Soil hydraulic material properties and layered architecture from time-lapse GPR

Stefan Jaumann[1,2] and Kurt Roth[1,3]

[1]Institute of Environmental Physics, Heidelberg University, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany
[2]HGSMathComp, Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany
[3]Interdisciplinary Center for Scientific Computing, Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

Correspondence to: Stefan Jaumann (stefan.jaumann@iup.uni-heidelberg.de)

Abstract. Quantitative knowledge of the subsurface material distribution and its effective soil hydraulic material properties is essential to predict soil water movement. Ground-penetrating radar (GPR) is a non-invasive and non-destructive geophysical measurement method that is suitable to monitor hydraulic processes. Previous studies showed that the GPR signal from a fluctuating groundwater table is sensitive to the soil water characteristic and the hydraulic conductivity function. In this work, we show that the GPR signal originating from both the subsurface architecture and the fluctuating groundwater table is suitable to estimate the position of layers within the subsurface architecture together with the associated effective soil hydraulic material properties with inversion methods. To that end, we parameterize the subsurface architecture, solve the Richards equation, convert the resulting water content to relative permittivity with the complex reflective index model (CRIM), and solve Maxwell's equations numerically. In order to analyze the GPR signal, we implemented a new heuristic algorithm that detects relevant signals in the radargram (events) and extracts the corresponding signal travel time and amplitude. This algorithm is applied to simulated as well as measured radargrams and the detected events are associated automatically. Using events instead of the full wave regularizes the inversion focussing on the relevant measurement signal. For optimization, we use a global-local approach with preconditioning. Starting from an ensemble of initial parameter sets drawn with a Latin hypercube algorithm, we sequentially couple a simulated annealing algorithm with a Levenberg–Marquardt algorithm. The method is applied to synthetic as well as measured data from the ASSESS test site. We show that the method yields reasonable estimates for the position of the layers as well as for the soil hydraulic material properties by comparing the results to references derived from ground truth data as well as from time domain reflectometry (TDR).

1 Introduction

Quantitative understanding of soil water movement is in particular based on accurate knowledge of the subsurface architecture and the hydraulic material properties. As direct measurements are time-consuming and near to impossible at larger scales, soil hydraulic material properties are typically determined with indirect identification methods, such as inversion (Hopmans et al., 2002; Vrugt et al., 2008). Time domain reflectometry (TDR, e.g., Robinson et al., 2003) is a standard method to acquire the required measurement data, because it is sensitive to hydraulic processes. Yet, being an invasive method, the TDR sensors

disturb the soil texture of interest and typically require the maintenance of a local measurement station. Hence, it is difficult to apply the method at larger scales or to transfer the sensors to another field site. Ground-penetrating radar (GPR, e.g., Daniels, 2004; Neal, 2004) is an established non-invasive method for subsurface characterization and has the potential to become a standard method for efficient, accurate, and precise determination soil hydraulic material properties.

5 Available research studies regarding the estimation of hydraulic properties from GPR measurements may be categorized according to the applied methods for the different components of the research study, such as (i) GPR measurement procedure, (ii) experiment type, (iii) GPR simulation method, (iv) optimization method, and (v) evaluation method of the GPR signal.

  Most of these studies either use on-ground, off-ground, or borehole GPR measurements. On-ground measurements (e.g., Buchner et al., 2012; Busch et al., 2012; Léger et al., 2015) offer the most flexible approach. They have the disadvantage, 10 however, that the antenna characteristics is influenced by the coupling to the ground. Off-ground measurements (e.g., Lambot et al., 2009; Jadoon et al., 2012; Jonard et al., 2015) avoid these effects, but the measurements are influenced by surface roughness. Cross-borehole measurements allow for high resolution tomography of the subsurface (e.g., Ernst et al., 2007; Looms et al., 2008; Scholer et al., 2011), but require boreholes which are destructive and expensive.

  The applied experiment types range from infiltration, fluctuating groundwater table, to evaporation. Infiltration experiments 15 (e.g., Léger et al., 2014; Thoma et al., 2014; Rossi et al., 2015) are fast (hours) and provide indirect information about the near surface material properties. Through its dependence on the form of the infiltration front or plume, the resulting GPR signal can get rather complicated to reproduce when used for quantitative evaluation. Difficulties arise from multiple reflections in the plume, waveguides in the infiltration front, and from noise originating for small-scale heterogeneity or fingering. In particular if the infiltration is done artificially, accurate knowledge of the spatial distribution of the infiltration flux is required. Also 20 simultaneous GPR measurements during the infiltration process are difficult, because the antenna coupling to the subsurface is influenced by the changing water content close to the surface. Fluctuating groundwater table experiments (e.g., Bradford et al., 2014; Léger et al., 2015) require intermediate time scales (hours to days) and provide information about the material properties close to the groundwater table. These experiments are typically limited to fluvial or coastal areas or are induced artificially in test sites. Evaporation experiments (e.g., Moghadas et al., 2014) demand long time scales (weeks), because the hydraulic 25 dynamics is slow at low water contents. Yet, this kind of experiment is important to understand the coupling of the pedosphere with the atmosphere.

  The applied models to simulate the GPR signal are faced by an inherent trade-off between performance and accuracy. Ray tracing (Léger et al., 2014, 2015) is fast but merely yields an approximate solution of Maxwell's equations. These equations can be solved analytically with a Green's function (e.g., Lambot et al., 2009; Busch et al., 2012; Jonard et al., 2015) assuming a 30 layered subsurface architecture. Alternatively, Maxwell's equations can be solved numerically with the finite differences time domain (FDTD) method (e.g., Taflove and Hagness, 2000). This method is computationally expensive, but grants full flexibility concerning the source wavelet and the subsurface architecture (e.g., Lampe et al., 2003; Giannopoulos, 2005; Buchner et al., 2012).

  Due to the inherent oscillating nature of the electromagnetic signal, inversion of GPR data generally demands globally 35 convergent and robust optimization techniques. Sequentially coupling a globally convergent search algorithm, e.g., the global

multilevel coordinate search algorithm (GMCS, Huyer and Neumaier, 1999) with the gradient-free locally convergent Nelder–Mead simplex algorithm (NMS, Nelder and Mead, 1965), was successfully applied to estimate hydraulic material properties from GPR measurements (e.g., Lambot et al., 2004; Busch et al., 2012; Moghadas et al., 2014). The NMS was further developed to the shuffled complex evolution (SCE-UA, Duan et al., 1992) which has become a standard tool in hydrology and was also
5   applied on GPR measurements (e.g., Jadoon et al., 2012; Léger et al., 2014, 2015). Additionally, Markov chain Monte Carlo (MCMC) methods (e.g., Scholer et al., 2011; Thoma et al., 2014; Jonard et al., 2015) and data assimilation approaches (e.g., Tran et al., 2014; Manoli et al., 2015; Rossi et al., 2015) have been successfully applied so far.

    The GPR signal has to be processed automatically for parameter estimation. Many full waveform inversion approaches directly use the resulting Green's function (e.g., Lambot et al., 2009; Busch et al., 2012; Jadoon et al., 2012) in the cost function.
10   Using the full antenna signal may lead to many local minima prohibiting a reliable identification of the global minimum (e.g., Bradford et al., 2014). In contrast, filtering the radargram with convolution approaches to determine travel time and amplitude of a limited number of events leads to better convergence and may even allow the application of efficient locally-convergent algorithms (e.g., Buchner et al., 2012).

    In homogeneous materials, the transition zone above the groundwater table exhibits a smooth variation of the relative per-
15   mittivity. Since the resulting GPR reflection is a superposition of a series of infinitesimal contributions along the transition zone, the detailed form of this reflection is sensitive to the variation of the relative permittivity. For simplicity, we refer to this reflection as transition zone reflection. Dagenbach et al. (2013) showed that this reflection is sensitive to the hydraulic material parameterization model. Bradford et al. (2014) measured the transition zone reflection of a drainage pumping test in a fluvial area with a antenna center frequency of 200 MHz and estimated hydraulic material properties. Klenk et al. (2015)
20   employed numerical forward simulations and experiments using GPR antennas with higher antenna center frequency (400 and 600 MHz) for a more detailed explanation of the transition zone reflection during imbibition, equilibration, and drainage. They also concluded that the transition zone reflection is sensitive to hydraulic material properties.

    This work builds upon previously published methods for simultaneous estimation of the subsurface architecture and the effective water content based on on-ground multi-offset GPR measurement data (e.g., Gerhards et al., 2008; Buchner et al.,
25   2012). In order to develop methods to additionally estimate hydraulic material properties, the ASSESS test site was forced with a fluctuating groundwater table ensuring a large hydraulic dynamics. In this work, we use the resulting transition zone reflection together with reflections originating from material interfaces to estimate the positions of layers within the subsurface architecture as well as their hydraulic material properties. Running 2D or even 3D measurements and inversions is a massive experimental and computational effort. Prior to embarking on this, the individual steps must be demonstrated. To this end,
30   the experiment was monitored with a stationary on-ground bistatic antenna that operates at a center frequency of 400 MHz leading to time-lapse GPR measurement data. The hydraulic dynamics below this antenna is modeled in 1D using the Richards equation. The resulting the water content distribution is extruded and converted to a relative relative permittivity distribution to solve Maxwell's equations in 2D. Similar to Buchner et al. (2012), we developed a new heuristic semi-automatic approach to extract the signal travel time and amplitude of relevant reflections in the radargram (events). This approach is applied to both
35   the simulated and measured radargram and the corresponding events are associated automatically. For optimization, a global-

local inversion approach with preconditioning is applied. To that end, we draw parameter sets with a Latin hypercube algorithm that serve as initial parameters for the preconditioning step. In this step, a simulated annealing algorithm and a Levenberg–Marquardt algorithm are sequentially coupled using a subsampled data set for a limited number of iterations. Subsequently, the resulting parameters of the preconditioning step serve as initial parameters for another run of the Levenberg–Marquardt algorithm using the full data set. We show that this procedure accurately estimates the layered subsurface architecture as well as the associated effective hydraulic material properties for synthetic and measurement data.

2 Methods

2.1 ASSESS

The measurement data for this work are acquired at an approximately $2\,\mathrm{m} \times 20\,\mathrm{m} \times 4\,\mathrm{m}$ large the test site (ASSESS) which is located near Heidelberg, Germany, and provides an effectively 2D subsurface architecture consisting of three kinds of sands (Fig. 1). Below the sands, an approximately $0.1\,\mathrm{m}$ thick gravel layer ensures rapid distribution of the water pressure at the lower boundary. This gravel layer is separated from the sand via a geotextile and is the only connection of the site to a groundwater well. The groundwater well is in particular used to manipulate the groundwater table by pumping water in and out of the well. The groundwater table in the test site is automatically determined with a tensiometer (UMS T4-191). The test site incorporates 32 soil temperature and TDR sensors which are operated via a weather station and a Campbell Scientific TDR100. The site is confined by a basement layer below the gravel layer and by a wall at each of the four sides. During the construction, the materials were compacted with a vibrating plate for stabilization.

2.2 Representation

Quantitative understanding of a system of interest requires its mathematical representation. Based on Bauser et al. (2016), we define the representation of a system as a set consisting of: (i) the dynamics corresponding to the mathematical propagation of the variable of interest at predefined scales in space and time, (ii) the coupling to the sub-scale physics through typically heuristic material properties, (iii) the coupling to the super-scale physics through the forcing in space and time, and (iv) the state corresponding to the variable of interest that is propagated by the dynamics.

2.2.1 Dynamics

The standard model to describe the volumetric water content θ $(-)$ and the matric head h_m (m) in space and time t (s) is the Richards equation (Richards, 1931),

$$\partial_t \theta - \nabla \cdot [K_\mathrm{w}(\theta)[\nabla h_\mathrm{m}(\theta) - \boldsymbol{e}_z]] = 0. \tag{1}$$

To solve this partial differential equation, the soil water characteristic $\theta(h_\mathrm{m})$ and the hydraulic conductivity function $K_\mathrm{w}(\theta)$ are required. The direction of gravity is indicated with the unit vector in z-direction $\boldsymbol{e}_z$.
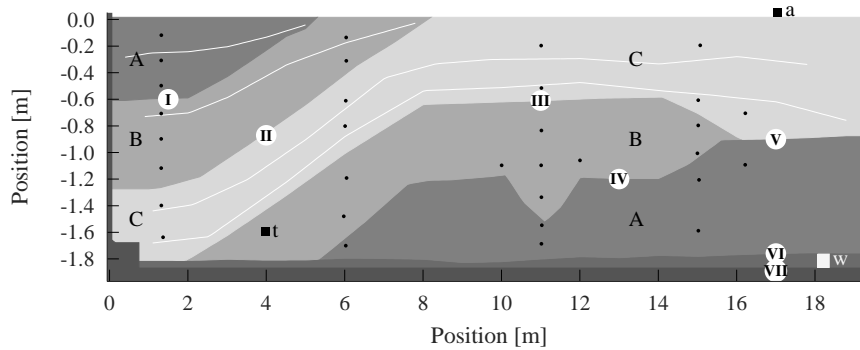
Figure 1. ASSESS provides an effectively 2D subsurface architecture consisting of three kinds of sand (A, B, and C). During the experiment that is evaluated in this work, the groundwater table was manipulated via a groundwater well (white square at $18.3\,\mathrm{m}$). The resulting hydraulic dynamics was monitored automatically with a stationary GPR antenna (black square at $17.05\,\mathrm{m}$), a tensiometer (black square, at $4.0\,\mathrm{m}$), and 32 TDR sensors (black dots). A gravel layer below the sands allows for a rapid water pressure distribution at the lower boundary of the site. An L-element (left wall, at $0.4\,\mathrm{m}$) and compaction interfaces (white lines) result from the construction of the site. Additionally to those visualized, GPR evidence indicates additional compaction interfaces (Fig. 14) that were not determined during construction. Roman numbers (I)–(VII) indicate material interfaces referred to in the text. Note the different scales on the horizontal and vertical axes.

2.2.2 Sub-scale physics

We choose the Brooks–Corey parameterization (Brooks and Corey, 1966) for the soil water characteristic $\theta(h_\mathrm{m})$, since it describes the materials in ASSESS appropriately (Dagenbach et al., 2013). Inverting this parameterization for $\theta_\mathrm{r} \leq \theta \leq \theta_\mathrm{s}$ leads to

$$h_\mathrm{m}(\theta) = h_0 \left(\frac{\theta - \theta_\mathrm{r}}{\theta_\mathrm{s} - \theta_\mathrm{r}} \right)^{-1/\lambda}, \tag{2}$$

with a saturated water content θ_s $(-)$, a residual water content θ_r $(-)$, a scaling parameter h_0 (m) related to the air entry pressure ($h_0 < 0\,\mathrm{m}$), and a shape parameter λ $(-)$ related to the pore size distribution ($\lambda > 0$).

The hydraulic conductivity function $K_\mathrm{w}(\theta)$ is parameterized combining the Brooks–Corey parameterization with the hydraulic conductivity model of Mualem (1976). This yields

$$K_\mathrm{w}(\theta) = K_s \left(\frac{\theta - \theta_\mathrm{r}}{\theta_\mathrm{s} - \theta_\mathrm{r}} \right)^{\tau + 2 + 2/\lambda}, \tag{3}$$

with the saturated hydraulic conductivity K_s $(\mathrm{m\,s^{-1}})$ and a turtuosity factor τ $(-)$.

2.2.3 Forcing

In order to provide the measurement data for this study, ASSESS was forced with a fluctuating groundwater table leading to two characteristic phases comprising an initial drainage phase and a multistep imbibition phase (Fig. 2). The experiment was
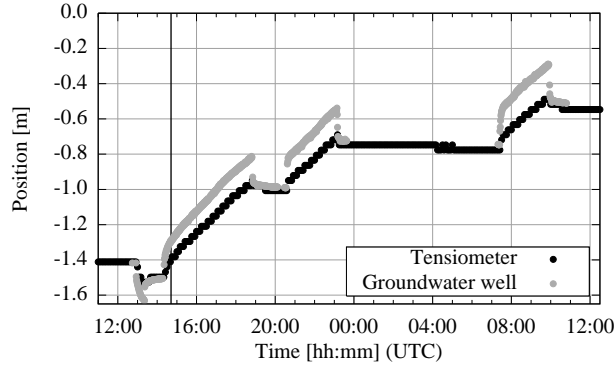
Figure 2. During the experiment with two distinct phases (initial drainage and multistep imbibition – separated by the vertical line), the position of the groundwater table was measured manually in the groundwater well and automatically with the tensiometer (Fig. 1). The difference between them is proportional to the driving force of water flow in the gravel layer.

realized at the end of November and the weather was cloudy with 2–7 °C air temperature. Hence, evaporation is neglected in this work. Further details about the experiment are given by Jaumann and Roth (2017).

2.2.4 State

During the experiment, the groundwater table was measured (i) automatically with a tensiometer and (ii) manually via the
5 groundwater well (Fig. 2). The hydraulic state was monitored with a stationary GPR antenna (Fig. 1). This shielded bistatic single-offset 400 MHz GPR antenna (Ingegneria dei Sistemi S.p.A., Italy) has an internal antenna separation of 0.14 m. The measurement resolution was set to 2048 samples for 60 ns. The acquired GPR data are analyzed in detail in Sect. 3.3. Additionally, a mean soil temperature ($T_\mathrm{s} = 8.5\,$°C) and a mean electrical conductivity ($\sigma = 0.003\,\mathrm{S\,m^{-1}}$) was estimated from TDR related measurements available in ASSESS. The electrical conductivity was evaluated from the TDR pulse shape and
10 thus includes the direct current conductivity as well as dielectric losses. For this evaluation and the corresponding temperature correction, we implemented the methods of Heimovaara et al. (1995).

The observation operator required to compare the simulated hydraulic state with the GPR measurement data involves the solution of the time-dependent Maxwell's equations in linear macroscopic isotropic media. These equations quantify the propagation of the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ (Jackson, 1999):

15 $$\nabla \times \frac{\boldsymbol{B}}{\mu} - \varepsilon \frac{\partial \boldsymbol{E}}{\partial t} = \sigma \boldsymbol{E} + \boldsymbol{J}, \tag{4}$$

$$\nabla \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0. \tag{5}$$

The dielectric permittivity $\varepsilon = \varepsilon_0 \varepsilon_r$, magnetic permeability $\mu = \mu_0 \mu_r$, and electrical conductivity σ are generally spatially variable and represent the electromagnetic properties of the subsurface. Here, we use the static relative permittivity $\varepsilon_r = \lim_{\omega \to 0} \varepsilon_r(\omega)$, neglect dispersive effects ($\partial \varepsilon_r / \partial \omega = 0$), and assume the permittivity to be real valued ($\varepsilon_r \in \mathbb{R}$). The relative

magnetic permeability is assumed to be that of vacuum ($\mu_r = 1$). The source current density $\boldsymbol{J}$ is applied at the position of the transmitter antenna.

The relative permittivity of the subsurface $\varepsilon_r = \varepsilon_{r,b}$ is calculated from the water content distribution using the complex refractive index model (CRIM, Birchak et al., 1974):

$$\sqrt{\varepsilon_{r,b}(\theta, T_s, \phi)} = \theta \cdot \sqrt{\varepsilon_{r,w}(T_s)} + (\phi - \theta) \cdot \sqrt{\varepsilon_{r,a}} + (1 - \phi) \cdot \sqrt{\varepsilon_{r,s}}. \tag{6}$$

The application of the CRIM requires knowledge of the porosity ϕ, the relative permittivity of water $\varepsilon_{r,w}$, the relative permittivity of air $\varepsilon_{r,a}$, and the relative permittivity of the soil matrix ε_s. The relative permittivity of air $\varepsilon_{r,a}$ was set to 1. Since we assume that the soil matrix consists mainly of Quartz (SiO_2) grains, the relative permittivity of the soil matrix $\varepsilon_{r,s}$ was set to 5 (Carmichael, 1989). We further assume that porosity ϕ is equal to the saturated water content θ_s (Eq. 2). The dependency of the static relative permittivity of water $\varepsilon_{r,w}$ on the soil temperature T_s (°C) is parameterized following Kaatze (1989)

$$\varepsilon_{r,w}(T_s) = 10^{1.94404 - T_s \cdot 1.991 \cdot 10^{-3}}. \tag{7}$$

The electrical conductivity σ of the subsurface is assumed to be constant in space and time. As for the relative permittivity, we neglect dispersive effects of the electrical conductivity ($\partial\sigma/\partial\omega = 0$) and assume it to be real valued ($\sigma \in \mathbb{R}$).

2.3 GPR Analysis

To analyze the GPR data, we follow Buchner et al. (2012) and extract the travel time t and the corresponding amplitude A for M samples of the GPR signal (events)

$$E_x \mapsto \left\{ \begin{pmatrix} t_1 \\ A_1 \end{pmatrix} \cdots \begin{pmatrix} t_M \\ A_M \end{pmatrix} \right\} \tag{8}$$

with a heuristic approach, because this allows us to focus on the phenomena that are represented in the model. However, to associate the events extracted from the measured signal with events extracted from the simulated signal, this procedure demands an automatic event association algorithm. Thus, the evaluation method consist of four steps: (i) signal processing, (ii) event detection, (iii) event selection, and (iv) event association (Fig. 3).

2.3.1 Signal processing

The processing of the GPR signal includes the following step: (i) time-zero correction, (ii) dewow filter, (iii) 2D to 3D conversion, (iv) removal of the direct signal (particularly including the direct wave and the ground wave) and the trailing signal (signal at the end of the trace which is disturbed by the dewow filter), and (v) normalization (Fig. 3).

Since the time-zero of the GPR antennas changes over time, we pick the direct signal and subtract the evaluated travel time from each trace of the radargram for time-zero correction. Subsequently, a dewow filter is applied to subtract inherent low-frequency wow noise of the GPR signal. Because the observation is in 3D and the simulation in 2D, we convert the simulated signal to 2.5D, meaning to 3D with translational symmetry perpendicular to the survey line and parallel to the ground surface
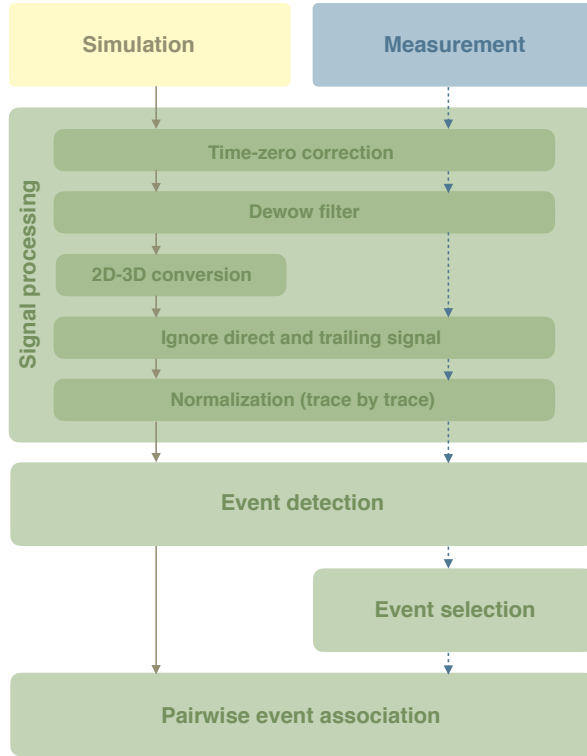
Figure 3. The GPR data evaluation method presented in this section consists of four main steps. In the first step, the signal is processed (Sect. 2.3.1). The 2D to 3D conversion in this step is applied to the simulated data. In the second step, extrema in the GPR signal are detected (Sect. 2.3.2). The detected events in the measurement data can be selected manually for the subsequent evaluation (Sect. 2.3.3). This ensures that the optimization focuses only on the most relevant information in the data. Finally, the most plausible association of simulated and measured events is determined (Sect. 2.3.4). Note that for each parameter set that is tested during the optimization procedure, the according simulation data is evaluated automatically (solid lines). In contrast, the measurement data is only evaluated once before starting the optimization procedure (dashed lines).

(Bleistein, 1986). The ASSESS site conforms to this 2D requirement (Sect. 2.1). For the conversion, each trace is transformed to the frequency domain with the fast Fourier transform (FFT, denoted by $\widehat{}$). Afterwards, the electric field is modified depending on the angular frequency ω:

$$\widehat{E}_x^{\mathrm{3D}} = \widehat{E}_x^{\mathrm{2D}} \cdot C_i \cdot \sqrt{\frac{|\omega|}{2\pi\sigma_c}} \exp\left(-\frac{i\pi}{4}\mathrm{sign}(\omega)\right), \tag{9}$$

where i is the complex unit, C_i is a constant (m), and σ_c denotes the integral of the velocity with respect to the length s of the ray trajectory. Assuming a direct ray path and a horizontal reflector with the reflector distance d as well as the mean square

root of dielectric permittivity $\overline{\sqrt{\varepsilon_r}}$ along the ray path, this is leads to

$$\sigma_c = \int c(s)\mathrm{d}s = \frac{c_0 d}{\overline{\sqrt{\varepsilon_r}}}. \tag{10}$$

Subsequently, all traces are transformed back to the time domain with the inverse FFT. Due to the frequency conversion and the manipulation, a high frequency noise remains on the signal which is smoothed with a fourth order Savitzky–Golay filter (e.g., Press, 2007, we employed the implementation of the 'signal' package for GNU Octave: https://octave.sourceforge.io/signal/) using a window width of 41 samples. Subsequently, the direct signal and the trailing signal of the dewow filter are set to zero. Finally, the each trace is normalized to its maximal absolute amplitude since the absolute power of the GPR source is typically unknown. Thus, also the value of the constant C_i in Eq. 9 becomes irrelevant.

2.3.2 Event detection

In this step, events are detected in each trace separately. To facilitate the detection of relevant events at large signal travel times, the amplitude of the processed signal (Sect. 2.3.1) is amplified quadratically with travel time using an arbitrary gain function that showed to work well. Subsequently, the extrema of the amplified amplitude are detected with a local neighborhood search. We keep a predefined number of events (15) with the largest amplified absolute amplitude. If the non-amplified amplitude of a detected extremum is below a predefined amplitude threshold (0.006), the event is discarded in any case. In order to correct the perturbation in travel time due to the amplification and to cope with the discrete measurement resolution, we fit a Gaussian centered at the travel time of the detected event with width of ± 5 samples to the non-amplified amplitude of the processed signal. The travel time of the resulting extremum of the Gaussian fit is directly used for the following evaluation. The according amplitude of the extremum is used as the amplitude of this event. This procedure makes the form of the previously applied gain function irrelevant. The amplitude all detected events is normalized with the absolute maximal amplitude of all detected events within the same trace.

2.3.3 Event selection

After the event detection, the measured signal and the detected events (Sect. 2.3.2) are inspected manually. In this one-time processing step events can either be deleted or added manually. Thus, it can be ensured, that only those events enter the parameter estimation that are also represented in the model. This step is skipped for the analysis of the simulated data. The resulting amplitude of the selected events is normalized with the absolute maximal amplitude of all selected events of each trace.

2.3.4 Pairwise event association

The selected events extracted from the measured data have to be associated with the detected extracted from the simulated data for the parameter estimation. To this end, Buchner et al. (2012) tested all possible combinations of events, using the one with the minimal summed absolute travel time difference. However, this is only feasible for a small number of events. As we
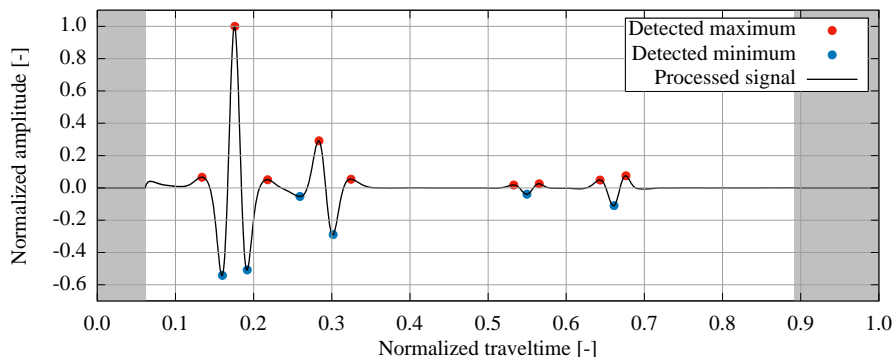
Figure 4. The detected events of the first trace of the synthetic radargram analyzed in Sect. 3.2. The amplitude of a trace is searched for extrema with a neighborhood search algorithm. For the subsequent evaluation, the amplitude of the detected events is normalized to the maximal absolute amplitude of all events detected in the trace. The direct signal and the trailing signal of the dewow filter with normalized travel times < 0.06 and > 0.89, respectively, are set to zero in a processing step (Sect. 2.3.1) and possible events close to these signals are ignored.

are not using a Gaussian convolution of the data but the data themselves, the number of events increases. Hence, testing all combinations is often prohibitively expensive.

In order to exclude combinations a priori, the detected events are aggregated in clusters (Fig. 5a). Then, these clusters are associated by testing all possible combinations and finally using the combination with the minimal summed absolute travel
5 time difference. Afterwards, those events, that are aggregated in the associated clusters, are associated themselves. The applied association procedure requires the events to have an identical amplitude sign and a consistent temporal order which reflects the principle of causality (Fig. 5b). After iterating over all allowed combinations, the association with maximal number of associated events and the minimal summed absolute travel time difference is used. It is critical to also consider combinations where some intermediate events (e.g., $(t_{s,2}, A_{s,2})$ in Fig. 5) can not be associated.
10 After the association of the events, outliers are detected by calculating the mean and standard deviation of the travel time differences. All associations that exhibit an absolute travel time difference larger than 3 standard deviations of all absolute travel time differences are discarded. Finally, the amplitude of the associated events is normalized to the maximal absolute amplitude of the associated events of each trace separately for the simulation and the measurement.

2.4 Parameter estimation

15 Inversion of GPR data typically requires globally convergent parameter estimation algorithms which are computationally expensive. In order to keep the parameter estimation procedure efficient, we use an iterative strategy (Fig. 6). We start the optimization procedure by drawing an ensemble of initial parameter sets with a Latin hypercube algorithm (implemented by the pyDOE package, https://github.com/tisimst/pyDOE).
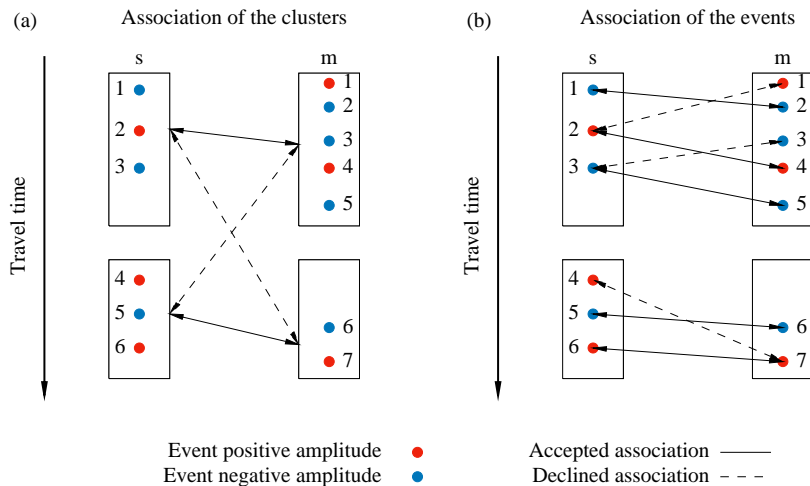
Figure 5. Exemplary association of simulated (s) and measured (m) events with indices 1–6 and 1–7, respectively. The color of the dots indicates the sign of amplitude of the events. (a) The detected events (Fig. 4) are aggregated in clusters to minimize the number of possible event combinations. The clusters are associated such that the summed absolute travel time difference of the mean travel time of the events in the cluster is minimal. (b) The events in the clusters are associated according to consistent temporal order and amplitude sign. Hence, if $(t_{s,1}, A_{s,1})$ is associated with event $(t_{m,2}, A_{m,2})$, event $(t_{s,2}, A_{s,2})$ can not be associated with event $(t_{m,1}, A_{m,1})$, if $t_{m,1} < t_{m,2}$ or $\text{sign}(A_{s,2}) \neq \text{sign}(A_{m,1})$. Solid (dashed) arrows indicate some of the accepted (declined) association combinations. The combination with maximal number of associations and minimal summed absolute travel time difference is used for evaluation. Thus, for example $(t_{s,1}, A_{s,1})$ is associated with $(t_{m,2}, A_{m,2})$ and not with $(t_{m,3}, A_{m,3})$.

The most expensive operation of the forward simulation is the calculation of the observation operator, which includes the solution of Maxwell's equations (Sect. 2.2.4) and the subsequent event association (Sect. 2.3). Since time-lapse GPR data are highly correlated in experiment time (e.g., Fig. 14b), we equidistantly subsample the number of traces of the time-lapse GPR radargram and generate a data set with lower temporal resolution. Those data are used to improve the distribution of the initial parameters (preconditioning). Therefore, the drawn parameter sets are used to initialize the simulated annealing algorithm (Sect. 2.4.2) which allows for a robust and fast parameter update. The resulting parameters then serve as initial parameters for the Levenberg–Marquardt algorithm (Sect. 2.4.3) which concludes the preconditioning step. The resulting parameters of the preconditioning step are used as the initial parameter sets for the more expensive optimization of high resolution data set with the Levenberg–Marquardt algorithm. The details of the setup and the analysis of the parameter estimation are given in Sect. 3.1.2.
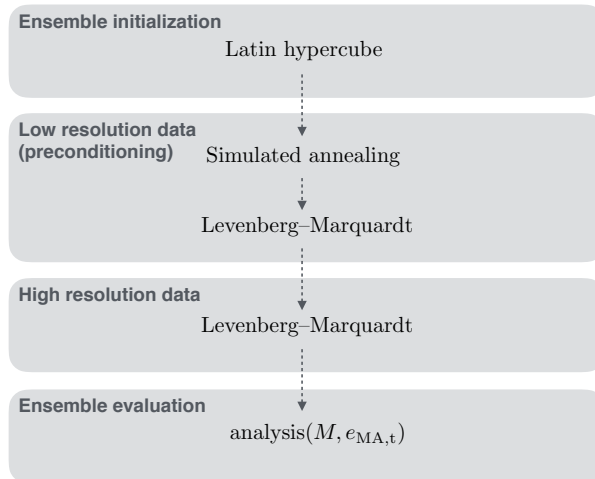
Figure 6. We choose a sequentially coupled parameter estimation procedure which (i) allows to minimize the computational cost and (ii) facilitates the implementation of tagging (Sect. 2.4.1). Therefore, Latin hypercube sampled parameter sets are preconditioned with a data set with reduced number of traces (low resolution data) by sequentially coupling the simulated annealing algorithm and the Levenberg–Marquardt algorithm. The preconditioned parameter sets for each ensemble member serve as initial parameters for the final parameter estimation based on high resolution data. The subsequent evaluation of the ensemble is based on the number of associated events M and the mean absolute error in travel time $e_{\mathrm{MA,t}}$ (Sect. 3.1.2).

2.4.1 Cost function

Assuming P parameters p_π $(1,\ldots,\pi,\ldots,P)$ and M associations $(1,\ldots,\mu,\ldots,M)$ of measured events $(t_{\mu,\mathrm{m}}, A_{\mu,\mathrm{m}})$ with simulated events $(t_{\mu,\mathrm{s}}(\boldsymbol{p}), A_{\mu,\mathrm{s}}(\boldsymbol{p}))$, the cost function $S(\boldsymbol{p})$ is given by

$$S(\boldsymbol{p}) = \frac{1}{2}\sum_{\mu=1}^{M}\left(\frac{t_{\mu,\mathrm{s}}(\boldsymbol{p})-t_{\mu,\mathrm{m}}}{\sigma_{t,\mu}}\right)^2 + \left(\frac{A_{\mu,\mathrm{s}}(\boldsymbol{p})-A_{\mu,\mathrm{m}}}{\sigma_{A,\mu}}\right)^2 = \frac{1}{2}\sum_{\mu=1}^{M}\left(r_{t,\mu}^2 + r_{A,\mu}^2\right) \tag{11}$$

5   with the constant standard deviation of the measured normalized travel times $\sigma_{t,\mu} = \sigma_t$ and of the measured normalized amplitudes $\sigma_{A,\mu} = \sigma_A$. This leads to the standardized residuals in travel time $r_{t,\mu}$ and amplitude $r_{A,\mu}$.

Due to the oscillating nature of the GPR signal and due to the applied GPR data evaluation (Sect. 2.3), the cost function is not necessarily convex and may even be discontinuous at some points, because the number of associated events M may change during the minimization process. Hence, adding and removing associations of events requires a compensation to prevent the

10   cost function from becoming discontinuous. To this end, Buchner et al. (2012) introduced *tagging*. If the number of measured events is smaller than the number of the simulated events, the simulated events that are not associated are excluded. Alternatively, if there are more measured events, measured events without partner are tagged as partnerless. If a reflection event has been tagged and becomes untagged after the parameter update, the contribution of the event and its new partner to the cost

function is added to the previous cost. If an event has not been tagged and becomes tagged after the parameter update, the contribution to the cost function is subtracted from the previous cost.

2.4.2 Simulated annealing

We choose the simulated annealing algorithm (Press, 2007) to start the minimization of the cost function (Eq. 11), because this algorithm is gradient-free and updates the parameters randomly. Besides the implementation of tagging (Sect. 2.4.1), which can be implemented easily for this algorithm, no further regularization is required due to its random parameter update. Hence, this algorithm is in particular suitable for initial iterations where the association of the single events may lead to an inconsistent association of the reflections. Once the reflections are associated consistently, the Levenberg–Marquardt algorithm is typically more efficient than the simulated annealing algorithm.

If the parameter update is drawn from the whole parameter space, the simulated annealing algorithm is globally convergent. However, this approach is typically inefficient. Hence, we search the neighborhood for better parameters starting from Latin hypercube sampled initial parameters $p_{\pi,0}$. For each iteration i $(1,\ldots,I)$, new parameters are proposed randomly via

$$p_{\pi,i+1} = p_{\pi,i} + m \cdot (p_{\pi,\max} - p_{\pi,\min}) \cdot u_{\mathrm{p}}, \tag{12}$$

with a mobility parameter $m = 0.1$, uniformly distributed random number $u_{\mathrm{p}} \sim \mathcal{U}(-1,1)$, and the parameter limits $p_{\pi,\max}$ and $p_{\pi,\min}$. In order provide the control parameter T, which is an analogue of temperature, we choose an exponential cooling schedule

$$T_{i+1} = T_0 \cdot \alpha^{i+1}, \tag{13}$$

with $\alpha = 0.85$ and initial temperature $T_0 = 10^3$ which is of the order of the initial cost. According to Metropolis et al. (1953), we draw an uniformly distributed random number $u_{\mathrm{d}} \sim \mathcal{U}(0,1)$, calculate the acceptance probability

$$P_{i+1} = \exp\left(-\frac{S(\boldsymbol{p}_{i+1}) - S(\boldsymbol{p}_i)}{k \cdot T_{i+1}}\right), \tag{14}$$

choosing parameter $k = 1$, and accept the proposed parameter set if $P_{i+1} > u_{\mathrm{d}}$ or else draw a new parameter set.

2.4.3 Levenberg–Marquardt

The Levenberg–Marquardt algorithm is implemented as described by Jaumann and Roth (2017). The application of this gradient-based algorithm on GPR data requires the implementation of tagging (Sect. 2.4.1) as well as additional regularization of the optimization. This regularization can be achieved by focussing in particular on the improvement of the small residuals, because if the small residuals improve, the larger residuals are likely to also improve in subsequent iterations due to the temporal correlation of the time-lapse data. Therefore, events with $r_{t,\mu} > 100$ or $r_{A,\mu} > 100$ are tagged. Tagged events are excluded from the optimization by setting the according entries in the Jacobi matrix ($J_{\mu,\pi} = \partial r_{\mu}/\partial p_{\pi}$) to zero. The event association may also change during the perturbation of the parameters for the numerical assembly of the Jacobi matrix. This can lead to large changes in the residuals, which in turn may lead to a disturbed parameter update. Hence, corresponding entries

of large changes in the residual $|r_\mu(\boldsymbol{p}_{\text{perturbed}}) - r_\mu(\boldsymbol{p})| > 50$ are also set to zero together with entries of the Jacobi matrix that are larger than 10^4.

We choose $\lambda_{\text{LM,initial}} = 5$ as initial value for λ_{LM} and apply the delayed gratification method by decreasing (increasing) λ_{LM} by a factor of 2 (3) if the parameter update is successful (not successful). This assures that the algorithm takes small steps such that association and the Jacobi matrix can adapt smoothly.

3 Application

In this section, the methods presented in the previous section are applied to GPR data. First, the setup of the case study, its implementation and the detailed setup of the parameter estimation procedure are explained (Sect. 3.1). Subsequently, the suitability of the presented methods to estimate the subsurface architecture and the corresponding soil hydraulic material properties is first tested with synthetic data (Sect. 3.2). Finally, the methods are applied to measurement data (Sect. 3.3).

3.1 Setup of the case study

3.1.1 Implementation

In this work, the subsurface architecture of ASSESS is represented with layers. The position of these layers is parameterized and can be estimated. For illustration, the setup is shown in Fig. 7. The Richards equation (Eq. 1) is solved numerically with $\mu\varphi$ (muPhi, Ippisch et al., 2006) which uses a cell centered finite volume scheme with full upwinding in space and an implicit Euler scheme in time. The nonlinear equations are linearized by an inexact Newton method with line search and the linear equations are solved with an algebraic multigrid solver. We solve the Richards equation in 1D (z-dimension) on a structured grid with a resolution of ≈ 0.005 m. The simulated water content is converted to relative permittivity via the CRIM using the mean soil temperature $T_\text{s} = 8.5$ °C (Sect. 2.2.4). Since Maxwell's equations are solved in 2D, the simulated 1D permittivity distribution is extruded in the y-dimension using the same spatial resolution. The total size of the represented subsurface architecture is $1.0\,\text{m} \times 1.9\,\text{m}$. Generally, the boundary condition is implemented with a Neumann no-flow condition. However, during the forcing phases, we prescribe the measured groundwater table as a Dirichlet boundary condition at the position of the groundwater well. We initialize the simulation with hydraulic equilibrium based on the measured groundwater table position.

To simulate the temporal propagation of the electromagnetic signal, we solve Maxwell's equations (Sect. 2.2.4) in 2D with the MIT electromagnetic equation propagation software (MEEP, Oskooi et al., 2010). The transmitter antenna is represented with an infinite dipole pointing in x-dimension (Fig. 7). Thus, we neglect any effects from the real antenna geometry (bow tie), cross coupling or antenna shielding. The antenna source current density $\boldsymbol{J}$ is given by the first derivative of a Gaussian–shaped excitation function leading to a Ricker wavelet with a center frequency of 400 MHz. The receiver antenna is not represented explicitly. Instead, E_x is read directly at the position of the receiver antenna. We use the antenna separation of the real GPR system (0.14 m) in the simulation. Perfectly matched layers (PML) of 0.15 m thickness serve as boundary condition. The electromagnetic fields in the domain are zero initially.

We use one tenth of the minimal wavelength $\lambda_{w,min}$ as upper limit for the spatial resolution Δz:

$$\Delta z \leq \frac{\lambda_{w,min}}{10} = \frac{\frac{c_0}{\sqrt{\varepsilon_{r,max}}}}{10 f_{max}} \approx 0.007 \text{ m}, \tag{15}$$

with the speed of light in vacuum c_0, maximal frequency $f_{max} = 2 \cdot 400$ MHz, and $\varepsilon_{r,max} = 31.25$ corresponding to $\theta_{s,max} = 0.5$. Hence, we choose the numerical resolution $\Delta z = 0.005$ m for the 2D isotropic, structured, and rectangular grid. The Courant number for the FDTD method is set to 0.5.

To avoid multiple reflections at the air-soil boundary, we set the relative permittivity above the soil to 3.5, which is typical for dry sand. This is justified, because the air wave and the ground wave are not evaluated and the amplitude is normalized according to the detected events. The permittivity of the basement below ASSESS is set to 23.0 based on previous simulations. The electrical conductivity of the subsurface σ is set to 0.003 Sm^{-1} (Sect. 2.2.4). All electromagnetic properties are smoothed by MEEP according to Farjadpour et al. (2006).

The GPR data are evaluated according to Sect. 2.3. The detection of the clusters is chosen to be identical for the simulated and the measured data. The characteristic shape of the time-lapse radargram allows to separate the clusters at a specific normalized travel time ($t_{split} = 0.5$) for all traces. Hence, all events with a travel time $t \leq t_{split}$ are in cluster 1 and the others are in cluster 2.

3.1.2 Setup of the parameter estimation

The general setup of the optimization is explained with Fig. 8. This setup is used in a sequential approach (Fig. 6) with a preconditioning step for which we subsampled the number of the traces of the time-lapse GPR data to generate a data set with decreased temporal resolution. The data set with high (low) resolution includes 86 (9) traces corresponding to one trace per 15 (150) min. Hence, both data sets subsample the actually measured number of traces (one trace per ≈ 30 s) equidistantly in time. Within the sample range given in Table 1, 60 initial parameter sets are drawn with the Latin hypercube algorithm and the data set with low temporal resolution is used to improve these parameter sets running 200 iterations of the simulated annealing algorithm. The parameter fit range given in Table 1 determines the parameter update via $p_{\pi,max}$ and $p_{\pi,max}$ according to Eq. 12. After the application of the simulated annealing algorithm, maximally 15 iterations of the Levenberg–Marquardt algorithm are run. This optimization completes the preconditioning step. The resulting parameter sets serve as initial parameters for the Levenberg–Marquardt algorithm which is applied to the data with high temporal resolution.

In order to evaluate the performance of the ensemble members, the mean absolute error in normalized travel time $e_{MA,t}$ is used since this statistical measure is independent of the number of associated events. The number of associated events is accounted for by evaluating only those 10 members with minimal $e_{MA,t}$ that associated at least 85% of the measured events. Each of these members has locally optimal parameters. However, the exact position of these local minima typically depends on the initial parameters and the random numbers drawn in the simulated annealing algorithm. There is also no guarantee that the global optimum was found by one of the ensemble members. However, the distribution of these 10 best ensemble members contains valuable information about the shape of the cost function. To account for this information, we (i) analyze the mean parameter set of the best members and (ii) use the according standard deviation to indicate the uncertainty of these parameters.
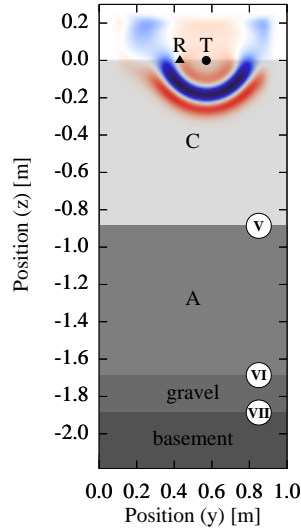
Figure 7. For the simulation of the GPR signal, we assume a layered subsurface architecture (Fig. 1). The transmitter of the antenna is represented with an infinitesimal dipole (T) and the electric field is read at the position of the receiver antenna (R). A PML is used as boundary condition. The x-component of the resulting electric field after 5 ns is shown. The markers for the material interfaces are used consistently in this work. The position of the interface of materials C and A (V) as well the position of the interface of materials A and gravel (VI) are parameterized (d^{V} and d^{VI}) and can be estimated.

Notice that the mean parameter set is not necessarily optimal. However, if uncertainty of the resulting parameters is small, the mean parameter set is typically more reliable than the parameter set of the best ensemble member.

The standard deviation of the measured data, $\sigma_t \approx 6 \cdot 10^{-4}$ and $\sigma_A \approx 5 \cdot 10^{-3}$ for normalized travel times and amplitudes, respectively, are used to standardize the residuals of the cost function (Eq. 11). These standard deviations were calculated from
5   travel time and amplitude data acquired by picking different reflections with approximately constant travel time. In order to perturb the travel time and amplitude of the selected events of the synthetic measurement data (Sect. 3.2), a realization of white Gaussian noise with the according standard deviation is added to the data.

3.2 Synthetic data

In this section, the synthetic data generated along the lines given in Sect. 3.1 are first analyzed qualitatively (Sect. 3.2.1).
10   Afterwards, these data are used to estimate a layered subsurface architecture and the corresponding soil hydraulic material properties using the methods proposed in Sect. 2. The results of this inversion are discussed in Sect. 3.2.2.
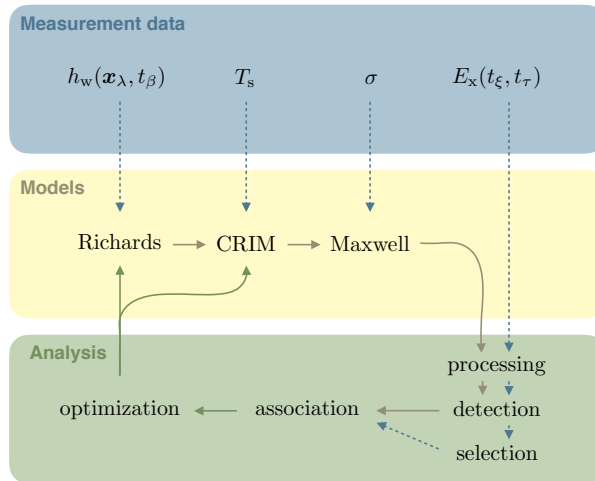
Figure 8. The general setup of the optimization is sketched with this figure. The available hydraulic potential h_{wt} is measured at the position of the groundwater well $\boldsymbol{x}_\lambda$ times t_β. These measurements are used as a boundary condition for the Richards equation (Sect. 2.2.1). Estimates for the soil temperature T_s and the electrical conductivity σ are derived from TDR related measurements. The actual signal of the GPR system is proportional to the x-component of the electric field E_x and measured discretely at experiment time t_ξ and signal travel time t_τ. This signal is processed (Sect. 2.3.1) and used for event detection (Sect. 2.3.2). Based on the detected events in the measurement data, other events can be either added or deleted in the subsequent event selection step (Sect. 2.3.3). The simulated water content distribution is converted to relative permittivity distribution with the CRIM and used to solve Maxwell's equations (Sect. 2.2.4 and 3.1.1). After the processing step and the event detection, the simulated events are assigned to measured events (Sect. 2.3.4). The resulting mapping of the events is used to calculate the cost in the optimization step (Sect. 2.4). Dashed arrows indicate initial processing steps, whereas solid arrows indicate iterative steps required for the optimization.

3.2.1 Phenomenology

The phenomenology of the transition zone reflection for characteristic times during imbibition, equilibration, and drainage was discussed by Klenk et al. (2015) for typical coarse sand. Here, we focus on the temporal development of this reflection during imbibition and equilibration. To this end, the water content distribution of the 1D profile located at $17.05 \ \mathrm{m}$ of the ASSESS site

5 (Fig. 1) was simulated using typical parameters for coarse-textured sandy soils. These parameters are given together with the estimated parameters in Table 2. The groundwater table measured in the well (Fig. 2) is used as Dirichlet boundary condition at the bottom boundary. The resulting simulation is visualized over time and over water content in Fig. 9.

Initialized with static hydraulic equilibrium, the simulation starts with an initial drainage step where the groundwater table is lowered. Hence, the material at the upper end of the capillary fringe with high initial water content is desaturated. After

10 the subsequent equilibration step, the groundwater table is raised during the subsequent imbibition step. The Brooks–Corey parameterization (Eq. 2) features a sharp kink where air enters the material at the upper end of the capillary fringe. Furthermore, the imbibition introduces an additional kink in the water content distribution (marker (2) in Fig. 9b), because the relaxation time
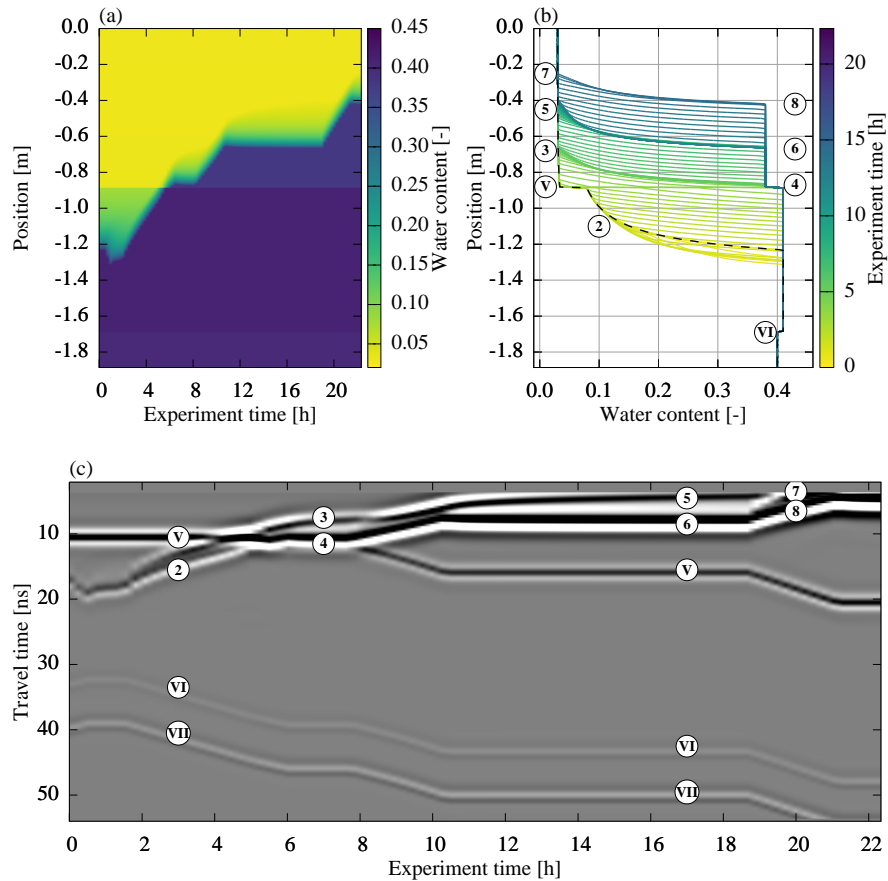
Figure 9. The true synthetic data are simulated with hydraulic parameters that represent coarse-textured sandy soils (Table 2). Subfigures (a) and (b) show different representations of the simulated water content. In subfigure (b) the initial water content distribution is marked with a black dashed line. Subfigure (c) shows the according simulation of the GPR signal. The imbibition leads to a characteristic transition zone reflection (marker (2)). The temporal evolution of this reflection is sensitive to the initial water content distribution, the soil water characteristic and the hydraulic conductivity function. The data shown are processed according to Sect. 2.3 except for the normalization. In contrast to the data that are evaluated, the shown radargram is normalized to the maximum absolute amplitude of all traces, facilitating the visual comparison of the traces. The markers are defined in Fig. 14.

from hydraulic non-equilibrium is much shorter for high water contents compared to the relaxation time for low water contents. This is due to the non-linear dependency of the hydraulic conductivity (Eq. 3) on the water content leading the differences in hydraulic conductivity of several orders of magnitude. Hence, the width of the transition zone is decreased during the imbibition phase.

During the equilibration step after the first imbibition, the additional kink smoothes. Thus, the water content increases in the material with low water content (3) and decreases in the material at the upper end of the capillary fringe (4). This smoothing depends on both the soil water characteristic and the hydraulic conductivity function. Sharpening and smoothing of the transition zone are repeated consistently for the other subsequent imbibition and equilibration phases ((5), (6) and (7), (8)).

According to the CRIM (Sect. 2.2.4), the relative permittivity distribution has the same shape as the water content distribution. Hence, kinks in the water content distribution directly induce partial reflections of the GPR signal (Fig. 9c). Shortly after starting the imbibition, the amplitude of the reflection at the additional kink (2) increases. After passing the material interface (V), the spatial distance of the kinks increases such that the two resulting reflection wavelets (3) and (4) are separable. Note that since the water content changes continuously, the signal in between these wavelets is a superposition of infinitesimal reflections which also contain information about the form of the transition zone. Additionally, the reflection (3) scans the initial water content distribution, which in steady state corresponds to the soil water characteristic. With progressing equilibration, the amplitude of reflection (3) decreases as the transition zone smoothes. The GPR signal of the subsequent imbibition and equilibration phases ((5), (6) and (7), (8)) show similar behavior and emphasize the relatively long time scale for hydraulic equilibration of sandy materials.

In summary, this numerical simulation confirms qualitatively (i) that the dynamics of the fluctuating groundwater table is sensitive to both the soil water characteristic and the hydraulic conductivity function and (ii) that the transition zone reflection leads to tractable reflections during the imbibition step.

3.2.2 Results and discussion

After the inversion of the synthetic data, we find that the resulting soil water characteristics for material A (Fig. 10a) exhibit a similar curvature but are shifted. Both the parameters h_0 and λ influence the shape of the desaturated transition zone. Hence, merely evaluating the shape of the desaturated part of the transition zone is not necessarily sufficient to uniquely identify both parameters leading to large correlation coefficients. However, parameter h_0 additionally determines the extent of the capillary fringe. If the evaluation is also sensitive to the extent of the capillary fringe, h_0 can be uniquely identified which significantly decreases the correlation between h_0 and λ. Hence, we conclude that the strong correlation of the parameters h_0^C and λ^C (-0.7, Fig. 11) indicates that the evaluation is more sensitive to the shape of the desaturated part of the transition zone than to the extent of the capillary fringe.

Since the architecture is a layered structure where material C is located above material A (Fig. 7), the water content in material C contributes to the travel time of the other reflections. This introduces correlations of θ_s^A with all the parameters associated with the soil water characteristic of material C. A high correlation of parameters indicates that the problem is not well-posed. This typically increases the number of local minima and thus the uncertainty of the parameters.
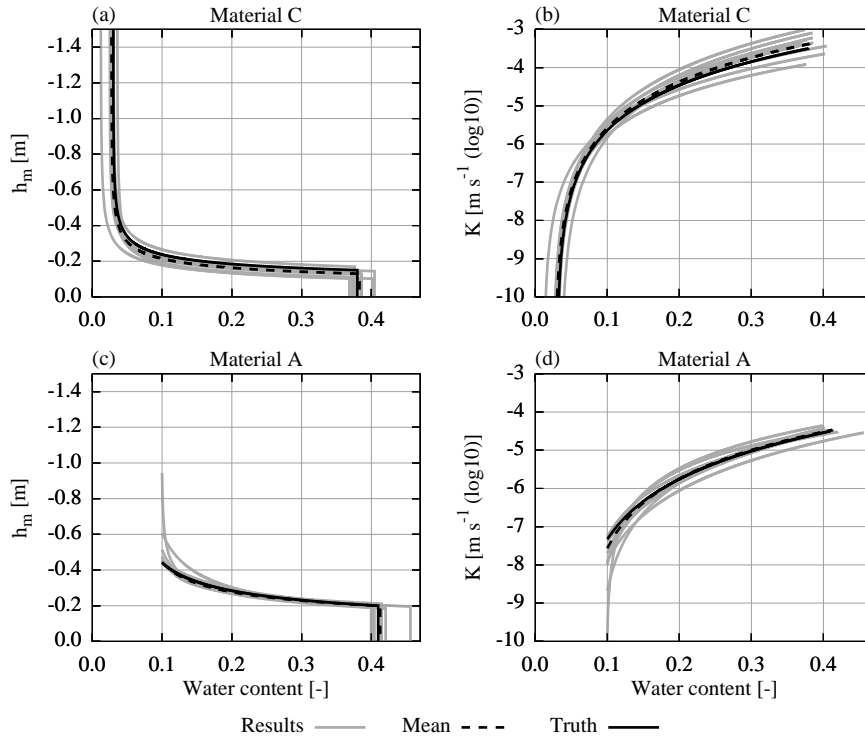
Figure 10. The resulting material parameters estimated from synthetic data are shown for the 10 best ensemble members (Sect. 3.1.2) together with the mean of these parameter sets and the true parameter set (Table 2). The plot range of the parameters is adjusted to the water content range of the data.

The saturated hydraulic conductivity of material A (Fig. 10b) is approximately one order of magnitude smaller than the saturated hydraulic conductivity of material C (Table 2). Since the 1D architecture is forced at the lower boundary, the hydraulic conductivity of material A limits the water flux into material C. Hence, the data are not sensitive to K_s^C. Consequently, the uncertainty of the hydraulic conductivity in material C decreases for low water contents as the reflection at the additional kink

5 (markers 3 and 5 in Fig. 9) is sensitive to the hydraulic conductivity. The hydraulic conductivity function (Eq. 3) is not unique if K_s is not fixed. This leads to a strong correlation of the parameters K_s^C and τ^C (0.6, Fig. 11). Note that the uncertainty of the saturated hydraulic conductivity of material A also influences the uncertainty of the hydraulic conductivity of material C.

The uncertainty of the soil water characteristic of material A (Fig. 10c) is largest for low water contents, because there are only few data points available. In particular, this increases the uncertainty of λ^A (± 0.7, Table 2). The material properties of

10 the unsaturated material A are only monitored during the first ≈ 5 h of the experiment and are independent of the largest part of the other data. This regularizes the optimization leading to fewer local minima. Similar to material C, the parameters h_0^A and λ^A are strongly correlated (-0.6). Yet, the uncertainty of h_0^A is relatively small (± 0.008, Table 2) mainly because it is essentially uncorrelated to other parameters. In contrast, the parameter θ_s^A is correlated to the parameters h_0^C, λ^C, θ_s^C, and θ_r^C,
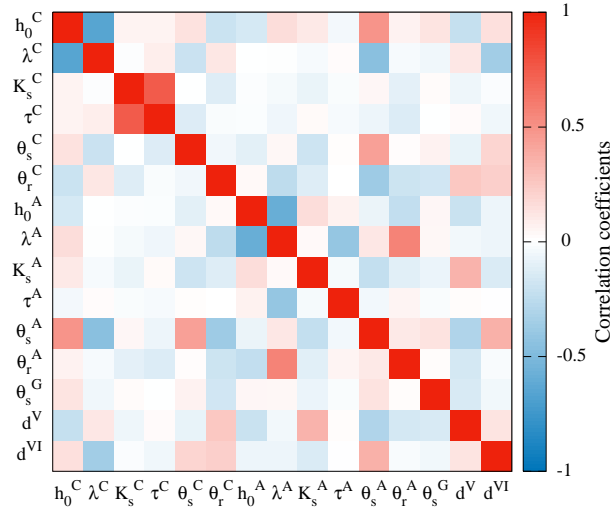
Figure 11. The correlation coefficients for the mean parameter set show in particular that the porosity of the gravel layer (θ_s^G) as well as the position of the material layers (d^V and d^{VI}) can be reliably estimated from single-offset GPR when evaluating both the signal travel time and amplitude.

because wrong parameters for material C introduce changes in the total water content which can be partially balanced out by adjusting θ_s^A.

The uncertainty of the saturated hydraulic conductivity of material A (Fig. 10d) is comparably small, because the largest fraction of the data are influenced by this parameter. Hence, the parameters τ^A and K_s^A are only very weakly correlated.

5   The correlation coefficients (Fig. 11) also show that both the permittivity and the thickness of the gravel layer can be estimated reliably with the presented evaluation method using travel time and amplitude information of a single-offset time-lapse radargram. Evaluation methods that merely exploit the signal travel time (e.g., Gerhards et al., 2008), require a multi-channel approach to achieve this goal.

In order to further investigate the quality of the mean parameter set, we simulated the resulting water content distribution
10   (Fig. 12a) and subtracted the true water content distribution (Fig. 12b). Due to the narrow pore size distribution of the sandy material, small deviations in the parameters h_0 and λ lead to large differences in the volumetric water content above the capillary fringe ($\approx \pm 0.04$). Combined with deviations in the position of the material interface, the largest differences in volumetric water content reach up to 0.17. Still, the mean absolute deviation of the volumetric water content is 0.004.

The remaining deviations in soil water content after the parameter estimation cause residuals in the GPR signal (Fig. 13).
15   These residuals are most evident for the reflection at the gravel layer (VI). The bias of its travel time shows that the total water content above the gravel layer is underestimated with the mean parameter set. This bias is essentially balanced out with the properties of the gravel layer. However, the reflection originating from the basement of ASSESS (VII) reveals residuals that
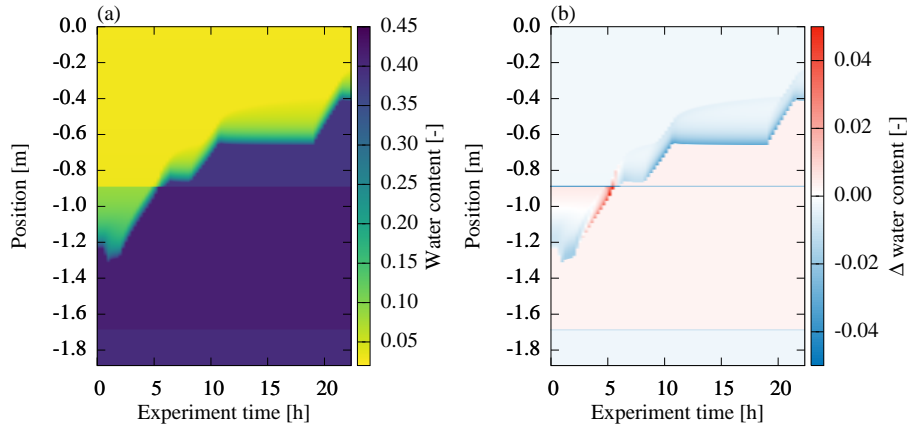
Figure 12. This figure shows (a) the water content distribution simulated with the resulting mean parameter set and (b) the difference to the true water content distribution (Fig. 9). The mean absolute deviation of the volumetric water content is 0.004. The overall balance of the volumetric water content can be characterized by calculating the mean of the summed difference per grid cell over all measurement times which yields -0.003. Hence, the mean parameter set generally underestimates the water content in the profile. The constant deviations above as well as below the groundwater table and in the gravel layer are due to small deviations in the estimated parameters θ_r^C, θ_s^C, θ_s^A, and θ_s^G. Still, the standard deviation of the estimated parameters contains the true parameter values (Table 2).

decrease as soon as the groundwater table is above the initial groundwater table. This indicates (i) deviations in the initial water content distribution and (ii) that the hydraulics during the initial drainage phase is not correctly represented.

Similar to the analysis of the deviation in water content (Fig. 12), the largest residuals in the unsaturated part of the domain are found where the groundwater table is crossing the interface of materials A and C. This indicates that the interference of the according reflections still contains information which could not be exploited in the parameter estimation procedure.

3.3 Measured data

In this section, the measured data (Sect. 2.2.4) are examined first qualitatively (Sect. 3.3.1). Afterwards, these data are used to estimate the subsurface architecture and the corresponding soil hydraulic material properties using the methods proposed in Sect. 2. The results of this inversion are discussed in Sect. 3.3.2.

3.3.1 Phenomenology

Before starting the experiment, a single-offset measurement was acquired to analyze the initial state of ASSESS revealing material interfaces as well as compaction interfaces (Fig. 14). Typically, it is difficult to associate the reflections based on an individual radargram. In particular the reflection of the compaction interface (iv) close to the reflection of the initial position of the groundwater table (1) is difficult to distinguish from reflections originating from material interfaces.
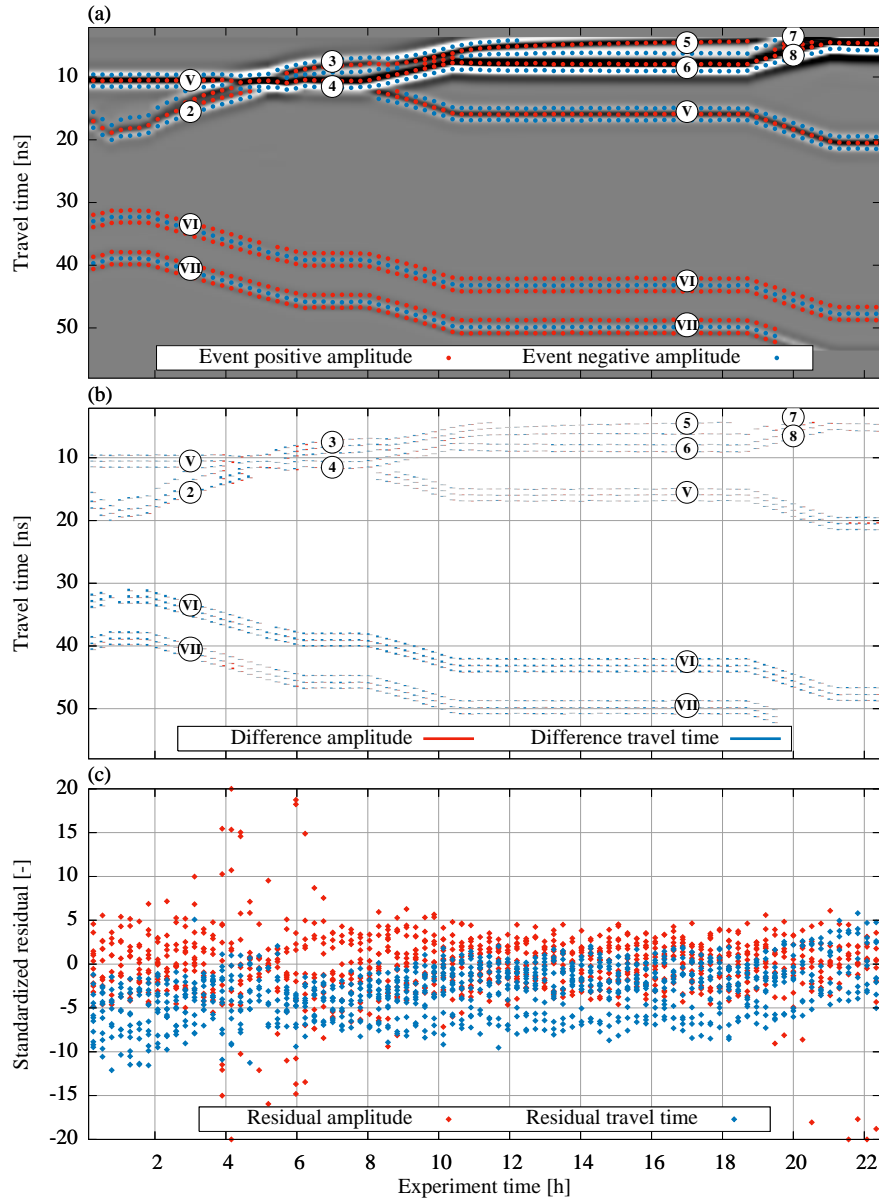
Figure 13. The evaluation of the synthetic GPR data is separated into three parts: Subfigure (a) shows the selected events (Sect. 2.3.3) which are evaluated with the optimization. The data shown are processed according to Sect. 2.3 except for the normalization. In contrast to the data that are evaluated, the shown radargram is normalized to its maximum absolute amplitude, facilitating the visual comparison of the traces. Subfigure (b) shows resulting differences in travel time and amplitude of the mean parameter set. The differences in amplitude are given in arbitrary units which are consistently used in this study. For the synthetic data, these differences are practically invisible. For the measured data, however, they are more clearly recognizable (Fig. 13). Subfigure (c) shows standardized residuals (Eq. 11), essentially zooming into the small differences given in subfigure (b). Note that outliers are set onto the boundary. The markers are defined in Fig. 14.

ASSESS is confined by walls at all four sides. Reflections from confining walls are most visible around 1 m (W) but influence the signal for more than 2 m. The walls parallel to the measurement direction are approximately 4 m apart from each other. Thus, it is assumed that the measurement is also influenced by reflections originating from those walls. The reflection of the edge of the L-element (L) is particularly prominent.

As an aside, closer scrutiny of the radargrams reveals that the single-offset and the time-lapse data were measured with different but structurally identical antennas. Thus, in particular the measured GPR signals of the direct wave and the ground wave are slightly different.

The time-lapse GPR measurement was recorded at 17.05 m (Fig. 14b). As the groundwater table is raised, the reflection originating from the groundwater table (2) separates from the reflection of the compaction interface (iv). After passing the material interface, the reflection of the groundwater table (2) splits in two separate reflections (3) and (4). This is due to the dependency of the hydraulic conductivity on the water content and was also identified for the synthetic data (Sect. 3.2.1). Since the transition zone is smoothing during the equilibration phase, the amplitude of reflection (3) decreases and the distance of the reflections (3) and (4) increases. During the subsequent imbibition step, the reflections are separated.

Corresponding to the analysis of the synthetic data (Sect. 3.2.1), the effects of the smoothing water content distribution are most clearly visible during the equilibration phase at the reflections (5) and (6). However, the associated measured signals interfere with the direct wave, the ground wave, and the reflection from the compaction interface (i) which makes the identification of these effects difficult. The reflections (7) and (8) measured during the final imbibition phase confirm the previous observations.

Together with the water content distribution, the time-lapse GPR data also contain information about the subsurface architecture. However, separating signal contribution from the subsurface architecture and the hydraulic dynamics is not always possible. Here, this is most prominent for the reflection of the material interface (V). Initially, the amplitude of this reflection is large, because the water content in material C is near the residual water content, whereas the water content in material A is significantly higher at the material interface. As soon as both materials are water saturated, the amplitude of the material interface reflection (V) is low since the effective porosities of the two materials are similar. Thus, the amplitude of the reflected signal originating from the material interfaces may change depending on the hydraulic state. Additional information about the subsurface architecture can be inferred from the reflection at the material interface between material A and the gravel layer (VI) and from the reflection at the material interface of the gravel layer and the concrete basement (VII). These reflections are in particular suitable to analyze the total change of water content over time.

In summary, we note that the characteristic properties of the transition zone reflection during the imbibition and equilibration steps that were identified in the simulation (Fig. 9) can also be identified in the measured data (Fig. 14).

3.3.2   Results and discussion

Since the GPR measurements cover only a small portion of the subsurface architecture of ASSESS, the hydraulic representation is restricted to 1D (Sect. 3.1.1). Hence, 2D effects such as lateral flow are neglected. This has to be considered during the event selection of measured data (Sect. 2.3.3). Therefore, the focus of this study is on evaluation of the imbibition phase of the
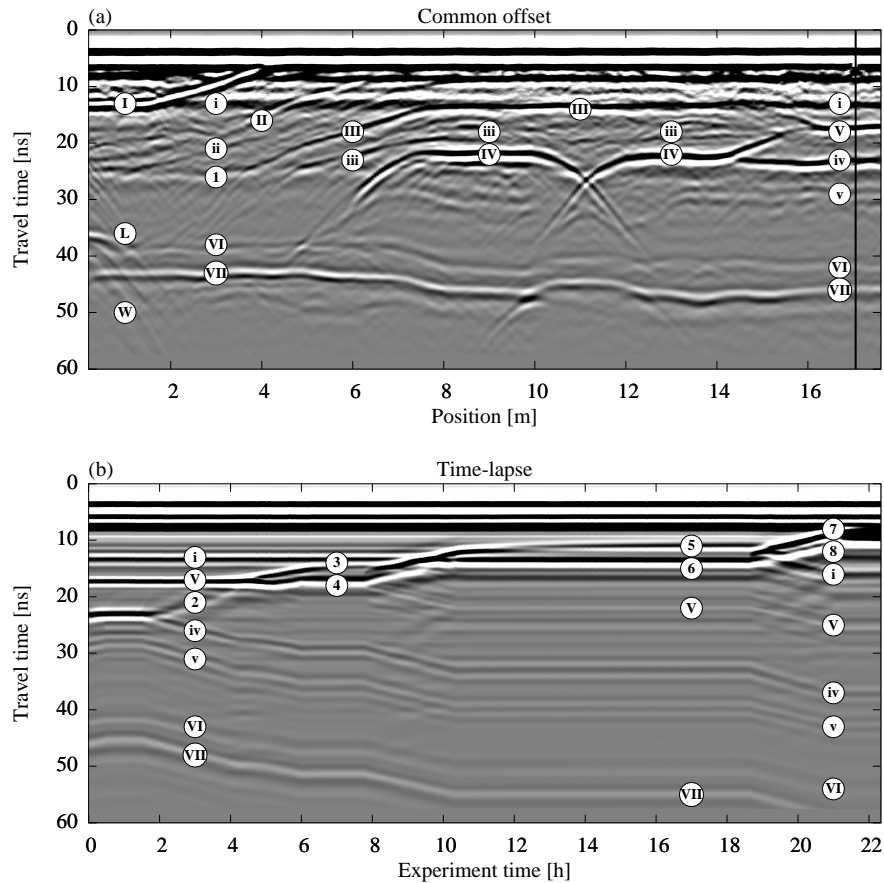
Figure 14. (a) A single-offset measurement of the hydraulic state of ASSESS (Fig. 1) at the beginning of the experiment is shown. For this measurement, the antenna was moved over the site at one point in time. The temporal evolution of the subsequent hydraulic dynamics was monitored with a stationary antenna at the position indicated with the vertical black line. The resulting time-lapse measurement data are shown in subfigure (b). Both radargrams are measured with internal channels with an antenna separation of 0.14 m. Except for the normalization, the data are processed according to Sect. 2.3. In contrast to the quantitative evaluation, the radargrams are normalized to their maximal absolute amplitude, facilitating the visual comparison of the traces. The markers – uppercase Roman markers for material interfaces (I, II, III, IV, V, VI, and VII), lowercase Roman markers for compaction interfaces (i, ii, iii, iv, and v), and markers with Arabic numbers for reflections originating from the water content distribution – are used consistently in this paper and are further explained in the text.
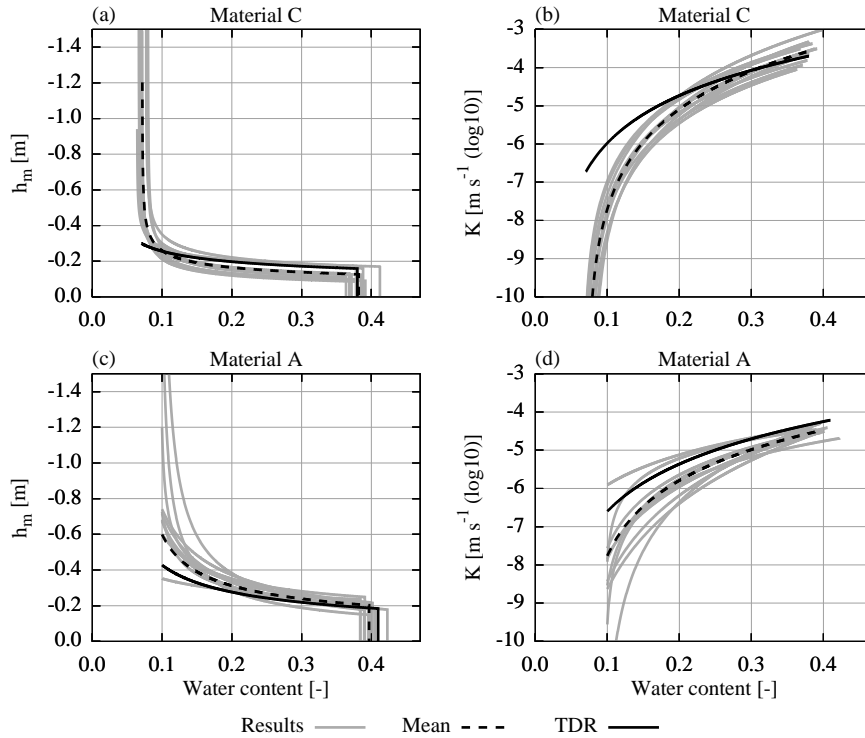
Figure 15. The resulting material parameters estimated from measured data are shown for the 10 best ensemble members (Sect. 3.1.2) together with the mean of these parameter sets (Table 3) and a reference parameter set determined from TDR data acquired during the same experiment (Jaumann and Roth, 2017). The deviation from the reference parameters can be explained with representation errors in GPR analysis (e.g., neglecting compaction interfaces) and missing data (e.g., for low water contents in material A). The plot range of the parameters is adjusted to the water content range of the corresponding data.

experiment, because the effect of lateral flow in fluctuating groundwater table experiments is largest during drainage and close to the capillary fringe (Jaumann and Roth, 2017).

Investigating the resulting material properties of the inversion (Fig. 15), the main findings, that were discussed previously for the mean parameters for the synthetic data (Sect. 3.2.2), can also be identified for the measured data. These findings comprise
5   (i) the shift in the soil water characteristic of material C, (ii) the large uncertainty of the saturated hydraulic conductivity of material C, (iii) the high uncertainty of the soil water characteristic of material A for low water contents, and (iv) the high sensitivity on K_s^A.

Compared to the uncertainties based on synthetic data (Table 2), the uncertainty of the resulting mean parameters (Table 3) mostly increased. Except for four parameters, the parameters estimated from TDR measurements (Jaumann and Roth, 2017)
10   are within one standard deviation of the mean parameter set. The deviations of the other four parameters are clearly visible in Fig. 15 and will be analyzed in the following.

The parameter θ_r^C estimated from the GPR data is approximately a factor of 3 larger than the estimated value based on the TDR data (Table 3). Essentially, there are three main reasons for this. First, by evaluating the travel time of reflection (V), integrated water content is included in the inversion. This also comprises the compaction interface (i) which is not represented in the model. At the beginning of the experiment, the amplitude of this reflection is comparable to the amplitude of the reflection

5 originating from the interface of material A and C (V). Notice that the amplitude of the reflection (i) does not vanish, but merely decreases when the material is saturated at the end of the experiment (Fig. 14). This indicates that this reflection originates from changes in both the small-scale texture of the material and the stored water content at the beginning of the experiment. Hence, since this compaction interface is not represented in the model, the resulting θ_r^C is increased coping for this representation error. Second, a deviation in the position of the groundwater table with reference to the antenna position at the surface can be

10 partially adapted by changing θ_r^C. As the position of the surface is subject to change over the years, the measurements of the groundwater table are referenced to a fixed point at the end of the groundwater well, leaving the exact position of the surface relative to groundwater table uncertain. According to Buchner et al. (2012), the accuracy of the ASSESS architecture when compared to GPR measurements is ± 0.05 m. The estimation of an offset to the Dirichlet boundary could mitigate this problem, but would in any case increase the number of local minima significantly making the optimization process less stable. Third,

15 analyzing the TDR data, we find that an underestimation of θ_r^C is likely due to the lack of TDR measurements at low hydraulic potential. Hence, the sensitivity of the TDR data to θ_r^C is low.

The resulting value for parameter τ^C for the GPR evaluation is a factor of 2 larger compared to the evaluation of the TDR data. This parameter adjusts the hydraulic conductivity for the unsaturated material and is mainly determined with the reflections (3) and (5) originating from the additional kink during imbibition (Fig. 16). These reflections exhibit a small amplitude for

20 low water contents. However, both reflections interfere with the rather prominent reflection originating from the compaction interface (i). Additionally, the reflection (5) also interferes with parts of the direct and the ground wave. Hence, the travel time of these reflections does hardly change leading to an underestimation of the hydraulic conductivity for low water contents resulting in an overestimation of parameter τ^C.

Similar to parameter θ_r^C, the parameter θ_r^A estimated from the GPR data is approximately a factor of 3 smaller than the

25 result from the TDR evaluation. However, this parameter can only be approximated evaluating the GPR data, because they lack events that are influenced by low water content.

The resulting value for parameter K_s^A is by a factor of 2 smaller for the GPR evaluation than for the TDR evaluation. This parameter limits the flux through the lower boundary, because the domain is forced with a Dirichlet boundary condition. Hence, the parameter can be used to cope with errors in the boundary condition. Forcing ASSESS with a groundwater well instantiates

30 a 3D flux (Jaumann and Roth, 2017). Since this 3D flux is not represented, the hydraulic potential at the bottom and thus also the water flux are too large. This is compensated by the parameter estimation with decreasing K_s^A.

Concerning the position of the material interfaces, we find that the estimated interface position of the material A and C (d^V) corresponds well the to ground truth measurements acquired during the construction of the ASSESS site (Table 3). In contrast, the estimated position of the gravel layer (d^{VI}) deviates from the according ground truth measurements. However, the estimates

35 are still within the uncertainty of the ground truth measurements when compared to GPR measurements.
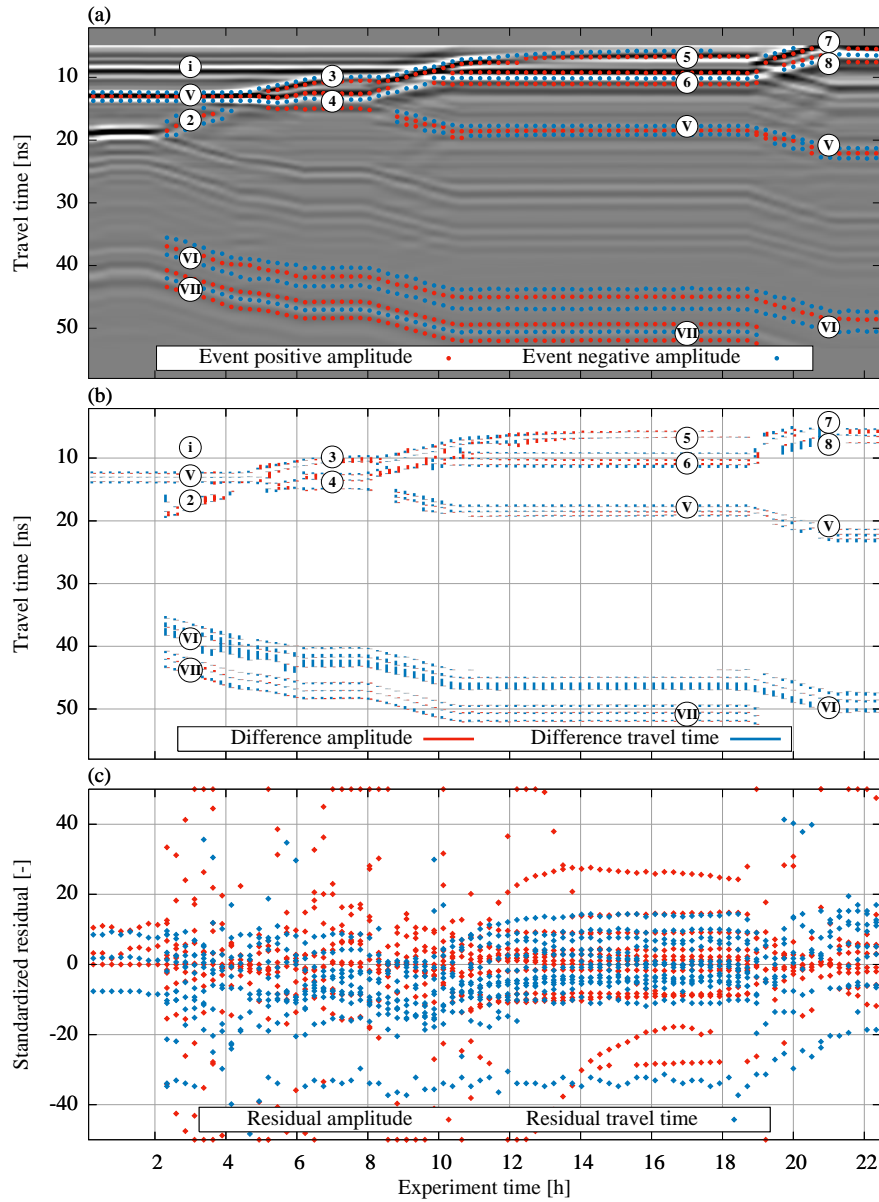
Figure 16. Analogous to Fig. 13 but for measured data. Since the measurement data acquired with a stationary antenna cannot spatially resolve the lateral flow present in the initial drainage phase, the measured events of the hydraulic dynamics during the first 2 h are neglected.

An analysis of the remaining residuals in travel time after the optimization (Fig. 16b) shows that deviations in the width of the reflected wavelet contribute to the residuum significantly, in particular for the reflections (V), (VI) and (VII). At the beginning of the experiment, the simulated wavelet is too broad for reflection (V) whereas it is too narrow for reflections (VI) and (VII). This is in particular noticeable for the residuals for reflection (VI) which are the major contribution to the cost function. Of all the events in this wavelet, the events with the longest travel time exhibit the largest residuals. The difference in the width of the reflected wavelet can be explained with the roughness of the material interfaces (Dagenbach et al., 2013). Due to the large grain size of the gravel, the real material interface is rougher than its representation. This also explains the partly non-symmetric broadening of the measured compared to the simulated wavelet of reflection (VI).

The measured reflection (6) interferes with the reflection of the compaction interface (i) leading to a compressed reflected wavelet in the measurement. Similarly, also the reflections (3) and (5) interfere with the compaction interface (i). Since interferences cannot be correctly evaluated if not all contributions are represented, this analysis shows that representing compaction interfaces is relevant in ASSESS.

As a side remark, note that the error originating from assuming a constant soil temperature for the calculation of the relative permittivity of water is relatively small regarding the total residuum. However, it is worth noting that the corresponding residuals easily exceed one standard deviation in signal travel time.

The distribution and the support of the measurement data (i) differs between the TDR sensors and GPR measurements (Fig. 1), (ii) relates directly to the applicability of the resulting parameters for the different evaluations, and (iii) influences the quantitative effect of different representation errors. In the reference analysis, the TDR sensors are distributed over a 2D slice of ASSESS measuring in all available materials (Fig. 1). Yet, their measurement volume is limited to the position of the sensors yielding the average permittivity along the central TDR rod of each sensor. Hence, these measurements are subject to representation errors such as small-scale heterogeneity or uncertainty of the sensor position (Jaumann and Roth, 2017). In contrast, the GPR measurements do not cover the whole ASSESS test site and their support is depending on the evaluated events of the wavelet. This includes the whole depth average (travel time) and the contrast (amplitude) of both the permittivity and electrical conductivity. Hence, these measurement data are subject to representation errors such as neglected compaction interfaces and the roughness the material interfaces. Hence, the previous analysis illustrates how GPR-determined parameters can differ from TDR-determined ones making joint evaluation procedures both challenging and promising since they open a window to the soils multi-scale nature.

4 Summary and conclusions

TDR measurements are a standard method that provides measurement data for the estimation of soil hydraulic material properties. However, this invasive measurement method yields point-scale measurements and typically requires a local measurement station. Hence, it is difficult to apply at larger scales or to transfer the sensors to another field site. In contrast, GPR is a non-invasive measurement method that is traditionally used for subsurface characterization including subsurface architecture and

effective water content. The analysis of GPR measurements is much more challenging than that of TDR and there is still a need for efficient quantitative evaluation methods.

In this study, we propose a new heuristic semi-automatic evaluation approach to identify, extract, and associate relevant information from GPR data. Focussing the optimization on this relevant information regularizes the parameter estimation. The suitability of the proposed methods to accurately identify the subsurface architecture and the soil hydraulic material properties was analyzed for synthetic and measured time-lapse GPR data.

The developed GPR data evaluation method first detects the most important extrema of the signal (events) in the measurement and in the simulation. Subsequently, the detected measured events are associated with the detected simulated events. All plausible combinations of simulated and measured events are analyzed to identify the optimal pair association of these events. To decrease the computational effort, the detected events are grouped in clusters. First, the clusters are associated. Then follows the association of the events contained in these clusters. In order to estimate the subsurface architecture and the corresponding soil hydraulic material properties, the difference in the signal travel time and amplitude of the associated events is minimized with inversion methods. Using events instead of the full GPR signal regularizes the optimization.

Synthetic and measured single-offset time-lapse GPR data are first analyzed qualitatively. It was confirmed that a fluctuating groundwater table experiment introduces characteristic transition zone reflections that are likely to provide valuable information for the parameter estimation. Subsequently, the subsurface architecture and soil hydraulic material properties are estimated based on the GPR data using a global-local parameter estimation approach with preconditioning. The preconditioning step starts from an ensemble of 60 Latin hypercube sampled initial parameters sets. They are used to initialize a preconditioning step in which a simulated annealing algorithm and a Levenberg–Marquardt algorithm are sequentially coupled. In this step, these algorithms optimize parameters based on a subsampled data set with a limited number of iterations. The resulting parameter sets are then used to initialize the Levenberg–Marquardt algorithm that operates on the full data set.

Employing the presented approaches on synthetic data shows that the true parameters are within one standard deviation of the resulting mean parameter set based on the ten best ensemble members. This mean parameter set describes the hydraulic dynamics with a mean absolute error in volumetric water content of 0.004. Additionally, we found that the parameter correlations are mostly specific to the experiment type and the subsurface architecture. Using travel time and amplitude information in the evaluation allowed to estimate effective permittivity and layer depth simultaneously with a single GPR channel.

The resulting parameters for the measured data are mostly consistent with results from reference TDR measurement data. We discussed the deviations of the parameters and basically associated them with representation errors and the lack of available measurement data. Relevant representation errors in the GPR data analysis comprise in particular the neglected (i) compaction interfaces and (ii) roughness of the material interfaces.

The three major drawbacks of the presented approach comprise (i) the computational effort which is required to solve Richards' and Maxwell's equations, (ii) the limited number of events that can be analyzed due to the pairwise event association which investigates all plausible combinations of simulated and measured events, and (iii) the fact that the hyperparameters for the GPR evaluation algorithm have to be determined a priori. The latter is difficult and requires expert knowledge, especially as the shape of the radargram is like to change considerably during the optimization procedure.

Although, the proposed methods have been shown in 1D, going to 2D and even to 3D is first and foremost a matter of computational effort with already 2D demanding significant time on a large compute cluster. No concepts or methods beyond of what we demonstrated in this paper are required, however.

Previous work showed that the location of moderately complicated layer interfaces and of the mean water content between them can be obtained from multi-offset measurements (Buchner et al., 2012). Together with the demonstration in this paper that the effective hydraulic material properties of layers can be estimated from single-offset time-lapse measurements, we now have the methods to determine the subsurface architecture and its hydraulic properties for moderately complicated situations. This obviously demands quite a significant experimental effort together with subsequent massive computations, as spatially resolved time-lapse measurements of the region of interest are required which then have to be inverted.

5 Data availability

The underlying measurement data are available at http://ts.iup.uni-heidelberg.de/data/jaumann-roth-2017-gpr-hess.zip

Author contributions. S. Jaumann designed and conducted the experiment, developed the main ideas, implemented the algorithms, and analyzed the measurement data. K. Roth contributed with guiding discussions. S. Jaumann prepared the manuscript with contributions of both authors.

Acknowledgements. We thank Jens S. Buchner for software to process (i) the raw data of the ASSESS architecture and (ii) GPR data according to the *Constructive Inversion* approach. We are grateful to Angelika Gassama for technical assistance with respect to ASSESS. We especially thank Patrick Klenk and Elwira Zur for assistance during the experiment. The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grants INST 35/1134-1 FUGG and RO 1080/12-1. We are also grateful to the editor Insa Neuweiler and to two anonymous reviewers, who helped to improve the manuscript significantly.

References

Bauser, H. H., Jaumann, S., Berg, D., and Roth, K.: EnKF with closed-eye period – towards a consistent aggregation of information in soil hydrology, Hydrology and Earth System Sciences, 20, 4999–5014, doi:10.5194/hess-20-4999-2016, 2016.

Birchak, J. R., Gardner, C. G., Hipp, J. E., and Victor, J. M.: High dielectric constant microwave probes for sensing soil moisture, Proceedings of the IEEE, 62, 93–98, doi:10.1109/PROC.1974.9388, 1974.

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

Bradford, J., Thoma, M., and Barrash, W.: Estimating hydrologic parameters from water table dynamics using coupled hydrologic and ground-penetrating radar inversion, in: Proceedings of the 15th International Conference on Ground Penetrating Radar, pp. 232–237, doi:10.1109/ICGPR.2014.6970420, 2014.

Brooks, R. H. and Corey, A. T.: Properties of porous media affecting fluid flow, Journal of the Irrigation and Drainage Division, 92, 61–90, 1966.

Buchner, J. S., Wollschläger, U., and Roth, K.: Inverting Surface GPR Data using FDTD Simulation and Automatic Detection of Reflections to Estimate Subsurface Water Content and Geometry, Geophysics, 77, H45–H55, doi:10.1190/geo2011-0467.1, 2012.

Busch, S., van der Kruk, J., Bikowski, J., and Vereecken, H.: Quantitative conductivity and permittivity estimation using full-waveform inversion of on-ground GPR data, Geophysics, 77, H79–H91, doi:10.1190/geo2012-0045.1, 2012.

Carmichael, R. S.: Physical properties of rocks and minerals, CRC press Boca Raton, 1989.

Dagenbach, A., Buchner, J. S., Klenk, P., and Roth, K.: Identifying a parameterisation of the soil water retention curve from on-ground GPR measurements, Hydrology and Earth System Sciences, 17, 611–618, doi:10.5194/hess-17-611-2013, 2013.

Daniels, D.: Ground Penetrating Radar, The Institution of Engineering and Technology, 2nd edn., doi:10.1049/PBRA015E, 2004.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015–1031, doi:10.1029/91WR02985, 1992.

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

Farjadpour, A., Roundy, D., Rodriguez, A., Ibanescu, M., Bermel, P., Joannopoulos, J. D., Johnson, S. G., and Burr, G. W.: Improving accuracy by subpixel smoothing in the finite-difference time domain, Opt. Lett., 31, 2972–2974, doi:10.1364/OL.31.002972, 2006.

Gerhards, H., Wollschläger, U., Yu, Q., Schiwek, P., Pan, X., and Roth, K.: Continuous and simultaneous measurement of reflector depth and average soil-water content with multichannel ground-penetrating radar, Geophysics, 73, J15–J23, doi:10.1190/1.2943669, 2008.

Giannopoulos, A.: Modelling ground penetrating radar by GprMax, Construction and Building Materials, 19, 755–762, doi:10.1016/j.conbuildmat.2005.06.007, 2005.

Heimovaara, T. J., Focke, A. G., Bouten, W., and Verstraten, J. M.: Assessing Temporal Variations in Soil Water Composition with Time Domain Reflectometry, Soil Science Society of America Journal, 59, 689–698, doi:10.2136/sssaj1995.03615995005900030009x, 1995.

Hopmans, J. W., Šimůnek, J., Nunzio, R., and Wolfgang, D.: Simultaneous determination of water transmission and retention properties. Inverse Methods., in: Methods of Soil Analysis. Part 4. Physical Methods, edited by Dane, J. and Topp, G. C., pp. 963–1008, Soil Science Society of America Book Series, 2002.

Huyer, W. and Neumaier, A.: Global Optimization by Multilevel Coordinate Search, Journal of Global Optimization, 14, 331–355, doi:10.1023/A:1008382309369, 1999.

Ippisch, O., Vogel, H.-J., and Bastian, P.: Validity Limits for the van Genuchten-Mualem Model and Implications for Parameter Estimation and Numerical Simulation., Advances in Water Resources, 29, 1780–1789, doi:10.1016/j.advwatres.2005.12.011, 2006.

Jackson, J. D.: Classical Electrodynamics, John Wiley & Sons, Inc., New Jersey, USA, 3 edn., 1999.

Jadoon, K. Z., Weihermüller, L., Scharnagl, B., Kowalsky, M. B., Bechtold, M., Hubbard, S. S., Vereecken, H., and Lambot, S.: Estimation of soil hydraulic parameters in the field by integrated hydrogeophysical inversion of time-lapse ground-penetrating radar data, Vadose Zone Journal, 11, doi:10.1190/1.3255105, 2012.

Jaumann, S. and Roth, K.: Effect of unrepresented model errors on estimated soil hydraulic material properties, Hydrology and Earth System Sciences, 21, 4301–4322, doi:10.5194/hess-21-4301-2017, 2017.

Jonard, F., Weihermüller, L., Schwank, M., Jadoon, K. Z., Vereecken, H., and Lambot, S.: Estimation of Hydraulic Properties of a Sandy Soil Using Ground-Based Active and Passive Microwave Remote Sensing, IEEE Transactions on Geoscience and Remote Sensing, 53, 3095–3109, doi:10.1109/TGRS.2014.2368831, 2015.

Kaatze, U.: Complex permittivity of water as a function of frequency and temperature, Journal of Chemical and Engineering Data, 34, 371–374, doi:10.1021/je00058a001, 1989.

Klenk, P., Jaumann, S., and Roth, K.: Quantitative high-resolution observations of soil water dynamics in a complicated architecture using time-lapse ground-penetrating radar, Hydrology and Earth System Sciences, 19, 1125–1139, doi:10.5194/hess-19-1125-2015, 2015.

Lambot, S., Antoine, M., Van den Bosch, I., Slob, E., and Vanclooster, M.: Electromagnetic inversion of GPR signals and subsequent hydrodynamic inversion to estimate effective vadose zone hydraulic properties, Vadose Zone Journal, 3, 1072–1081, doi:10.2113/3.4.1072, 2004.

Lambot, S., Slob, E., Rhebergen, J., Lopera, O., Jadoon, K. Z., and Vereecken, H.: Remote estimation of the hydraulic properties of a sand using full-waveform integrated hydrogeophysical inversion of time-lapse, off-ground GPR data, Vadose Zone Journal, 8, 743–754, doi:10.2136/vzj2008.0058, 2009.

Lampe, B., Holliger, K., and Green, A. G.: A finite-difference time-domain simulation tool for ground-penetrating radar antennas, Geophysics, 68, 971–987, doi:10.1190/1.1581069, 2003.

Léger, E., Saintenoy, A., and Coquet, Y.: Estimating saturated hydraulic conductivity from ground-based GPR monitoring Porchet infiltration in sandy soil, in: Proceedings of the 15th International Conference on Ground Penetrating Radar, pp. 124–130, doi:10.1109/ICGPR.2014.6970399, 2014.

Léger, E., Saintenoy, A., Tucholka, P., and Coquet, Y.: Inverting surface GPR data to estimate wetting and drainage water retention curves in laboratory, in: Advanced Ground Penetrating Radar (IWAGPR), 2015 8th International Workshop on, pp. 1–5, IEEE, doi:10.1109/IWAGPR.2015.7292672, 2015.

Looms, M. C., Binley, A., Jensen, K. H., Nielsen, L., and Hansen, T. M.: Identifying Unsaturated Hydraulic Parameters Using an Integrated Data Fusion Approach on Cross-Borehole Geophysical Data, Vadose Zone Journal, 7, 238–248, doi:10.2136/vzj2007.0087, 2008.

Manoli, G., Rossi, M., Pasetto, D., Deiana, R., Ferraris, S., Cassiani, G., and Putti, M.: An iterative particle filter approach for coupled hydro-geophysical inversion of a controlled infiltration experiment, Journal of Computational Physics, 283, 37–51, doi:10.1016/j.jcp.2014.11.035, 2015.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, The Journal of Chemical Physics, 21, 1087–1092, 1953.

Moghadas, D., Jadoon, K. Z., Vanderborght, J., Lambot, S., and Vereecken, H.: Estimation of the near surface soil water content during evaporation using air-launched ground-penetrating radar, Near Surface Geophysics, 12, 623–633, doi:10.3997/1873-0604.2014017, 2014.

Mualem, Y.: A new Model for Predicting the Hydraulic Conductivity of Unsaturated Porous Media, Water Resources Research, 12, 513–522, doi:10.1029/WR012i003p00513, 1976.

Neal, A.: Ground-penetrating radar and its use in sedimentology: principles, problems and progress, Earth-Science Reviews, 66, 261–330, doi:10.1016/j.earscirev.2004.01.004, 2004.

5 Nelder, J. A. and Mead, R.: A Simplex Method for Function Minimization, The Computer Journal, 7, 308, doi:10.1093/comjnl/7.4.308, 1965.

Oskooi, A. F., Roundy, D., Ibanescu, M., Bermel, P., Joannopoulos, J. D., and Johnson, S. G.: MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method, Computer Physics Communications, 181, 687–702, doi:10.1016/j.cpc.2009.11.008, 2010.

Press, W. H.: Numerical recipes 3rd edition: The art of scientific computing, Cambridge University Press, 2007.

10 Richards, L. A.: Capillary Conduction of Liquids through Porous Mediums, Physics, 1, 318–333, doi:10.1063/1.1745010, 1931.

Robinson, D., Jones, S. B., Wraith, J., Or, D., and Friedman, S.: A review of advances in dielectric and electrical conductivity measurement in soils using time domain reflectometry, Vadose Zone Journal, 2, 444–475, doi:10.2136/vzj2003.4440, 2003.

Rossi, M., Manoli, G., Pasetto, D., Deiana, R., Ferraris, S., Strobbia, C., Putti, M., and Cassiani, G.: Coupled inverse modeling of a controlled irrigation experiment using multiple hydro-geophysical data, Advances in Water Resources, 82, 150–165,

15 doi:10.1016/j.advwatres.2015.03.008, 2015.

Scholer, M., Irving, J., Binley, A., and Holliger, K.: Estimating vadose zone hydraulic properties using ground penetrating radar: The impact of prior information, Water Resources Research, 47, doi:10.1029/2011WR010409, w10512, 2011.

Taflove, A. and Hagness, S.: Computational Electrodynamics: The Finite-Difference Time-Domain Method, Artech House, 2nd edn., 2000.

Thoma, M. J., Barrash, W., Cardiff, M., Bradford, J., and Mead, J.: Estimating Unsaturated Hydraulic Functions for Coarse Sediment from

20 a Field-Scale Infiltration Experiment, Vadose Zone Journal, 13, doi:10.2136/vzj2013.05.0096, 2014.

Tran, A. P., Vanclooster, M., Zupanski, M., and Lambot, S.: Joint estimation of soil moisture profile and hydraulic parameters by ground-penetrating radar data assimilation with maximum likelihood ensemble filter, Water Resources Research, 50, 3131–3146, doi:10.1002/2013WR014583, 2014.

Vrugt, J. A., Stauffer, P. H., Wöhling, T., Robinson, B. A., and Vesselinov, V. V.: Inverse modeling of subsurface flow and transport properties:

25 A review with new developments, Vadose Zone Journal, 7, 843–864, doi:10.2136/vzj2007.0078, 2008.

Table 1. The fit range limits the parameter space available for parameter estimation and is in particular used by the simulated annealing algorithm to draw parameter updates (Sect. 2.4.2). The sample range is used to generate an ensemble of initial parameter sets with a Latin hypercube algorithm.

Material	Parameter	Fit range		Sample range	
		min	max	min	max
C	h_0 (m)	-0.25	-0.05	-0.20	-0.10
	λ $(-)$	1.0	5.0	2.0	4.0
	K_s (m s^{-1})	$10^{-4.1}$	$10^{-2.9}$	10^{-4}	10^{-3}
	τ $(-)$	-1.0	2.0	0.0	1.0
	θ_s $(-)$	0.33	0.43	0.36	0.40
	θ_r $(-)$	0.00	0.10	0.02	0.08
A	h_0 (m)	-0.30	-0.10	-0.25	-0.15
	λ $(-)$	1.0	5.0	2.0	4.0
	K_s (m s^{-1})	$10^{-5.1}$	$10^{-3.9}$	10^{-5}	10^{-4}
	τ $(-)$	-1.0	2.0	0.0	1.0
	θ_s $(-)$	0.36	0.46	0.39	0.43
	θ_r $(-)$	0.00	0.10	0.02	0.08
Gravel	θ_s $(-)$	0.30	0.50	0.38	0.42
Architecture	d^{V} (m)	0.90	1.10	0.95	1.05
	d^{VI} (m)	0.10	0.30	0.15	0.25

Table 2. The mean and the standard deviation are calculated using the resulting parameters from the 10 best ensemble members (Sect. 3.1.2) estimated from synthetic data. The corresponding material functions are given in Fig. 10. Notice that the true parameter set lies within the standard deviation of the mean parameter set.

Material	Parameter	Truth	Mean results
C	h_0 (m)	-0.15	-0.13 ± 0.02
	λ $(-)$	3.5	3.2 ± 0.3
	K_s (m s^{-1})	$10^{-3.5}$	$10^{-3.4 \pm 0.2}$
	τ $(-)$	0.5	0.6 ± 0.2
	θ_s $(-)$	0.38	0.38 ± 0.01
	θ_r $(-)$	0.03	0.027 ± 0.006
A	h_0 (m)	-0.20	-0.199 ± 0.008
	λ $(-)$	2.5	2.8 ± 0.7
	K_s (m s^{-1})	$10^{-4.5}$	$10^{-4.47 \pm 0.05}$
	τ $(-)$	0.5	0.4 ± 0.5
	θ_s $(-)$	0.41	0.41 ± 0.02
	θ_r $(-)$	0.05	0.06 ± 0.02
Gravel	θ_s $(-)$	0.40	0.40 ± 0.03
Architecture	d^{V} (m)	1.00	0.99 ± 0.02
	d^{VI} (m)	0.20	0.20 ± 0.01

Table 3. The mean and the standard deviation are calculated using the resulting parameters from the ten best ensemble members (Sect. 3.1.2) estimated from measured data. The corresponding material functions are shown in Fig. 15. The reference parameters for the materials A and C are determined from TDR data acquired during the same experiment (Jaumann and Roth, 2017). Note that the according standard deviations for these reference parameters are determined from a single Levenberg–Marquardt run and thus are only representative for one local minimum. Also, these standard deviations are given with the understanding that they are specific to the applied algorithm and will change for different algorithm parameters. Hence, these standard deviations are in particular not suitable to compare the precision of the TDR and GPR evaluation. For the TDR evaluation, the porosity of the materials is assumed to be known from core samples. The reference parameters for the subsurface architecture are calculated from ground truth measurements acquired during the construction of ASSESS. The corresponding standard deviations are given according to Buchner et al. (2012).

Material	Parameter	Reference	Mean results
C	h_0 (m)	-0.159 ± 0.004	-0.13 ± 0.03
	λ $(-)$	3.28 ± 0.02	3.3 ± 0.7
	K_s (m s^{-1})	$10^{-3.70 \pm 0.02}$	$10^{-3.6 \pm 0.3}$
	τ $(-)$	0.74 ± 0.06	1.4 ± 0.4
	θ_s $(-)$	0.38	0.38 ± 0.01
	θ_r $(-)$	0.026 ± 0.002	0.071 ± 0.005
A	h_0 (m)	-0.184 ± 0.005	-0.20 ± 0.03
	λ $(-)$	1.94 ± 0.07	2.1 ± 0.7
	K_s (m s^{-1})	$10^{-4.212 \pm 0.004}$	$10^{-4.5 \pm 0.1}$
	τ $(-)$	0.33 ± 0.07	0.4 ± 1.0
	θ_s $(-)$	0.41	0.40 ± 0.01
	θ_r $(-)$	0.025 ± 0.004	0.07 ± 0.03
Gravel	θ_s $(-)$		0.41 ± 0.02
Architecture	d^{V} (m)	0.99 ± 0.05	0.97 ± 0.02
	d^{VI} (m)	0.13 ± 0.05	0.17 ± 0.02