Near Surface Geophysics

EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

Near Surface Geophysics, 2021 doi: 10.1002/nsg.12154

Improving crosshole ground-penetrating radar full-waveform inversion results by using progressively expanded bandwidths of the data

Zhen Zhou[1,2,3]*, Anja Klotzsche[1,2]* and Harry Vereecken[1,2]

[1]*Agrosphere (IBG-3), Institute of Bio- and Geosciences, Forschungszentrum Jülich GmbH, Jülich, Germany,* [2]*Centre for High-Performance Scientific Computing in Terrestrial Systems, HPSC TerrSys, Geoverbund ABC/J, Jülich, Germany, and* [3]*School of Information Engineering, ZheJiang A&F University, Hangzhou, China*

Received September 2020, revision accepted February 2021

ABSTRACT

In the last decade, time-domain crosshole ground-penetrating radar full-waveform inversion has been applied to several different test sites and has improved the resolution and reconstruction of subsurface properties. The full-waveform inversion requires several diligent executed pre-processing steps to guarantee a successful inversion and to minimize the risk of being trapped in a local minimum. Thereby, one important aspect is the starting models of the full-waveform inversion. Generally, adequate starting models need to fulfil the half-wavelength criterion, which means that the modelled data based on the starting models need to be within half of the wavelength of the measured data in the entire investigation area. Ray-based approaches can provide such starting models, but in the presence of high contrast layers, such results do not always fulfil this criterion and need to be improved and updated. Therefore, precise and detailed data processing and a good understanding of experimental ground-penetrating radar data are necessary to avoid erroneous full-waveform inversion results. Here, we introduce a new approach, which improves the starting model problem and is able to enhance the reconstruction of the subsurface medium properties. The new approach tames the non-linearity issue caused by high contrast complex media, by applying bandpass filters with different passband ranges during the inversion to the modelled and measured ground-penetrating radar data. Thereby, these bandpass filters are considered for a certain number of iterations and are progressively expanded to the selected maximum frequency bandwidth. The resulting permittivity full-waveform inversion model is applied to update the effective source wavelet and is used as an updated starting model in the full-waveform inversion with the full bandwidth data. This full-waveform inversion is able to enhance the reconstruction of the permittivity and electrical conductivity results in contrast to the standard full-waveform inversion results. The new approach has been applied and tested on two synthetic case studies and an experimental data set. The field data were additionally compared with cone penetration test data for validation.

Key words: 2D, Data processing, Ground-penetrating radar, Inversion.

*E-mails: z.zhou@fz-juelich.de and a.klotzsche@fz-juelich.de

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.
This is an open access article under the terms of the CreativeCommonsAttribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Detailed and high-resolution characterization of variable saturated soil-aquifer systems is critical and full of challenges, but highly important to improve the understandings of flow and transport processes of subsurface water. Small-scale soil heterogeneities can have a significant influence on the plant water uptake process, the transport process of pollutants or greenhouse gas emissions from peatland and permafrost soils (e.g., Simmer *et al*., 2015; Vereecken *et al*., 2016; Klotzsche *et al*., 2019a). In addition, the precision of large-scale atmospheric models is influenced by recharging and transport of subsurface water (Becker, 2006). To enhance our understandings of such processes, high-resolution images and accurately described subsurface properties, especially for small-scale heterogeneities of aquifers, are highly important (e.g., Klotzsche *et al*., 2019b). Thereby, some key parameters of geophysical and hydrological investigations, such as relative dielectric permittivity, electrical conductivity, hydraulic conductivity and porosity, play an important role (e.g., Binley *et al*., 2015). Traditional approaches to characterize aquifer properties either capture a small spatial sampling with a high vertical but poor lateral resolution, such as drillings and core sampling (Tillmann *et al*., 2008), or, have an average response over a large volume with a lack of detailed characterization at smaller scale, such as pumping test and remote sensing methods (Landon *et al*., 2001). In the last decades, geophysical methods have been developed and applied to obtain images of the near subsurface and to improve the characterization of hydrogeological properties (Binley *et al*., 2015), such as seismics (e.g., Doetsch *et al*., 2010), electrical resistivity tomography (ERT; e.g., Coscia *et al*., 2011) and ground-penetrating radar (GPR; e.g., Klotzsche *et al*., 2018). Especially, crosshole GPR has shown a great potential to characterize aquifers and has matured because of the opportunity to derive the highest possible resolution by using high-frequency electromagnetic (EM) pulses (e.g., Paz *et al*., 2017), and the opportunity to simultaneously derive electromagnetic wave velocities and amplitudes. The estimated EM velocities and attenuations can be transformed into relative dielectric permittivity ε_r and electrical conductivity σ (e.g., Ernst *et al*., 2007b; Dafflon *et al*., 2011), which are related for example to water and clay content, respectively (e.g., Kowalsky *et al*., 2005; Cassidy, 2007; Klotzsche *et al*., 2014). For crosshole applications, the EM pulses are emitted from a dipole-type antenna in a borehole and received by an antenna in another borehole. Within the scope of hydrogeological site characterization, several studies have shown the potential of crosshole GPR by comparing GPR permittivity results with independent measurements of soil water content (e.g., Looms *et al*., 2008; Kuroda *et al*., 2009). The good consistency illustrates crosshole GPR is able to close the gap between small-scale investigations with high-resolution (e.g., coring) and large-scale zones mapping (e.g., flowmeter tests).

Traditionally, ray-based approaches based on the geometrical ray theory are used for crosshole GPR data to derive tomographic images of the subsurface, which exploit first-arrival travel times and maximum first-circle amplitudes of the GPR wave (e.g., Maurer et al., 2004; Dafflon et al., 2012). For such approaches, damping and smoothing constraints are necessary to stabilize the inversion (e.g., Holliger *et al*., 2001; Maurer and Musil, 2004). Thereby, most applications consider only a limited angular coverage of the rays to avoid an increasing apparent velocity for increasing ray path angles (Peterson, 2001). Ray-based approaches are often not able to resolve targets smaller than the dominant wavelength and relatively smooth images are obtained, with a resolution that scales approximately with the diameter of the first Fresnel zone. In comparison to ray-based approaches, full-waveform inversion (FWI) exploits the entire information of the data (or significant parts of it), and is therefore able to provide higher resolution images within the sub-wavelength scale. Inspired by the FWI approach that was first developed and applied in the seismic community (e.g., Tarantola, 1984; Shin and Cha, 2008; Virieux and Operto, 2009), a 2D time-domain FWI for crosshole GPR data was implemented by Ernst *et al*. (2007a, b) and Kuroda *et al*. (2007), based on solving Maxwell's equations. Meles *et al*. (2010) improved the method of Ernst *et al*. (2007a) by including the vector characteristics of the EM fields, and introduced a simultaneous update of the permittivity and electrical conductivity parameters. Klotzsche *et al*. (2019b) provides an overview of the current developments, applications and corresponding challenges for the 2D crosshole time-domain GPR FWI. Several applications of the FWI to crosshole data sets from different environments, ranging from various aquifers to karst environments, demonstrated that the FWI was able to provide higher resolution images than the ray-based approaches, and provided images within decimetre scale.

Generally, FWI approaches can be implemented in the time and frequency domains. Both methods have pros and cons. While frequency approaches can highly minimize the calculation costs and improve the cycle skipping problem, the time-domain approaches provide the most flexible framework to apply time windowing of arbitrary geometries (Virieux and Operto, 2009). Similar to seismics, frequency-domain FWI approaches for GPR data have also been implemented. Such approaches allow, for example, to use only a few

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

discrete frequencies of the data and allow the possibility to implement a wider range of misfit functions, which are beneficial for frequency-dependent medium properties (e.g., Lavoué *et al.*, 2014). Ellefsen *et al.* (2011) inverted measured crosshole GPR data from a laboratory tank by using a 2.5D frequency-domain FWI approach. Furthermore, a 2D frequency-domain quasi-Newton approach for multioffset GPR data was implemented and tested on a synthetic model (Lavoué *et al.*, 2014), and on GPR data acquired at carbonate rocks (Pinard *et al.*, 2016). Although frequency-domain FWI seems to have certain benefits compared with time-domain FWI, for most experimental GPR data, low-frequency data are missing or show a low signal-to-noise ratio. Until now, almost all successful applications to experimental data have been performed using the time-domain FWI approach for GPR data (Klotzsche *et al.*, 2019b).

The 2D time-domain crosshole GPR FWI uses a conjugate-gradient algorithm (Polak et al., 1969) to minimize the misfit function between the measured and modelled data. This minimization can be achieved by updating the model parameters ε_r and σ by calculating updating directions and corresponding step lengths for both parameters according to the conjugate-gradient algorithm. It is well known that the non-linearity problem of the forward modelling is associated mainly with multiple scattering of propagating waves (Mora, 1987; Meles *et al.*, 2012). As with most of the inversion approaches, the FWI with conjugate-gradient method is also ill posed, which means that we can have data residuals of multiple models which fit the data equally well (e.g., Backus and Gilbert, 1968; Brittan *et al.*, 2013). In addition, the FWI results can be trapped in local minima if certain criteria are not fulfilled. One of the most important criteria is that the forward modelled data of the starting model are within half of a wavelength of the measured data in the entire inversion domain. Normally, ray-based starting models fulfil this criterion (Virieux and Operto, 2009). However, in the presence of specific subsurface structures such as high contrast layers related to, for example, the presence of the water table or small-scale heterogeneities, the ray-based starting models can be inaccurate and adaptations are required (e.g., Klotzsche *et al.*, 2012 and 2013). An experienced user and other tools, such as the amplitude analysis approach, are needed to ensure that the measured data are fully understood before starting the FWI and in the entire domain of interest the starting models fulfil the half-wavelength criterion (Klotzsche *et al.*, 2014; Zhou *et al.*, 2020).

To solve the problem of being trapped in a local minimum in the FWI, and to avoid a detailed pre-processing of the data, some solutions were proposed in microwave imaging and seismic inversion. For example, the frequency hopping method was used in frequency domain for microwave applications (Chew and Lin, 1995; Dubois *et al.*, 2009) and for seismic data (Pratt *et al.*, 1998; Maurer *et al.*, 2009). Thereby, the inversion is first performed with a low-frequency bandwidth. The results of this low-frequency inversion are then used as the next starting models for the FWI, with a progressively increased frequency bandwidth. These steps are repeated until the maximum boundary of the frequency bandwidth reaches the centre frequency of the data. One problem of this approach is that, for many field experimental data sets, not enough low-frequency data are present or that the low-frequency data are highly contaminated with noise. Therefore, several researchers tried to generate artificial low-frequency information for seismic data, for example, by using mathematical transformations based on the angle difference identity for cosine (e.g., Wang, *et al.*, 2019) or using the modulation signal approach introduced by Wu *et al.* (2014). To tame the non-linearity issue of GPR inversion, also for the time-domain FWI, Meles *et al.* (2012) presented an inversion scheme that incorporated a frequency-dependent effective source wavelet. Thereby, the modelled data are progressively bandwidth expanded (PBED) in an iterative process based on a stepwise increased frequency bandwidth of the source wavelet, while the observed data kept the full bandwidth. This approach worked very well for synthetic data, but for experimental data, the choice of the frequency bandwidths and the inversion parameters such as the perturbation factors hindered a successful application so far.

In this paper, we introduce a novel approach that continues the work of Meles *et al.* (2012), by improving the permittivity starting model and the effective source wavelet, based on progressively expanded bandwidths of the modelled and observed data (PEBDD). In contrast to the approach of Meles *et al.* (2012), we applied tapered bandpass filters to the effective source wavelet (used for the modelled data) and the observed data. The new FWI scheme can be divided into two steps: (1) construction of the new permittivity starting model by using PEBDD and (2) performing the FWI with the new starting model and an updated effective source wavelet using the full bandwidth data. To evaluate our new inversion scheme, we performed two synthetic case studies. One study is conducted using standard ray-based starting models and the second considers an enforced smaller starting ε_r model, which provides modelled data more than half a wavelength away from the measured data. After the validation of this new approach, we used the crosshole GPR FWI with PEBDD scheme to characterize the Krauthausen test site, using four inline crosshole GPR sections.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

METHODS

Ground-penetrating radar full-waveform inversion

Standard full-waveform inversion scheme

In this section, we only discuss the most important data processing and inversion steps of the standard full-waveform inversion (FWI) approach. For more detail, we refer the reader to Klotzsche *et al.* (2019b). One critical aspect to guarantee reliable and stable crosshole ground-penetrating radar (GPR) FWI results is to define adequate starting models. These starting models should return modelled data that are within half a wavelength ($\lambda/2$) of the measured traces in the entire inversion domain to avoid cycle clipping and trapping of the inversion in a local minimum of the misfit function (Meles *et al.*, 2012). In addition, an effective source wavelet needs to be estimated, because the source wavelet emitted into the earth system is unknown for experimental GPR data and accounts additionally for coupling effects. Such an effective source wavelet is estimated with two steps. First, an initial source wavelet is estimated using only horizontally travelling rays. The initial wavelet only considers the shape of the wavelet and the amplitude is normalized to one. This initial wavelet is in a second step updated using the deconvolution approach (e.g., Ernst *et al.*, 2007b; Klotzsche *et al.*, 2010) in the frequency domain, which can be described by:

$$\hat{\mathbf{G}}\ (f) = \hat{\mathbf{E}}^{\mathrm{syn}}(f)\left[\hat{s}_k\ (f) + \eta_D\right]^{-1}, \tag{1}$$

and

$$\hat{s}_{k+1}\ (f) = \left[\hat{\mathbf{G}}\ (f) + \eta_I\right]^{-1}\ \hat{\mathbf{E}}^{\mathrm{obs}}\ (f), \tag{2}$$

where s_k represents the initial source wavelet. $\mathbf{E}^{\mathrm{syn}}$ is the forward modelled data based on the ray-based starting models and s_k. $\mathbf{G}$ is Green's function and $\mathbf{E}^{\mathrm{obs}}$ is the observed GPR data. η_D and η_I are prewhitening factors that are applied to stabilize the solution and avoid dividing by zero. ˆ indicates frequency domain. s_{k+1} describes the updated effective source wavelet with an optimized phase and amplitude. If needed, this wavelet can be updated after a certain number of FWI iterations. Note that for most experimental data applications, the effective source wavelet is based on the ray-based models.

The forward modelling is based on a 2D finite-difference time-domain (FDTD) solution that solves Maxwell's equations in the Cartesian coordinates (Ernst *et al.*, 2007a; Meles *et al.*, 2010). In the process of the inversion, the misfit function $C\ (\varepsilon,\ \sigma)$ is defined by the differences between the observed data $\mathbf{E}^{\mathrm{obs}}$ and model-predicted data $\mathbf{E}^{\mathrm{syn}}(\varepsilon,\ \sigma)$:

$$C(\varepsilon,\sigma) = \frac{1}{2}\sum_s\sum_r\sum_\tau\left[\mathbf{E}^{\mathrm{syn}}\ (\varepsilon,\sigma) - \mathbf{E}^{\mathrm{obs}}\right]^T_{r,\tau} \delta\ (\mathbf{x}-\mathbf{x}_r, t-\tau)\left[\mathbf{E}^{\mathrm{syn}}\ (\varepsilon,\sigma) - \mathbf{E}^{\mathrm{obs}}\right]_{r,\tau}, \tag{3}$$

where s, r and τ are transmitters, receivers and the observation time, respectively. T denotes the transpose operator. Each of the fields is locally defined at any point of space $\mathbf{x}$ and time t. The multiplication with the Dirac delta δ function selects from the entire wavefield the used 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:

$$\begin{bmatrix}\nabla C_\varepsilon\ (\mathbf{x}')\\ \nabla C_\sigma\ (\mathbf{x}')\end{bmatrix} = \sum_s(\delta(\mathbf{x}-\mathbf{x}')\partial_t\mathbf{E}^{\mathrm{syn}})^T\widehat{\mathbf{G}}^T\mathbf{R}^{\mathrm{S}} (\delta(\mathbf{x}-\mathbf{x}')\mathbf{E}^{\mathrm{syn}})^T\widehat{\mathbf{G}}^T\mathbf{R}^{\mathrm{S}} \tag{4}$$

with

$$\mathbf{R}^{\mathrm{S}} = \sum_r\sum_\tau\delta\ (\mathbf{x}-\mathbf{x}_r,\ t-\tau)\ \left[\mathbf{E}^{\mathrm{syn}}\ (\varepsilon,\ \sigma) - \mathbf{E}^{\mathrm{obs}}\right]_{r,\tau} = \sum_r\sum_\tau\Delta\ [\mathbf{E}^{\mathrm{syn}}]_{r,\tau}\ . \tag{5}$$

where $\mathbf{R}^S$ can be interpreted as the back-propagated residual wavefield in the same medium as the incident wavefield $\mathbf{E}^{\mathrm{syn}}$. The spatial delta function $\delta\ (\mathbf{x}-\mathbf{x}')$ in equation (4) 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 two different step-lengths are considered in the process of estimating how far the permittivity and conductivity models need to be updated in the gradient direction. The medium parameters are updated to reduce the misfit function equation (3) by updating all gradient values in equations (4) and (5). In this paper, we employed the gradient normalization approach as proposed by van der Kruk *et al.* (2015) to minimize inversion artefacts close to the transmitter and receiver positions in the vicinity of the boreholes. As stopping criterion for the inversion, we consider that the root-mean-squared (RMS) error between observed and modelled data changes less than 0.5% between two subsequent iterations. Besides the stopping criterion, reliable FWI results also include the absence of a remaining gradient for the final models and a match between the measured and modelled data with a correlation coefficient greater than 0.8 (e.g., Klotzsche *et al.*, 2019b). For synthetic case studies, the mean absolute

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

error (MAE) in the 2D domain can be described as (Zhou *et al.*, 2021):

$$\mathrm{MAE} = \sum_{i=1}^{k}\sum_{j=1}^{m}(|\mathrm{RM}_{i,j} - \mathrm{FWI}_{i,j}|)/(k \times m), \tag{6}$$

where $\mathrm{RM}_{i,j}$ and $\mathrm{FWI}_{i,j}$ represent the input and the FWI model value located at the cell of i, j. k and m show the number of cells in the 2D domain along horizontal and vertical directions, respectively.

Novel progressively expanded bandwidths of the modelled and observed data full-waveform inversion scheme

Different from the approach of Meles *et al.* (2012) that is using the expanded bandwidths for the effective source wavelet only, we introduce a new scheme that applies the progressive increase bandwidths on both measured data and effective source wavelet. Our new updated approach proposes to tame the non-linearity issue of the time-domain FWI, which can be considered as an extension of the standard FWI procedure to improve the characterization of small-scale subsurface structures. Therefore, we apply the idea of a frequency-domain approach that uses longer wavelengths with lower frequencies in the beginning of the inversion to avoid the cycle skipping problem and that defines the bandpass filters according to the centre frequency of an effective source wavelet. First, we construct a series of bandpass filters with different high cut frequencies, while keeping the low cut frequency constant. Figure 1(a) shows an example for an effective source wavelet with a centre frequency of 65 MHz and a bandwidth of 12–140 MHz. For such a wavelet, we would select a low cut frequency of 12 MHz, while the maximum cut frequency is considered larger than the centre frequency of the effective source wavelet, which in our case is 68 MHz. To smooth the bandpass, we assigned two tapers with lengths of 12 MHz and 10 MHz for the starting and ending frequency points, respectively (Fig. 1a). These tapered bandpass filters are applied in the next step to the observed GPR data and the effective source wavelet, which results after applying the FDTD in sub-modelled data with a similar frequency spectrum as the sub-observed data (Fig. 1b, green box). The comparison of the effective source wavelets with the expanded bandwidths with the different bandpass filters shows a clear effect on the phase and amplitude of the wavelets (Fig. 2).

Second, the FWI with the progressively expanded bandwidths of the modelled data and observed data (PEBDD) is performed (Fig. 1b, green box loops 1 and 2). The first starting models for the FWI are based on the ray-based results $\mathrm{Model}_{\mathrm{Ray}}$ ('Ray (ε_r) and Homo (σ)' in Fig. 1b, green box). Traditionally, for the conductivity a homogeneous starting model is used based on the results of the first cycle amplitude inversion of the data (Klotzsche *et al.*, 2019b). Using the starting models with the sub-source wavelet and sub-observed data with the same smallest bandwidth, we perform a certain number of iterations of the FWI. Note that we chose for our studies five iterations for each bandwidth expansion n (Fig. 1b, green box loop 1). Several synthetic tests indicated that this choice is the best compromise between computational efficiency and stability of the inversion. The perturbation factors for the inversion, which are necessary to define the steplengths for the gradient approach, are kept the same as for the standard FWI. The ε_r and σ FWI results after this fifth iteration are considered as the next new starting models for the next sub-data with progressively expanded bandwidth, while the maximum cut frequency is increased by 4 MHz in our case (Fig. 1b, green box loop 2). These steps are repeated until the selected maximum cut frequency is reached (Fig. 1b green box loop 1: $n = n_{\max}$). Until this point all data (including source wavelet and observed data) used for the inversion are progressive bandwidth expanded.

Third, the FWI with the full bandwidth data (FBD) is calculated (Fig. 1b, red box). From these results, only the PEBDD permittivity FWI results with the maximum cut frequency are considered as new permittivity starting model in the next step (see, 'low-freq FWI (ε_r)' in Fig. 1b). Together with the conductivity starting model, which is the same as for the traditional FWI (Fig. 1b, 'Homo (σ)'), we construct the new starting models $\mathrm{Model}_{\mathrm{New}}$. Note that tests indicated that using the conductivity results with the maximum cut frequency as starting models does not improve the final results. The new starting models $\mathrm{Model}_{\mathrm{New}}$ and the traditional effective source wavelet $\mathrm{SW}_{\mathrm{Ray}}$ are used in the next step to generate an updated effective source wavelet $\mathrm{SW}_{\mathrm{Low}}$ with the full bandwidth data using the deconvolution approach (equation (2), Fig. 1b, red box). After obtaining the updated source wavelet $\mathrm{SW}_{\mathrm{Low}}$, an updated FWI is performed with the new starting models $\mathrm{Model}_{\mathrm{New}}$ and the wavelet $\mathrm{SW}_{\mathrm{Low}}$. Generally, a second-updated source wavelet $\mathrm{SW}_{\mathrm{New}}$ is necessary to further improve the FWI results. Thereby, the effective source wavelet $\mathrm{SW}_{\mathrm{Low}}$ is updated to $\mathrm{SW}_{\mathrm{New}}$ using the deconvolution method with the wavelet $\mathrm{SW}_{\mathrm{Low}}$ and the permittivity model from the updated FWI results with $\mathrm{SW}_{\mathrm{Low}}$. The final updated FWI results with FBD are solved based on the second-updated source wavelet $\mathrm{SW}_{\mathrm{New}}$ and the new starting models $\mathrm{Model}_{\mathrm{New}}$ (details can be found in Fig. 1b).

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

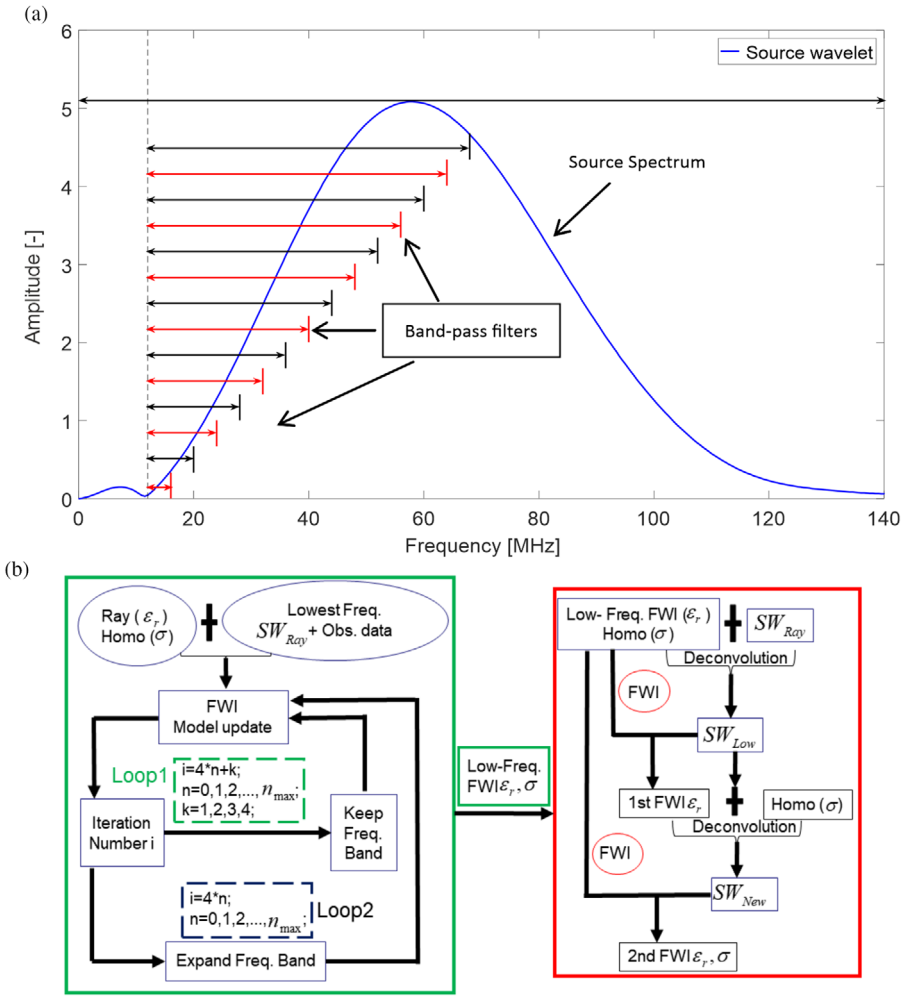

Figure 1 (a) Example for an effective source wavelet amplitude spectrum and corresponding stepwise increased bandpass filters (shown schematically by the black and red horizontal bars). These filters are progressively expanded after the centre frequency is reached and applied to the observed data and the effective source wavelet. The bandpass filters start at the lowest frequency of 12 MHz (taper length 12 MHz) indicated by a vertical dashed line. The high cut frequency is stepwise expanded every five iterations of the FWI and has a taper of 10 MHz. After the highest corner frequency of 68 MHz is reached (centre frequency of 65 MHz), all subsequent iterations use the full bandwidth of an updated source wavelet and the observed data. (b) Flow chart of the new proposed approach (adapted from Meles *et al.*, 2012). In the first part (green box) the progressively expanded effective source wavelet and the observed data are used, in which i, n and k represent all iterations using all sub-data, the numbers of bandwidth expansions, and individual iterations of each group sub-data, respectively. n_{max} is depending on the centre frequency of the effective source wavelet. In this case, we select $n_{max}=$ 13. In the second part (red box), the full bandwidth data (FBD) FWI is performed, where two effective source wavelet corrections are performed and used during FWI. The final result is indicated by the second FWI with the SW_{New} as an effective source wavelet.

RESULTS AND DISCUSSION

Synthetic case studies

Synthetic study I: Stochastic input models and ray-based starting models

To verify the aforementioned new full-waveform inversion (FWI) scheme for improving the ground-penetrating radar (GPR) FWI results, we construct realistic synthetic models of relative dielectric permittivity and electrical conductivity using a stochastic simulation (sequential Gaussian simulation). For the simulation, existing data set of the Krauthausen aquifer test site in Germany is used to generate a facies model with certain hydrological and geophysical parameters (based on Tillmann *et al* 2008, and more details can be found in Zhou *et al* 2021). The construed aquifer consisted of a three-layered

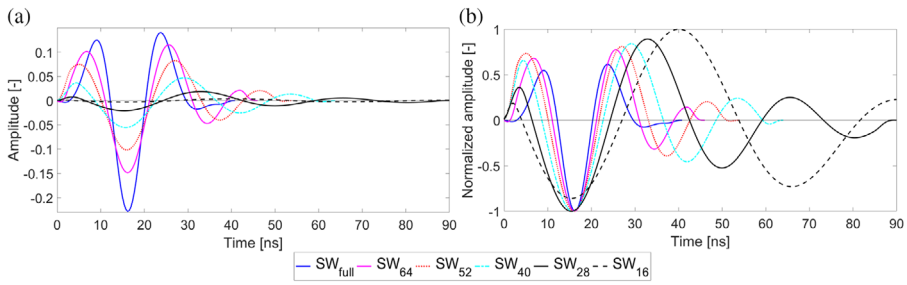

Figure 2 Comparisons of the selected sub-source wavelets with the different bandpass filters of the new PBEDD approach used for synthetic case study I in time domain. (a) Shows the source wavelets with the absolute amplitudes and (b) the corresponding wavelets normalized to the corresponding minimum (or maximum) for a better comparison of the phase.

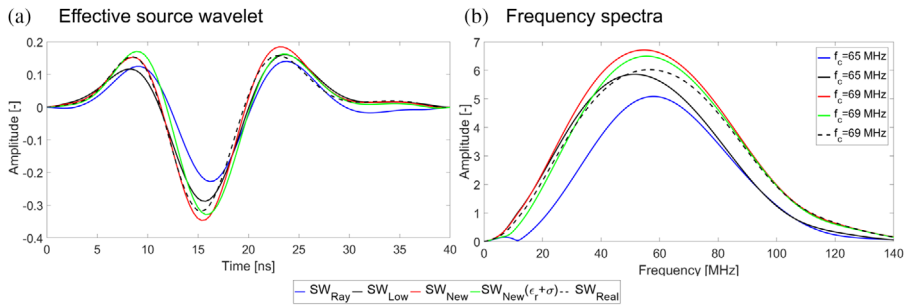

Figure 3 Comparisons of the (a) effective source wavelets and the (b) corresponding frequency spectra of the different steps of the new PBEDD approach used for synthetic case study I. The legend values in (b) indicate centre frequencies for different source wavelets. The source wavelet used to generate the 'observed data' is indicated with a dashed black line. Note that SW_{New} ($\varepsilon_r + \sigma$) is the effective source wavelet based on the results shown in the Appendix.

structure similar to the Krauthausen aquifer: sandy layer between 1.0 m and 4.0 m, sandy gravel between 4.0 m and 5.5 m and below coarse gravel (left side in Fig. 4 and Fig. 8). For the unsaturated zone above water table at 2.0 m, we chose a homogeneous layer with a relative permittivity of $\varepsilon_r = 4.4$ (not shown, same for all following inversions). To generate the realistic synthetic GPR data (called observed data), we used a source wavelet (SW_{Real}; dashed curves in Fig. 3a and b) based on available experimental data from the cross-section B38-31 of the Krauthausen test site (Gueting *et al.*, 2015). Similar to the acquisition of experimental data, a semi-reciprocal acquisition set-up was used with transmitter and receiver spacing of 0.5 m and 0.1 m, respectively. The realistic synthetic crosshole GPR data are modelled with the same time-domain 2D FDTD approach as used for the FWI. Similar to experimental data applications, we defined the ε_r starting model for the FWI by picking the first-arrival times of the synthetic data and performed a ray-based inversion (first column of Fig. 4a; e.g., Maurer and Musil, 2004). Similar to previous studies, we choose for a homogeneous σ starting model of 13 mS/m (not shown) that was defined by testing various different homogeneous models and used the results of the first cycle amplitude inversion of the experimental data.

To keep it as realistic as for experimental data applications, we used the ray-based starting models to estimate an effective source wavelet (SW_{Ray}; blue curves in Fig. 3a and b) and performed the standard FWI (first column of Fig. 4b and c). As expected, the FWI results show higher resolution images as the ray-based results within decimetre-scale resolution. Because of the known input models, we can calculate the mean absolute error (MAE) according to equation (6) between the resolved FWI models and the true input models based on the stochastic simulation (shown in the first column titles of Fig. 4d and e). The final results were obtained after 28 iterations, and the inverted results fulfilled the criteria for a reliable inversion. The corresponding RMS curve that is calculated between the observed and modelled radar data is indicated by the blue curve in Fig. 5 with the final RMS of 6.9945×10^{-7}. Generally, most of the distinct features of the stochastic input models are resolved, and a good fit between the observed and the modelled data is obtained (see, Fig. 6a and b). By comparing the differences between observed data and modelled data for one

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

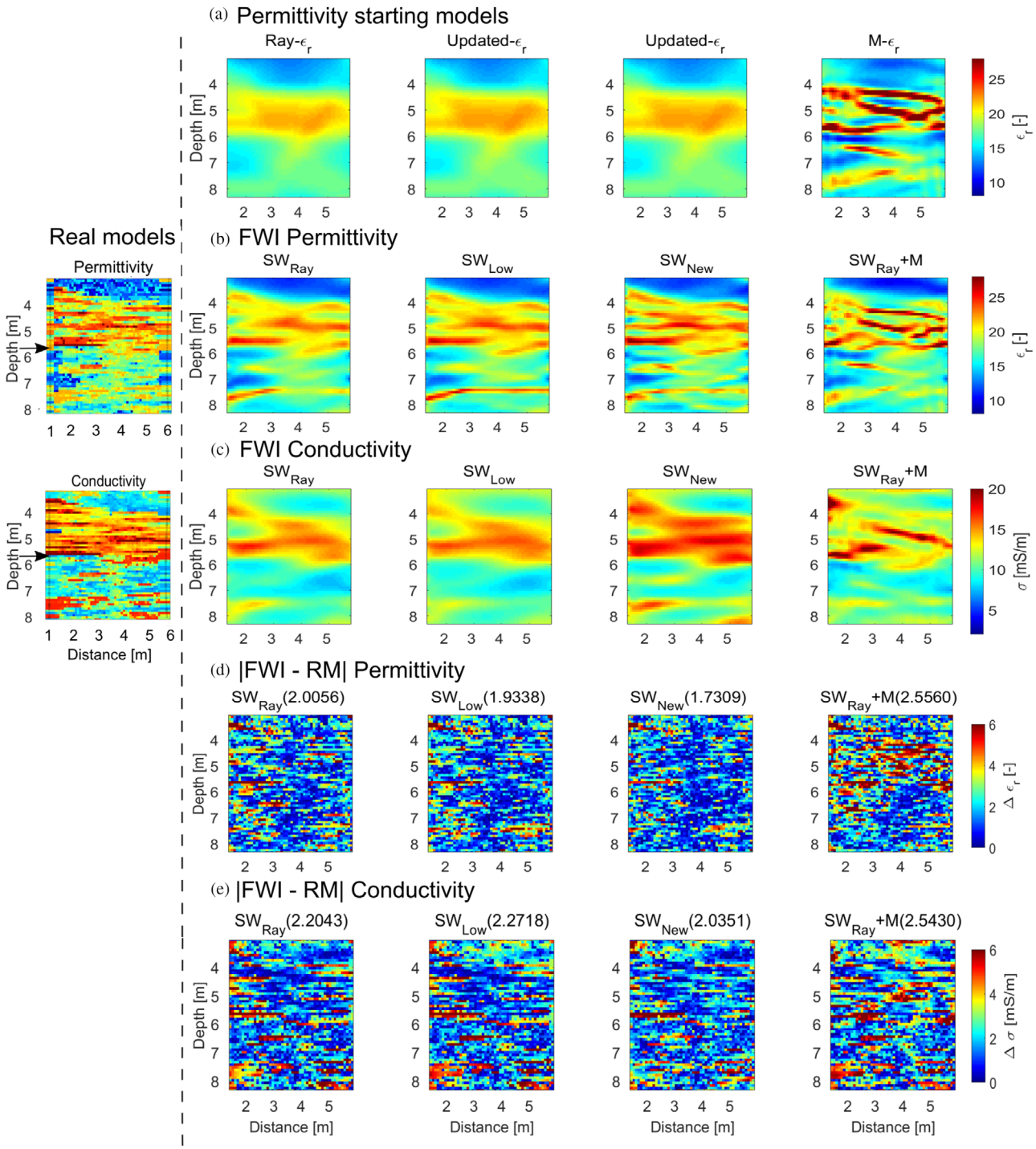

Figure 4 Overview of the FWI results using the different approaches and starting models. Left, the real input models based on the stochastic simulation (Zhou *et al.*, 2021). (a) Permittivity starting models and corresponding FWI results for (b) ε_r and (c) σ for the different FWI approaches. The applied effective source wavelets are named in the legends. Image plots of the absolute error between the real input models and final FWI results are shown for (d) ε_r and (e) σ for the different approaches. The mean absolute errors of the entire 2D domain are shown in parentheses.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

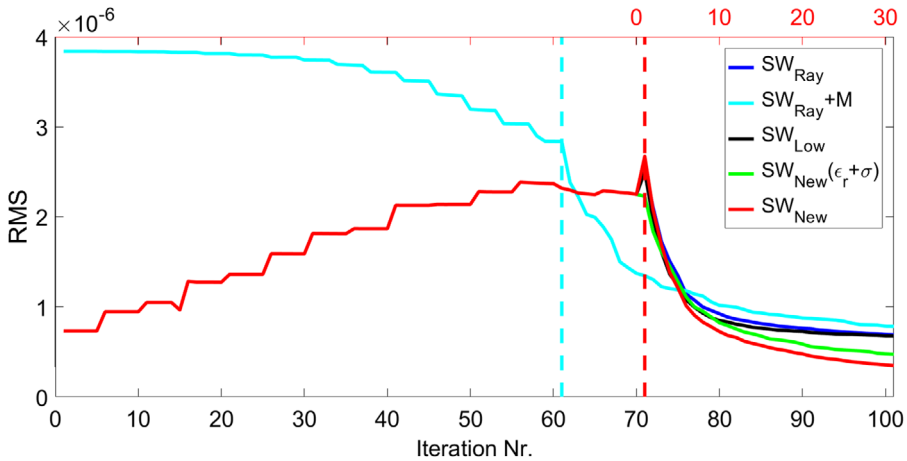

Figure 5 RMS misfit curves of the different FWI approaches of the synthetic case study I. Blue graph represents the RMS of the standard FWI using the ray-based starting models (iteration numbers are shown at the top with a red label). The cyan graph represents the RMS behaviour for the PBED inversion scheme based on Meles *et al.* (2012). Black and red graphs represent the new PEBDD inversion scheme using the first updated and second updated source wavelets, respectively. The green graph shows the new FWI RMS according to the updated starting models including ε_r and σ (Appendix) by using PEBDD inversion scheme. Note that on the left side of the dashed lines, the progressive expansion of the bandwidths of effective source wavelets and observed data was used, while on the right side the FWI was performed considering the full bandwidth of all data. New PEBDD inversion RMS values are increased before 70 iterations (red and black lines), because of the larger amplitude values of the source wavelet and the observed data.

exemplary data set (Fig. 6c), we find most of the regions in the tomograms a small misfit is visible, but for some domains a increased difference can still be noticed. Considering that the real effective source wavelet has a centre frequency of 69 MHz and the stochastic model has an approximate average permittivity value of 18, the corresponding wavelength of the GPR signal is 1.03 m. Comparing the model cell size of 0.09 m with the wavelength 1.03 m, it is hard to match all the features of the stochastic model especially when the contrast is relatively high.

Following our new introduced approach of the progressively expanded bandwidths of the modelled and observed data (PEBDD), we applied the different bandpass filters as shown in Fig. 1(a) to the standard effective source wavelet SW_{Ray} (different effective source wavelets shown in Fig. 2) and the observed data. For this study, we choose the optimal five iterations for the sub-data FWI by comparing with other iterations values. The first sub-data FWI started from the filtered sub-source wavelet and filtered sub-observed data with bandwidth 12–16 MHz and the ray-based starting models $Model_{Ray}$. Every five iterations the bandpass is increased by 4 MHz until the final cut-off frequency of 68 MHz is reached. The final permittivity FWI results using the maximum bandwidth sub-data are shown in Fig. 4a (title is 'Updated – ε_r'). Combining the conductivity starting model with a homogeneous value of 13 mS/m, we construct the new starting models $Model_{New}$ for the following FWI results and the updated source wavelet. Following the flowchart (Fig. 1b, red box), the effective source wavelet SW_{Low} is updated using the new starting models $Model_{New}$ and the wavelet SW_{Ray}. The effective source wavelet SW_{Low} together with the starting models $Model_{New}$ is used to calculate the FWI as shown in Fig. 4 (a and b) (second column) using the full data set. By comparing the RMS curve behaviour of this inversion (black graph in Fig. 5; black graph is covered by red graph before 71 iterations), we can notice that first the RMS is stepwise increased until the full data are used and the RMS curve with the full data are decreased after 23 iterations to a final value of 6.9749 $\times$ 10^{-7}. The misfits between the input and resolved tomograms (Fig. 4d and e) and the data misfit (Fig. 6c) are slightly better than using the standard FWI approach.

These updated FWI results are in the last step considered to update the effective source wavelet a last time (SW_{New} in Fig. 3a and b, red source wavelet) by using the new FWI ε_r results with SW_{Low} in the deconvolution approach. The corresponding final FWI results (Fig. 4, third column) are derived with the SW_{New} and $Model_{New}$ as the starting models. Note that we also tested the new approach with the both permittivity and conductivity starting models based on the maximum bandwidth sub-data (Fig. A.1a). The obtained FWI results (Fig. A.1b) indicate the results with SW_{New} ($\varepsilon_r + \sigma$) did not improve significantly than the FWI results with the SW_{New}

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers., *Near Surface Geophysics*, 1–23

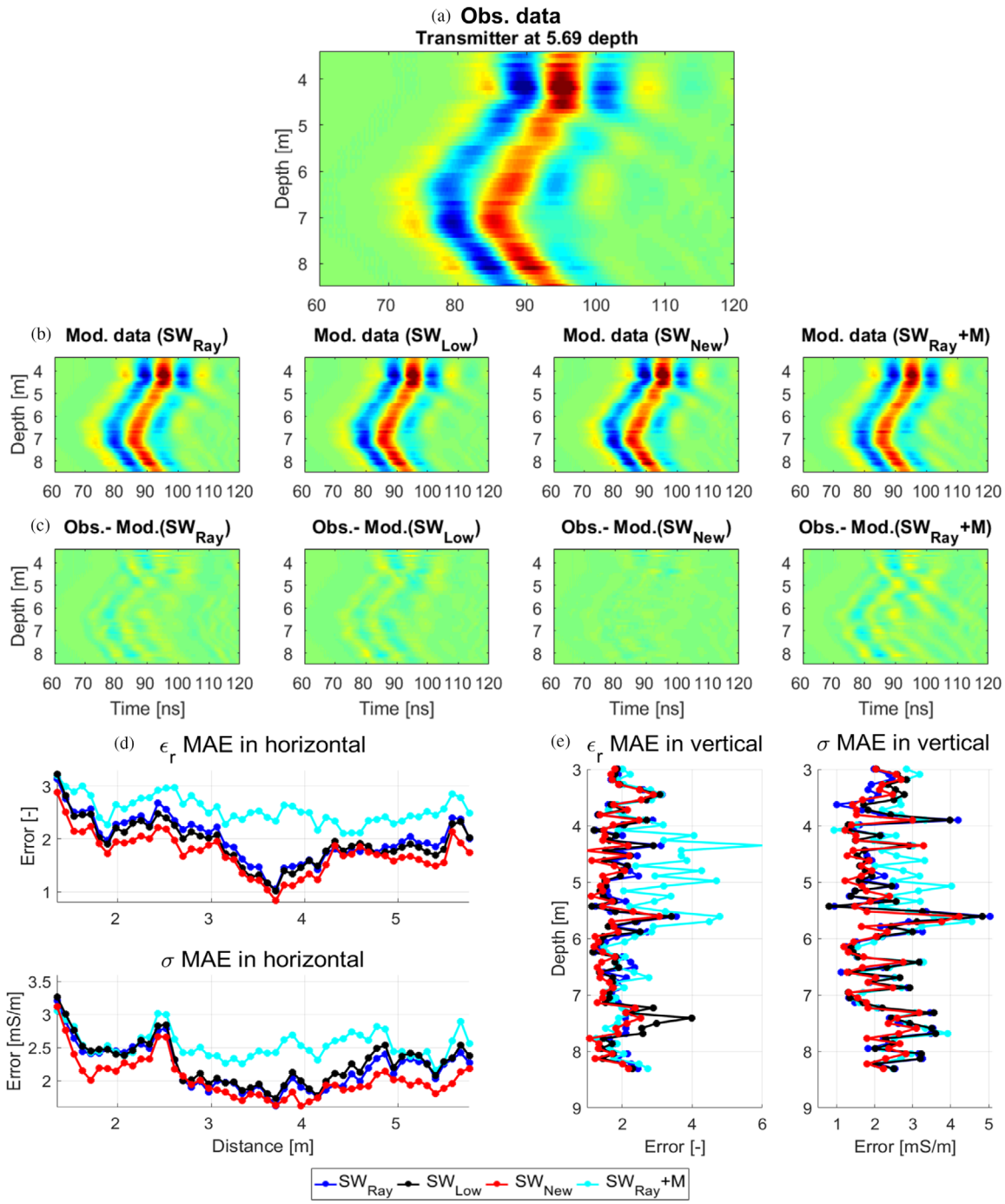

Figure 6 (a) Observed data, (b) modelled data based on the FWI results and (c) differences between the observed and modelled data for one exemplary data set of transmitter location at 5.69 m depth (see input models and black arrow for the transmitter location in Fig. 4). Note the amplitudes in (a–c) are normalized to the maxima amplitude of the real observed data (shown range from -7×10^{-1} to 7×10^{-1}). The mean absolute errors of the permittivity and conductivity between the real input and the different FWI results are shown along (d) the horizontal cross-section and (e) vertical direction.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

and $\text{Model}_{\text{New}}$ as the starting models; therefore, we consider in further applications only the permittivity updated starting model. For comparisons, we performed the approach of Meles *et al.* (2012) that only uses the sub-source wavelets in the FWI, which means using the progressively bandwidth expanded modelled data (PBED), while the observed data include the full bandwidth (Fig. 4, fourth column). Note that the generated FWI results (including ε_r and σ) with the maximum bandwidth sub-source wavelet are the new starting models (only show the ε_r starting model with the title $M - \varepsilon_r$ in Fig. 4a) for the following FWI with full bandwidth data.

The FWI results using the SW_{New} show the smallest misfit of the resolved tomograms and the smallest final RMS, indicating that this approach is able to resolve the input tomograms best. Note that we are able to improve the FWI results for the permittivity and the conductivity by 13.7% and 7.7% in comparison to the standard FWI models using SW_{Ray} by comparing the MAE values of the two approaches, respectively (Table 1). Especially, the small-scale structures close to the boreholes are clearer and more accurately resolved as by the standard method. By comparing the final RMS curves (only show the max iterations until 101 or 31; the same for Fig. 9) of the five different FWI results (green graph based on results shown in the Appendix), the second-updated FWI results with the wavelet SW_{New} indicate the best convergence behaviour resulting in the smallest residual value of 2.8996 $\times 10^{-7}$ after 50 iterations (red curve in Fig. 5). Comparisons of four different FWI ε_r and σ results in 2D domain (Fig. 4b and c) show the FWI results with the approach of Meles *et al.* (2012) are unrealistic. By computing the absolute errors between these different FWI results and the real input models in 2D domain (Fig. 4d and e), we can find that the second-updated FWI results with SW_{New} are close to the input models indicated by the smallest MAE values (Table 1). By investigating the fit between the observed and the FWI modelled GPR data, generally a good fit can be observed for all the FWI results (one example shown for transmitter depth 5.69 m in Fig. 6a and b), but analysing the differences (Fig. 6c) in more detail, we notice that the FWI using SW_{New} shows the smallest misfit.

Table 1 Comparison of the different FWI approaches for synthetic case study I using the mean absolute error between the real input models and the different FWI results for the entire 2D domain, and the root-mean-squared RMS error between observed and modelled radar traces represents residual values. Percentages in parentheses indicate the ratio of the single FWI RMS to the standard FWI RMS with SW_{Ray}, while a decreased value means the higher improvement efficiency

Real models	MAE (ε_r)	MAE (σ)	RMS (10^{-7})
FWI (SW_{Ray})	2.0056	2.2043	6.9945 (100%)
FWI (SW_{Low})	1.9338	2.2718	6.9749 (99.7%)
FWI (SW_{New})	1.7309	2.0351	2.8996 (41.5%)
FWI ($\text{SW}_{\text{Ray}} + M$)	2.5560	2.5430	8.6499 (123.7%)
FWI ($\text{SW}_{\text{New}}(\varepsilon_r + \sigma)$)	1.8500	2.1851	3.8620 (55.2%)

In order to describe the regional differences of different FWI models, we computed the MAE between the input stochastic models and the FWI model results along the

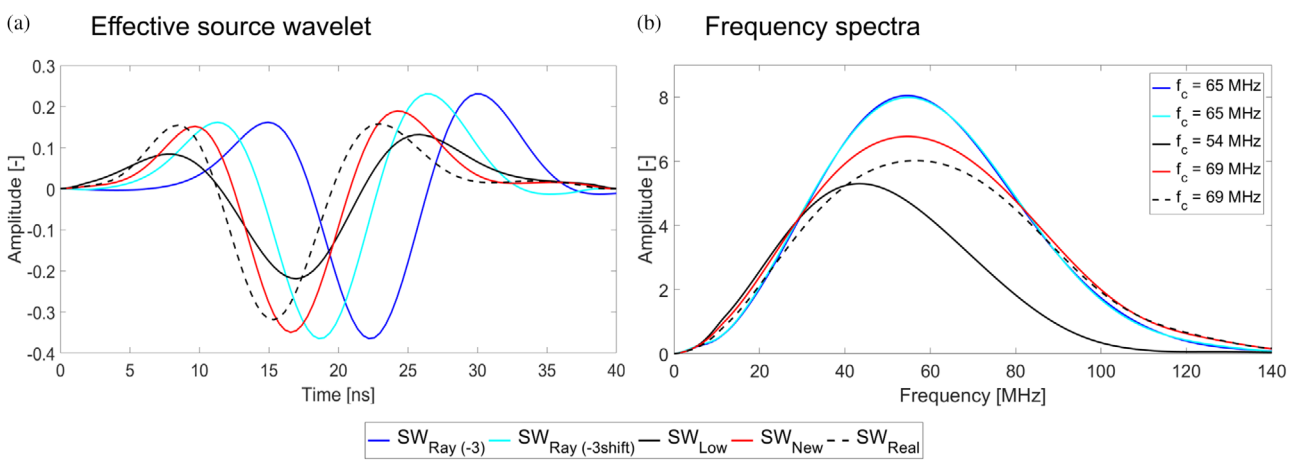

Figure 7 (a) Comparisons of the different effective source wavelets used for synthetic case study II. The source wavelet used to generate the 'observed data' is indicated with a dashed black graph. (b) Frequency spectra comparisons from 0 to 140 MHz for the five different effective source wavelets. The legend values in (b) indicate these effective centre frequencies for different steps.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

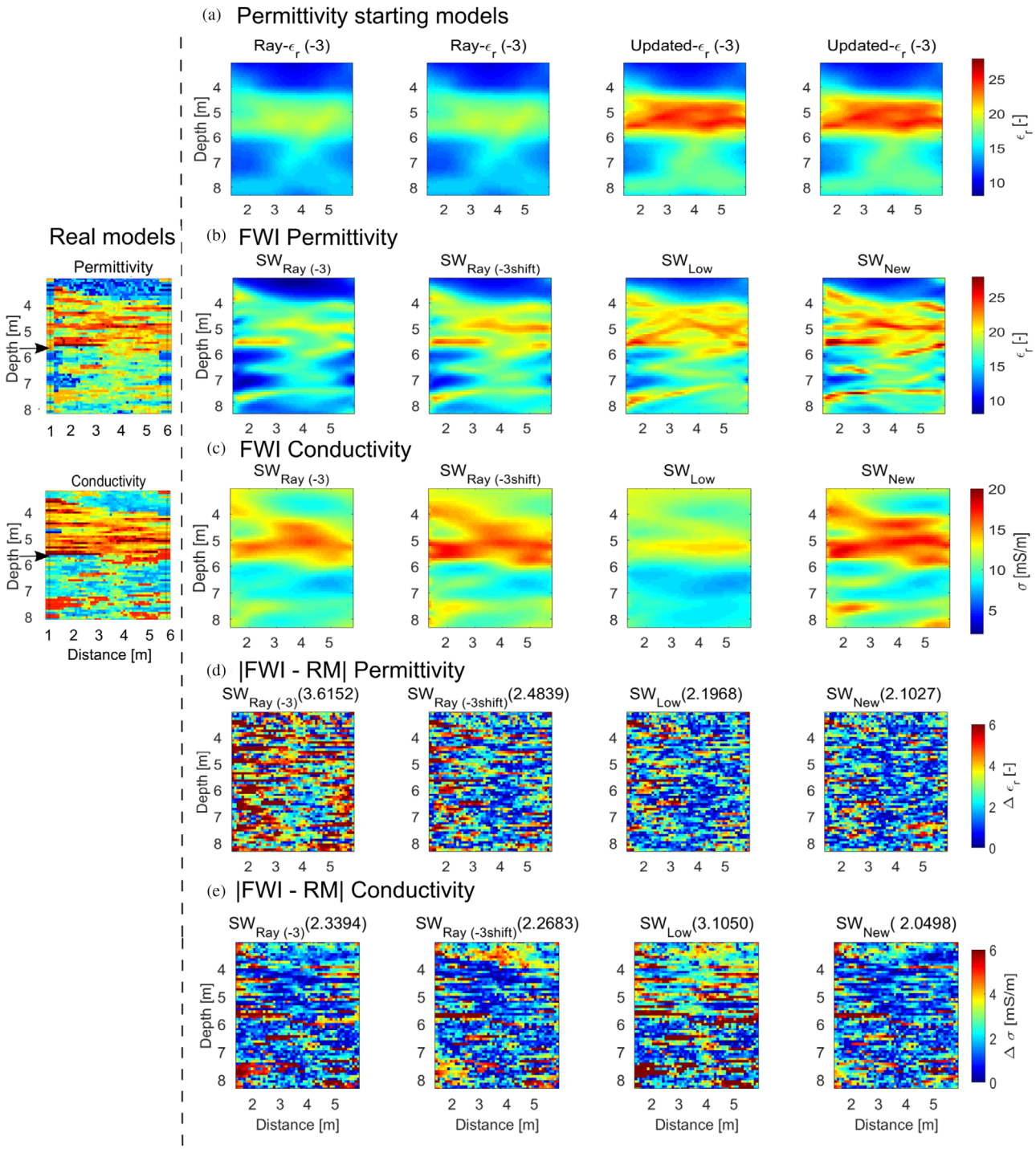

Figure 8 Overview of the FWI results using different approaches. Left, real input models based on the stochastic models. (a) Permittivity starting models and corresponding FWI results for (b) ε_r and (c) σ for the different FWI approaches. The applied effective source wavelets are named in titles of the Fig. 8. Image plots of the absolute error between the input models and final FWI results for (d) ε_r and (e) σ for the different approaches. The mean absolute errors of the entire 2D domain are shown in parentheses.

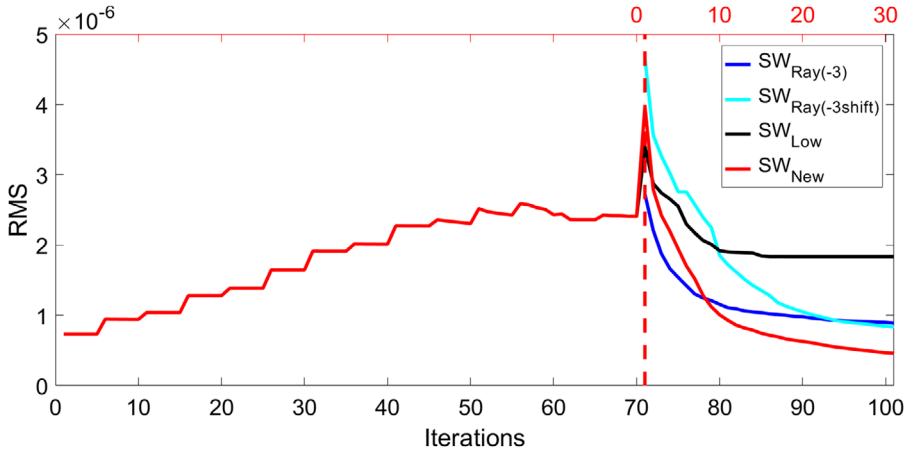

Figure 9 RMS misfit curves of the different FWIs for synthetic case study II. The blue and cyan graphs represent the FWI RMS convergence behaviours using the source wavelets $SW_{Ray\ (-3)}$ and $SW_{Ray\ (-3\ shift)}$ based on enforcing a smaller ε_r starting model (iterations are from 1 to 31 at red label on the top), respectively. Black and red graphs represent the new PEBDD inversion scheme using the first updated and second updated effective source wavelets, respectively. Note that before the dashed lines, the progressively expanded of the bandwidths of source wavelets and observed data are used, while on the right side the FWI with the full bandwidth of all data are performed.

Table 2 Comparison of the different FWI approaches of the mean absolute error between the real input models and the different FWI models for the entire 2D domain for synthetic case study II. RMS represents residual values between observed and modelled radar traces. Percentages in parentheses indicate the ratio of the other FWI RMS to the standard FWI RMS with $SW_{Ray\ (-3)}$, the lower value means the higher improvement efficiency

Real models	MAE (ε_r)	MAE (σ)	RMS (10^{-7})
FWI ($SW_{Ray\ (-3)}$)	3.6152	2.3394	9.1843 (100%)
FWI ($SW_{Ray\ (-3shift)}$)	2.4839	2.2683	8.3326 (90.7%)
FWI (SW_{Low})	2.1968	3.1050	18.366 (200.0%)
FWI (SW_{New})	2.1027	2.0498	3.6288 (39.5%)

horizontal direction (Fig. 6d) and the vertical direction (Fig. 6e) for ε_r and σ, respectively. The smallest MAE of the horizontal direction can be observed in the central part of the tomograms between 3.0 m and 5.0 m (Fig. 6d). For all results, the MAE of the horizontal direction increases towards the boundaries of the inversions domain, meaning the boreholes of the cross-section (Fig. 6d). The MAE in the horizontal direction and an increase in MAE towards the boreholes are related to the acquisition geometry and the related distribution of the ray coverage between the boreholes. Oberröhrmann *et al.* (2013) showed already that the resolution is highly affected by the acquisition geometry and depends on the ray coverage. The MAE along the vertical direction shows more fluctuations around 2, while the results in horizontal direction are smoother. Interestingly, all the FWI results (except the approach based on Meles *et al.*, 2012) show improved results with a smaller MAE in the vertical direction between 4.0 m and 5.5 m depth in comparison to other depths, where the small-scale structures are located. A small increase of the MAE at 5.5 m depth can be noticed at the lower boundary of the small-scale high permittivity zone. For both horizontal and vertical directions, the FWI results based on the SW_{New} show the lowest MAE, in contrast to the other FWI results. Comparing the FWI results and the related MAE, we can conclude that our newly updated PEBDD scheme is effective to enhance the complicated synthetic models' FWI results for permittivity and conductivity, in contrast to the standard FWI approach.

Synthetic case study II: Permittivity starting model beyond the half-wavelength criteria

In the presence of high contrast layers, ray-based results are often erroneous and need to be updated to fulfill the half-wavelength starting model criteria for the standard FWI. Here, we want to demonstrate the potential of the PEBDD scheme, which also allows starting models that are beyond the half-wavelength criteria. Therefore, we perform a second synthetic case study with the same stochastic input models as before and enforce the permittivity starting model to provide modelled data more than half a wavelength away from the measured data by using a smaller ε_r model. The changed ε_r starting model was obtained by subtracting three from the normal ray-based ε_r model in the entire domain (Fig. 8a, title is 'Ray $-\varepsilon_r$ (-3)'). The conductivity starting model is unchanged with a homogeneous value of 13 mS/m. We follow the same

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers., *Near Surface Geophysics*, 1–23

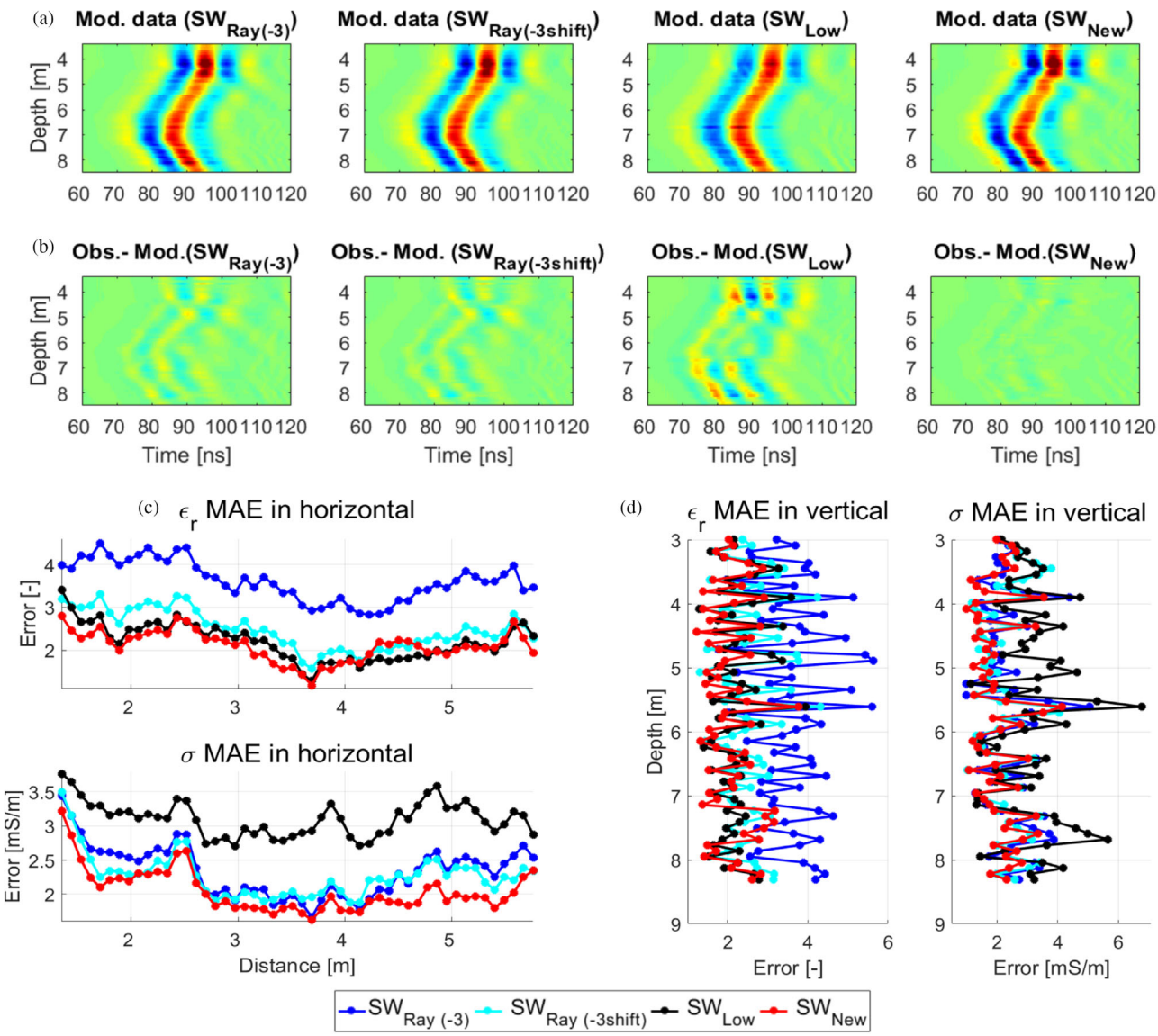

Figure 10 (a) Modelled data based on the FWI results and (b) differences between the observed and modelled data for the transmitter at 5.69 m depth (see input models and black arrow for the transmitter location in Fig. 8). Note the amplitudes in (a) and (b) are normalized to the maxima amplitude of the real observed data (shown range from -7×10^{-1} to 7×10^{-1}). The mean absolute errors of the permittivity and conductivity between the input models and the different FWI results are shown along (c) horizontal cross-section and (d) vertical direction.

approach as in the synthetic study case I and obtained the different updated effective source wavelets (Fig. 7) and FWI results (Fig. 8).

First, we estimate the effective source wavelet $SW_{Ray\ (-3)}$ based on the erroneous starting models $Model_{Ray-3}$. We can clearly notice that the $SW_{Ray\ (-3)}$ based on the erroneous starting models is shifted in time to the right (Fig. 7a). This is indicating that the permittivity starting model is currently too far away from the input models and that the wavelet is compensating for this by starting later in time (blue curves in Fig. 7a and b). Note that a good effective source wavelet needs to start at 0 ns (Klotzsche *et al.*, 2019b). The FWI results (first column in Fig. 8b and c) based on these starting models and the effective source wavelet $SW_{Ray\ (-3)}$ fulfil the stopping criteria after 24 iterations and no remaining gradients are present. The FWI RMS curve by using the source wavelet $SW_{Ray\ (-3)}$ is shown with the blue curve in Fig. 9. Generally, the resulting FWI results show a lower ε_r than the input model,

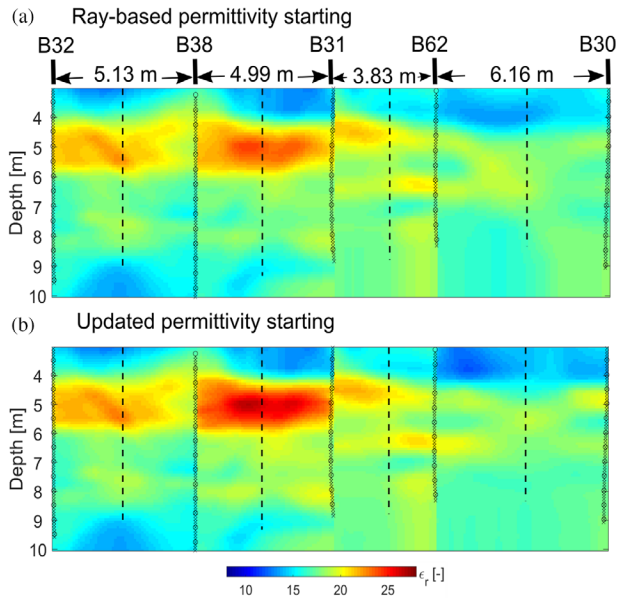

Figure 11 (a) Ray-based $\varepsilon_r$results and (b) FWI ε_r results using the selected maximum frequency bandwidth sub-data. Note that results shown in (b) are used as starting models for the full bandwidth inversion of the experimental Krauthausen data. And a homogeneous conductivity starting model of 13 mS/m was used for all inversions.

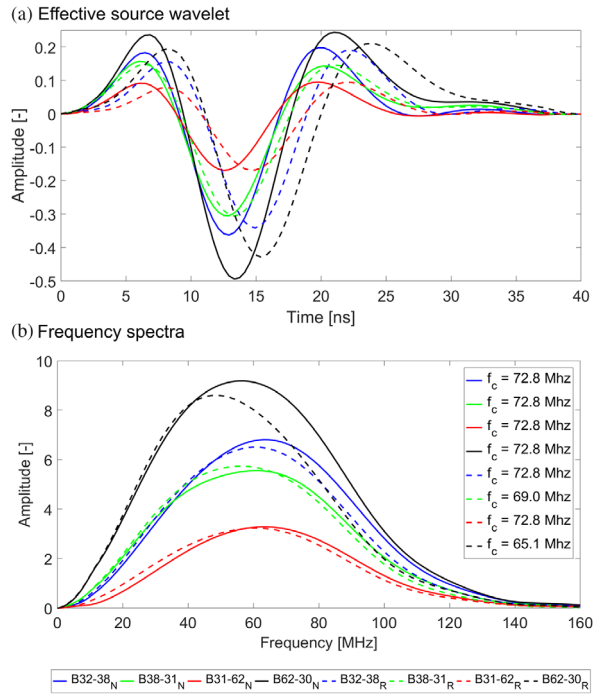

Figure 12 (a) Comparisons of the effective source wavelets estimated based on ray-based models (dashed graphs) and the updated source wavelets of the PEBBD scheme (solid graphs) for the four cross-sections in time domain for the experimental data of the Krauthausen test site. (b) Corresponding frequency spectra of standard and updated effective source wavelets. The legend values in (b) indicate these effective centre frequencies for different cross-sections.

demonstrating that the ε_r result gets trapped in a local minimum of the inversion process (Fig. 8b, first column). This is also indicated by the differences between the input model and FWI results in Fig. 8(d), where many regions are present that show differences with more than three in permittivity. Note that this inversion still fulfils most of the criteria for a good inversion process. The only indicators that the inversion is too far away from the input model are provided by the effective source wavelet that is shifted in time and the high MAE (not available for experimental data applications). To fulfil the criteria that the effective source wavelet needs to start at 0 ns, we shifted in the next step the source wavelet by −4 ns in the time domain $SW_{Ray\ (-3shift)}$ (cyan curves in Fig. 7a and b) to ensure that it starts at 0 ns. The corresponding FWI results (Fig. 8b and c, second column) show a better reconstruction of the permittivity results than the previous inversion, and the final RMS value after 30 iterations is 8.3326×10^{-7} (cyan curve in Fig. 9). Meanwhile the FWI σ results are better than the FWI σ results with $SW_{Ray\ (-3)}$ (Fig. 8c, columns first and second).

Similar to the previous synthetic case I, we apply the same bandpass filters for the same number of iterations to the shifted effective source wavelet $SW_{Ray\ (-3shift)}$ to enhance the reconstruction of the FWI results with the PEBDD approach. For the first sub-data FWI, the first starting models $Model_{Ray-3}$ are used. The final FWI ε_r results using the maximum bandwidth sub-data (Fig. 8a, updated – ε_r (−3)) and the homogeneous σ model with 13 mS/m are used as the new updated starting models $Model_{New-3}$. It can be noticed that the updated permittivity starting model shows generally higher values in the entire domain and the model is much closer to the ray-based model used in the synthetic study I, indicating that the PEBDD approach is able to enhance the erroneous starting model (compare Figs 4a and 8a). The updated starting models are considered to derive SW_{Low} before performing the FWI.

The updated FWI results using SW_{Low} show improved permittivity results and more continuous structures, although the conductivity performs less good indicated by the image differences (Fig. 8b and c, third column) and the higher RMS value for the final iteration (Fig. 9, black graph). The final permittivity results of this inversion and SW_{Low} are used to update the effective wavelet a last time to SW_{New} (Fig. 7a and b, red source wavelet). The corresponding FWI results (Fig. 8b and c,

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

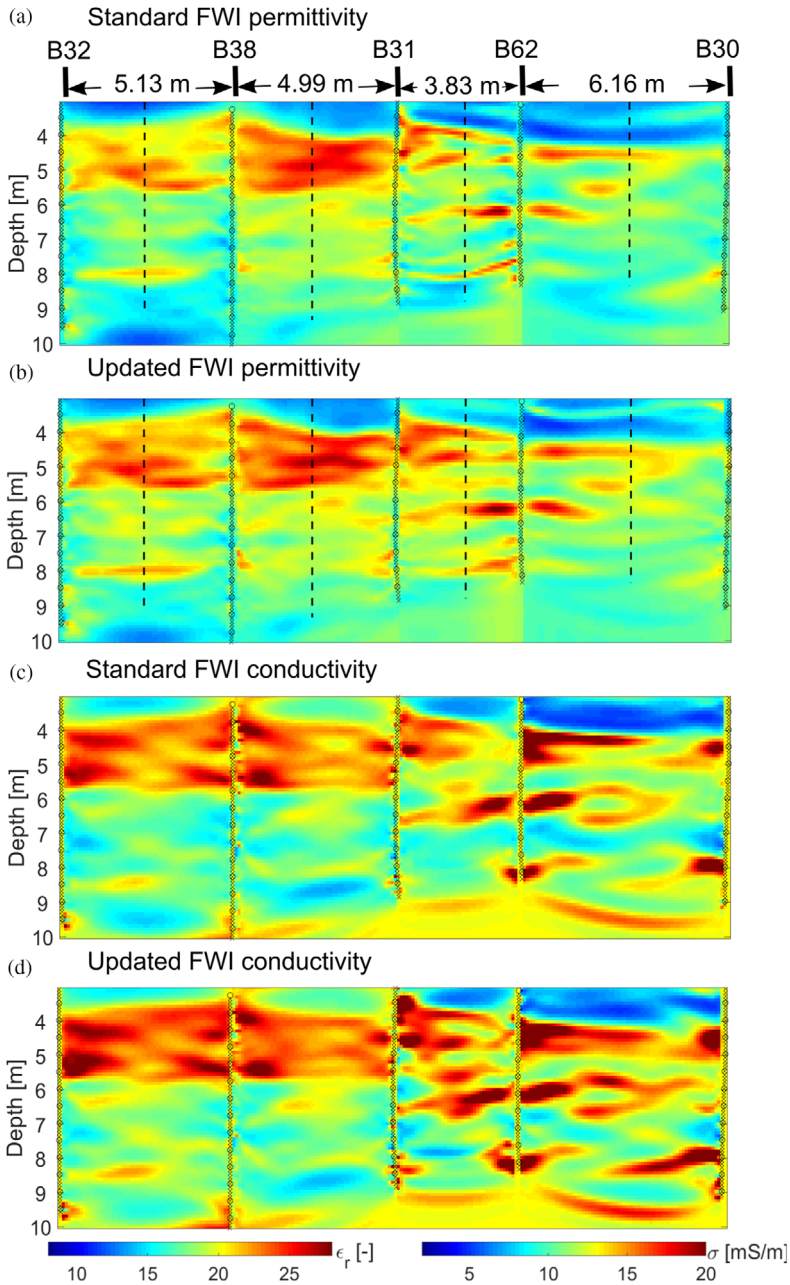

Figure 13 Standard FWI results of (a) ε_r and (c) σ of the four cross-sections. Updated FWI results of (b) ε_r and (d) σ based on the new ε_r starting models (Fig. 11b) and corresponding updated new effective source wavelets (solid graphs in Fig. 12). Dashed lines mark the CPT data locations between the boreholes.

fourth column) after 50 iterations are derived with the SW_{New} and $Model_{New-3}$ as the starting models. These results show an improved reconstruction of the permittivity and the conductivity results in comparison to the previous FWI results. Furthermore, comparisons of the absolute error tomograms for different FWI ε_r and σ results in 2D domain (Fig. 8d and e) show that the second-updated FWI results with the wavelet SW_{New} are the most accurate reconstruction of the input models.

The MAE values (shown in titles of Fig. 8d and e) and behaviour of the RMS curves (Fig. 9) are best for the second-updated FWI results with the wavelet SW_{New} and indicate the best FWI results (more details in Table 2). Analysing the misfit between the observed and FWI modelled data shows that the second-updated FWI results provide the best fit and indicates that this inversion obtained a model that describe the data well and best, while for the other three FWI results a significant misfit can be observed (Fig. 10a and b, exemplary

Table 3 Comparison of the different FWI approaches for the experimental data set of the Krauthausen test site. RMS represents residual values between observed and modelled radar traces. Percentages in parentheses indicate the ratio of the new FWI RMS to the standard FWI RMS with SW_{Ray} , the lower value means the higher improvement efficiency. Correlation coefficient R and RMS between 1D ε_r FWI and the CPT data (CPT porosity has been converted into ε_r) represent the reliability of the FWI results, the larger R and the lower RMS values mean the FWI results are closer to the real values

Borehole No. (distance between boreholes)	32–38 (5.13 m)	38–31 (4.99 m)	31–62 (3.83 m)	62–30 (6.16 m)
Standard FWI RMS (10^{-7}) (SW_{Ray})	8.5055	9.4156	9.1763	7.0538
New FWI RMS (10^{-7}) (SW_{New})	6.8210	7.1789	7.8503	5.4136
	(80.2%)	(76.2%)	(85.5%)	(76.7%)
R (Sta-FWI:CPT)	0.8105	0.9032	0.8301	0.8830
R (New FWI:CPT)	0.8409	0.8973	0.8408	0.8723
RMS (Sta-FWI:CPT)	2.6807	1.9307	2.5131	2.1026
RMS (New FWI:CPT)	1.9054	2.1503	1.5224	1.7333

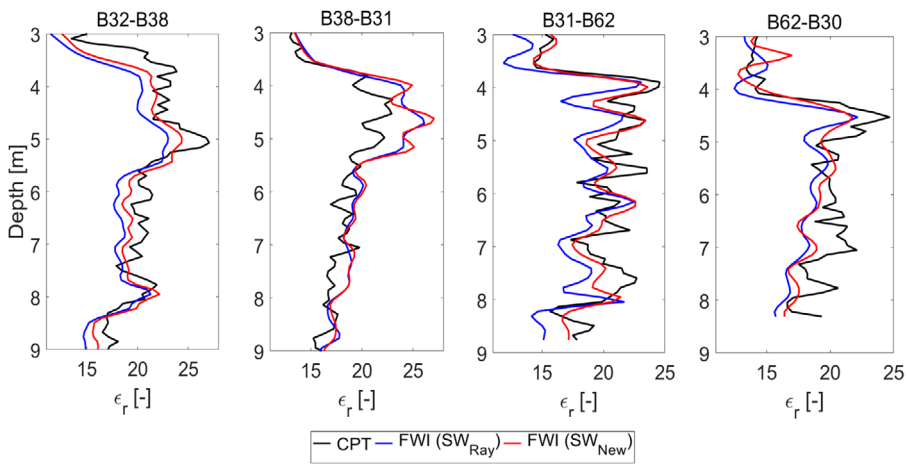

Figure 14 Permittivity comparisons derived from cone penetration test (CPT) data (black), standard FWI ε_r results (blue) and the updated new FWI ε_r results (red) along CPT profiles between boreholes (see Fig. 13 for locations).

for transmitter at 5.69 m depth). To describe the regional differences of different FWI models, we compute the MAE between the stochastic input models and the FWI model results along the horizontal direction (Fig. 10c) and the vertical direction (Fig. 10d) for ε_r and σ, respectively. Thereby, we can notice that the FWI ε_r results with $SW_{Ray\ (-3)}$ display the largest MAE values for both directions. The FWI σ results with SW_{Low} (black curves) have a large MAE, which implies that they cannot be resolved by only using the first-updated effective source wavelet in time when the starting models are too far away from the real models. Finally, we can conclude that the progressively expanded bandwidth scheme can not only improve the FWI results, but also that it is able to retrieve accurate FWI results for starting models more than a half-wavelength away from the measured data. Therefore, a lot of previous detailed work to construct good starting models for experimental GPR data can be reduced and the application to field data could be much easier.

Experimental ground-penetrating radar data studies

As a final test, we applied the new PEBDD approach to an experimental data set from the Krauthausen test site in Germany. The Krauthausen study site is located approximately 10 km northwest of the city Düren in Germany and a detailed description of the site can be found in Vereecken *et al.* (2000). In the last decades, many hydrological and geophysical field techniques have been applied on this site to study the aquifer spatial distribution and flow characteristics, including flowmeter tests (Li *et al.*, 2008), tracer experiments (Vereecken *et al.*, 2000; Vanderborght and Vereecken, 2001), cone penetration tests (Tillmann *et al.*, 2008) and GPR

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers., *Near Surface Geophysics*, 1–23

measurements (e.g., Gueting *et al.*, 2015). Here, we utilized the measured crosshole GPR data of Gueting *et al.* (2015) and Zhou *et al.* (2021) using 200 MHz antennae (Sensors and Software) to test the PEBDD FWI scheme for four cross-sections between five boreholes. The GPR data were acquired using a semi-reciprocal acquisition set-up with a transmitter and receiver spacing of 0.5 m and 0.1 m, respectively.

Before the FWI of experimental GPR data can be performed, some pre-processing steps are necessary (more details can be found in Klotzsche *et al.*, 2019b). After applying a standard dewow filter and defining the borehole coordinates for the antennae positions, we need to determine an accurate time zero of the received radar signals. Here, we employed an improved time-zero correction method based on the ZOP-MOG cross-correlation method (Oberröhrmann *et al.*, 2013). The ray-based method was applied for the four GPR cross-sections to obtain the ray-based permittivity starting models (Fig. 11a). Similar to previous studies, the electric conductivity starting models were chosen to be homogeneous with 13 mS/m in the entire domain (not shown). In addition, to reduce the influence of 3D wave propagation phenomena, the 3D data need to be transformed into 2D according to the approach by Bleistein (1986). The effective source wavelet SW_{Ray} (Fig. 12, dashed curves) is estimated based on the ray-based starting models according to the standard procedure. Using the wavelet SW_{Ray} and the corresponding starting models, the standard FWI is performed (Fig. 13a and c). Note that in contrast to Gueting *et al.* (2015) the time-zero correction of the GPR data was improved and corrected (more details in the Corrigendum; Gueting *et al.*, 2020). Therefore, the final standard FWI results show generally higher permittivity of about 3–4 and lower electrical conductivity results, while the structures are similar.

Similar to the previous synthetic case studies, we filtered the effective source wavelets and corresponding measured GPR data by using the PEBDD scheme. The smallest filtered sub-data bandwidths are 12–16 MHz for four cross-sections. The bandpass widths are increased every five iterations by 4 MHz until the maximum bandwidth sub-data are reached. Note that we selected different highest cut corner frequencies according to centre frequencies of the different cross-section data sets. For the cross-sections B38-31 and B62-30, we used 68 MHz, and for the B32-38 and B31-62 data set, we applied 72 MHz. The final FWI ε_r results using the maximum bandwidth sub-data (Fig. 11b) and the homogeneous σ model were considered as the updated starting models and applied in a following FWI under the full bandwidth data. Following the flowchart in Fig. 1(b), the effective source wavelet SW_{Low} is updated using the new starting models and the effective source wavelet SW_{Ray}. The updated source wavelets SW_{Low} together with the new starting models are used to calculate the FWI results (not shown). Finally, the second-updated effective source wavelets SW_{New} are derived using the new FWI ε_r results and SW_{Low} in the deconvolution approach (Fig. 12, solid curves). The corresponding updated final FWI results (Fig. 13b and d) show very similar structures as the standard FWI results, but the structures are more continuous closer to the boreholes. The RMS values between the observed and modelled data based on selected different FWI models show that the updated FWI results are smaller than those of the standard FWI (Table 3), indicating improved results of the updated FWI.

To validate the different FWI results, we compared them with porosity data based on cone-penetrations test (CPT) measurements (Yang et al., 2013; Zhou *et al.*, 2021). Note that the CPT data were acquired mostly in the centre of the cross-sections. First of all, we converted the porosity values ϕ of the CPT to the relatively permittivity using the three phase complex refractive index model (CRIM) in saturated zone (e.g., Gueting *et al.*, 2015),

$$\varepsilon_r = \left[\phi \times \sqrt{\varepsilon_f} - \sqrt{\varepsilon_s} + \sqrt{\varepsilon_s}\right]^2 , \tag{7}$$

where ε_s and ε_f are the permittivity of the solid and the fluid, respectively.

Considering a water temperature of 10°C we use a ε_f of 84 and we apply a ε_s value with 4.5 based on literature values of quartz (e.g., Eisenberg and Kauzmann *et al.*, 2005). We compared the two FWI ε_r results with the CPT data along the vertical profile locations (dashed lines in Fig. 13). Comparisons of these 1D vertical results indicate that the updated FWI ε_r results (Fig. 14, red lines) are closer to the CPT data (Fig. 14, black lines) than the standard FWI ε_r results (Fig. 14, blue lines). Quantitative comparisons are described in Table 3, where we computed RMS and correlation coefficient R with larger than 0.8 between the CPT and two FWI results. In general, the updated FWI results with the PEBDD scheme are better than the standard FWI results, indicating that the new PEBDD approach improved the experimental GPR FWI results. Note that the RMS value (for 1D CPT) of the updated FWI for boreholes 38–31 is larger than the standard FWI (green value in Table 3). A possible explanation could be that the standard FWI was already very good because many studies have been already performed with this data set, and many optimizations have been performed with the standard approach. For the other data sets, less intense tests have been applied for the standard FWI.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

CONCLUSIONS AND OUTLOOK

We have presented a new crosshole ground-penetrating radar (GPR) full-waveform inversion (FWI) approach using progressively expanded bandwidths of the measured and modelled data (PEBDD) with frequency bandpass filters to improve the reconstruction of the subsurface properties. The new PEBDD scheme was applied to two synthetic case studies and an experimental data set from the Krauthausen test site in Germany. Thereby, we have demonstrated that the new scheme is able to improve the FWI results by taming the problem of the inversion becoming easily trapped in a local minima caused by non-linear and ill-posed problems. The introduced PEBDD scheme applies designed bandpass filters to the effective source wavelet and the observed data. Thereby, the different sub-data sets with various bandwidths are progressively expanded after the centre frequency of the data is reached. The FWI results of this PEBDD provide an updated ε_r starting model, which is used to update an effective source wavelet in the deconvolution approach and to perform the FWI with full bandwidth data. Further, we update the effective source wavelet SW_{New}. Considering this updated effective source wavelet and the starting models based on the PEBDD scheme, the resulting FWI results show a better reconstruction of the medium properties and the final root-mean-square values are decreased in contrast to the standard FWI.

Two synthetic cases have proven that the PEBDD scheme is robust and can also reconstruct the permittivity and electrical conductivity results for starting models that are more than half a wavelength away from the measured data. To fulfil the starting model criteria, the modelled data need to be within half a wavelength of the measured data and normally need an accurate processing of the data and a detailed understanding of the trained user. Problems arising from erroneous starting models are sometimes hard to notice, and therefore an improved approach such as the PEBDD can significantly improve the applicability of the FWI. For the experimental GPR data, we have computed the standard and the updated FWI results using the PEBDD scheme for four cross-sections. We compared the different FWI results with CPT data from the centre of the cross-sections, where the updated FWI results provided the best correlation to the CPT data. Finally, we concluded that not only the progressively expanded bandwidth approach can improve the FWI results, but it is also able to retrieve more accurate FWI results for starting models more than a half wavelength away from the measured data. Therefore, a lot of previous detailed work to construct good starting models for experimental GPR data can be reduced and the application to field data will be much easier. In future work, we will also investigate the possibility to apply the new PEBDD approach to other frequencies such as for 100 MHz antennae and to different test sites.

ACKNOWLEDGEMENTS

The first author acknowledges the financial support from China Scholarship Council (Project No. 201506340136). We gratefully acknowledge Peleg Haruzi for providing help in the process of constructing the synthetic stochastic models. Especially, we are grateful to Jessica Schmäck for her constructive suggestions. Further, 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).

Open access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Zhen Zhou https://orcid.org/0000-0003-0839-8802
Anja Klotzsche https://orcid.org/0000-0002-7021-5045
Harry Vereecken https://orcid.org/0000-0002-8051-8517

REFERENCES

Backus, G. and Gilbert, F. (1968) The resolving power of gross earth data. *Geophysical Journal International*, 16(2), 169–205. https://doi.org/10.1111/j.1365-246X.1968.tb00216.x.

Becker, M.W. (2006) Potential for satellite remote sensing of ground water. *Groundwater*, 44(2), 306–318. https://doi.org/10.1111/j.1745-6584.2005.00123.x.

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. https://doi.org/10.1002/2015WR017016.

Bleistein, N. (1986) Two-and-one-half dimensional in-plane wave propagation. *Geophysical Prospecting*, 34(5), 686–703. https://doi.org/10.1111/j.1365-2478.1986.tb00488.x.

Brittan, J., Bai, J., Delome, H., Wang, C. and Yingst, D. (2013) Full waveform inversion–the state of the art. *First Break*, 31(10), 75–82. https://doi.org/10.3997/1365-2397.31.10.71541.

Cassidy, N.J. (2007) Evaluating LNAPL contamination using GPR signal attenuation analysis and dielectric property measurements: practical implications for hydrological

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

studies. *Journal of Contaminant Hydrology*, 94(1–2), 49–75. https://doi.org/10.1016/j.jconhyd.2007.05.002.

Chew, W.C. and Lin, J.H. (1995) A frequency-hopping approach for microwave imaging of large inhomogeneous bodies. *IEEE Microwave and Guided Wave Letters*, 5(12), 439–441. https://doi.org/10.1109/75.481854.

Coscia, I., Greenhalgh, S.A., Linde, N., Doetsch, J., Marescot, L., Günther, T., *et al*. (2011) 3D crosshole ERT for aquifer characterization and monitoring of infiltrating river water. *Geophysics*, 76(2), G49–G59. https://doi.org/10.1190/1.3553003.

Dafflon, B., Irving, J. and 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. https://doi.org/10.1016/j.jappgeo.2011.02.001.

Dafflon, B. and 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 Resources Research*, 48(5), W05553. https://doi.org/10.1029/2011WR010916.

Doetsch, J., Linde, N., Coscia, I., Greenhalgh, S.A. and Green, A.G. (2010) Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data. *Geophysics*, 75(6), G53–G64. https://doi.org/10.1190/1.3496476.

Dubois, A., Belkebir, K., Catapano, I. and Saillard, M., (2009) Iterative solution of the electromagnetic inverse scattering problem from the transient scattered field. *Radio Science*, 44(01), 1–13. https://doi.org/10.1029/2007RS003765.

Eisenberg, D., Kauzmann, W. and Kauzmann, W. (2005) *The Structure and Properties of Water*. Oxford University Press.

Ellefsen, K.J., Mazzella, A.T., Horton, R.J. and McKenna, J.R. (2011) Phase and amplitude inversion of crosswell radar data. *Geophysics*, 76(3), J1–J12. https://doi.org/10.1190/1.3554412.

Ernst, J.R., Green, A.G., Maurer, H. and Holliger, K. (2007b) Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data. *Geophysics*, 72(5), J53–J64. https://doi.org/10.1190/1.2761848.

Ernst, J.R., Maurer, H., Green, A.G. and Holliger, K. (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(9), 2807–2828. https://doi.org/10.1109/TGRS.2007.901048

Gueting, N., Klotzsche, A., van der Kruk, J., Vanderborght, J., Vereecken, H. and 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. https://doi.org/10.1016/j.jhydrol.2015.03.030.

Gueting, N., Klotzsche, A., van der Kruk, J., Vanderborght, J., Vereecken, H. and Englert, A., (2020) Corrigendum to 'Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests' [*J. Hydrol.* (2015) 680-695]. *Journal of Hydrology*, 590, 125483. https://doi.org/10.1016/j.jhydrol.2020.125483.

Holliger, K., Musil, M. and Maurer, H.R. (2001) Ray-based amplitude tomography for crosshole georadar data: a numerical assessment. *Journal of Applied Geophysics*, 47(3–4), 285–298. https://doi.org/10.1016/S0926-9851(01)00072-6.

Klotzsche, A., Jonard, F., Looms, M.C., van der Kruk, J. and Huisman, J.A. (2018) Measuring soil water content with ground penetrating radar: a decade of progress. *Vadose Zone Journal*, 17(1), 1–9. https://doi.org/10.2136/vzj2018.03.0052.

Klotzsche, A., van der Kruk, J., Bradford, J. and Vereecken, H. (2014) Detection of spatially limited high-porosity layers using crosshole GPR signal analysis and full-waveform inversion. *Water Resources Research*, 50(8), 6966–6985. https://doi.org/10.1002/2013WR015177.

Klotzsche, A., van der Kruk, J., Linde, N., Doetsch, J. and Vereecken, H. (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(2), 932–944. https://doi.org/10.1093/gji/ggt275

Klotzsche, A., van der Kruk, J., Meles, G.A., Doetsch, J., Maurer, H. and Linde, N. (2010) Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the Thur River, Switzerland. *Near Surface Geophysics*, 8(6), 635–649. https://doi.org/10.3997/1873-0604.2010054

Klotzsche, A., van der Kruk, J., Meles, G. and Vereecken, H. (2012) Crosshole GPR full-waveform inversion of waveguides acting as preferential flow paths within aquifer systems. *Geophysics*, 77(4), H57–H62. https://doi.org/10.1190/geo2011-0458.1.

Klotzsche, A., Vereecken, H. and van der Kruk, J. (2019a) GPR full-waveform inversion of a variably saturated soil-aquifer system. *Journal of Applied Geophysics*, 170, 103823. https://doi.org/10.1016/j.jappgeo.2019.103823.

Klotzsche, A., Vereecken, H. and van der Kruk, J. (2019b) Review of crosshole ground-penetrating radar full-waveform inversion of experimental data: Recent developments, challenges, and pitfalls. *Geophysics*, 84(6), H13–H28. https://doi.org/10.1190/geo2018-0597.1.

Kowalsky, M.B., Finsterle, S., Peterson, J., Hubbard, S., Rubin, Y., Majer, E., *et al*. (2005) Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data. *Water Resources Research*, 41(11), WR004237. https://doi.org/10.1029/2005WR004237.

Kuroda, S., Takeuchi, M. and Kim, H.J. (2007) Full-waveform inversion algorithm for interpreting crosshole radar data: a theoretical approach. *Geosciences Journal*, 11(3), 211–217. https://doi.org/10.1007/BF02913934.

Kuroda, S., Jang, H. and Kim, H.J. (2009) Time-lapse borehole radar monitoring of an infiltration experiment in the vadose zone. *Journal of Applied Geophysics*, 67(4), 361–366. https://doi.org/10.1016/j.jappgeo.2008.07.005.

Landon, M.K., Rus, D.L. and Harvey, F.E. (2001) Comparison of instream methods for measuring hydraulic conductivity in sandy streambeds. *Groundwater*, 39(6), 870–885. https://doi.org/10.1111/j.1745-6584.2001.tb02475.x.

Lavoué, F., Brossier, R., Métivier, L., Garambois, S. and Virieux, J. (2014) Two-dimensional permittivity and conductivity imaging by full waveform inversion of multioffset GPR data: A frequency-domain quasi-Newton approach. *Geophysical Journal International*, 197(1), 248–268. https://doi.org/10.1093/gji/ggt528.

Li, W., Englert, A., Cirpka, O.A. and Vereecken, H. (2008) Three-dimensional geostatistical inversion of flowmeter and pumping test data. *Groundwater*, 46(2), 193–201. https://doi.org/10.1111/j.1745-6584.2007.00419.x.

Looms, M.C., Jensen, K.H., Binley, A. and Nielsen, L. (2008) Monitoring unsaturated flow and transport using cross-borehole geophysical methods. *Vadose Zone Journal*, 7(1), 227–237. https://doi.org/10.2136/vzj2006.0129.

Maurer, H. and Musil, M. (2004) Effects and removal of systematic errors in crosshole georadar attenuation tomography. *Journal of Applied Geophysics*, 55(3–4), 261–270. https://doi.org/10.1016/j.jappgeo.2004.02.003.

Maurer, H., Greenhalgh, S. and Latzel, S. (2009) Frequency and spatial sampling strategies for crosshole seismic waveform spectral inversion experiments. *Geophysics*, 74(6), WCC79–WCC89. https://doi.org/10.1190/1.3157252.

Meles, G., Greenhalgh, S., Van der Kruk, J., Green, A. and Maurer, H. (2012) Taming the non-linearity problem in GPR full-waveform inversion for high contrast media. *Journal of Applied Geophysics*, 78, 31–43. https://doi.org/10.1016/j.jappgeo.2011.12.001.

Meles, G.A., Van der Kruk, J., Greenhalgh, S.A., Ernst, J.R., Maurer, H. and Green, A.G. (2010) A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination crosshole/borehole-to-surface GPR data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(9), 3391–3407. https://doi.org/10.1109/TGRS.2010.2046670.

Mora, P. (1987) Nonlinear two-dimensional elastic inversion of multioffset seismic data. *Geophysics*, 52(9), 1211–1228. https://doi.org/10.1190/1.1442384.

Oberröhrmann, M., Klotzsche, A., Vereecken, H. and van der Krak, J. (2013) Optimization of acquisition setup for cross-hole: GPR full-waveform inversion using checkerboard analysis. *Near Surface Geophysics*, 11(2), 197–209. https://doi.org/10.3997/1873-0604.2012045.

Paz, C., Alcalá, F.J., Carvalho, J.M. and Ribeiro, L. (2017) Current uses of ground penetrating radar in groundwater-dependent ecosystems research. *Science of the Total Environment*, 595, 868–885. https://doi.org/10.1016/j.scitotenv.2017.03.210.

Peterson Jr, J.E. (2001) Pre-inversion corrections and analysis of radar tomographic data. *Journal of Environmental and Engineering Geophysics*, 6(1), 1–18. https://doi.org/10.4133/JEEG6.1.1.

Pinard, H., Garambois, S., Métivier, L., Dietrich, M., Sénéchal, G. and Rousset, D. (2016) Full-waveform inversion of GPR data acquired between boreholes in Rustrel carbonates. *E3S Web of Conferences*, 12, 01002). https://doi.org/10.1051/e3sconf/20161201002.

Polak, E. and Ribiere, G. (1969) Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 3(R1), 35–43.

Pratt, R.G., Shin, C. and Hick, G.J. (1998) Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2), 341–362. https://doi.org/10.1046/j.1365-246X.1998.00498.x.

Shin, C. and Cha, Y.H. (2008) Waveform inversion in the Laplace domain. *Geophysical Journal International*, 173(3), 922–931. https://doi.org/10.1111/j.1365-246X.2008.03768.x.

Simmer C., Thiele-Eich I., Masbou M., Amelung W., Bogena H., Crewell S., Diekkrüger B., Ewert F., Hendricks Franssen H. J., Huisman J. A., Kemna A. (2015) Monitoring and Modeling the Terrestrial System from Pores to Catchments: The Transregional Collaborative Research Center on Patterns in the Soil–Vegetation–Atmosphere System. *Bulletin of the American Meteorological Society*, 96 (10), 1765–1787. https://doi.org/10.1175/bams-d-13-00134.1.

Tarantola, A. (1984) Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8), 1259–1266. https://doi.org/10.1190/1.1441754.

Tillmann, A., Englert, A., Nyari, Z., Fejes, I., Vanderborght, J. and Vereecken, H. (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(1–2), 57–75. https://doi.org/10.1016/j.jconhyd.2007.07.013.

Vanderborght, J. and Vereecken, H. (2001) Analyses of locally measured bromide breakthrough curves from a natural gradient tracer experiment at Krauthausen. *Journal of Contaminant Hydrology*, 48(1–2), 23–43. https://doi.org/10.1016/S0169-7722(00)00176-5.

van der Kruk, J., Gueting, N., Klotzsche, A., He, G., Rudolph, S., von Hebel, C. and Vereecken, H. (2015) Quantitative multi-layer electromagnetic induction inversion and full-waveform inversion of crosshole ground penetrating radar data. *Journal of Earth Science*, 26(6), 844–850. https://doi.org/10.1007/s12583-015-0610-3.

Vereecken, H., Döring, U., Hardelauf, H., Jaekel, U., Hashagen, U., Neuendorf, O., *et al*. (2000) Analysis of solute transport in a heterogeneous aquifer: the Krauthausen field experiment. *Journal of Contaminant Hydrology*, 45(3–4), 329–358. https://doi.org/10.1016/S0169-7722(00)00107-8.

Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T., *et al*. (2016) Modeling soil processes: review, key challenges, and new perspectives. *Vadose zone Journal*, 15(5). https://doi.org/10.2136/vzj2015.09.0131.

Virieux, J. and Operto, S. (2009) An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6), WCC1–WCC26. https://doi.org/10.1190/1.3238367.

Wang, G., Yuan, S. and Wang, S. (2019) Retrieving low-wavenumber information in FWI: an efficient solution for cycle skipping. *IEEE Geoscience and Remote Sensing Letters*, 16(7), 1125–1129. https://doi.org/10.1109/LGRS.2019.2892998.

Wu, R.S., Luo, J. and Wu, B. (2014) Seismic envelope inversion and modulation signal model. *Geophysics*, 79(3), WA13–WA24. https://doi.org/10.1190/geo2013-0294.1.

Yang, X., Klotzsche, A., Meles, G., Vereecken, H. and Van Der Kruk, J. (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. https://doi.org/10.1016/j.jappgeo.2013.08.007.

Zhou, Z., Klotzsche, A., Hermans, T., Nguyen, F., Schmäck, J., Haruzi, P., *et al*., (2020) 3D aquifer characterization of the Hermalle-sous-Argenteau test site using crosshole ground-penetrating radar amplitude analysis and full-waveform inversion. *Geophysics*, 85(6), H133–H148. https://doi.org/10.1190/geo2020-0067.1.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers., *Near Surface Geophysics*, 1–23

Zhou, Z., Klotzsche, A., Schmäck, J., Vereecken, H. and van der Kruk, J., (2021) Improvement of GPR full-waveform inversion images using cone penetration test data. *Geophysics*, 86(3), 1–74. https://doi.org/10.1190/geo2020-0283.1.

APPENDIX

For the synthetic case study I, we constructed a second set of starting models $\text{Model}_{\text{New}}$ ($\varepsilon_r + \sigma$), which were derived from the final FWI ε_r and σ results using the maximum bandwidth sub-data (Fig. A.1a). Following the flowchart (Fig. 1b), an effective source wavelet SW_{Low} ($\varepsilon_r + \sigma$) (not shown) is updated using the new starting models $\text{Model}_{\text{New}}$ ($\varepsilon_r + \sigma$) and the wavelet SW_{Ray}. This updated wavelet SW_{Low} ($\varepsilon_r + \sigma$) together with the starting models $\text{Model}_{\text{New}}$ ($\varepsilon_r + \sigma$) are used in the next step to calculate the FWI results using the full data set (not shown). These updated FWI results (including ε_r and σ) are in the last step considered to update the effective source wavelet SW_{New} ($\varepsilon_r + \sigma$) (Fig. 3a and b, green source wavelet). The corresponding final FWI results (Fig. A.1b) are derived with the SW_{New} ($\varepsilon_r + \sigma$) and $\text{Model}_{\text{New}}$ ($\varepsilon_r + \sigma$) as the starting models. By comparing the final RMS curves after 45 iterations (green graph in Fig. 5), the second-updated FWI results with the wavelet SW_{New} (RMS = 2.8996×10^{-7}; red curve) indicate a better convergence behaviour and a smaller RMS than the new updated results with the SW_{New} ($\varepsilon_r + \sigma$) (RMS = 3.8620×10^{-7}; green curve). Computing the absolute

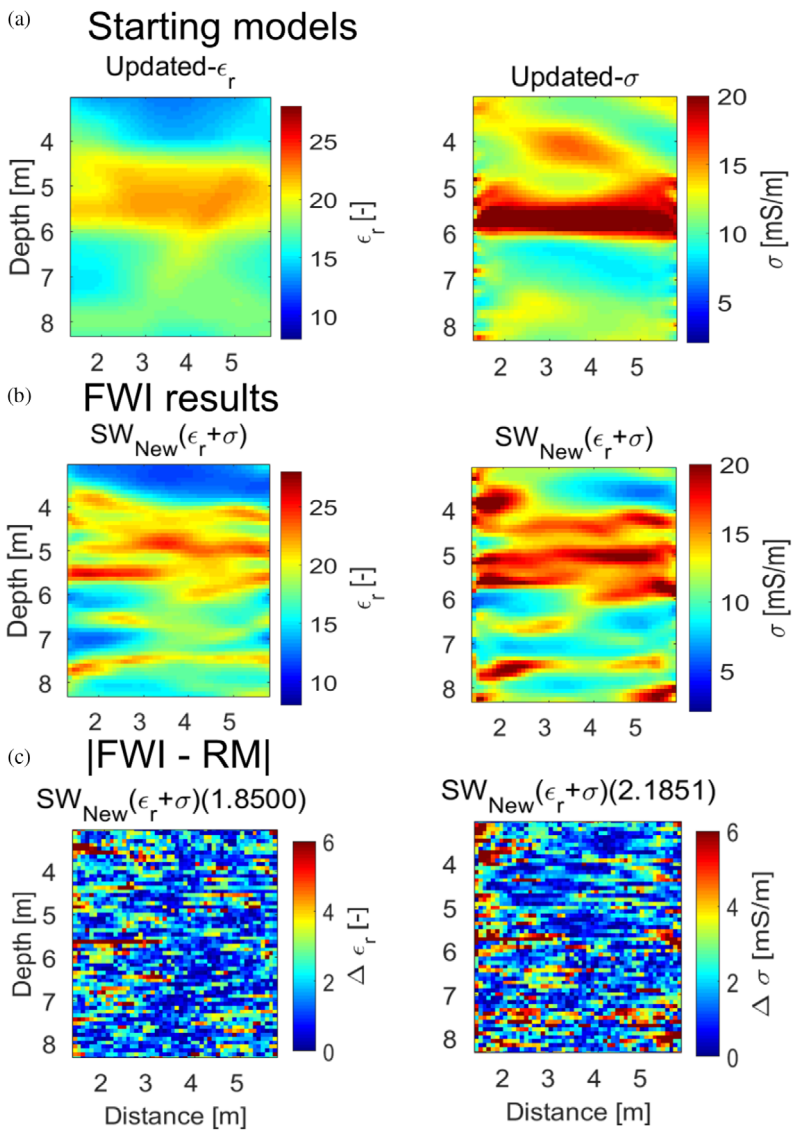

Figure A.1 (a) The permittivity and conductivity starting models based on the FWI results using the maximum expended bandwidth sub-data. (b) Corresponding FWI results using the updated starting models and the updated effective source wavelet. Effective source wavelet is named in titles. (c) Image plots of the absolute error between the stochastic input models and the final FWI ε_r and σ results. The mean absolute errors MAE of the entire 2D domain are shown in parentheses of titles.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,

errors between the FWI results and the input models of the 2D domain (Fig. A.1c), it can be noticed that the new updated FWI results with SW_{New} ($\varepsilon_r + \sigma$) show a less good fit than the second-updated FWI results with SW_{New} (Table 1). Therefore, we abandon this approach of using both permittivity and conductivity results derived by the FWI using the maximum bandwidth sub-data and consider for all other studies the homogeneous σ starting model with 13 mS/m.

© 2021 The Authors. *Near Surface Geophysics* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers.,