Improvement of GPR Full-waveform inversion images using Cone Penetration Test data

Right Running Head: Improvement of GPR FWI images using CPT

Zhen Zhou [1, 2], Anja Klotzsche [1], Jessica Schmäck [1], Harry Vereecken [1], and Jan van der Kruk [1]

[1] Agrosphere (IBG-3), Institute of Bio- and Geosciences, Forschungszentrum Jülich, 52428, Germany, and Centre for High-Performance Scientific Computing in Terrestrial Systems, HPSC TerrSys, Geoverbund ABC/J, Jülich, Germany. Email: z.zhou@fz-juelich.de, a.klotzsche@fz-juelich.de, j.schmaeck@fz-juelich.de, h.vereecken@fz-juelich.de, j.van.der.kruk@fz-juelich.de

[2] School of Information Engineering, ZheJiang A&F University, Hangzhou 311300, China

Original paper date of submission: 30th April 2020

Revised paper (R1) date of submission: 1st September 2020

Revised paper (R2) date of submission: 10th November 2020

Revised paper (R3) date of submission: 9th December 2020

ABSTRACT

Detailed characterization of aquifers is critical and challenging due to the existence of heterogeneous small-scale high-contrast layers. For an improved characterization of subsurface hydrological characteristics, crosshole ground penetrating radar (GPR) and Cone Penetration Test (CPT) measurements are performed. In comparison to the CPT approach that can only provide 1D high resolution data along vertical profiles, crosshole GPR enables measuring 2D cross-sections between two boreholes. Generally, a standard inversion method for GPR data is the ray-based approach that considers only a small amount of information and can therefore only provide limited resolution. In the last decade, full-waveform inversion (FWI) of crosshole GPR data in time domain has matured, and provides inversion results with higher resolution by exploiting the full recorded waveform information. However, the FWI results are limited due to complex underground structures and the non-linear nature of the method. A new approach that uses CPT data in the inversion process is applied to enhance the resolution of the final relative permittivity FWI results by updating the effective source wavelet. The updated effective source wavelet possesses a priori CPT information and a larger bandwidth. Using the same starting models, a synthetic model comparison between the conventional and updated FWI results demonstrates that the updated FWI method provides reliable and more consistent structures. To test the method, five experimental GPR cross-section results are analyzed with the standard FWI and the new proposed updated approach. Both synthetic and experimental results indicate the potential of improving the reconstruction of subsurface aquifer structures by combining conventional 2D FWI results and 1D CPT data.

INTRODUCTION

The spatial variability in the subsurface aquifers has a great influence on the prediction of groundwater storage, the determination of priority flow direction, and the spread of contaminants, therefore, it is critical and challenging to predict subsurface flow and transport because of the complex small-scale heterogeneities in subsurface aquifers. In recent years, crosshole geophysical methods have been widely applied for subsurface imaging, including seismic (e.g., Doetsch et al., 2010), electrical resistivity tomography (ERT; e.g., Coscia et al., 2012) and ground penetrating radar (GPR; e.g., Klotzsche et al., 2013). GPR has shown great potential to map and characterize aquifers due to the high imaging resolution of the method (e.g., Hubbard et al., 2001; Garambois et al., 2002; Klotzsche et al., 2018). GPR provides lateral distributions of velocity and attenuation information of electromagnetic waves, which can be used to calculate relative dielectric permittivity ε_r and electrical conductivity σ. These two parameters are for example related to the subsurface connectivity, soil water content, and clay content. In particular, the dielectric permittivity is mainly influenced by the porosity and pore structure, and the electrical conductivity is influenced by the ion concentration, soil texture, and clay content (e.g., Busch et al., 2012).

Crosshole (borehole to borehole) GPR, which uses high-frequency electromagnetic pulses that are emitted from a dipole-type antenna in one borehole and received by a second antenna in the other borehole, is well suited to derive high-resolution images due to the known distance between the antennae and the possibility to use advanced inversion schemes. Conventional crosshole GPR tomographic inversion is based on geometrical ray theory (e.g., Maurer et al., 2004; Irving et al., 2007; Dafflon et al., 2011 and 2012), which uses only the first-arrival time and the maximum first-cycle amplitude of the measured data. Such approaches consider damping and smoothing constraints, and the resulting tomographic results are limited in resolution (relatively smooth images with a resolution on the order of the diameter of the first Fresnel zone). In contrast, full-waveform inversion (FWI) takes the entire waveform of GPR data into account including secondary events like scattered and refracted waves (e.g., Klotzsche et al., 2010; Meles et al., 2010). One of the first time-domain FWI approaches for crosshole GPR was presented for synthetic and experimental data sets by Ernst et al. (2007a; 2007b). This time-domain FWI has been improved by including a simultaneous update of the permittivity and conductivity, and by incorporating the vectorial behavior of electromagnetic fields (Meles et al., 2010). Since the first applications of this crosshole GPR FWI approach, the method has been further improved and successfully applied to several different aquifers. The tomographic results show great potential for high-resolution characterization of aquifers. Often, results of FWI are compared with independently measured porosity and hydraulic conductivity logging data (e.g., Klotzsche et al., 2013; Yang et al., 2013). Klotzsche et al. (2019a) provides a broad overview of the current state of the art of crosshole GPR FWI and its applications.

GPR FWI has shown benefit to bridge the gap in terms of resolution and coverage that exists between traditional hydrogeological methods, such as small-scale core analysis, and large-scale pumping tests. Yang et al. (2013) compare crosshole GPR FWI results with Neutron-Neutron logging data and observe a high goodness-of-fit.

In addition, they apply a low-pass filter (e.g., Ristau and Moon, 2001) to remove the high-wavenumber information of the Neutron-Neutron logging data and obtain an improved similarity between porosity estimates derived from the GPR FWI and the logging data. Similar to logging data, Cone Penetration Test (CPT) data can provide high spatial resolution at the point-scale along the profile depth and have the advantage to provide directly hydrological parameters such as porosity and electrical resistivity. Therefore, the CPT approach is popular to investigate shallow unconsolidated sediments in a fast, accurate, and minimally invasive manner (e.g., Fejes et al., 1990; Lunne et al., 1997; Tillmann et al., 2008). Gueting et al. (2015) employ CPT measurements to verify GPR FWI results and introduce a clustering of the data to identify lithological structures of the aquifer. This study indicates that combining the 2D crosshole GPR FWI results with the 1D CPT data can help to improve FWI imaging results along the GPR cross-section.

Although crosshole GPR FWI shows several benefits comparted to standard methods, the approach also has some limiting factors and requires challenging data processing steps (Klotzsche et al., 2019a). Firstly, the electromagnetic properties derived from GPR data are only indirectly related to hydrogeological parameters. Secondly, the effective source wavelet used for forward modeling contains approximations, especially if when estimated from standard ray-based models. The effective source wavelet compensates for all the missing information that are not included in the current models. Therefore, if errors are present in the models, they will directly propagate into the effective source wavelet and hence in the inversion results. In addition, the optimization method used in FWI is often a conjugate gradient technique (Polak et al., 1969), which relies on an appropriate initial model for the inversion and may converge to a local minimum. Therefore, one crucial criterion to successfully perform FWI is that the starting models need to provide modeled data within half a wavelength of the measured data; otherwise, cycle skipping can occur and the inversion may get trapped in a local minimum (Meles et al., 2012).

In this study, we propose a new method that improves the accuracy of the effective source wavelet and enhances the relative permittivity FWI results by incorporating additional logging information. This new method combines two different types of geophysical data sets in FWI: 1D CPT profiles and 2D crosshole GPR sections. As shown by Gueting et al. (2015), the 1D CPT data shows very high resolution along a 1D profile and by combining this information with the GPR data measured around such a CPT location, we intend to increase the resolution and reconstruction of the medium properties derived by the GPR FWI. First, we construct a low-pass filter based on the vertical locations where CPT data coincide with GPR FWI results. Secondly, the obtained 1D filter is applied to the 2D GPR crosshole FWI. Because of white noise in the measured GPR data and inversion artifacts close to the boreholes, the filter amplifies inconsistent high wavenumber information at locations farthest away from the CPT profile (mostly close to the boreholes). Therefore, we propose to remove this inconsistent information by estimating an updated effective source wavelet inspired by work on spectral whitening deconvolution (Li et al., 2009). The updated source wavelet contains lower-wavenumber information of CPT data than the standard source wavelet. Finally, a FWI using the updated source wavelet is performed while the starting models remain the same. To quantitatively evaluate the efficiency of the proposed updated source wavelet based on the CPT data, a synthetic model is generated using a stochastic simulation based on measured parameters at the Krauthausen test site in Germany. After successful validation of this new approach, the wavenumber amplifying filter is applied to experimental data from the same test site.

EFFECTIVE SOURCE WAVELET UPDATE USING AMPLIFIED FWI

Conventional FWI

To apply the conventional crosshole 2D time-domain FWI to experimental GPR data, some pre-processing steps are necessary. Details about the different inversion steps and developing stages of FWI can be found in Klotzsche et al. (2019b). Here we will only concentrate on the main steps that are important for the application of our new approach. First, the experimental data need to be converted from 3D to 2D using the approach introduced by Bleistein (1986) to reduce the influence of 3D wave propagation phenomena. This step is necessary because the forward modeling part of the FWI is based on 2D finite-difference time-domain (FDTD) solutions of Maxwell's equations (Ernst et al., 2007a; Meles et al., 2010). Traditional ray-based methods are applied to estimate the initial models of relative dielectric permittivity ε_r and electrical conductivity σ for the FWI. Such ray-based starting models need to provide modeled data that are within half a wavelength of the measured traces. In the presence of high contrast layers, the starting models need to be updated, for example by using an amplitude analysis approach (Klotzsche et al., 2014; Zhou at al., 2020), to fulfill this criterion. Finally, an effective source wavelet is determined by first using mainly horizontally traveling rays of the observed GPR data to estimate an initial source wavelet, and, second this initial wavelet is updated using the deconvolution method (e.g., Ernst et al., 2007b). We will refer to such an estimated wavelet as the “standard effective source wavelet”.

The inversion process is based on a conjugate gradient type method and according to Meles et al. (2010) and Klotzsche et al. (2019b) the cost function C is defined as the summation of the difference between the observed $\mathbf{E}^{\mathrm{obs}}$ and synthetic $\mathbf{E}^{\mathrm{syn}}$ (ε,σ) data over the number of transmitters s, receivers r, and the observation time τ as follows:

$$C(\varepsilon,\sigma) = \frac{1}{2}\sum_{s}\sum_{r}\sum_{\tau}\left[\mathbf{E}^{\mathrm{syn}}(\varepsilon,\sigma) - \mathbf{E}^{\mathrm{obs}}\right]_{r,\tau}^{T}\delta(\mathbf{x}-\mathbf{x}_r,t-\tau)\left[\mathbf{E}^{\mathrm{syn}}(\varepsilon,\sigma) - \mathbf{E}^{\mathrm{obs}}\right]_{r,\tau}, \quad (1)$$

here T denotes the transpose operator. The fields are locally defined at any point of space $\mathbf{x}$ and time t. The Dirac delta δ function extracts the wavefield at the receiver locations and observation times. Additionally, the gradients of the misfit function with respect to permittivity ∇C_ε and conductivity ∇C_σ are calculated by a zero-lag cross-correlation of the synthetic wavefield with the back-propagated residual wavefield as follows (more details in Meles et al., 2010):

$$\begin{bmatrix}\nabla C_{\varepsilon}(\mathbf{x}')\\ \nabla C_{\sigma}(\mathbf{x}')\end{bmatrix} = \sum_{s} \begin{matrix}\left(\delta(\mathbf{x}-\mathbf{x}')\partial_t \mathbf{E}^{\mathrm{syn}}\right)^T \hat{\mathbf{G}}^T \mathbf{R}^S \\ \left(\delta(\mathbf{x}-\mathbf{x}')\mathbf{E}^{\mathrm{syn}}\right)^T \hat{\mathbf{G}}^T \mathbf{R}^S\end{matrix} \tag{2}$$

with

$$\mathbf{R}^S = \sum_{r}\sum_{\tau} \delta(\mathbf{x}-\mathbf{x}_r, t-\tau)\left[\mathbf{E}^{\mathrm{syn}}(\varepsilon,\sigma) - \mathbf{E}^{\mathrm{obs}}\right]_{r,\tau} = \sum_{r}\sum_{\tau}\left[\Delta \mathbf{E}^{\mathrm{syn}}\right]_{r,\tau} . \tag{3}$$

$\hat{\mathbf{G}}^T$ indicates the Green's function and $\hat{\mathbf{G}}^T \mathbf{R}^S$ represents the back propagated residual wavefield in the same medium as the incident wavefield $\mathbf{E}^{\mathrm{syn}}$. The spatial delta function δ $(\mathbf{x}-\mathbf{x}')$ in equation 2 corresponds to the spatial components of the gradients and reduces the inner product to a zero-lag cross-correlation in time (Meles et al., 2010). Note that the only difference between the gradients of ε and σ is a time derivative.

Generally, to avoid overfitting of the observed data, the inversion of experimental data is stopped if the change of the misfit function value is less than 0.5% between two subsequent iterations (Klotzsche et al., 2019b). To evaluate the performance of the FWI, we analyzed the behavior of the root mean squared error *RMSE* and computed the correlation coefficient *R* between the modeled and observed data. Further, we calculated the mean of the remaining gradient of the final FWI permittivity ε_r results. The FWI results with optimal number of iterations should have the smallest summation value of the normalized ε_r gradients and of the normalized *RMSE*s. Considering that the gradients are highly sensitive close to the transmitter and receiver positions, inversion artifacts can arise close to the boreholes. To minimize these inversion artifacts, the approach of Kurzmann et al. (2013) is applied using a gradient preconditioning technique for both permittivity and conductivity (van der Kruk et al., 2015).

Wavenumber filter

The CPT data present high spatial resolution along a vertical 1D profile. Transforming the spatial porosity CPT data and the 1D FWI permittivity results to the amplitude-wavenumber domain by a fast Fourier transform (FFT), we can obtain a broader bandwidth of the CPT data than the FWI permittivity models bandwidth (e.g., Yang et al., 2013). Therefore, it is feasible to improve the FWI resolution by expanding the bandwidth of the FWI amplitude values using the CPT data in the amplitude-wavenumber domain. In a first step, we convert the relative dielectric permittivity ε_r (with

$\varepsilon_r = \varepsilon \cdot \varepsilon_0$, where ε is the real part of the bulk dielectric permittivity in natural medium and ε_0=8.8542 $\cdot 10^{-12}$ F/m is as permittivity of the free space) of the FWI results into porosity ∅ by using the three-phase complex refractive index model (CRIM) for the saturated zone (e.g., Birchak et al., 1974) with

$$\emptyset = \frac{\sqrt{\varepsilon_r} - \sqrt{\varepsilon_s}}{\sqrt{\varepsilon_f} - \sqrt{\varepsilon_s}} . \tag{4}$$

Similar to Gueting et al. (2015), we consider the fluid permittivity ε_f to be 84 (for a water temperature of 10℃) and the solid permittivity ε_s to be 4.5 (based on literature values of quartz, e.g., Eisenberg and Kauzmann, 2005; Carmichael, 2017).

In the second step, the selected 1D vertical porosity-profile of the CPT data (located close by or at the GPR cross-section) and transformed FWI porosity results are interpolated in spatial domain to obtain enough data points to generate the filter in the wavenumber domain. Here, we use a mean value of the selected data and two cosine functions (represented as tapers) to expand the initial data. The resampled process can be expressed by

$$\mathbf{d}_{exp} = p \times mean(\mathbf{d}_{org}), \quad \text{for} \quad \begin{cases} x < \mathbf{x}_{org}(0) - TL, \\ x > \mathbf{x}_{org}(end) + TL, \end{cases} \text{or} \tag{5}$$

$$\mathbf{d}_{exp} = C_1 \times \left\{1 - \cos\left[\frac{\pi}{TL} \times (x - (\mathbf{x}_{org}(0) - 1 - TL))\right]\right\} + p \times mean(\mathbf{d}_{org}), \quad \text{for}$$

$$\mathbf{x}_{org}(0) - TL \le x < \mathbf{x}_{org}(0), \tag{6}$$

$$\mathbf{d}_{exp} = C_2 \times \left\{1 - \cos\left[\frac{\pi}{TL} \times ((\mathbf{x}_{org}(end) + 1 + TL) - x)\right]\right\} + p \times mean(\mathbf{d}_{org}), \quad \text{for}$$

$$\mathbf{x}_{org}(end) < x \le \mathbf{x}_{org}(end) + TL, \tag{7}$$

where $\mathbf{d}_{org}$ and $\mathbf{d}_{exp}$ represent the original and the final expanded data points, respectively, p is a selected parameter of 0.8 following Yang et al. (2013), x and $\mathbf{x}_{org}$ represent point positions of the space vector of the expanded and original data, $\mathbf{x}_{org}(0)$ and $\mathbf{x}_{org}(end)$ indicate the start and end positions of original data in the expanded data domain, TL is the taper length that is 0.3 times of the original data length (18 in this case), and C_1 and C_2 are selected parameters which are used to adjust the connection points between the taper start position and the original two data points. The final interpolated data (512 data points in this case) with the corresponding tapers are transformed into the wavenumber domain using the 1D FFT.

In a third step, a smooth function is applied to flatten the highly fluctuating amplitudes of both interpolated data sets, which are caused by the interference of the real and imaginary parts of the data in the wavenumber domain (Yang et al., 2013). These smooth results are estimated by

$$SA(k) = \begin{cases} A(k), & k = 1 \\ smooth(A(k), span), & 1 < k \leq k_{max} \end{cases} \quad (8)$$

where $A(k)$ and $SA(k)$ represent amplitude and smooth amplitude values in the wavenumber domain, respectively, k indicates the wavenumber sample up to the selected maximum wavenumber threshold k_{max}, and $span$ represents the number of data points needed for calculating the smoothed value. For the smoothing of amplitude values in amplitude wavenumber domain, we apply the standard MATLAB function $smooth$, which applies a lowpass filter with filter coefficients equal to the reciprocal of the span (MathWorks, Inc. 2016). To obtain suitable smooth results, the $span$ value should be chosen carefully. Note that the smooth function starts with the second sample (equation 8) because of unusual zero-frequency values and ends with an appropriate k_{max}. In general, the selected maximum threshold value is determined using an empirical rule that keeps the generated filter to be monotonically increasing or to be fluctuating around an amplitude value of one (Zhou et al., 2019). Finally, a filter is designed with a ratio factor between the smooth CPT data and the smooth FWI results that is calculated in the wavenumber domain. This 1D filter is implemented with:

$$Filter(k) = \frac{SA_{CPT}(k)}{SA_{FWI}(k)}\ , \qquad 1 \le k \le k_{max} \tag{9}$$

where $SA_{CPT}(k)$ and $SA_{FWI}(k)$ represent the smooth CPT data and the smooth FWI results from one to the maximum threshold in the wavenumber domain, respectively. After this filter has been calculated, it is multiplied with the 2D conventional FWI results along each vertical profile in the wavenumber domain. In the next step, we generate the 2D wavenumber amplified FWI (WA-FWI) permittivity results in the spatial domain using an inverse fast Fourier transform (IFFT).

Since the real emitted source wavelet of experimental GPR data cannot be directly obtained, it is important to estimate an effective source wavelet for the FWI. Different from the traditional deconvolution approach that uses the ray-based starting models or later iterations of the FWI results, we employ the 2D WA-FWI ε_r results to replace the ray-based ε_r model. Therefore, similar to the standard procedure, synthetic data are generated with forward modeling using the standard effective source wavelet and the WA-FWI ε_r model (σ model is the same as the standard procedure). Using the Green's function (synthetic data divided by the conventional source wavelet in frequency domain) and the observed data, an updated effective source wavelet SW_{WA-FWI} can be obtained that contains the high wavenumber information. After obtaining the updated source wavelet SW_{WA-FWI}, an updated FWI is performed using the same starting models as the ones used in the conventional FWI. Generally, a second-updated source wavelet $SW_{New-FWI}$ is necessary that can be computed based on the deconvolution method approach (equations 1 and 2 in Zhou et al, 2019). In the process of updating the wavelet $SW_{New-FWI}$, we use the first-updated source wavelet SW_{WA-FWI} to replace the standard source wavelet and use the new FWI ε_r results to replace the WA-FWI ε_r models. Finally, we perform the updated FWI by using the ray-based starting models and the wavelet $SW_{New-FWI}$. The updated processing sequence including generating the filter, updating the effective wavelet and performing the updated FWI, is summarized in Figure 1.

SYNTHETIC CASE STUDIES

Stochastic aquifer models

To verify the approach of improving the resolution of GPR FWI results using the CPT data, a hydrological model based on experimental hydrological and geophysical data of the well-known Krauthausen test site (Figure 2) is used to derive synthetic GPR data (Haruzi et al., 2018; Zhou et al., 2019). We construct realistic synthetic models of relative dielectric permittivity and electrical conductivity using a stochastic simulation called sequential Gaussian simulation (e.g., Bortoli et al., 1993). For the simulation, the aquifer facies model (Gueting et al., 2017) is divided into three facies based on Tillmann et al. (2008): sand, sandy gravel and gravel (Figure 2a). The simulation of each facies is performed separately. The mean and variance values for permittivity and conductivity are calculated from the traditional GPR FWI results of the Krauthausen test site. Correlation lengths of both ε_r and σ are the same and are adapted from hydraulic conductivity values estimated from high spatial resolution CPT analysis (Tillman et al., 2008). The input parameters (mean, variance, horizontal and vertical correlation lengths) for the variogram model are summarized in Table 1.

Before computing the forward synthetic modeling results, the boundaries of the stochastic models should be enlarged to use the same borehole geometries as experimental GPR boreholes (B38-31 in Figure 2d) and to avoid interactions with the inversion domain boundaries. Here, we employ a uniform value, which is close to the boundaries within the stochastic models (shadowed areas in Figure 3a). For the unsaturated zone above the water table, we choose a homogenous layer with a relative permittivity of ε_r= 4.4 (not shown, same for all following inversions). A semi-reciprocal acquisition setup is used for the models with transmitter TRN and receiver REC spacing of 0.5 m and 0.1 m, respectively. Black circles (TRN=27) and crosses (REC=129) show the exact transmitter and receiver positions within the boreholes. The effective source wavelet used to generate synthetic data is similar to the effective source wavelet of previous measurements performed in the borehole pair B38-31 of the Krauthausen test site (Figure 2d, Gueting et al., 2015). Realistic synthetic GPR trace data (called observed data) hereafter without noise based on the stochastic models are generated using 2D FDTD modeling. The vertical dashed line (Figure 3a) indicates the selected locations of stochastic CPT (Sto-CPT) data that are used to calculate the wavenumber filter.

Conventional FWI results

First, we apply the ray-based method to generate the relative permittivity starting model for the FWI (Figure 3c). For the electrical conductivity starting model, a homogeneous model with a value of 13 mS/m is used. This is consistent with previous inversion results of experimental GPR data from this test site. The homogenous value for the conductivity starting model is based on averaging the first cycle amplitude inversion result. In the work of Gueting et al. (2015), several different starting model values for the conductivity were tested, while 13 mS/m showed the best FWI results and convergence. Using the ray-based starting models, the standard effective source wavelet SW_{Ray} is computed using the deconvolution approach based on Klotzsche et al. (2010). Thereby, the Green's function **G** based on the initial wavelet and forward modeled $\mathbf{E}^{\text{syn}}$ is calculated, and is used in the next step to obtain an effective source wavelet by deconvolving the measured data $\mathbf{E}^{\text{obs}}$ with **G**. To determine the optimal number of iterations for the final FWI results, the normalized ε_r remaining gradient values and the normalized *RMSE*s are analyzed (Figure 4a). Thereby, iteration 28 is selected as the optimal FWI iteration, and the FWI ε_r and σ results are shown in Figures 4b and 4c. A comparison of the ray-based results (Figure 3c), the FWI results (Figure 4b) and the real stochastic models (Figure 3a) indicates that the FWI results show higher resolution images and more details in the tomograms than the ray-based results. However, a certain mismatch with the real models can still be observed. Note that a good fit between the modeled traces based on the FWI results and observed data is achieved and almost no remaining gradient is present (not shown in this paper) for the chosen number of iterations.

Construction of the wavenumber filter

To obtain a generalized filter in the wavenumber domain for the synthetic data set, we apply equation 8 to smooth the highly fluctuating amplitudes of the selected and interpolated 1D FWI permittivity (dashed line in Figure 4b) and stochastic CPT (Sto-

CPT) data (Figure 3a). Note that both data sets are transformed into porosity using equation 4. To find the optimal span value of the smoothing function, the 1D wavenumber-amplified FWI results and the filtered 1D stochastic CPT data are compared in the spatial domain by computing the *RMSE* and the *R* for different span values (Figure 5a). Note that the span value needs to be chosen carefully. If it is too large, the solved filter is too smooth, creating a lower resolution result. While if the span value is too small (e.g., span=1), the solved filter is only valid for the 1D profile. A final span value of 21 is selected because it provides a high *R* and a low *RMSE* value. Using this span value, we transform and smooth the three different results (ray-based, FWI and Sto-CPT) in the wavenumber domain (Figure 5b). Here, the selected maximum threshold wavenumber is k_{max}= 2.00 m^{-1} (vertical dashed line) so that the generated filter still provides approximately monotonically increasing results. Finally, the filter is calculated by dividing the smooth Sto-CPT by the smooth FWI results (equation 9). To intuitively show the resolution differences of the different methods along the selected vertical profile, we analyze the porosity value distribution along the depth direction from 3.00 m to 8.28 m (Figure 6). The comparison of the full wavenumber information of the three different results along this vertical profile indicates that the resolution is different of the three approaches (Figure 6a). In addition, the comparison of the low wavenumber parts of the different results indicates that the wavenumber-amplified FWI (WA-FWI) results are better fitting the filtered Sto-CPT data, which means the calculated filter is valid along the 1D profile (Figure 6b). A quantitative comparison of the results can be found in Table 2, which supports these findings.

Updating the effective source wavelet and deriving new FWIs

Although the developed filter is based on 1D vertical information, it is employed for the entire 2D domain of the conventional FWI permittivity model. Thereby, for some locations, especially those that are far away from the CPT profile location, higher-wavenumber information that is not consistent with the true model appears. To remove this inconsistent noise, we use an approach inspired by spectral whitening deconvolution (Li et al., 2009). In particular, we replace the traditional ray-based permittivity model with the 2D wavenumber-amplified FWI (WA-FWI) results and use the deconvolution method to generate an updated effective source wavelet. To analyze and investigate which source wavelet strategy provides the most accurate final FWI results, we test six different effective source wavelets based on different input models in the deconvolution approach.

The effective source wavelet used to generate the observed data is named real source wavelet SW_{Real}. The effective source wavelet SW_{Ray} is used to generate the conventional FWI results, which is based on the ray-based ε_r and a homogeneous σ equal to 13 mS/m. For a comparison to the standard procedure without CPT data, this effective source wavelet SW_{Ray} is updated with the final conventional FWI results providing SW_{FWI}. As mentioned before, the source wavelet SW_{WA-FWI} is based on the WA-FWI permittivity results and SW_{Ray}. Similar to the conventional approach, this wavelet is updated once with the final FWI results using SW_{WA-FWI}, which provides $SW_{New-FWI}$. For a complete comparison of all cases, an ideal source wavelet SW_{Sto} is estimated based on the real subsurface structures of the stochastic ε_r model and the homogeneous σ model equal to 13 mS/m. Note that for a better comparison of these effective source wavelets, all source wavelets are normalized to their minimum in the provided figure (Figure 7). Comparing the six different effective source wavelets, a similar shape can be observed although a minimal time difference of the pulses is visible. Except for SW_{Ray} (blue line), all the wavelets show similar amplitude spectra in the frequency domain (Figure 7b). Note that the bandwidth for SW_{Ray} is smaller compared to the other wavelets suggesting a lower resolution of the FWI results using SW_{Ray}. The bandwidth of $SW_{New-FWI}$ (cyan line) is slightly larger than the bandwidth of SW_{WA-FWI} (red line). As expected the bandwidth of SW_{Sto} and SW_{Real} are showing the largest bandwidth. By analyzing the unwrapped phases, we find that the phase of $SW_{New-FWI}$ is closest to the phase of SW_{Real}, especially for high frequency parts (Figure 7c) indicating that $SW_{New-FWI}$ should provide the most optimal effective source wavelet when the real models are unknown.

All source wavelets are tested with FWI using the same starting models based on the ray-based ε_r results and a homogenous σ model equal to 13 mS/m to verify the relationship between source wavelet bandwidth and the accuracy of FWI results. The final FWI results for the five different source wavelets (except for SW_{Real}) are shown in

Figures 8a and 8b. *RMSE* values are computed based on the filtered stochastic permittivity model and the filtered FWI permittivity model in 2D domain to keep the same wavenumber information as the WA-FWI model. Notice that all FWI results show more details and structures than the ray-based results. Further, it is interesting to observe that although the final *RMSE* for the FWI results of SW_{FWI} and $SW_{New-FWI}$ are similar, more consistent structures close to the boreholes can be seen for the FWI results of $SW_{New-FWI}$, which better match the input model. As expected the FWI conductivity tomograms are very similar (except FWI conductivity results with SW_{Ray}), which relies on the fact that the wavenumber filter is based on porosity values in CPT data and should only change the reconstruction of the FWI permittivities. Finally, using the stochastic permittivity model as starting model for SW_{Sto} cannot significantly improve the FWI results compared to $SW_{New-FWI}$. A similar behavior can be observed by analyzing the vertical distribution of the *R* and the *RMSE*s for the filtered 2D permittivity models (Figure 8c and Table 3). Comparisons between the WA-FWI and the other FWI results show that the WA-FWI results have larger differences between x=1 m and x=3 m, which indicate that the filter is not valid in these zones due to over amplification. As expected, while all the FWI results are slightly better resolved in the middle regions of the tomograms, FWI results are degraded in the vicinity of the boreholes due to the acquisition strategy in crosshole applications. Furthermore, the FWI results of $SW_{New-FWI}$ show a higher *R* and a smaller *RMSE* value than the results of SW_{WA-FWI} and SW_{Ray}, especially in the vicinity of the left borehole. Only when the synthetic GPR trace data are noise-free, are the FWI results of SW_{FWI} better than those of $SW_{New-FWI}$ (Zhou et al., 2019). As expected, the optimal FWI results are obtained using SW_{Sto}, which can only be obtained in synthetic model studies. In the absence of complete knowledge of the subsurface, the FWI results based on $SW_{New-FWI}$ show the best results.

EXPERIMENTAL GPR DATA STUDIES

At the Krauthausen test site in Germany (Figure 2c), we measured crosshole GPR data in the saturated aquifer using 200 MHz borehole antennae (Gueting et al., 2015) between several boreholes (red lines in Figure 2d). A detailed description of the site is provided by Vereecken et al. (2000). The measured aquifer can be broadly divided into three layers (Figure 2a): A poorly sorted gravel layer extending from 1 m to 4 m in depth; the middle sand layer extending from 4 m to 6 m in depth; and a bottom layer including sandy and gravely grains extending from 6 m to 11.5 m depth (Tillmann et al., 2008). For the acquisition of the experimental data, a semi-reciprocal acquisition setup (Figure 2b) was used with a transmitter and receiver spacing of 0.5 m and 0.1 m, respectively. The water table was approximately at a 2 m depth during the measurements. Therefore, GPR measurements started below 3 m in depth. The CPT profiles that are closest to the crosshole sections are shown in Figure 2d (red asterisk). To improve the crosshole GPR FWI results with our new approach, we analyze five GPR cross-sections and the corresponding CPT profiles. For five CPT locations, the

CPT probe was pushed into the subsurface to measure cone resistance, electrical resistivity, natural gamma, gamma-gamma and neutron activity values every 10 cm (Gueting et al., 2015). The neutron log data was transformed to soil water content using the proposed calibration of Tillmann et al. (2008). In contrast to Gueting et al. (2015), we reanalyze the FWI results following the suggestion given by the Corrigendum of the Gueting et al. (2020) paper. A reanalysis of the zero-time correction of the GPR data showed that there was an error in the automatic picking routine which is now updated. Therefore, the conventional FWI results are different to the results of Gueting et al. (2015) and show generally higher permittivities and lower electrical conductivities, while the structures are similar.

In the first step, the porosity information of five 1D vertical CPT profiles is compared to the corresponding FWI porosities, and the wavenumber filter for each borehole pair is estimated separately (Figure 9). Note that the original CPT data are used (Tillmann et al., 2008) without applying a shift as proposed by Gueting et al. (2015). For the experimental GPR data, a span value of 27 in the smooth function is selected for all cross-sections and the maximum threshold wavenumber of the filters k_{max} is 2.31 m^{-1}. The 1D porosity amplitude values along the CPT profile locations in the amplitude-wavenumber domain clearly show that the CPT values contain the largest bandwidth, whereas the FWI results have a reduced bandwidth, and the ray-based data have the lowest bandwidth for all five borehole pairs (Figures 9a to 9e). By comparing the five obtained filters (Figure 9f) in the wavenumber range of 0 - 2.31 m^{-1}, we can observe differences of the filters near 0.5 m^{-1} for the filter of profile 103 between boreholes B62-30. Note that the cross-section distance between the boreholes B62 and B30 is 6.16 m, which is the largest between any pair (Figure 2d).

In the next step, these five wavenumber filters are applied to derive WA-FWI results between each borehole pair and the corresponding updated effective source wavelets $SW_{New-FWI}$ (Figure 10). Similar to the synthetic case study, by using the deconvolution approach, we update the standard effective source wavelets SW_{Ray} based on the WA-FWI results and a homogenous σ model equal to 13 mS/m to generate SW_{WA-FWI} and, then, update these wavelets SW_{WA-FWI} to obtain $SW_{New-FWI}$

(Figure 1). Note that we only show the permittivity FWI results based on $SW_{New-FWI}$, since this source wavelet has provided satisfying results in the synthetic study. In addition, in the experimental GPR tests, we have observed that inverted conductivity FWI results with $SW_{New-FWI}$ are not always better than the conductivity FWI results with SW_{Ray}. A possible reason is that the designed filter based on the CPT data is not always effective to improve the conductivity FWI results, especially if noise is present the experimental GPR data. Therefore, we do not show the comparison for conductivity FWI results in this study.

The final effective source wavelets show similar shapes with slight shifts in time (Figure 10a) and similar bandwidth in the frequency spectra (Figures 10b and 10c). Note that the effective source wavelet SW_{WA-FWI} for the cross-section between boreholes B62 and B30 is solved based on “WA-FWI subtract 1”, which is necessary because the WA-FWI permittivity values results in modeled data outside of the half wavelength criteria and hence it is not possible to solve an adequate effective source wavelet. One possible reason is that the largest borehole distance lower the FWI resolution and then the derived filter over amplifies the WA-FWI values. We also consider the ray-based ε_r results and the homogenous σ model equal to 13 mS/m as the starting models in the FWI process. The traditional FWI ε_r results (Figure 11a) using SW_{Ray} are used to derive the WA-FWI results (Figure 11b). The updated FWI results (Figure 11c) are derived using the corresponding updated source wavelets $SW_{New-FWI}$. Similar to the synthetic studies, the WA-FWI results show over amplified features close to the boreholes. The vertical dashed lines indicate the locations of CPT data for each pair of boreholes. The updated FWI results show more consistent structures in the individual planes and at the crossings of the boreholes in comparison to the conventional FWI permittivity results. Generally, improved *RMSE* values and *R* factors are obtained for the updated FWI results than for the conventional FWI results (Table 5).

Finally, to verify the updated FWI results, we compute and compare the FWI porosity results using equation 4 with the CPT porosity values (Figure 12). Thereby, we first compare the wavenumber-amplified FWI results with the filtered CPT (Figure 12a) similar to the synthetic case study (Figure 6b). Note that we select the same depth of the CPT data from 3.00 m to 8.28 m for five different measurements to calculate the filters and compare with different FWI results. Both Figure 12a and Table 4 show the comparison of the filtered porosity values along their respective 1D CPT profiles. Comparisons of the full wavenumber information for CPT (blue), ray-based results (green), conventional FWI (red) and updated FWI (black) along each CPT profile are shown in Figure 12b (quantitative comparison in Table 5). An improved fit between the CPT and updated FWI results in contrast to the conventional FWI results is visible. By comparing the computed *R* and *RMSE* between the CPT data and the 1D different FWI results, we conclude that the updated effective source wavelets, which incorporate the CPT information, improve the FWI permittivity results for all planes.

CONCLUSION

We demonstrate a new approach to improve the permittivity FWI results by incorporating additional information from CPT data. By updating the effective source wavelet with the amplified FWI results, we include the 1D CPT information into the effective source wavelet. Therefore, this updated wavelet is able to provide improved

FWI results. The novel method is tested and verified on a realistic synthetic case study and applied to an experimental data set from the Krauthausen test site in Germany. To improve the FWI results, we propose to design a 1D wavenumber filter based on CPT porosity data and to apply this filter to the 2D conventional FWI results. To verify the approach of updating the source wavelet based on the CPT data, we generate a stochastic model of the Krauthausen test site. Combining the conventional FWI permittivity results and Sto-CPT data, we generate a filter that is applied to the 2D FWI results to yield the WA-FWI results. Note that the FWI permittivity amplification is only performed once using the convention final FWI permittivity results and the derived filter. To remove the inconsistent high wavenumber data present in the wavenumber-amplified FWI results, we estimate an effective source wavelet SW_{WA-FWI} based on the WA-FWI results. Further, we use five different effective source wavelets to perform FWI to determine the best effective source wavelet. The synthetic studies indicate that we can obtain an enhanced source wavelet $SW_{New-FWI}$ by applying an additional source wavelet correction cycle with the deconvolution approach. Although the new approach is not significantly improving FWI results, more consistent structures, especially close to the boreholes, are obtained and an enhanced data correlation is achieved.

The new approach for optimizing the effective source wavelet with the CPT data is tested at experimental GPR datasets of five cross-boreholes sections. Comparisons of the final updated FWI results and the CPT porosities confirm the improvement compared to the conventional FWI results. In future research, we will try to tame the non-linearity problem by gradually expanding the bandwidth of the updated effective source wavelet, as it is traditionally done with seismic data FWI.

ACKNOWLEDGMENTS

The first author acknowledges the financial support from China Scholarship Council (Project No. 201506340136). We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC). We also want to thank T. Liu, Y. Yu, and P. Haruzi for providing help and useful suggestions in the process of constructing the synthetic stochastic models and dealing with data. Especially, we are grateful to J.A. Huisman for his constructive suggestions. This work was supported in part by the TERrestrial Environmental Observatories (TERENO). Further, we would like to thank the assistant editor A. Guitton, the associate editor M. Bano, the editor-in-chief J. Shragge and two anonymous reviewers for suggestions and comments that greatly improved the paper.

REFERENCES

Birchak, J.R., C.G. Gardner, J.E. Hipp, and J.M. Victor, 1974, High dielectric constant microwave probes for sensing soil moisture: Proceedings of the IEEE, **62**, no.

1, 93-98.

Bleinstein, N., 1986, Two-and-one-half dimensional in-plane wave-propagation: Geophysical Prospecting, **34**, 686–703.

Bortoli, L. J., F. Alabert, A. Haas, and A. G. Journel, 1993, Constraining stochastic images to seismic data, in A. Soares, ed., Geostatistics Tróia 1992, Proceedings of the 4th International Geostatistics Congress: Kluwer Academic Publishers, 325–337.

Busch, S., J. van der Kruk, J. Bikowski, and H. Vereecken, 2012, Quantitative conductivity and permittivity estimation using full-waveform inversion of on-ground GPR data: Geophysics, **77**, no. 6, H79–H91.

Carmichael, R.S. ed., 2017, Handbook of Physical Properties of Rocks (1984), Vol.3: CRC press.

Coscia, I., N. Linde, S. A. Greenhalgh, T. Vogt, and A. G. Green, 2012, Estimating traveltimes and groundwater flow patterns using 3D time-lapse crosshole

ERT imaging of electrical resistivity fluctuations induced by infiltrating river water: Geophysics, **77**, no. 4, E239-E250.

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

Dafflon, B., and W. Barrash, 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 Resources Research, **48**, W05553.

Doetsch, J., N. Linde, I. Coscia, S. A. Greenhalgh, and A. G. Green, 2010, Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data: Geophysics, **75**, no. 6, G53–G64.

Eisenberg, D. and W. Kauzmann, 2005, The structure and properties of water: Oxford University Press on Demand.

Ernst, J. R., H. Maurer, A. G. Green, and K. Holliger, 2007a, Full-waveform inversion of crosshole radar data based on 2-D finite-difference time-domain solutions of Maxwell's equations: IEEE Transactions on Geoscience and Remote Sensing, **45**, 2807–2828.

Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007b, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, no. 5, J53–J64.

Fejes, I., and E. Jósa, 1990, The engineering geophysical sounding method. Principles, instrumentation, and computerised interpretation, in S. H. Ward, ed., Geotechnical and environmental geophysics, Vol. 2, Environmental and groundwater: SEG, 321–331.

Garambois, S., P. Sénéchal, and H. Perroud, 2002, On the use of combined geophysical methods to assess water content and water conductivity of near-surface formations: Journal of Hydrology, **259**, 32–48.

Gueting, N., A. Klotzsche, J. van der Kruk, J. Vanderborght, H. Vereecken, and

A. Englert, 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.

Gueting, N., T. Vienken, A. Klotzsche, J. van der Kruk, J. Vanderborght, J. Caers, H. Vereecken, and A. Englert, 2017, High resolution aquifer characterization using crosshole GPR full-waveform tomography: Comparison with direct‐push and tracer test data: Water Resources Research, **53**, no. 1, 49-72.

Gueting, N., A. Klotzsche, J. van der Kruk, J. Vanderborght, H. Vereecken, and A. Englert, 2020, Corrigendum to" Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests": Journal of Hydrology, **590**, 125483. DOI: 10.1016/j.jhydrol.2020.125483.

Haruzi, P., N. Gueting, A. Klotzsche, J. Vanderborght, H. Vereecken, and J. van Kruk, 2018, Time-lapse ground-penetrating radar full-waveform inversion to detect tracer plumes: A numerical study: 88th Annual International Meeting, SEG, Expanded Abstracts, 2486–2490.

Hubbard, S. S., J. Chen, J. E. Peterson, E. L. Majer, K. H. Williams, D. J. Swift, B. Mailloux, and Y. Rubin, 2001, Hydrogeological characterization of the South Oyster bacterial transport site using geophysical data: Water Resources Research, **37**, 2431–2456.

Irving, J. D., M. D. Knoll, and R. J. Knight, 2007, Improving crosshole radar velocity tomograms: A new approach to incorporating high-angle traveltime data: Geophysics, **72**, no. 4, J31–J41.

Klotzsche, A., J. van der Kruk, G. A. Meles, J. Doetsch, H. Maurer, and N. Linde, 2010, Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the Thur River, Switzerland: Near Surface Geophysics, **8**, 635–649.

Klotzsche, A., J. van der Kruk, N. Linde, J. Doetsch, and H. Vereecken, 2013, 3-D characterization of high-permeability zones in a gravel aquifer using 2-D crosshole GPR full-waveform inversion and waveguide detection: Geophysical Journal International, **195**, 932–944.

Klotzsche, A., J. van der Kruk, J. Bradford, and H. Vereecken, 2014, Detection of spatially limited high-porosity layers using crosshole GPR signal analysis and full-waveform inversion: Water Resources Research, **50**, 6966–6985.

Klotzsche, A., F. Jonard, M. C. Looms, J. van der Kruk, and J. A. Huisman, 2018, Measuring soil water content with ground penetrating radar: A decade of progress: Vadose Zone Journal, **17**, no. 1, 180052.

Klotzsche, A., H. Vereecken, and J. van der Kruk, 2019a, GPR full-waveform inversion of a variably saturated soil-aquifer system: Journal of Applied Geophysics, **170**, 103823.

Klotzsche, A., H. Vereecken, and J. van der Kruk, 2019b, Review of Crosshole GPR Full-waveform Inversion of Experimental Data: Recent Developments, Challenges and Pitfalls: Geophysics, **84**, no. 6, H13-H28.

Kurzmann, A., A. Przebindowska, D. Köhn, and T. Bohlen, 2013, Acoustic full waveform tomography in the presence of attenuation: A sensitivity analysis: Geophysical Journal International, **195**, 985–1000.

Li, G., H. Zhou, and C. Zhao, 2009, Potential risks of spectrum whitening deconvolution compared with well-driven deconvolution: Petroleum Science, 146–152. DOI: 10.1007/s12182-009-0023-y.

Lunne, T., P. Robertson, and J. Powell, 1997, CPT in geotechnical practice: Blackie Academic.

MathWorks, Inc., 2016, Signal processing toolbox: for use with MATLAB: User's Guide, the MathWorks; https://www.mathworks.com/help/curvefit/smooth.html#d122e45185.

Maurer, H., and M. Musil, 2004, Effects and removal of systematic errors in crosshole georadar attenuation tomography: Journal of Applied Geophysics, **55**, 261-270.

Meles, G., J. Van der Kruk, S. A. Greenhalgh, J. R. Ernst, H. Maurer, and A. G. Green, 2010, A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination crosshole/borehole-to-surface GPR data: IEEE Transactions on Geoscience and Remote Sensing, **48**, 3391–

3407.

Polak, E., and G. Ribière, 1969, Note on convergence of conjugate direction methods: Revue Française d'Informatique de Recherche Opérationnelle, **3**, 35–43.

Ristau, J. P., and W. M. Moon, 2001, Adaptive filtering of random noise in 2-D geophysical data: Geophysics, **66**, 342–349.

Tillmann, A., A. Englert, Z. Nyari, I. Fejes, J. Vanderborght, and H. Vereecken, 2008, Characterization of subsoil heterogeneity, estimation of grain size distribution and hydraulic conductivity at the Krauthausen test site using cone penetration test: Journal of Contaminant Hydrology, **95**, 57–75.

van der Kruk, J., N. Gueting, A. Klotzsche, G. W. He, S. Rudolph, C. von Hebel, X. Yang, L. Weihermuller, A. Mester, and H. Vereecken, 2015, Quantitative multi-layer electromagnetic induction inversion and full waveform inversion of crosshole ground penetrating radar data: Journal of Earth Science, **26**, 844–850.

Vereecken, H., U. Döring, H. Hardelauf, U. Jaekel, U. Hashagen, O. Neuendorf,

H. Schwarze, and R. Seidemann, 2000, Analysis of solute transport in a heterogeneous aquifer, The Krauthausen field experiment: Journal of Contaminant Hydrology, **45**, 329–358.

Yang, X., A. Klotzsche, G. Meles, H. Vereecken, and J. van der Kruk, 2013, Improvements in crosshole GPR full-waveform inversion and application on data measured at the Boise Hydrogeophysics Research Site: Journal of Applied Geophysics, **99**, 114–124.

Zhou, Z., A. Klotzsche, N. Güting, P. Haruzi, H. Vereecken, and J. van der Kruk, 2019, Improved resolution of ground penetrating radar full-waveform inversion by using cone penetration test data: A synthetic study: 89th Annual International Meeting, SEG, Expanded Abstracts, 2898–2902.

Zhou, Z., A. Klotzsche, T. Hermans, F. Nguyen, J. Schmäck, P. Haruzi, H. Vereecken, and J. van der Kruk, 2020, 3D aquifer characterization of the Hermalle-sous-Argenteau test site using crosshole GPR amplitude analysis and full-waveform inversion: Geophysics, **85**, no. 6, H133-H148.

LIST OF FIGURES

Figure 1. Illustration of the updating strategy of the effective source wavelet based on WA-FWI results and of the performance of the new FWI. The red boxes, which represent data in the wavenumber domain, show the process of constructing the filter. The green boxes indicate generating WA-FWI and updating the effective source wavelet. The blue boxes show the FWI process. Homo (σ) represents the homogenous σ starting model equal to 13 mS/m, which combines ray-based ε_r as the starting models used for the updating source wavelet and the FWI in this study.

Figure 2. (a) Generalized cross-section of the uppermost aquifer based on Tillmann et al. (2008). (b) Schematic of the crosshole GPR acquisition setup, in which the green arrow indicates the location of CPT data. (c) Picture of the Krauthausen test site and (d) location of boreholes (circles) and cone penetration tests (asterisk), in which the distance from the CPT 144 to the corresponding cross-section is about 0.5 m. (a) and (b) are adapted from Gueting et al. (2015).

Figure 3. The (a) ε_r and (b) σ models based on the stochastic simulation used to generate the realistic synthetic GPR data. The shadow zones at the boundaries indicate the extended domain of the inversion. The vertical dashed line indicates the selected Sto-CPT location used to compute the filter and to amplify the wavenumber of the FWI results. (c) Ray-based result for ε_r using the GPR data based on (a) and b) a uniform starting model for σ, which are FWI starting models.

Figure 4. (a) Evolution of the FWI *RMSE* misfit (black line) and the remaining absolute mean ε_r gradient (blue line) with iterations. The black line indicates the average value between the normalized remaining gradient values and the normalized *RMSE*. The red circle shows the FWI iteration with the optimal value. (b) The standard FWI permittivity and (c) conductivity results after 28 iterations. The dashed vertical line indicates the selected FWI profile used to generate the amplifying filter.

Figure 5. (a) The distributions of *RMSE* and *R* values for porosity results as a function of

the smooth function span values for the selected range of 0 to 71. The dashed line indicates the optimal span value of 21. (b) A comparison of the spatial wavenumber spectra of Sto-CPT data (blue), FWI (red) and ray-based (green) results. The filter is indicated by the black solid line, which is derived from the ratio between the smooth Sto-CPT and the smooth FWI (smooth span is 21). The dashed black line shows the maximum wavenumber for the filter.

Figure 6. Comparisons of the (a) full and (b) low wavenumber information for Sto-CPT (blue), ray-based (green) and FWI (red) porosity results.

Figure 7. Comparisons of different effective source wavelets in (a) time domain, (b) corresponding frequency spectra, and (c) phase spectra based on the different processing steps indicated in Figure 1. Note that all source wavelets are estimated for different ε_r models, while σ models are the same for all steps with a homogenous model of 13 mS/m. Amplitudes of (a) and (b) are normalized to their corresponding minimum and maximum for a better comparison.

Figure 8. Comparisons of FWI (a) permittivity and (b) conductivity results using different

effective source wavelets (Figure 7). Values in parentheses indicate the mean *RMSE* between filtered FWI permittivity models and the filtered stochastic permittivity model in the entire 2D domain (see Table 3 for more details). (c) Quantitative comparisons of the *RMSE* and *R* between filtered stochastic permittivity model and different filtered FWI permittivity results (same wavenumber as WA-FWI) along the vertical profile.

Figure 9. (a) To (e) comparisons of spatial frequency spectra of the CPT data (blue), the ray-based (green) and the conventional FWI (red) results in the wavenumber domain for different profiles (see Figure 2d for the locations of the profiles). The wavenumber filter is indicated by the black solid line for each profile. (f) Comparisons of the five filters, where a marked difference of profile 103 to the other profiles near 0.5 m^{-1} is noticeable.

Figure 10. Comparisons of the updated effective source wavelets of the five cross-sections used for the experimental study in (a) time domain, (b) frequency and (c) phase spectra. Amplitudes of (a) and (b) are normalized to their

corresponding minimum and maximum for a better comparison.

Figure 11. (a) Traditional permittivity FWI results using $\mathrm{SW}_{\mathrm{Ray}}$ for the five cross-sections. Circles and crosses indicate the transmitter and receiver locations, respectively. Dashed lines present the locations of the CPT profiles. (b) Permittivity images of the wavenumber-amplified FWI using the filters shown in Figure 9. (c) Updated FWI results using the updated effective source wavelets as shown in Figure 10.

Figure 12. (a) Porosity comparisons of the filtered CPT (blue), the ray-based results (green), the filtered FWI results (red) and the wavenumber-amplified FWI (black) along each vertical profile. (b) Full wavenumber porosity results comparison of the CPT, ray-based, the FWI results (using $\mathrm{SW}_{\mathrm{Ray}}$) and the updated FWI results (using $\mathrm{SW}_{New-FWI}$).

LIST OF TABLES

Table 3. Mean *RMSE* and *R* of different ε_r model comparisons for the entire 2D domain. F-Stochastic and F-FWI (to keep the same wavenumber information as WA-FWI) are filtered Stochastic and filtered FWI permittivity models, respectively. The F-FWI results with $SW_{New-FWI}$ are the optimal choice because of the lower *RMSE* and the higher *R* value.

Table 4. Comparisons between filtered CPT, filtered FWI and wavenumber amplified FWI (WA-FWI) porosity results of the experimental data set from the Krauthausen site. Percentages in parentheses indicate the improvement of the WA-FWI *RMSE* to the filtered FWI *RMSE*.

Table 5. Comparisons of the full wavenumber CPT and FWI porosity results using different effective source wavelets. *R* and *RMSE* are calculated based on 1D full wavenumber profile data. Percentages in parentheses indicate the improvement of the New-FWI *RMSE* to the traditional FWI *RMSE*.

Table 1. Parameters for stochastic simulation of permittivity and conductivity based on data at the Krauthausen test site (Tillman et al., 2008). Parameters ε and σ are mean values for different facies. s^2_ε and s^2_σ represent variance values for permittivity and conductivity, respectively. Parameters $\lambda_{\varepsilon,h}$ and $\lambda_{\varepsilon,v}$ are the horizontal and vertical correlation lengths fitted with an exponential model for permittivity. And the horizontal and vertical correlation lengths of conductivity are shown by $\lambda_{\sigma,h}$ and $\lambda_{\sigma,v}$, respectively.

		Sand (1)	Sandy gravel (2)	Gravel (3)
Permittivity	ε	21.52	17.82	13.89
	s^2_ε	9.83	8.71	8.68
	$\lambda_{\varepsilon,h}[\mathrm{m}]$	5	1.75	0.3
	$\lambda_{\varepsilon,v}[\mathrm{m}]$	0.19	0.2	0.41
Electrical conductivity	$\sigma\left[\frac{\mathrm{mS}}{\mathrm{m}}\right]$	15	10.4	9.6
	$s^2_\sigma\left[\left(\frac{\mathrm{mS}}{\mathrm{m}}\right)^2\right]$	4.32	17.68	4.48
	$\lambda_{\sigma,h}[\mathrm{m}]$	5	1.75	0.3
	$\lambda_{\sigma,v}[\mathrm{m}]$	0.19	0.2	0.41

Table 2. Comparisons of the correlation coefficient *R* and the root mean squared error *RMSE* of the filtered Sto-CPT, filtered FWI and wavenumber amplified FWI (WA-FWI) results given the maximum wavenumber, the suitable span value and the optimal FWI iteration value. *R* is Pearson's Correlation Coefficient between two variables (same for all following tables). The percentage in parentheses indicates the improvement of the WA-FWI *RMSE* to the filtered FWI *RMSE*.

Considered parameter	ε_r
Max. wavenumber for filter (m^{-1})	2.00
Span value of smooth function	21
Optimal iteration of FWI	28
R (Filtered FWI: Filtered Sto-CPT)	0.9562
R (WA-FWI: Filtered Sto-CPT)	0.9655
RMSE (Filtered FWI: Filtered Sto-CPT)	1.0907
RMSE (WA-FWI: Filtered Sto-CPT)	0.8830
Improvement	19.0%

Table 3. Mean *RMSE* and *R* of different ε_r model comparisons for the entire 2D domain. F-Stochastic and F-FWI (to keep the same wavenumber information as WA-FWI) are filtered Stochastic and filtered FWI permittivity models, respectively. The F-FWI results with $SW_{New-FWI}$ are the optimal choice because of the lower *RMSE* and the higher *R* value.

Compared models ($\boldsymbol{\varepsilon_r}$)	Mean *RMSE*	Mean *R*
F-Stochastic and Ray-based	2.4836	0.7237
F-Stochastic and F-FWI ($\boldsymbol{SW_{Ray}}$)	1.4886	0.9222
F-Stochastic and F-FWI ($\boldsymbol{SW_{FWI}}$)	1.3234	0.9409
F-Stochastic and WA-FWI	2.1754	0.8713
F-Stochastic and F-FWI ($\boldsymbol{SW_{WA-FWI}}$)	1.6465	0.8907
F-Stochastic and F-FWI ($\boldsymbol{SW_{New-FWI}}$)	1.3660	0.9315
F-Stochastic and F-FWI ($\boldsymbol{SW_{Sto}}$)	1.2029	0.9445

Table 4. Comparisons between filtered CPT, filtered FWI and wavenumber amplified FWI (WA-FWI) porosity results of the experimental data set from the Krauthausen site. Percentages in parentheses indicate the improvement of the WA-FWI *RMSE* to the filtered FWI *RMSE*.

Borehole # **Distance between boreholes**	**32-38** **(5.13m)**	**38-31** **(4.99m)**	**31-62** **(3.83m)**	**62-30** **(6.16m)**	**75-76** **(4.96m)**
Profiles of CPT	100	101	102	103	144
Max. wavenumber for filter ($\mathbf{m}^{-1}$)	2.31	2.31	2.31	2.31	2.31
Span value of smooth function	27	27	27	27	27
Optimal iteration of FWI	30	22	30	15	26
***R* (Filtered FWI: Filtered CPT)**	0.7851	0.9278	0.8414	0.8711	0.8340
***R* (WA- FWI: Filtered CPT)**	0.7604	0.9031	0.8771	0.9054	0.9062
***RMSE* (Filtered FWI: Filtered CPT)**	0.0436	0.0308	0.0386	0.0349	0.0269
***RMSE* (WA- FWI: Filtered CPT)**	0.0291	0.0197	0.0210	0.0210	0.0207
Improvement	33.3%	36.0%	45.6%	39.8%	23.0%

Table 5. Comparisons of the full wavenumber CPT and FWI porosity results using different effective source wavelets. *R* and *RMSE* are calculated based on 1D full wavenumber profile data. Percentages in parentheses indicate the improvement of the New-FWI *RMSE* to the traditional FWI *RMSE*.

Borehole # **Distance between boreholes**	**32-38** **(5.13 m)**	**38-31** **(4.99 m)**	**31-62** **(3.83 m)**	**62-30** **(6.16 m)**	**75-76** **(4.96 m)**
***R* (FWI: CPT)**	0.7576	0.9153	0.8049	0.8564	0.8149
***R* (New-FWI: CPT)**	0.7701	0.9189	0.8312	0.8569	0.8636
***RMSE* (FWI: CPT)**	0.0448	0.0316	0.0410	0.0360	0.0285
***RMSE* (New-FWI : CPT)**	0.0296	0.0249	0.0249	0.0272	0.0255
Improvement	33.9%	21.2%	39.3%	24.4%	10.5%

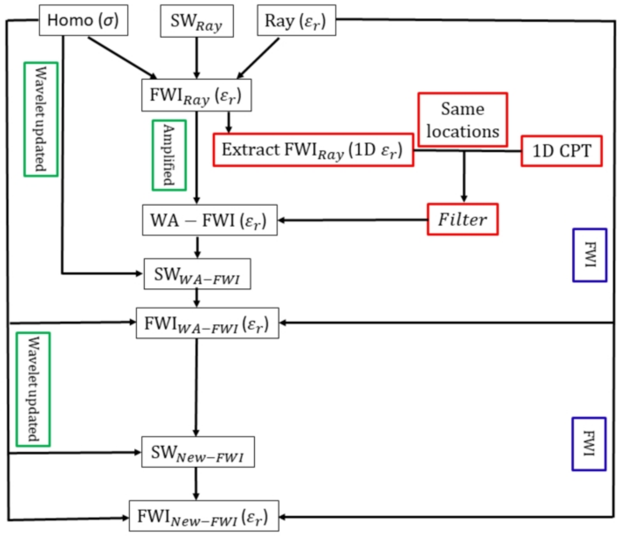

Figure 1. Illustration of the updating strategy of the effective source wavelet based on WA-FWI results and of the performance of the new FWI. The red boxes, which represent data in the wavenumber domain, show the process of constructing the filter. The green boxes indicate generating WA-FWI and updating the effective source wavelet. The blue boxes show the FWI process. Homo (σ) represents the homogenous σ starting model equal to 13 mS/m, which combines ray-based ε_r as the starting models used for the updating source wavelet and the FWI in this study.

338x190mm (300 x 300 DPI)

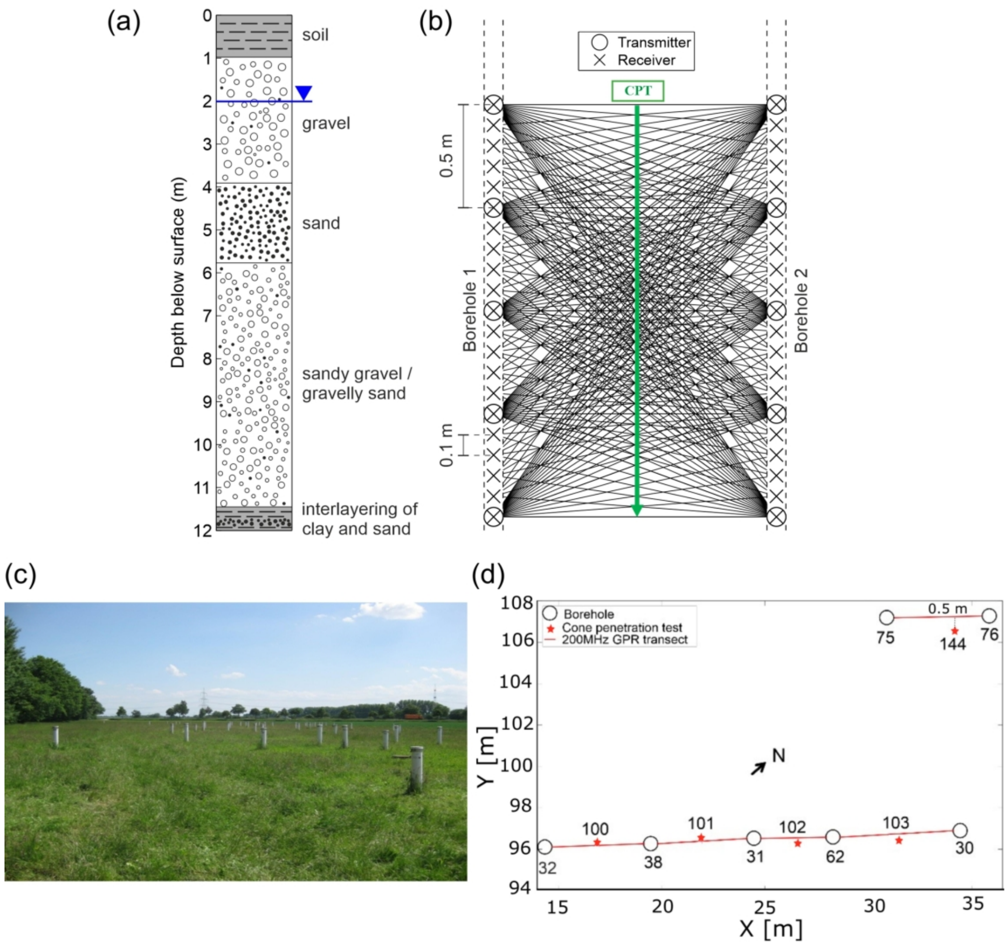

Figure 2. (a) Generalized cross-section of the uppermost aquifer based on Tillmann et al. (2008). (b) Schematic of the crosshole GPR acquisition setup, in which the green arrow indicates the location of CPT data. (c) Picture of the Krauthausen test site and (d) location of boreholes (circles) and cone penetration tests (asterisk), in which the distance from the CPT 144 to the corresponding cross-section is about 0.5 m. (a) and (b) are adapted from Gueting et al. (2015).

203x190mm (300 x 300 DPI)

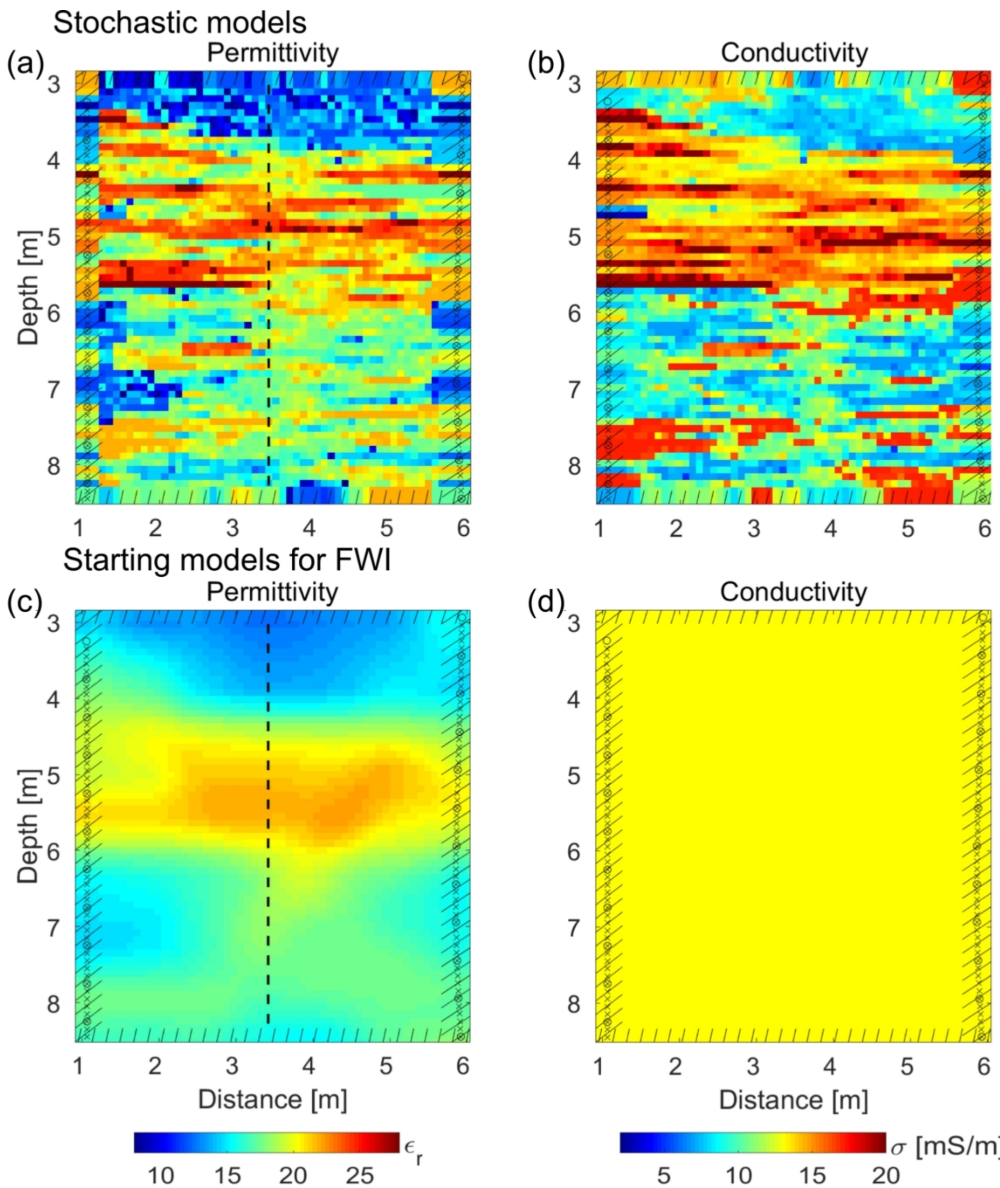

Figure 3. The (a) ε_r and (b) σ models based on the stochastic simulation used to generate the realistic synthetic GPR data. The shadow zones at the boundaries indicate the extended domain of the inversion. The vertical dashed line indicates the selected Sto-CPT location used to compute the filter and to amplify the wavenumber of the FWI results. (c) Ray-based result for ε_r using the GPR data based on (a) and b) a uniform starting model for σ, which are FWI starting models.

179x210mm (300 x 300 DPI)

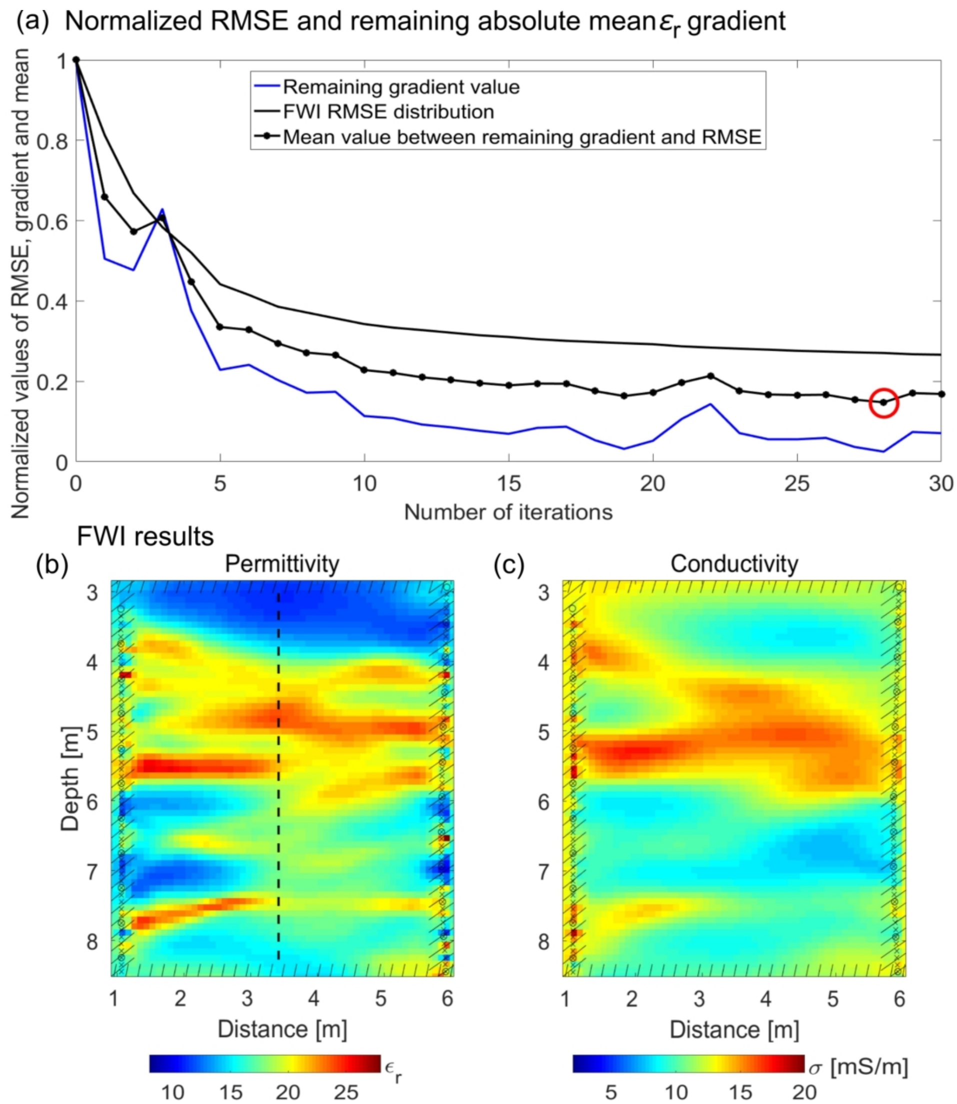

Figure 4. (a) Evolution of the FWI *RMSE* misfit (black line) and the remaining absolute mean ε_r gradient (blue line) with iterations. The black line indicates the average value between the normalized remaining gradient values and the normalized *RMSE*. The red circle shows the FWI iteration with the optimal value. (b) The standard FWI permittivity and (c) conductivity results after 28 iterations. The dashed vertical line indicates the selected FWI profile used to generate the amplifying filter.

206x225mm (300 x 300 DPI)

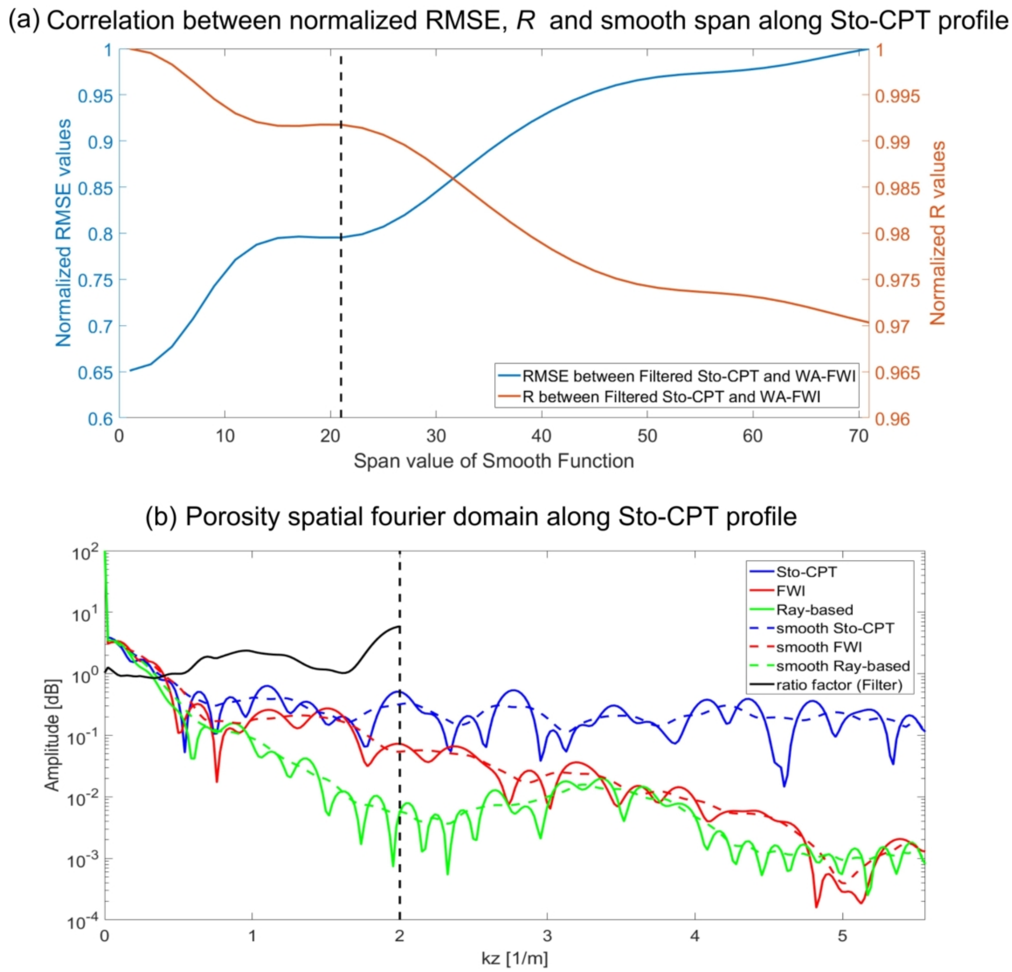

Figure 5. (a) The distributions of *RMSE* and *R* values for porosity results as a function of the smooth function span values for the selected range of 0 to 71. The dashed line indicates the optimal span value of 21. (b) A comparison of the spatial wavenumber spectra of Sto-CPT data (blue), FWI (red) and ray-based (green) results. The filter is indicated by the black solid line, which is derived from the ratio between the smooth Sto-CPT and the smooth FWI (smooth span is 21). The dashed black line shows the maximum wavenumber for the filter.

214x207mm (300 x 300 DPI)

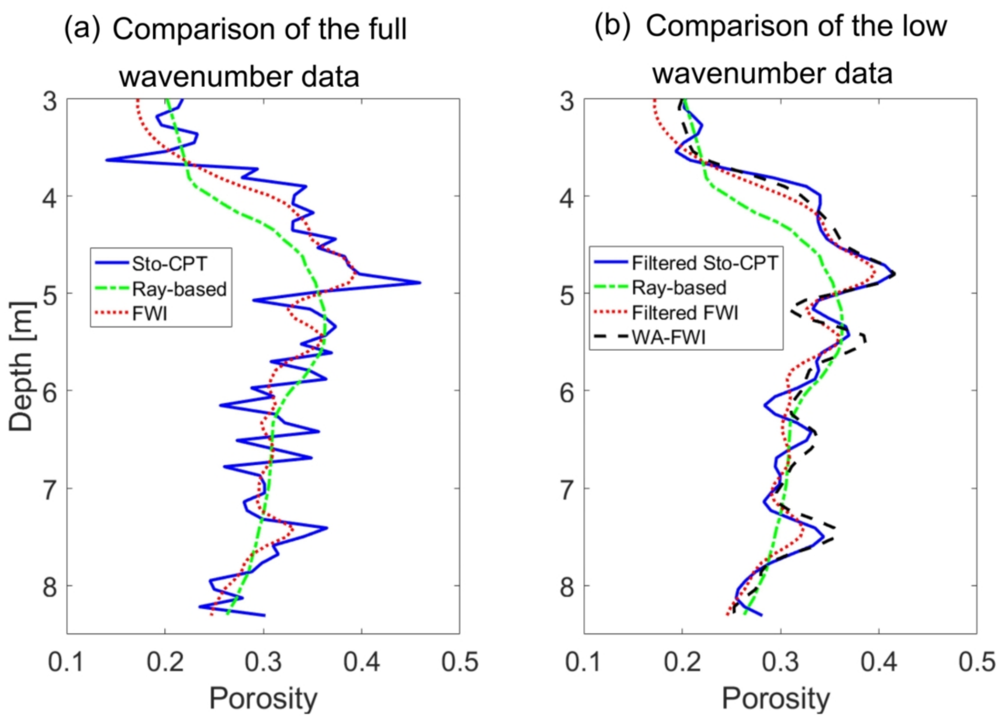

Figure 6. Comparisons of the (a) full and (b) low wavenumber information for Sto-CPT (blue), ray-based (green) and FWI (red) porosity results.

185x131mm (300 x 300 DPI)

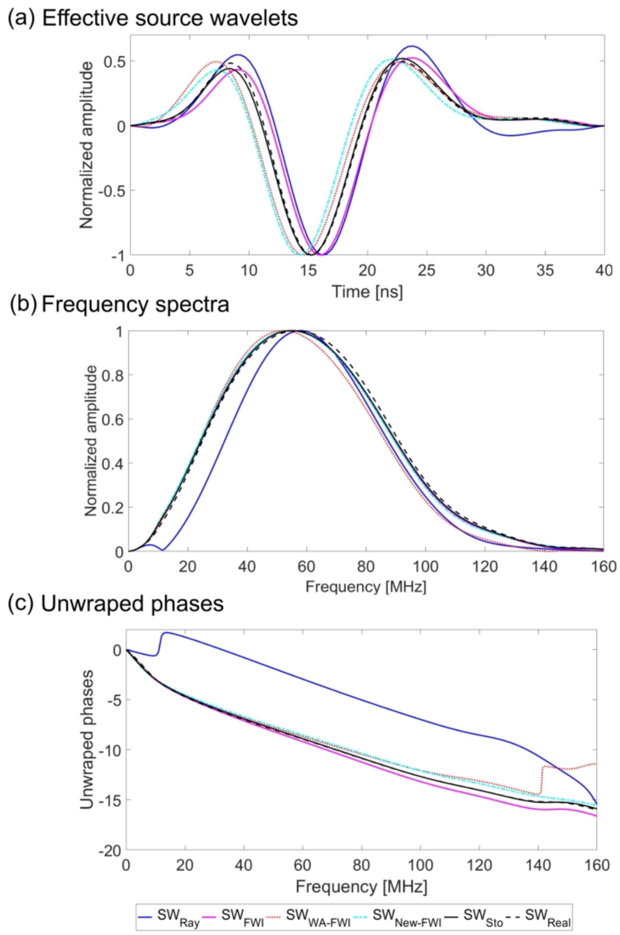

Figure 7. Comparisons of different effective source wavelets in (a) time domain, (b) corresponding frequency spectra, and (c) phase spectra based on the different processing steps indicated in Figure 1. Note that all source wavelets are estimated for different ε_r models, while σ models are the same for all steps with a homogenous model of 13 mS/m. Amplitudes of (a) and (b) are normalized to their corresponding minimum and maximum for a better comparison.

210x297mm (300 x 300 DPI)

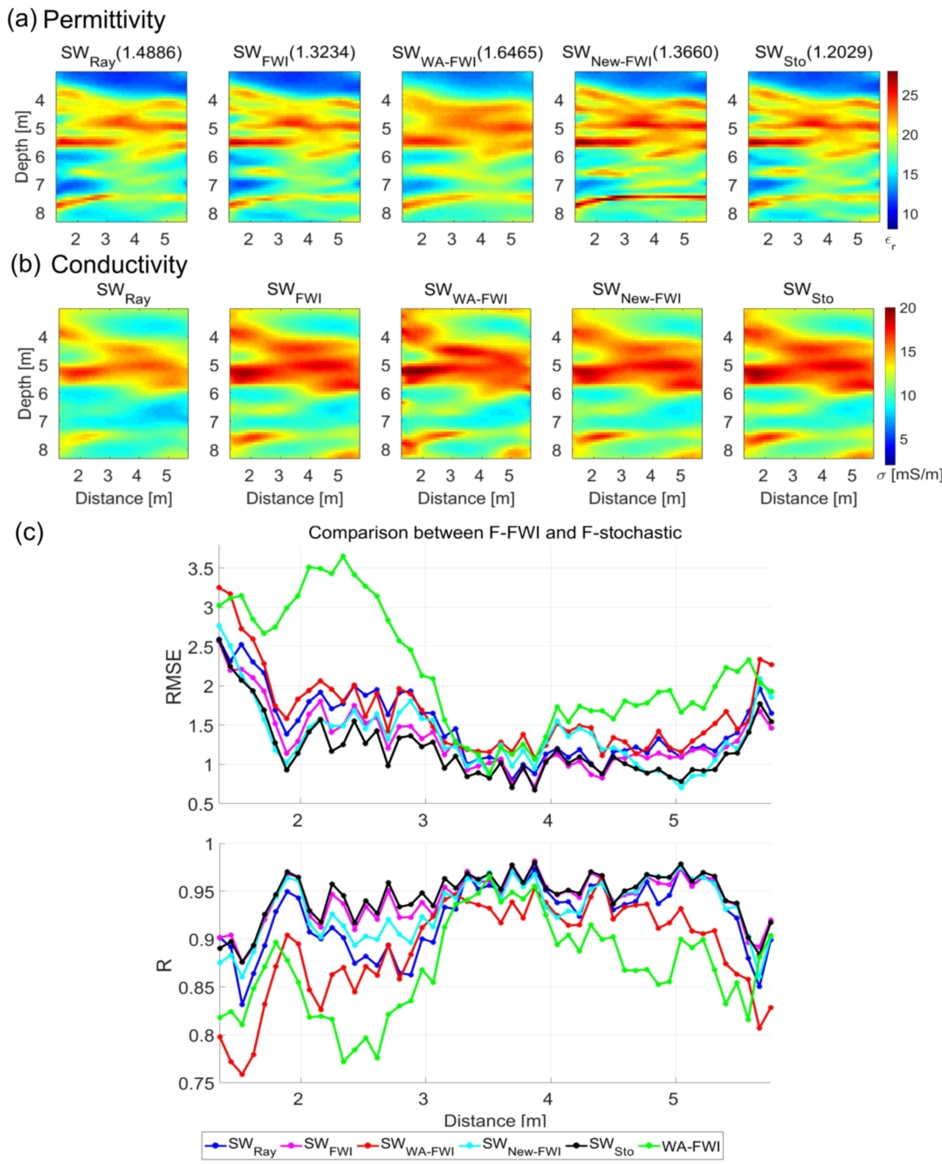

Figure 8. Comparisons of FWI (a) permittivity and (b) conductivity results using different effective source wavelets (Figure 7). Values in parentheses indicate the mean *RMSE* between filtered FWI permittivity models and the filtered stochastic permittivity model in the entire 2D domain (see Table 3 for more details). (c) Quantitative comparisons of the *RMSE* and *R* between filtered stochastic permittivity model and different filtered FWI permittivity results (same wavenumber as WA-FWI) along the vertical profile.

211x269mm (300 x 300 DPI)

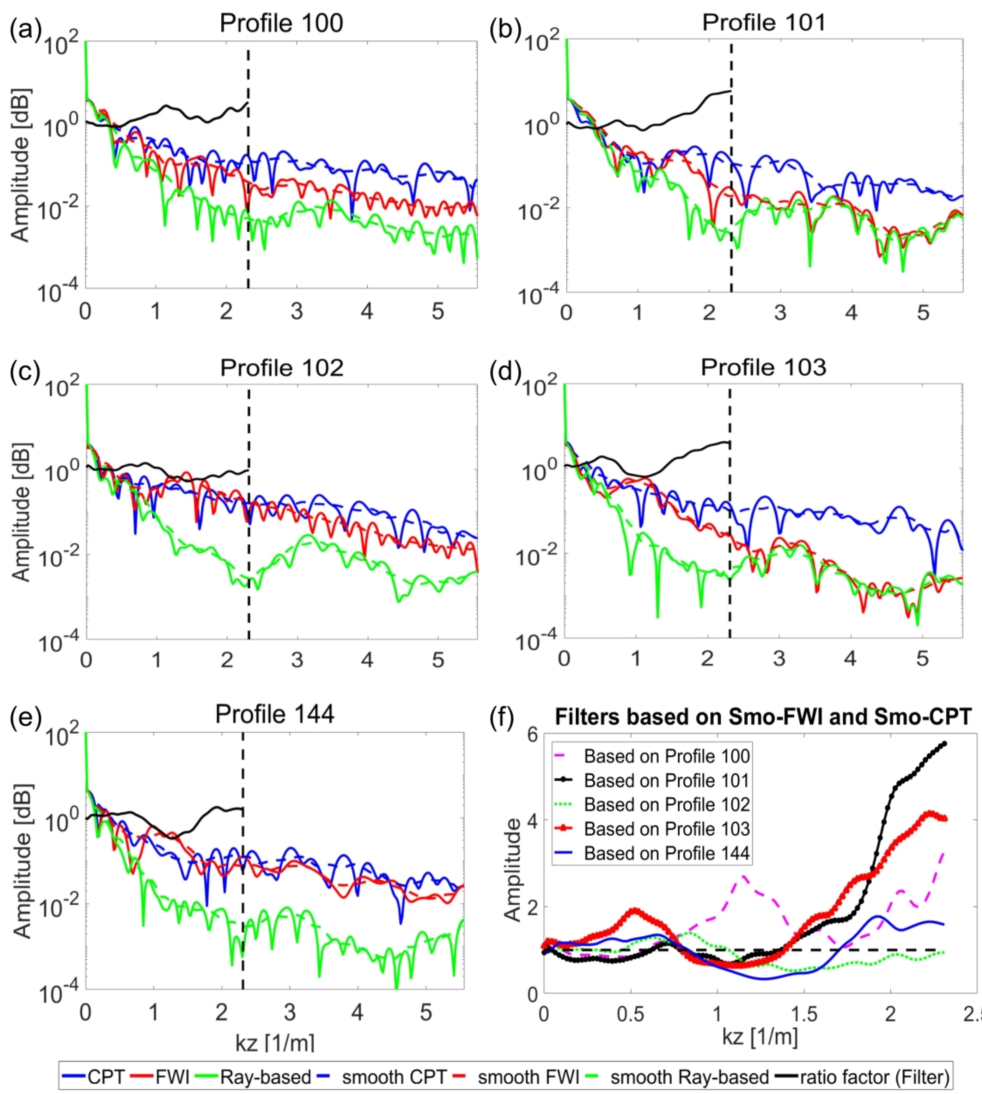

Figure 9. (a) To (e) comparisons of spatial frequency spectra of the CPT data (blue), the ray-based (green) and the conventional FWI (red) results in the wavenumber domain for different profiles (see Figure 2d for the locations of the profiles). The wavenumber filter is indicated by the black solid line for each profile. (f) Comparisons of the five filters, where a marked difference of profile 103 to the other profiles near 0.5 m^{-1} is noticeable.

207x238mm (300 x 300 DPI)

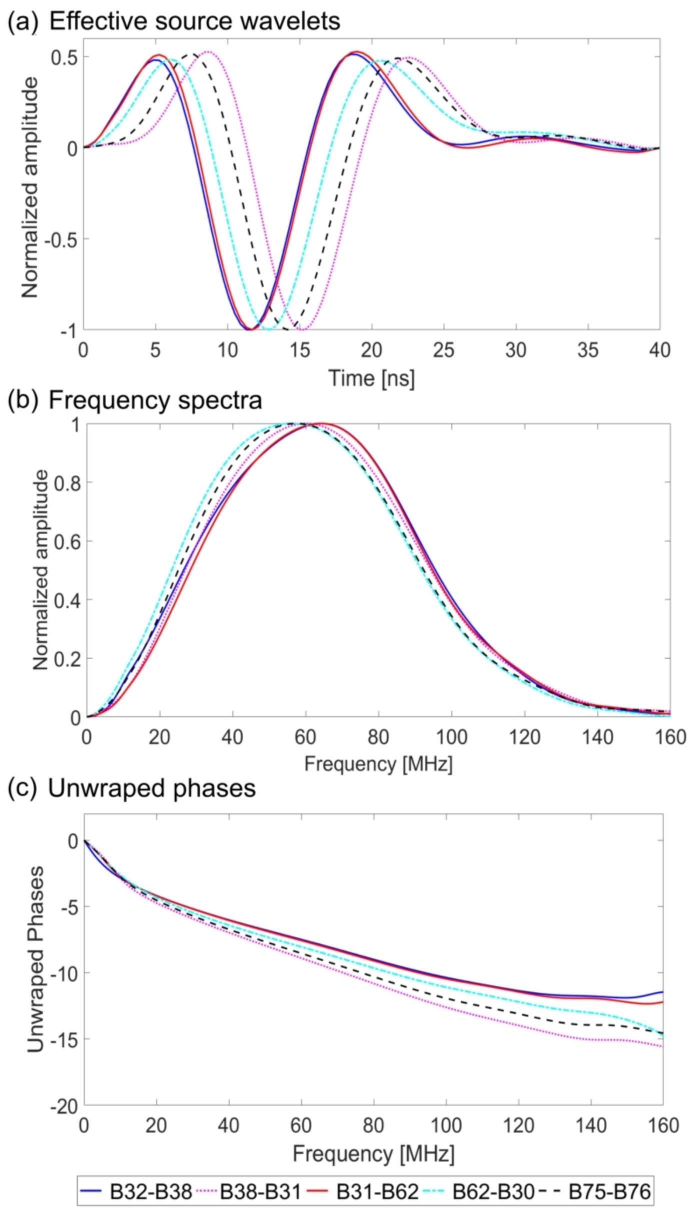

Figure 10. Comparisons of the updated effective source wavelets of the five cross-sections used for the experimental study in (a) time domain, (b) frequency and (c) phase spectra. Amplitudes of (a) and (b) are normalized to their corresponding minimum and maximum for a better comparison.

157x261mm (300 x 300 DPI)

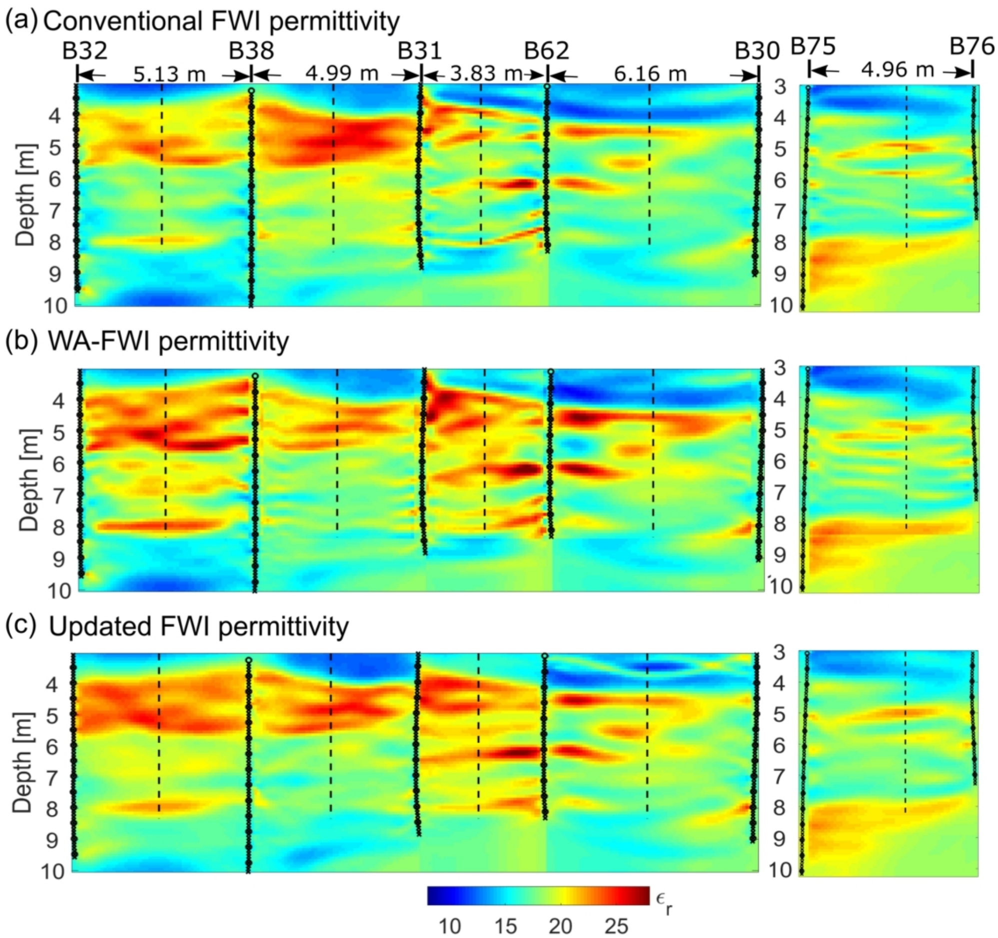

Figure 11. (a) Traditional permittivity FWI results using SW_{Ray} for the five cross-sections. Circles and crosses indicate the transmitter and receiver locations, respectively. Dashed lines present the locations of the CPT profiles. (b) Permittivity images of the wavenumber-amplified FWI using the filters shown in Figure 9. (c) Updated FWI results using the updated effective source wavelets as shown in Figure 10.

205x192mm (300 x 300 DPI)

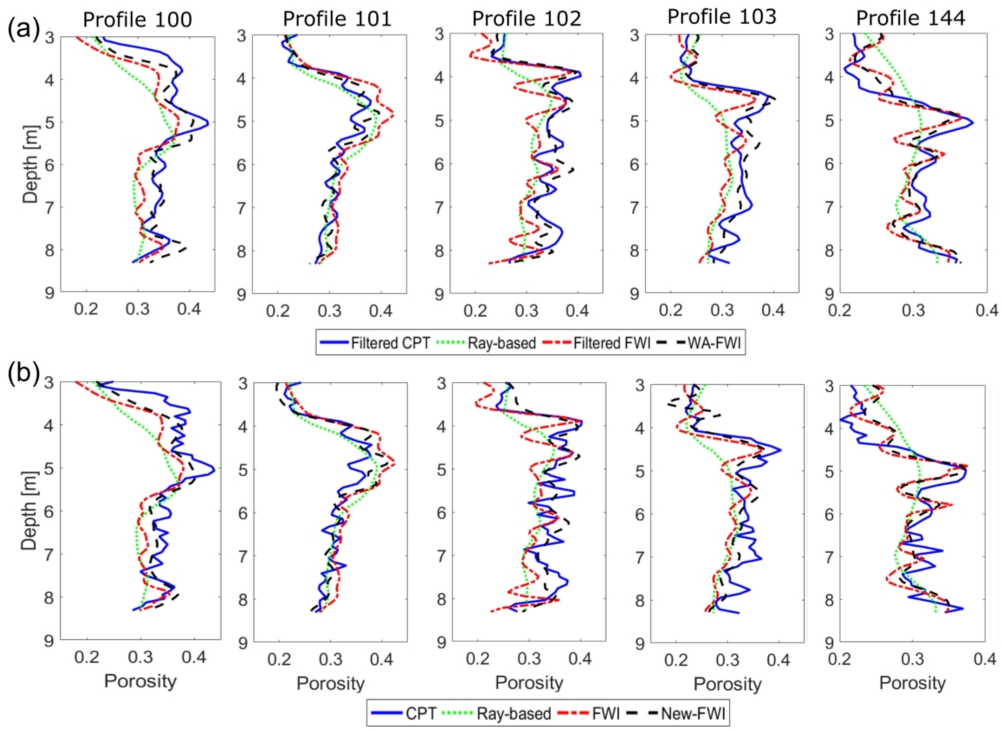

Figure 12. (a) Porosity comparisons of the filtered CPT (blue), the ray-based results (green), the filtered FWI results (red) and the wavenumber-amplified FWI (black) along each vertical profile. (b) Full wavenumber porosity results comparison of the CPT, ray-based, the FWI results (using SW_{Ray}) and the updated FWI results (using $SW_{New\text{-}FWI}$).

207x153mm (300 x 300 DPI)