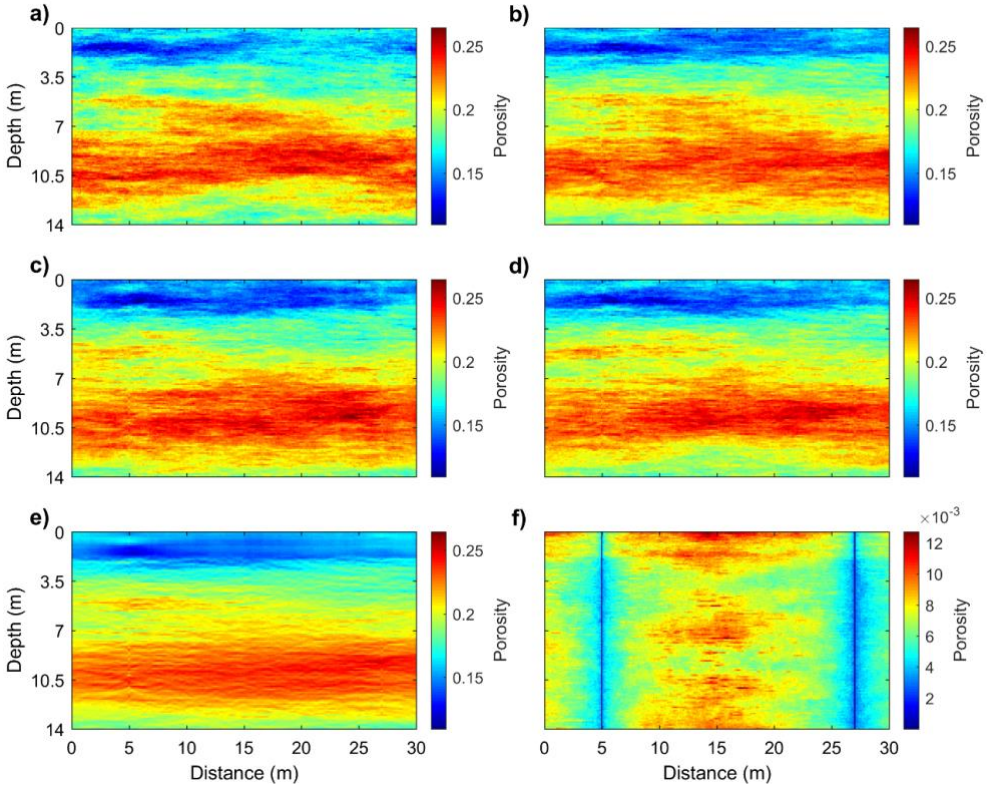


Figure A.5: Comparison of a) the true porosity model from Figure A.2 with b), c), and d) three stochastic realizations obtained using our inversion methodology. Also shown are e) the mean and f) the standard deviation obtained from 100 of such realizations.

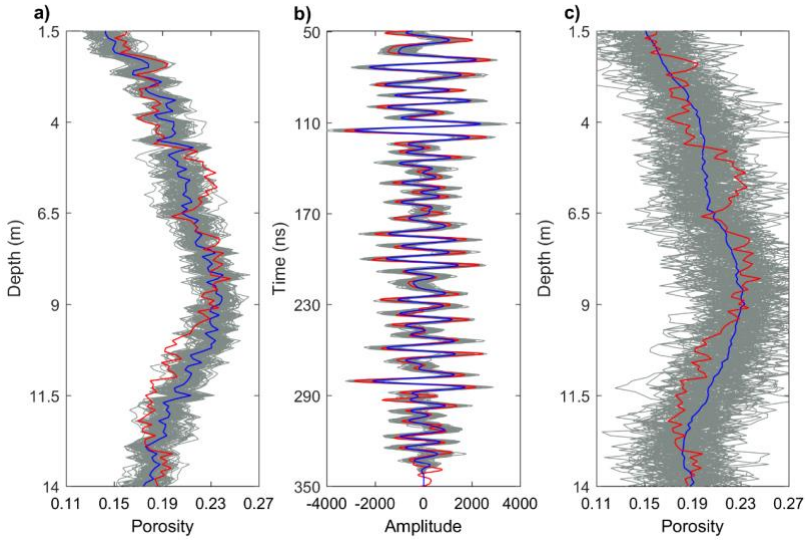


Figure A.6: Comparison of a) porosity profiles at the central borehole location shown in Figure A.2, and b) corresponding GPR traces. The red lines show the observed data, whereas the grey and blue lines show the results for 100 inverted realizations and their mean, respectively. Also plotted in c) are the "prior" porosity curves at the central borehole location corresponding to 100 conditional stochastic realizations that were not constrained to fit the GPR data (grey), along with their mean (blue) and the true porosity values (red).

A.4.2 Application to field data

Our proposed stochastic inversion method was applied to GPR reflection measurements from the Boise Hydrogeophysical Research Site (BHRS), which is located on a gravel bar adjacent to the Boise River, near Boise, Idaho, USA (Figure A.7). The site contains 13 boreholes in a central area, which has a diameter of ~20 m, and five boreholes near its borders located at distances of ~10 to ~35 m from this central area. The underlying braided-river-type aquifer consists of late Quaternary fluvial deposits dominated by coarse cobbles and sand. These are followed by a layer of red clay, which is situated at ~20 m depth (e.g., Barrash and Clemo, 2002). The depth to the groundwater table varies seasonally between ~2 and ~4 m. Over the past two decades, the site has been extensively used for the testing, validation, and improvement of a wide variety of geophysical and hydrogeological methods for characterizing heterogeneous aquifers (e.g., Tronicke et al., 2004; Bradford et al., 2009; Nichols et al., 2010; Dafflon et al., 2011; Dafflon and Barrash, 2012; Cardiff et al., 2013; Hochstetler et al., 2016).

The considered GPR reflection profile is part of a 3D survey, which was performed in the summer of 1998 using a PulseEkko Pro 100 system (Sensors & Software Inc.) with a nominal antenna center frequency of 100 MHz. The data were collected in common-offset mode using a transmitter-receiver antenna spacing of 1 m. Traces were recorded every 0.2 m along the profile, which is 18 m long and aligned with boreholes C6, A1, and C3 (Figure A.7). The time sampling interval was 0.8 ns and, for each recorded GPR trace, 32 vertical stacks were performed to improve the signal-to-noise ratio.

Figure A.7: Location of BHRS wellfield in relation to the Boise River. The dashed blue line joining wells C6, A1, and C3 corresponds to the considered surface-based GPR reflection survey.

The BHRS GPR data were subjected to a similar processing flow as their synthetic counterparts, which included time-zero and near-offset corrections, dewow filter, smooth time-varying gain based on the observed amplitude decay along each trace, and 2D f-k time migration using a constant velocity of 0.08 m/ns. The latter value was found to optimally focus diffraction events in the data below the water table, and is consistent with the average velocity of saturated sediments at the BHRS determined by Bradford (2009). The processed GPR image, along with the estimated source wavelet, are shown in Figure A.8, whereas Figure A.9 shows neutron-neutron porosity logs acquired along boreholes C6, A1, and C3, which are located 0.6 m, 9 m, and 16 m from the left edge the GPR profile, respectively (Barrash and Clemo, 2002). Given that the porosity measurements

are only available below the groundwater table, we limit our analysis to the saturated part of the probed subsurface region starting at ~2.8 m depth.

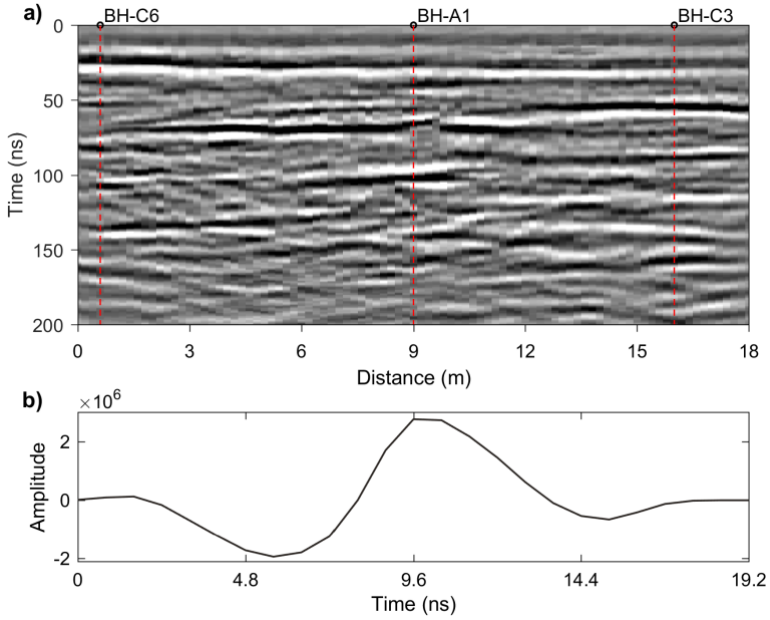


Figure A.8: a) Processed GPR section from the BHRS and b) the corresponding estimated source wavelet. The dashed vertical red lines in a) show the borehole locations.

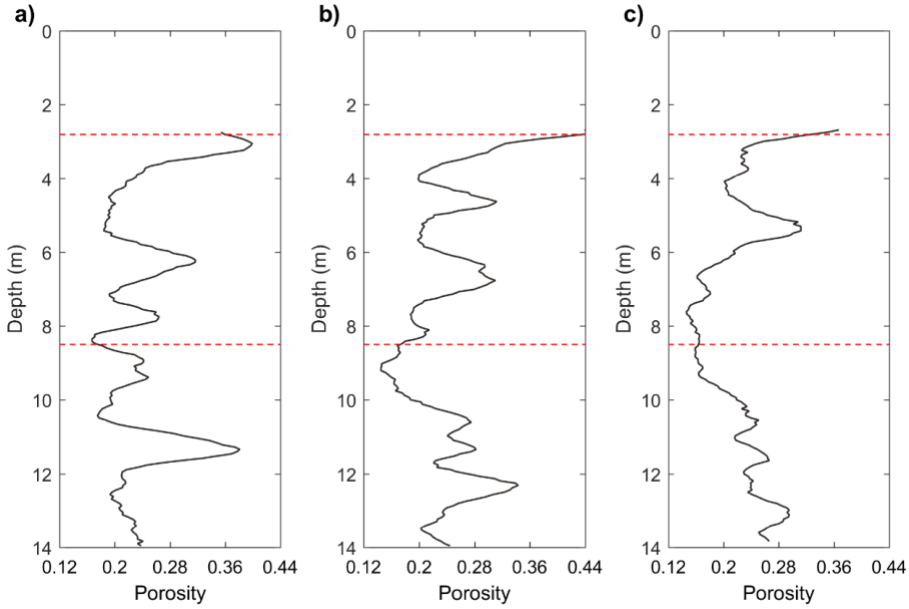


Figure A.9: Porosity logs obtained along BHRS boreholes a) C6, b) A1, and c) C3. The upper and lower dashed red lines indicate the depth of the groundwater table and penetration limit of the GPR data, respectively.

As in the synthetic case study, the left- and right-hand borehole logs (C6 and C3) were used for conditioning the stochastic inversion procedure, whereas the central log (A1) was reserved for validation. In this regard, geostatistical analysis of the porosity log data led to an estimated mean and standard deviation of 0.23 and 0.055, respectively, a Hurst number of $\nu = 0.35$, and a vertical correlation length of $a_z = 0.75$ m. Along the direction of the considered GPR profile, the stochastic analysis of 3D GPR data from the BHRS of Xu et al.

(2020) using the method of Irving et al. (2009) suggests that a horizontal-to-vertical aspect ratio of 12 is most likely, which implies a lateral correlation length of a_x = 9 m. All of these parameters were used to generate conditional stochastic porosity realizations that were then optimized to fit the GPR measurements.

Two example realizations obtained with our inversion procedure, along with the mean and standard deviation of an ensemble of 100 such realizations, are shown in Figure A.10. As observed previously in the synthetic case study (Figure A.5), the inverted models are consistent with each other as well as with the ensemble mean. The values of the ensemble standard deviation are also seen to follow the same overall spatial pattern as those in the synthetic case study. Note, however, that they are higher by approximately a factor of three, indicating greater uncertainty in the subsurface porosity distribution given the provided data.

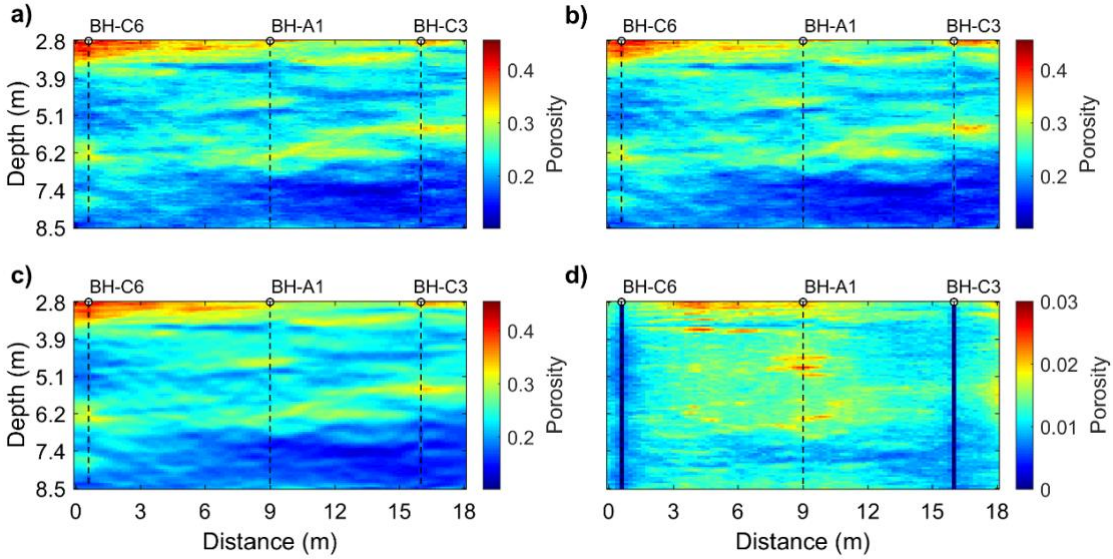


Figure A.10: a) and b) Two stochastic realizations of porosity along the BHRS profile obtained using our inversion methodology; c) and d) mean and standard deviation inferred from 100 of such stochastic realizations, respectively.

In Figure A.11 we show a comparison of the inverted and observed porosity profiles and corresponding GPR traces along the central borehole A1, as well as porosity curves along this borehole corresponding to the prior distribution assumed in the inversion procedure. We see that, overall, the proposed inversion approach provides a good fit to the observed GPR data, allows for a substantial reduction of uncertainty in porosity compared to the assumed prior distribution, and allows us to adequately reproduce the observed porosity profile to ~6.2 m depth. Between ~6.2 m and ~7.4 m depth, however, we observe in Figure A.11a a systematic mismatch between the observed and inverted porosity curves, which finds its clear expression in the fact that the range of the porosity values of the 100 accepted realizations shows no overlap with the observed porosity data. In this context, it is interesting to note that this region is characterized by an unusually low standard deviation (Figure A.10d), which suggests that our inversion procedure was not able to find any other means of fitting the observed data.

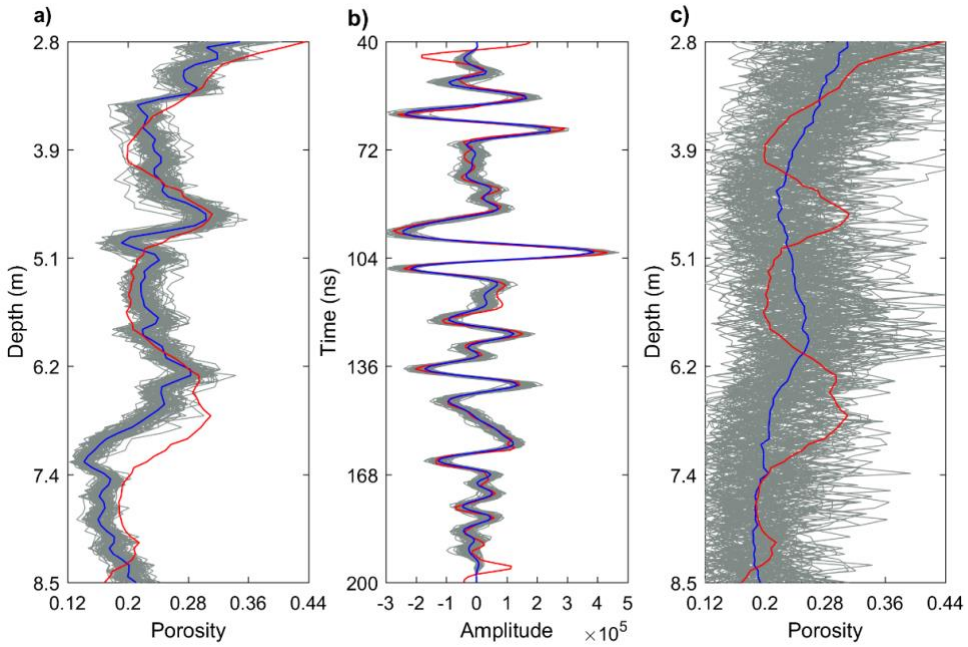


Figure A.11: Comparison of a) porosity profiles at the central borehole location shown in Figure A.8, and b) corresponding GPR traces. The red lines show the observed data, whereas the grey and blue lines show the results for 100 inverted realizations and their mean, respectively. Also plotted in c) are the "prior" porosity curves at the central borehole location corresponding to 100 conditional stochastic realizations that were not constrained to fit the GPR data (grey), along with their mean (blue) and the measured porosity-log data (red).

A.5 Discussion

The systematic bias towards too low estimated porosity values in the lower part of the profile in Figure A.11a could potentially be related to (i) inadequacies of the estimated GPR source wavelet; (ii) local variations in amplitude decay of the observed GPR data that have not been adequately compensated; (iii) problems with the neutron-neutron porosity log in central borehole A1, for example due to borehole enlargements and/or incomplete backfill behind the slotted PVC well casing that locally lead to values that are too high; or (iv) local violations of our inherent assumption of statistical stationarity.

While it is quite likely that the estimated source wavelet is to some degree sub-optimal, we would nevertheless expect the resulting mismatches between the inversion result and control data to be spatially more uniform and/or more gradual in their onset than those observed in Figure A.11a if the wavelet were the primary problem. Visual inspection of the observed GPR reflection data in Figure A.8a does, however, suggest that the amplitudes in the corresponding central region below ~150 ns are systematically weaker than elsewhere in the profile. Preliminary analysis indicates that this local amplitude deficit is not associated with a pronounced increase in signal dispersion and, thus, is unlikely to be indicative of stronger local attenuation, for example due to increased clay content. As such, it may be related to acquisition effects, such as variations in antenna coupling and/or system performance, which were incompletely compensated in the course of the data processing flow.

To explore the latter possibility, we completely reprocessed the GPR data presented in Figure A.8 using a variety of alternative gain functions. While in some cases this helped to raise the amplitudes in the lower parts

of the GPR section, it did not allow us to remove the bias in the inferred porosity distribution. In this context, it is important to note that, in the process, we also explored the impact of uncertainties in the source wavelet estimation as well as in the estimated correlation lengths and ν-value. None of these efforts allowed us to significantly reduce the bias. Indeed, the inferred porosity distributions remained remarkably stable, which in turn points to the inherent robustness of our inversion approach.

The peak at ~6.8 m depth in the central A1 borehole porosity log corresponds to an unusually high value, which exceeds the local mean by more than one standard deviation (Barrash and Clemo, 2002). The associated high-porosity region between ~6.2 m and ~7.4 m depth, where our stochastic inversion procedure consistently provides porosity estimates that are too low with regard to the log data, may therefore be a local anomaly. Such an anomaly could, as mentioned above, either be related to borehole enlargements and/or incomplete backfill behind the PVC casing, or it could represent an actual geological feature such as an isolated lens of open-frame gravels. Based on the available data, we are unable to distinguish between these two potential explanations. However, it is interesting to note that the laterally consistent transition to significantly lower porosities beyond ~6.8 m depth in our inversion results is consistent with Barrash and Clemo's (2002) interpretation of a transition from a high-porosity layer (Unit 4: mean porosity = 0.22, standard deviation = 0.05) to a low-porosity layer (Unit 3: mean porosity = 0.17, standard deviation = 0.02). This, in turn, illustrates that, while the proposed stochastic inversion approach is unable to account for local statistical non-stationarity, such as the unusually high porosity in the central region between ~6.2 m and ~7.4 m depth, it is robust with regard to laterally consistent changes in the medium properties.

A.6 Conclusions

We have presented a novel conditional stochastic inversion method for surface-based common-offset GPR reflection data. The associated workflow has been validated on a pertinent synthetic data set and applied to field data from the BHRS. While the synthetic test case illustrates the potential of the proposed approach to faithfully infer strongly heterogeneous porosity structures from surface-based GPR reflection measurements, the application to field data shows some local misfit with regard to the control data, the potential origins of which have been discussed. The current applications of our inversion methodology are 2D and consider full water saturation. However, generalization to 3D and partial water saturation are conceptually straightforward. In this regard, our proposed method has significant potential as the acquisition of 3D multi-offset is too time consuming for most geological applications.