Reply to referee comment 1

Jaumann and Roth consider the problem of inferring for soil hydraulic properties for the situation when (1) hydraulic and electrical properties only vary with depth; (2) interface locations are known and they are horizontal; (3) the initial position of the water table is known. The data at hand are GPR traces acquired over time during a controlled variation of the water table. The experiments are carried out at the ASSESS test site, a large-scale facility for studying the use of GPR data in vadose zone hydrology with known layering, control of water table and supplementary data, such as, time-domain reflectometry data. The inversion methodology uses a 1-D solver to the Richards equation, a 2-D electromagnetic forward, and a rather elaborate (but also somewhat convoluted) optimization algorithm with little demonstration that it enables the localisation of global minima and that the ensemble members used provide a reasonable estimate about parameter uncertainty. One of the key aspects of this algorithm is the automatic detection of events (typically negative and positive peaks) in the GPR traces. The writing is overall good, but I also feel that the manuscript could be simplified and shortened.

Reply: We thank the reviewer for the constructive comments and suggestions. We revised the manuscript accordingly and refer to the revised version in the following.

1. The authors state in title and abstract that they estimate the subsurface architecture. Reading this, I expected the retrieval of 2-D or 3-D geometry of lithofacies. Instead, the authors assume a known layered system and "simply" infer for parameters in the Mualem model (and allow for some very small variations in interface locations). I find this terminology to be inappropriate, and it should be stated that the authors infer for hydraulic properties of multiple known layers. From what I understand, this is the main novelty of this work: using water table fluctuations to infer hydraulic material properties for more than one layer (2 in this case; much less significant than the statement on line 5 in abstract).

Reply:

Previous work showed that the location of moderately complicated layer interfaces and of the mean water content between them can be obtained from multi-channel common 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-channel 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 time-lapse common offset measurements of the region of interest are required, which then have to be inverted.
The main novelty of this paper is the developed evaluation method that (i) is conceptionally not limited to layers but allows to evaluate moderately complicated subsurface architectures, (ii) does not require to estimate the full wave form but focuses on specific reflections chosen by the user, and (iii) can extract the information provided by changing shapes of the reflected wavelet. Thus, although the approach is demonstrated for a layered subsurface architecture, it is not limited to 1D or to simple layers.

2. The authors need to put this work into context. How does the presented work improve understanding about vadose zone processes, how can the used method be used for actual field applications (not that easy given that it is assumed that everything is 1-D and that interfaces are known, which is seldom the case)? Many readers are likely to question why to go through all this trouble instead of doing the same inversion using a few TDR probes. The answer is related to larger-scale applications, but this is not handled here (only one GPR position). A clear motivation is needed in introduction and before the conclusions. In short, why should someone that is not working at ASSESS read this work and how can it advance hydrology or the use of GPR to characterise hydrology. This is not clear reading the present version of the manuscript.
Reply:
We revised the manuscript accordingly (P1 L23ff, P3 L23ff, P29 L21ff, P30 L27ff, P30 L30ff).

3. It is disappointing to only see applications in 1-D. What is the reason for not modeling flow in 2-D, to use all three monitoring locations, and the full extent of the water table fluctuations? Is this something you plan to do in the future? Also, how to deal with the fact that the 1-D representation is unsuitable for significant drainage? After seeing Figure 1 and 2, it is easy to be a bit disappointed when only seeing results that consider GPR position 3 and no significant drainage. The tank has a nice 2-D subsurface architecture, but it is here simplified to a known 1-D layered system.
Reply:
As mentioned in the reply to remark 1, running the 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. This is the aim of the current paper, in conjunction with the earlier work of Buchner et al. (2012) that supplies all the prerequisites needed here. As the methods demonstrated here are capable of analyzing a number of measured radargrams simultaneously, they extend naturally to time-lapse common offset measurements acquired for more complicated subsurface architectures or during drainage conditions.

4. Figure 1: Make it very clear in figure and caption that you only consider data from GPR antenna 3. It is somewhat confusing to see this spatial representation, while all the treatment relates to one GPR position. I would not use the term "radargram" to

represent time-series of the GPR traces, as radargrams (e.g., page 6, line 7) are often thought of as a time-distance plot. Make it clear in the text that all the GPR results and simulations only model the trace at a given location over time and that no spatial information is treated (except depth).

Reply:

We agree and revised the manuscript accordingly (Fig. 1, Fig. 2, caption of Fig. 14). However, we did not separate the radargrams, because especially for people that are not used to time-lapse radargrams, having a corresponding common offset radargram of the initial state helps to associate the reflections and to understand their temporal evolution.

5. Explain clearly what is meant by subscale physics. These are all macroscopic representations, so why call them "subscale".

Reply:

The dynamics of the system is represented with a physics-based mathematical description for a predefined scale in space and time. In contrast, the physics below these scales is not represented explicitly. Instead, the macroscopic effects of this sub-scale physics are typically described heuristically. We clarified the manuscript (P4 L19ff).

6. I recommend that some pseudo-code is added for the algorithm used in 2.3.4. Is the method practical for 2-D and 3-D applications?

Reply:

We added a flowchart to explain the algorithm further (Fig. 3).
Going to 2D even to 3D is first and foremost a matter of computational effort with already 2D demanding significant time on a large computer cluster. No concepts or methods beyond of what we demonstrated in this paper are required, however.

7. 2.4 is called parameter estimation, but it is never written that the parameterization in terms of geology is assumed to be known +- epsilon. This is a very strong assumption and it would be much more difficult to solve the inverse problem if one would actually infer the "subsurface architecture".

Reply:

We agree but would like to point out, that estimating the architecture as demonstrated by Buchner et al. (2012) is a step that can (and should) be done prior to the actual estimation of the parameters because the architecture is invariant during the hydraulic experiment. Fine-tuning the position, as is done here, then ascertains that the entire inversion is self-consistent.

8. Equation 11. How are the standard deviations estimated in practice (see also page 15, line 11)? Are they due to observational errors (estimated how), modeling errors (estimated how) or purely ad hoc? Page 3, line 9: What is the implication of excluding data events for the global optimizer (simulated annealing) used?

Reply:

We clarified the manuscript (P17 L6ff).
The simulated annealing algorithm, as described in Sect. 2.4.2 is no global optimizer,

since the parameter update is only drawn from the neighborhood instead of the full parameter space. As the data set for low resolution is merely used for preconditioning, there are no significant implications. We also ran the inversions using random traces for preconditioning. This did not lead to significant changes in the results.
9, lines 5-6: I don't understand this statement at all. Is this simply related to the fact that you damp the update size or is it something else?
Reply:
We assume that the comment refers to page 11. The difference of the signal travel time and amplitude of associated events enters the cost function. Hence, if the number of associated events changes during the optimization, then the cost function becomes discontinuous. This happens, e.g., if the porosity of two saturated materials are similar during the optimization process. Then, the associated reflection will vanish and the measured events can not be associated anymore.

10, page 13, line 5: Why not estimate the source wavelet as a part of the inversion (frequency and shape)?
Reply:
We agree that the source wavelet does influence the reflected signal. Additionally, Dagenbach et. al. (2013), for example, showed that roughness of the material interfaces also influences the shape of the reflected wavelet. In particular to investigate the necessity to address these higher order uncertainties in a quantitative analysis, we did not represent them and investigated the structural residuals after the inversion. We propose that the effect of relevant representation errors on the estimated properties should be analyzed in a next step, similar as has been done for TDR data by Jaumann and Roth (2017). This is a significant effort well beyond the scope of this paper.

11. A transition is really needed when starting 3.2.1. Write explicitly that you first will consider a synthetic test case to gain knowledge about the information content in the data and the ability of the inversion to provide a reasonable model. Explain the geometry of this model, explain how noise was added to the generated data. Similarly, a transition is needed when starting 3.2.2 (e.g., After inversion, we find that the. . .)". In Figure 9, add estimated "by inversion".
Reply:
We agree and added transitions (P17 L12ff, P19 L17, and P23 L20ff).

12. Work a bit on the definitions of paragraphs (e.g., page 26).
Reply:
We revised the section 3.3.2 accordingly.

13. Reconsider the use of phenomenology in favour of more common language in hydrology: "the science of phenomena as distinct from that of the nature of being. An approach that concentrates on the study of consciousness and the objects of direct experience."
Reply:
We agree that this is one definition. However, the word is also used commonly in a range

of natural sciences from particle physics to meteorology (`https://en.wikipedia.org/wiki/Phenomenology`).

Smaller comments (suggestions):

Page 1, line 2: Should be "Ground. . ." not "ground. . ." Replace "to" with "that is suitable to" to clarify that the GPR method was not built explicitly for this application. There are many more applications of GPR.
Page 1, line 3: Remove "precisely". It is clear that a quantitative method (pretty clear) is used, so what is precisely supposed to mean? Especially given the rather low agreement with TDR estimates for the field data.
Page 1, line 8: Perhaps explain what an "association algorithm" is.
Page 1, line 20: Replace "monitors the hydraulic processes accurately" to "is sensitive to hydraulic processes". What is monitored with a TDR is essentially the dielectric constant, which indirectly is related to hydraulic processes.
Page 2, line 6: Remove "are the easiest and" with "offer". Maybe the measurement procedure is easier, but the real work is in the modeling and inversion. Not clear to me why this mode is easier than say borehole data.
Page 2, line 12: Add "indirect" before "information".
Page 2, line 14: add "to reproduce when used" before "for".
Reply:
We revised the manuscript accordingly.

Page 2, line 15: Replace "in" with "for", remove ")". Sentence starting on line 15 is not clear. Why is this information not as important when considering precipitation or flooding events?
Reply:
We revised the manuscript accordingly.
Generally, precipitation is spatially more homogeneously distributed compared to artificial irrigation.

Page 2, line 23: Remove "quantitatively", this statement does not add anything. Page 2, line 23: Remove "balance" with "are faced by an inherent trade-off between"
Reply:
We revised the manuscript accordingly.

Page 2, line 27: A fair bit of self-referencing throughout. Why not cite some of the many other works related to GPR modeling.
Reply:
Instead of repeating an extensive list of available literature on the topic, we tried to keep the number of references concise. Hence, we focused on those works that deal with estimation of subsurface properties and that influenced the manuscript.
We added some references (P2 L30ff).

Page 3, line 7: Replace "may even" with "lead to better convergence and may even"
Page 3, line 16 (and many other places. It should be "sensitive to", not "sensitive on".
Page 3, line 19: I know that both uses are correct, but I prefer to treat data in plural form: One datum, several data.
Page 3, line 12: Add "for porous media" before "the standard".
Page 4, line 18: Explain already here that the reason for ignoring the large drainage event is that Richards equation is solved in 1-D.
Page 6, line 8: Why only one antenna? Why not model this as a 2-D system?
Reply:
We revised the manuscript accordingly.

Page 6, line 15: Why conductivity at dc conditions. It should be the conductivity at around 400 MHz, typically 50% or so higher than the DC value.
Reply:
Please refer to the reply to comment number 2 of referee 2.

Page 6, line 20: Clarify that one normally has no idea about the power of the GPR source, only some basic idea about its shape. This implies that some sort of normalization of observed and simulated traces are needed.
Reply:
We revised the manuscript accordingly (P9 L5f).

Page 6, equation 7. Here, the dependence of temperature is included, but it is written later that this was not done and it led to errors on the scale of one standard deviation in time. Also, how are boundary conditions in the tank modeled in the EM code?
Reply:
In a field study, the distribution of soil temperature distribution is typically unknown. Hence, we inserted an estimate for the mean soil temperature. However, the error in the signal travel time can be calculated for a given travel path length, e.g., depending on the water content and the error in soil temperature. For the given model, the resulting error in the signal travel time exceeds one standard deviation of the signal travel time already for a deviation of a few Kelvin.
Perfectly matched layers are used as boundary conditions for the electromagnetic model (Sect. 3.1.1).

Page 6, line 31: Use Archie to explain how big this approximation is. Depends on the differences in porosity of the sands used.
Reply:
Besides the soil porosity, Archie's law (e.g., Friedman, 2005) depends on further parameters that are unknown a priori and may also vary in the subsurface architecture. We merely have estimates from TDR measurements for the electrical conductivity of the bulk. These vary during the experiment approximately between $1 \cdot 10^{-3}$ and $2 \cdot 10^{-2}$ $\mathrm{S\,m^{-1}}$.

Page 7, line 4: Should be "corresponding" instead of "corresponding".
Page 7, lines 13-14: Confusing as treatment to simulated and observed data are mixed. For example, (ii) only important for real data and (iii) only needed for simulated data. Please clarify what is done for (1) simulated data and (2) observed data.
Reply:
We revised the section 2.3 accordingly and added Fig. 3 for clarification.

Page 7, line 18: Is this correction valid for a dipole radiation pattern. Page 7, line 23: "i" in italics.
Reply:
Yes, this correction is valid for a dipole radiation pattern (see, e.g., Buchner (2012), p. 44ff).
We revised the manuscript accordingly (P8 L3).

Page 27, line 19: Replace "favourably" with "reasonably well".
Reply:
We revised the section 4 and don't use that sentence anymore.

References:

Buchner, J. S., Wollschläger, U., Roth, K. (2012). Inverting surface GPR data using FDTD simulation and automatic detection of reflections to estimate subsurface water content and geometry. Geophysics, 77(4), H45-H55.

Buchner, J. S. (2012). Constructive Inversion of Vadose Zone GPR Observations (Doctoral dissertation), Heidelberg University, `http://archiv.ub.uni-heidelberg.de/volltextserver/14171/4/Buchner_JS_Dissertation_2012.pdf`.

Dagenbach, A., Buchner, J. S., Klenk, P., Roth, K. (2013). Identifying a parameterisation of the soil water retention curve from on-ground GPR measurements. Hydrology and Earth System Sciences, 17(2), 611.

Friedman, S. P. (2005). Soil properties influencing apparent electrical conductivity: a review. Computers and electronics in agriculture, 46(1), 45-70.

Jaumann, S., Roth, K. (2017). Effect of unrepresented model errors on estimated soil hydraulic material properties. Hydrology and Earth System Sciences, 21(9), 4301.

Reply to referee comment 2

The authors deduce subsurface hydraulic properties by an inversion of time-lapse surface GPR measurements during an imbibition experiment of an artificial test site. The coupled inversion process includes a hydraulic simulation by solving the 1D Richards equation and a simulation of radar wave propagation by 2D finite-differences calculation. Water content distribution and electromagnetic soil properties are coupled by a petrophysical relation (CRIM). During the inversion, the misfit between events, i.e. traveltimes and amplitudes of selected reflections, in experimental and synthetic GPR data is minimised. The authors use an inversion scheme that combines several optimisation steps including global and gradient techniques. The approach is first demonstrated for synthetic data and later for experimental data. The result is a 1D subsurface model and for both predefined layers the characteristic hydraulic properties of a Brooks-Corey parameterisation of the water retention function is fitted. The presented work is a relevant contribution towards a non-destructive hydraulic characterisation of the subsurface, which is still an unsolved problem for the unsaturated zone.

Reply: We thank the reviewer for the constructive comments and suggestions. We revised the manuscript accordingly and refer to the revised version in the following.

However, the manuscript has to be overworked as the whole analysis including GPR data processing and inversion is somehow nebulous and difficult to follow. I would also suggest to shorten the text by writing more tersely, avoiding repetitions and possibly moving some parts into an appendix as e.g. GPR data conversion due to Bleistein, details on event detection/association and inversion. Besides this, some major points have to be clarified: 1. Amplitude handling: The formula used for spherical divergence correction for 3D data (P 7 Eq. 10) seems not correct. Correcting with square root of distance is used for 2D data. Also the dimensions of Eq. 10 do not fit. Various formulations of adequate gain functions for 3D (experimental) data is given in Yilmaz: Seismic Data Analysis (2001), e.g. Eq. (1-8a) $g(t) = \frac{v^2(t)t}{v_0^2 t_0}$*. The whole amplitude balancing in the manuscript is not clear to me. The radar traces are normalised several times (P8 L1-2) and normalisation is done relative to the maximal absolute amplitude, which is the first reflection. Is the reflector characteristics constant during the entire experiment? What is the advantage of the complicated amplitude adaption due to Bleistein (1986) compared to a simple correction of 2D circular divergence with the square root of the distance? I would suggest*

to provide a flow chart of amplitude handling for both experimental (3D) and synthetic (2D) radar data. It seems you use different amplitude handling for event detection and the inversion process?

Reply:

We revised the manuscript, however we decided not to move the GPR evaluation and optimization methods to the appendix since these are actually the essential parts of the manuscript.

Besides an amplitude correction, Bleistein (1986) also provides a correction for the signal frequency. Hence, we use this correction method.

The wave equation is typically transferred to the Helmholtz equation which may be solved with a Green's function approach. To yield the electric field, the resulting Green's function G is convoluted with the function f which is essentially the temporal partial derivative of the source current density ($f \propto \mu \partial_t J$):

$$\widehat{E}(\vec{x},\omega) = \int \mathrm{d}\vec{x}' \; G(\vec{x},\vec{x}',\omega) \; f(\vec{x}',\omega). \tag{1}$$

Similar to Bleistein (1986), also other authors, e.g., Miksat et al. (2008), propose Green's functions for a 3D point source and a 3D line source (in x-direction) which corresponds to a 2D point source. Theses Green's functions may be transferred into each other in the frequency domain using a correction factor $C_\omega(\omega)$ via

$$\widehat{G}^{\mathrm{3D}}(\vec{x},\vec{x}',\omega) = \widehat{G}^{\mathrm{2D}}(\vec{x},\vec{x}',\omega) \cdot C_\omega(\omega). \tag{2}$$

This correction factor is given by $C_\omega(\omega) = \sqrt{\frac{|\omega|}{2\pi\sigma_c}} \exp\left(-\frac{\mathrm{i}\pi}{4}\mathrm{sign}(\omega)\right)$, where σ_c denotes the integral of the velocity with respect to the length s of the ray trajectory $\sigma_c = \int c(s)\mathrm{d}s$. Since this correction factor is spatially constant, it may also be used to directly scale the Fourier transform of the electric field:

$$\begin{aligned}
\widehat{E}^{\mathrm{3D}} &= \int \mathrm{d}x\mathrm{d}y\mathrm{d}z \; \widehat{G}^{\mathrm{3D}} \; f && (3)\\
&= \int \mathrm{d}x\mathrm{d}y\mathrm{d}z \; \widehat{G}^{\mathrm{2D}} \; C_\omega(\omega) \; f && (4)\\
&= C_\omega(\omega) \int \mathrm{d}x \int \mathrm{d}y\mathrm{d}z \; \widehat{G}^{\mathrm{2D}} \; f && (5)\\
&= C_\omega(\omega) \int \mathrm{d}x \; \widehat{E}^{\mathrm{2D}} && (6)\\
&= C_\omega(\omega) \; C_i \; \widehat{E}^{\mathrm{2D}}. && (7)
\end{aligned}$$

This exploits that (i) the wave propagation is a linear problem in order to separate $C_\omega(\omega)$ and (ii) that the shape of the wave does not change in x-direction due to symmetry. Thus, the integration over the x-direction leads to the constant C_i (m). This the constant is independent of the frequency and hence does not change the electric field in the inverse Fourier transformation. Therefore, it is possible to directly scale the Fourier transform of the electric field with C_ω and to use the normalized amplitude of

the electrical field in the space domain. Thus, the value of the constant C_i is irrelevant. By separating the frequency and amplitude correction in the previous version of the manuscript, we set $\sigma_c = 1$ in the frequency correction and applied the correct value in the amplitude correction. We clarified section 2.1 as well as the normalization in the revised version of the manuscript (P9 L4ff, P9 L15f, P10 L4f, P10 L21f).
The first reflection is not always the one with the highest amplitude (see Figs. 13b and 16b – if the event has the maximal amplitude in both the simulation and the measurement, it has no error as both values are equal to 1, thus you can check that the event with maximal absolute amplitude is not always in the same reflection). The characteristics of the first reflector does change over the course of the experiment, due to the hydraulic dynamics (e.g., see marker (3) at Fig. 9).
A flowchart was added in the revised version of the manuscript (Fig. 3). Besides the 2D to 3D conversion and the event selection, simulated and measured data are treated the same.

2. Neglecting dielectric losses. I'm wondering if at frequencies of about 400 MHz, the impact of free water relaxation can be neglected and whether the imaginary part of permittivity has to be taken into account. When using complex permittivity of water according to Kaatze et al. (1989) and the CRIM formula and a DC conductivity of 0.003 S/m this results in: 3 dB/m (2 dB/m from free water relaxation, 1 dB/m from DC conductivity) for 10vol% water content and 5 dB/m (4 dB/m from free water relaxation and 1 dB/m from DC conductivity) for full saturation (40vol% water content). This means that up to 80% of total loss is caused by polarisation effects of free water. Neglecting these effects results in wrong amplitudes of the simulated data and I'm wondering how they can fit to the field data. At the end of the imbibition experiment (water table at -0.6 m) the amplitudes of the lower reflections (1 m saturated material above, i.e. 2 m two-way travel path) should appear to be approximately 8 dB (2.5 times) higher in the synthetic data than in the field data. I suggest to use the true complex permittivity of water at 400 MHz or, if the FDTD code cannot handle complex property values, an effective HF conductivity including both, DC conduction losses and HF polarisation losses.
Reply:
Using the parameters of Kaatze et. al. (1989) and the measurements of Light et. al. (2005) for the direct current conductivity yields the temperature and frequency dependency of the electrical conductivity of pure water shown in Fig. 1 of this reply. Due to the finite measurement time of the TDR traces, they yield an effective estimate of the electrical conductivity which is larger than the direct current conductivity. Hence, we corrected the notation in the revised version of the manuscript (e.g., P6 L13).

3. GPR forward calculation: Why is a 2D FDTD code used for a horizontally layered model? A 1D reflectivity method as e.g. used by Bradford et al., (2014) or a 1D FDTD code would be much more efficient. The power of FDTD is certainly that it can be used for complicated 2D/3D subsurface models and thus for inverting 2D/3D data with an according hydraulic simulation. However, in the presented study only 1D data are used and no outlook is given how to adopt the strategy to 2D or 3D problems. From this it is

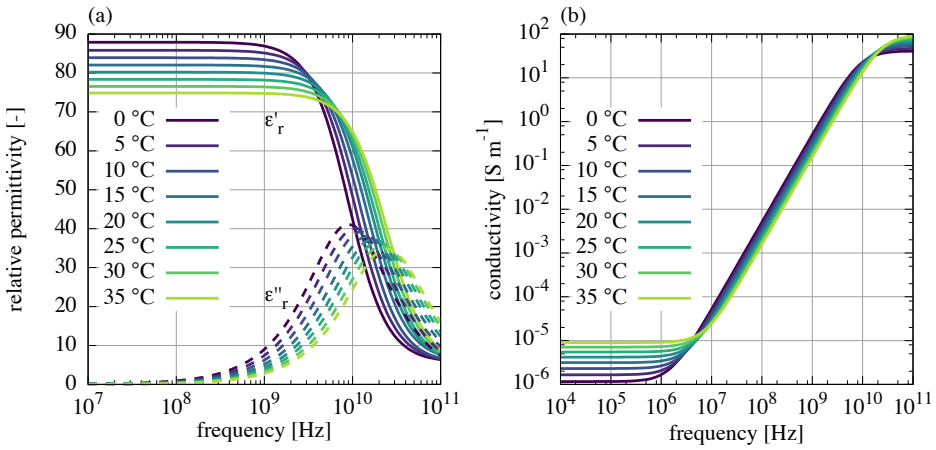

Figure 1: The temperature and frequency dependency of pure water using the parameters of Kaatze et. al. (1989) and the measurements of Light et. al. (2005) for the direct current conductivity.

not clear why the expensive 2D FDTD algorithm is used. The source wavelet of the simulation is different to the wavelet of the experimental data. When dealing with gradient interfaces as the capillary transition zone, the wavelet shape may have a big impact on the maximal amplitude of the reflected signal. Why not using the first reflected signal, which is used for normalisation, as source wavelet in the simulation?

Reply:

In a 1D model, the shape of the wavelet would change, increasing the deviation to the measured 3D antenna signal. Also, using a 1D model would lead to a depth-dependent error in the ray travel path (Fig. 2 of this reply).

We added an outlook on the further application of the proposed algorithm in the revised version of the manuscript (P30 L27ff and P30 L30ff).

Concerning the estimation of the source wavelet, please also note the reply to referee comment 1, point 10.

Inversion. The complex inversion scheme is a nesting of global and gradient methods. It is somewhat nebulous and it's difficult to get an impression of the quality of original data fit. Why do you use different but relatively narrow boundaries (fit ranges) for the inversion parameters of the two layers? By doing this, the inversion result is biased by a-priori information that is usually not known but the actual aim of the investigation. The inversion should work with the same (broader) fit range for both layers. If not, it cannot be adopted to the field. The fit ranges should be used to provide outer boundaries of the deduced material properties in Fig. 9 and Fig. 14. I'm also missing a figure showing experimental GPR data traces and synthetic traces based on the inversion models to prove that the experimental data are well described. This figure should include the resulting synthetic radar traces of the ten best inversion results to get an idea of the

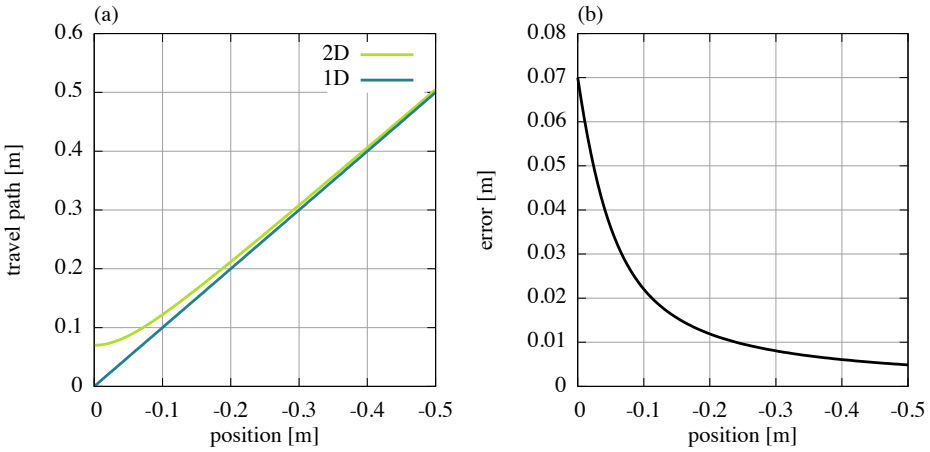

Figure 2: (a) The difference in travel path for 1D and 2D, and (b) the error in travelpath

fluctuations and an idea of the fitting quality.

Reply:

The global-local approach is a common method. It is also typical that larger problems are approached with a preconditioning step.

The deviation in amplitude and travel time as well as the residuals are given in the Figs. 13 and 16.

The fit ranges cover sandy materials. In a field application of the method, the material type of the subsurface can be sampled with a geological drill, e.g., using a "Pürckhauer".

The proposed method is intended to be used in combination with published multichannel method Gerhards et. al (2008) and Buchner et. al. (2012) which provide the architecture structure, layer depth, and average water content. These methods have been shown in 2D.

The applied optimization methods use a uniformly distributed prior information. Hence, even if the prior information was included in the cost function, the fit parameter range would not bias the parameter estimation. If the parameter range was too small, the resulting parameters would be close to the boundary. This is not the case in this study.

Choosing single traces out of the time-lapse radargram with many traces does not suffice to proof the quality of the fit, because, while the fit might be good for one trace, it could be very bad for other traces. At least in the synthetic study, where there are no additional reflections from the walls and compaction interfaces, the true and the estimated radargrams are very similar. It would be difficult to discern them visually. Thus, we show the evaluated events, their deviations in signal travel time and amplitude as well as the according residuals. This approach allows to pinpoint deviations of the simulation and the measurement very precisely. Hence, showing the 10 best members would require to show 10 plots such as Fig. 13 per study.

5. In the analysis, the reflection of the compaction layer is excluded. If this interface

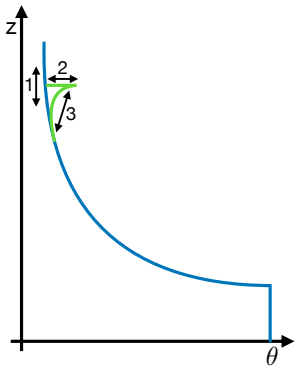

Figure 3: Sketch of the influence of a compaction layer (green) on the water content distribution (blue). Main uncertainties about the shape of the influence are indicated with arrows (1, 2, 3).

causes a GPR reflection, this must be caused by different water contents on both sides and hence, there must be significant differences in the material hydraulic properties (see P24 L25ff). So why should I ignore an interface that is present in the subsurface and reflects changes in hydraulic properties? Please explain.

Reply:

In order to describe the influence of the compaction interface on the water content distribution quantitatively, at least three uncertainties would have to be estimated (Fig. 3 in this reply): The vertical position of the compaction interface (1), the change of the pore-size distribution at the compaction interface (2) as well as the change of the pore-size distribution with increasing distance from the compaction interface (3).

In vertical soil samples taken at ASSESS, e.g., with a Pürckhauer, we could only visually discern the different sands but no compaction interfaces. Since the quantitative influence of the compaction interfaces on the hydraulic dynamics is unknown a priori, we assume homogeneous material properties in this first step, in particular to investigate the necessity for a detailed quantitative analysis of compaction interfaces. Relevance of this representation error is indicated by the structural residuals after the inversion. The results of this study, i.e. the effect on the estimated parameters and the remaining residuals, suggest that the representation of the compaction interfaces in ASSESS is relevant. Hence, we propose that the effect of all relevant representation errors on the estimated properties should be analyzed in a next step, similar as has been done for TDR data by Jaumann and Roth (2017). This is a significant effort well beyond the scope of this paper.

6. The title is misleading, I would suggest to delete "...and subsurface architecture. . . " as this would imply at least a 2D subsurface model. The section headings of 2.2 sound unusual to me. From a geophysical perspective the following headings would give a better description: 2.2.1 Water dynamics, 2.2.2 Hydraulic material characterisation 2.2.3

Time lapse experiment 2.2.4 GPR investigation and electromagnetic material characterisation.
Reply:
We changed the title to "... and layered architecture..." making it more precise. We clarified the introduction of the representation and thus the titles in the revised version of the manuscript (P4 L19ff).

Further comments:

(P2 L27) References are a bit biased by the own workgroup. E.g., when introducing the FDTD method I would expect the basic work of Yee, Taflove. . . and, e.g., the former ETHZ geophysics group or from the gprMax developers.
Reply:
Instead of repeating an extensive list of available literature on the topic, we tried to keep the number of references concise. Hence, we focused on those works that deal with estimation of subsurface properties and that influenced the manuscript. Still, we agree that classical work on methods should be acknowledged and added the references accordingly (P2 L30ff).

(P6 L23ff) The CRIM formula uses the square root of permittivities (see your original reference: Birchak et al., 1974). There is no need to first define a general formulation with an exponent α, which is not the original CRIM formula, and then fix the exponent $\alpha = 0.5$. Keep it simple and use the square root from the beginning.
Reply:
We revised the manuscript accordingly (P7 L3).

Further comments *(P6 L9) You should describe that you use the static permittivity of water (which is acceptable for 400 MHz, at least for the real part of permittivity).*
Reply:
We revised the manuscript accordingly (P6 L16f and P7 L7ff).

(P7 L14) "...removal of the direct and trailing signal". What is the trailing signal? Is this the interference of ground wave, crosstalk, reflection at the ground surface and the antenna metal shielding? In Fig. 3, a part of this trailing signal is remaining, which is confusing. Why not muting this part?
Reply:
We clarified the paragraph accordingly (P7 L21ff).

(P7 L15) ". . . we pick the direct signal and subtract it from the radargram" is confusing. Not the signal is subtracted but the travel time.
Reply:
We revised the manuscript accordingly (P7 L24f).

(P8 L5) "normalized amplitude (original amplitude)". Rephrase, as the amplitude is

either normalized or original.
Reply:
We rephrased the section 2.3.2 accordingly.

(P8 L6) "amplitude is amplified quadratically with travel time" means they are corrected for spherical divergence twice consecutively? Is this just an arbitrary gain function that showed to work well and to correct for spherical and intrinsic attenuation at the specific site? Please explain.
Reply:
This is an arbitrary gain function that showed to work well for the detection of events at lower travel times. This gain function is merely used for the detection, the travel time and amplitude. We clarified the manuscript accordingly (P9 L8ff).

(P11 L3) Eq. 11: I think the expression has to be divided by M to get the classical χ^2 *with* $\chi^2 = 1$ *if the data are described within the error.*
Reply:
Since the number of events (M) changes over the course of the optimization, this would lead to a balancing of the number of associated events and the associated residuals. Hence, the association of two events would only be added, if their residual is smaller than the average residual (what is very unlikely). Thus, the optimization algorithm would tend to decrease the number of associated events in order to decrease the cost.

(P11 L4) How is the standard deviation of the normalised travel times and amplitudes calculated, i.e. what are the input data?
Reply:
We added this information to the revised manuscript (P17 L6ff).

(P13 L33) ". . . infinite dipole pointing in x dimension". This should be y dimension (into the plane of projection). Please provide x,y direction in Fig. 6.
Reply:
We clarified the dimensions by adding them to the labels of Fig. 7.

(P13 L34) a Ricker function is the second derivative of a Gauss-fct, not the first derivative
Reply:
We improved the sentence (P14 L22f).

(P18 L26) What is the meaning of amplitude information of a single channel? You are using a single channel GPR system and only one antenna, so this expression is confusing.
Reply:
We clarified the paragraph (P23 L1ff).

(P24 L30) Couldn't the uncertainty of the groundwater table relative to the ground sur-

face be overcome by simple levelling the ground surface?
Reply:
In principle, this was possible. However, the bottom and the surface ASSESS site is inclined relative to the groundwater level (approx. 0.1 m over the length of the site (Jaumann and Roth, 2017)). Yet, when applying the method in the field, the uncertainty of the position of the groundwater level is also likely to increase with the distance from the well.

(P26 L20ff) is a partial repetition of (P22 L5ff (the lower line 5)).
(P26 L26) ". . . and that (ii) the direct electric conductivity can be assessed with GPR measurements". I cannot understand the context.
(P27 L13ff) This is again a partial repetition of (P22 L5ff) and (P26 L20ff)
Reply:
We revised the section by deleting the short summary (P27, L27ff).

(P27 L18) Better use "constant offset" (CO) instead of "single-channel" GPR data.
Reply:
In this sentence, we differentiate between single-channel and multi-channel approaches, e.g., used by Buchner et. al. (2012).

(Fig. 3, caption) Are these synthetic or experimental data?
Reply:
These are simulated data. We clarified the caption of Fig. 4 accordingly.

Fig. 6: Which E-field component is shown, what is the x and y direction?
Reply:
The x-component of the E-field is shown. We directions are now given in the labels of Fig. 7.

Fig. 12, bottom: y-axis label: standardized residual: Does 10 mean that the residual is 10 times the STD or should it be 10
Reply:
We clarified the caption of Figs. 13 and 16. It is 10 times the standard deviation, which is for the signal travel time $6 \cdot 10^{-4} \cdot 60$ ns $= 0.36$ ns. Hence, 10 times the standard deviation corresponds to 3.6 ns.

Fig. 13: I suggest to split the figure into two individual figures as it might be very confusing to mix the 2D radar section with the time lapse data at one location. The label for the groundwater table reflection is "I" in the upper radar section and "2" in the lower time-lapse data. Actually, it's very hard to distinguish the label "I" from the label "1" in the upper radar section. Please use the same and distinct labels for the GWT reflection in all figures.
Reply:
Especially for people that are not used to time-lapse radargrams, having a correspond-

ing common offset radargram of the initial state helps to associate the reflections and to understand their temporal evolution.
We clarified the caption of Fig. 14 and increased the size of the markers. Since the groundwater table is fluctuating, Arabic numbers indicate the water induced reflections at different times. Thus, the labels are used consistently in all the figures.

References:

Bleistein, N. (1986). Two-and-one-half dimensional in-plane wave propagation. Geophysical Prospecting, 34(5), 686-703.

Buchner, J. S., Wollschläger, U., Roth, K. (2012). Inverting surface GPR data using FDTD simulation and automatic detection of reflections to estimate subsurface water content and geometry. Geophysics, 77(4), H45-H55.

Holger Gerhards, Ute Wollschläger, Qihao Yu, Philip Schiwek, Xicai Pan, and Kurt Roth (2008). "Continuous and simultaneous measurement of reflector depth and average soil-water content with multichannel ground-penetrating radar." GEOPHYSICS, 73(4), J15-J23.

Jaumann, S., Roth, K. (2017). Effect of unrepresented model errors on estimated soil hydraulic material properties. Hydrology and Earth System Sciences, 21(9), 4301.

Kaatze, U. (1989). Complex permittivity of water as a function of frequency and temperature. Journal of Chemical and Engineering Data, 34(4), 371-374.

Light, T. S., Licht, S., Bevilacqua, A. C., Morash, K. R. (2005). The fundamental conductivity and resistivity of water. Electrochemical and Solid-State Letters, 8(1), E16-E19.

Miksat, J., Müller, T. M., Wenzel, F. (2008). Simulating three-dimensional seismograms in 2.5-dimensional structures by combining two-dimensional finite difference modelling and ray tracing. Geophysical Journal International, 174(1), 309-315.

Yilmaz, O. (1987). Seismic data processing, volume 2 of Investigations in Geophysics. Society of Exploration Geophysicists.

Soil hydraulic material properties and subsurface 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 effective soil hydraulic material properties is essential to predict soil water movement. ground-penetrating Ground-penetrating radar (GPR) is a non-invasive and non-destructive geophysical measurement method that is suitable to monitor the hydraulic processesprecisely. 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 this signal the GPR signal originating from a fluctuating groundwater table is suitable to accurately estimate the subsurface architecture and the associated effective soil hydraulic material properties with inversion methods. Therefore, 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 event detection and association algorithm algorithm which detects relevant signals in the radargram (events) and extracts the corresponding the signal travel time and amplitude. This algorithm is applied to simulated and measured radargrams and the detected events are associated automatically. Using events instead of the full wave regularizes the inversion as it allows to focus on the relevant measurement signal. For optimization, we use a global-local approach with preconditioning. Starting from an ensemble of Latin hypercube drawn initial parameter sets, we sequentially couple the simulated annealing algorithm with the Levenberg–Marquardt algorithm. We apply the method The method is applied to synthetic as well as measured data from the ASSESS test siteand . We show that the method yields accurate reasonable estimates for the soil hydraulic material properties as well as for the subsurface architecture by comparing the results to references derived from time domain reflectometry (TDR) and subsurface architecture ground truth data.

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 monitors the hydraulic processesaccuratelyis 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.

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) are the easiest and the offer the most flexible approach. They have the disadvantage, 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 (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 in for small-scale heterogeneity or fingering). If the infiltration is done artificially, accurate knowledge of the spatial distribution of the infiltration flux is required. Also simultaneous GPR measurements during the infiltration process are difficult as 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) as the hydraulic 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 balance 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 layered subsurface architecture. Alternatively, Maxwell's equations can be solved numerically with the finite differences

time domain (FDTD) method (e.g., Buchner et al., 2012)(e.g., Taflove and Hagness, 2000). This method is computationally expensive, but grants full flexibility concerning the source wavelet and the subsurface architecture .

Due to the inherent oscillating nature of the electromagnetic signal, inversion of GPR data generally demands globally 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 applied on GPR measurements (e.g., Léger et al., 2014, 2015; Jadoon et al., 2012)(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. 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 material, the transition zone above the groundwater table exhibits a smooth variation of the relative permittivity. As 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) 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, relaxationequilibration, and drainage. They also concluded that the transition zone reflection is sensitive on to hydraulic material properties.

In

This work builds upon previously published methods for simultaneous estimation of the subsurface architecture and the effective water content based on on-ground multi-channel common offset GPR measurement data (e.g., Gerhards et al., 2008; Buchner et al., 2012). In order to develop methods that improve these estimates, the ASSESS test site was forced with a fluctuating groundwater table ensuring large hydraulic dynamics. In this work, we use the resulting transition zone reflection together

with reflections at material interfaces to determine estimate the subsurface architecture and the corresponding hydraulic material properties. Therefore, the ASSESS test site was forced with a fluctuating groundwater table ensuring large hydraulic dynamics. The time-lapse measurement data was acquired with a single channel 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, the experiment was monitored with a stationary on-ground bistatic antenna operating bistatic antenna that operates at a center frequency of 400 MHz. Similar to Buchner et al. (2012), we This leads 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 2Dand employ a new semi-automatic heuristic . 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 . This allows the optimization procedure to focus on the relevant information in the radargram and decreases the number of local minima. We draw an ensemble of initial in the radargram (events). This approach is applied to both the simulated and measured radargram and the corresponding events are associated automatically. Using the association of the events regularizes the optimization. We apply a global-local optimization approach with preconditioning. For this purpose, we draw parameter sets with the Latin hypercube algorithm . These parameter sets that serve as initial parameters for the preconditioning step. In this step, the simulated annealing algorithm which is sequentially coupled with the and Levenberg–Marquardt algorithm are sequentially coupled allowing only a limited number of iterations working on a subsampled data set. Subsequently, the resulting parameters of the preconditioning step serve as initial parameters for another run of the Levenberg–Marquardt algorithm, now working on the full data set. We show that this procedure allows to accurately estimate the subsurface architecture and the associated effective hydraulic material properties for synthetic and measurement data.

2 Methods

2.1 ASSESS

The ASSESS test site 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 consists of three different kinds of sand (materials A, B, and C). Its effective provides an effectively 2D subsurface architecture is visualized in subsurface architecture consisting of three kinds of sands (Fig. 1. The approximately $2\ \mathrm{m} \times 20\ \mathrm{m} \times 4\ \mathrm{m}$ large site is equipped with a well to monitor and manipulate the groundwater table, a weatherstation, a tensiometer (UMS T4-191), as well as 32 soil temperature and TDR sensors. A geotextile separates the sand from an approximately). Below the sands, a $0.1\ \mathrm{m}$ thick gravel layer below, which ensures a rapid water pressure distribution 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 groundwater well to the rest of the test site. Below this gravel layer, 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 is measured automatically with a tensiometer (UMS T4-191)

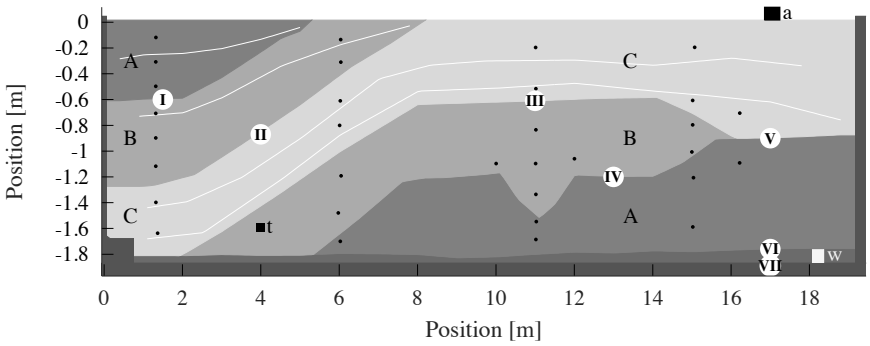

Figure 1. ASSESS emulates provides an effective effectively 2D geometry with subsurface architecture consisting of three distinct kinds of sand (A, B, and C). The hydraulic state can be During the experiment that is evaluated in this work, the groundwater table was manipulated with via a groundwater well (white square at 18.3 m18.3 m)and is . The resulting hydraulic dynamics was monitored with three a stationary GPR antennas antenna (1, 2, and 3black square at 17.05 m), a tensiometer (black square, at 4.0 m), and 32 TDR sensors (black dots). The A gravel layer at below the bottom ensures sands allows for a rapid water pressure distribution over at the lower boundary of the site. An L element (left wall, at 0.4 m) and compaction interfaces (white lines) were introduced during the construction of the site. Additionally to those visualized, GPR evidence indicates additional compaction interfaces (Fig. 14). Roman numbers (I)–(VII) indicate material interfaces referred to in the text. Note the different scales on the horizontal and vertical axes.

and manually in the groundwater well. 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 walls at each side at least partially consisting of reinforced concreteconfines the site. As the test site is built into a former fodder silo, a concrete L element serves as additional wall. In order to stabilize the material during the construction, it was compacted . During the construction, the materials were compacted with a vibrating plate for stabilization introducing compaction interfaces. In addition to those shown in Fig. 1, GPR measurements indicate even more compaction interfaces (Fig. 14).

2.2 Representation

We follow Bauser et al. (2016) and define the *representation of a system* 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: dynamics (mathematical description) (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, subscale physics (material properties), forcing (superscale physics) , and states(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 Richards equation (Richards, 1931),

is the standard model to describe the propagation of the volumetric water content θ $(-)$ and the matric head h_m (m) in space and time t (s) . The solution of 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 equationrequires the specification of material properties, namely , the soil water characteristic $\theta(h_\mathrm{m})$ and the hydraulic conductivity function $K_\mathrm{w}(\theta)$, which are (i) highly non–linear, (ii) varying over many orders of magnitude, (iii) showing hysteretic behavior, (iv) impossible to determine a priori, and (v) very expensive to measure directly. The are required as material properties. The direction of gravity is indicated with the unit vector in z-direction $\boldsymbol{e}_z$indicates the direction of gravity, typically pointing downwards.

2.2.2 Subscale Sub-scale physics

We choose the Brooks–Corey parameterization (Brooks and Corey, 1966) for the soil water characteristic $\theta(h_\mathrm{m})$, because since it describes the materials in ASSESS appropriately (Dagenbach et al., 2013). Neglecting hysteresis, this parameterization may be inverted Inverting this parameterization for $\theta_\mathrm{r} \leq \theta \leq \theta_\mathrm{s}$, leading to including parameters representing 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 $(\lambda > 0)$.

Inserting

The hydraulic conductivity function $K_\mathrm{w}(\theta)$ is parameterized combining the Brooks–Corey parameterization into with the hydraulic conductivity model of Mualem (1976)yields the parameterization for the hydraulic conductivity function where K_0 $(\mathrm{m\ s^{-1}})$ is the . 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 and τ $(-)$ K_s $(\mathrm{m\ s^{-1}})$ and a fudge factor τ $(-)$.

2.2.3 Forcing

The ASSESS site In order to provide the measurement data for this study, ASSESS was forced with a fluctuating groundwater table leading to three characteristic phases (Fig. 2): (i) two characteristic phases comprising an initial drainage phase , (ii)

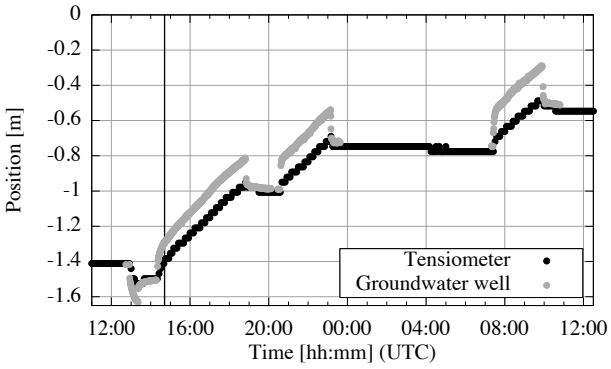

Figure 2. During the experiment with three two distinct phases (initial drainage , multistep imbibition, and multistep drainage imbibition – separated by the vertical linesline), the position of the groundwater table was measured manually in the groundwater well and automatically with the tensiometer (Fig. 1). Notice that the difference between them is proportional to the driving force of water flow in the gravel layer.

and a multistep imbibition phase , and (iii)multistep drainage phase. We neglect evaporation in the following, because the experiment took place (Fig. 2). The experiment was realized at the end of November and the weather was cloudy with 2–7 °C air temperature. The last precipitation was measured approximately 10 days before the experiment. More Hence, evaporation is neglected in this work. Further details about the experiment are given

In this work, we only focus on the initial drainage and multistep imbibition phase.

2.2.4 State

The During the experiment, the hydraulic state was monitored with GPR as well as with measurements of the position of the groundwater table in the groundwater well and at the position of the tensiometer . We used three shielded bistatic single channel as well as with a stationary GPR antenna (Fig. 1). This shielded bistatic single-channel 400 MHz GPR antenna pairs (Ingegneria dei Sistemi S.p.A., Italy) . These antennas are referred to as antenna 1, 2, and 3, respectivelyhas an internal antenna separation of 0.14 m. The measurement resolution was set to 2048 samples for 60 ns. In order to analyze the initial state of the test site, a one multi-channel common offset measurement was acquired with antennas 1 and 2. The internal separation of the transmitter and receiver of these antennas is 0.14 m. During the experiment, the antennas were used to measure three time-lapse radargrams. In this work, we focus on the quantitative evaluation of the time-lapse data from GPR antenna 3 (Fig. 1). These two different but structurally identical antennas. The acquired GPR data are analyzed in detail in Sect. 3.3. Additionally, a mean soil temperature ($T_s = 8.5$ °C) and a mean direct current conductivity ($\sigma_{dc} = 0.003$ S m^{-1}electrical conductivity ($\sigma = 0.003$ S m^{-1}) was estimated from TDR related measurements available in ASSESS.

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 electromagnetic field consisting of the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ (Jackson, 1999):

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

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

The dielectric permittivity $\varepsilon = \varepsilon_0 \varepsilon_r$, magnetic permeability $\mu = \mu_0 \mu_r$, and direct current conductivity σ_{dc} electrical conductivity σ are generally spatially variable and represent the electromagnetic properties of the subsurface. Here, we use the static relative permittivity $\varepsilon_r = \lim_{\omega \to 0} \varepsilon_r(\omega)$ and hence neglect dispersive effects ($\partial \varepsilon_r / \partial \omega = 0$) as well as the imaginary part of the dielectric permittivity its imaginary part ($\varepsilon_r \in \mathbb{R}$). 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 $\varepsilon_{\mathrm{r}} = \varepsilon_{\mathrm{r,b}}$ is calculated from the water content distribution θ resulting from the Richards equation using the complex refractive index model (CRIM) (Birchak et al., 1974):

with the geometry parameter $\alpha = 0.5$ (**?**). In order to apply the CRIM , (CRIM, Birchak et al., 1974):

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

The application of the CRIM requires knowledge of the porosity ϕ, the relative permittivity of water $\varepsilon_{\mathrm{r,w}}$, the relative permittivity of air $\varepsilon_{\mathrm{r,a}}$, and the relative permittivity of the soil matrix ε_{s}have to be known. The relative permittivity of air $\varepsilon_{\mathrm{r,a}}$ was set to 1. Assuming that the sand Since we assume that the soil matrix consists mainly of Quartz (SiO_2) grains, the relative permittivity of the soil matrix $\varepsilon_{\mathrm{r,s}}$ was set to 5 (Carmichael, 1989). The We further assume that porosity ϕ is assumed to be equal to the saturated water content θ_s (Eq. 2)which is estimated from the data. Following Kaatze (1989), we parameterize the . The dependency of the static relative permittivity of water $\varepsilon_{\mathrm{r,w}}$ on the soil temperature T_{s} (°C) with is parameterized following Kaatze (1989)

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

In this work, the required direct current conductivity σ_{dc} The electrical conductivity σ of the subsurface is assumed to be constant in the whole architecturespace and time. As for the relative permittivity, we neglect dispersive effects of the electrical conductivity ($\partial \sigma / \partial \omega = 0$).

2.3 GPR Analysis

Similar to Buchner et al. (2012), we extract the signal

Following Buchner et al. (2012), the travel time t and the according corresponding amplitude A at for M samples of the GPR signal (events)

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

are extracted with a heuristic approach. This Hence, this allows us to focus on the phenomena that are represented in the model and to exclude events of, e. g., reflections originating from compaction interfaces or confining walls. . However, this procedure demands an automatic event association algorithm which associates to associate the events extracted from the measured signal with events extracted from the simulated signal. ThusAs outlined in Fig. 3, the evaluation method presented in this section consists consist of four steps: (i) signal processing, (ii) event detection, (iii) event selection, and (iv) event association.

2.3.1 Signal processing

The GPR signal is processed for further evaluation according to the following steps: (i) time-zero correction, (ii) dewow filter, (iii) 2D to 3D conversion, (iv) removal of the direct and trailing signalsignal (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 .

As the time-zero of the GPR antennas changes over time, we pick the direct signal and subtract it from the evaluated travel time from each trace of the radargram for time-zero correction. ThenSubsequently, a dewow filter is applied to subtract inherent low frequency low-frequency wow noise of the GPR signal. Since 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 (Bleistein, 1986). Note that ASSESS is built accordingly (Sect. 2.1). In this work, the conversion is done for the frequency and the amplitude separately. First, each For the conversion trace is transformed to the frequency domain with the fast Fourier transform (FFT, denoted by $\widehat{\ }$). Afterwards, the amplitude electric field is modified depending on the angular frequency ω: where i

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

where i is the complex unit, $\omega_k = \Delta\omega \cdot (k' - \frac{K'}{2})$ ($k' \in \{1, \dots, K'\}$, K' is number of samples per trace enlarged to the next power of two). C_i is a constant (m), and σ_c denotes the integral of the velocity with respect to the arclength s of the ray trajectory. Assuming a direct ray path and horizontal reflector with the reflector distance d and mean square root of dielectric permittivity $\sqrt{\varepsilon_r}$ along the ray path, this is leads to

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

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.,

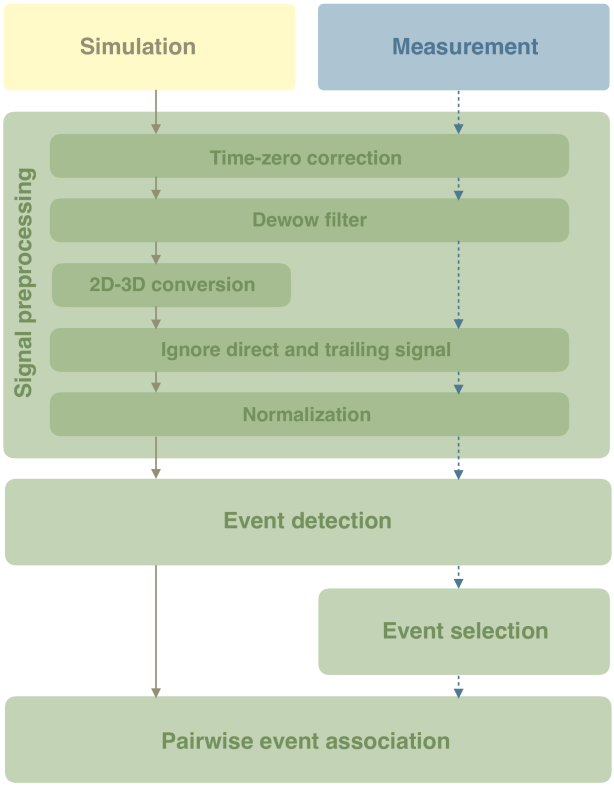

Figure 3. The GPR data evaluation method presented in this section consists of four main steps. In the first step, the signal is preprocessed (Sect. 2.3.1). The 2D to 3D conversion in this step is applied to the simulated data. In the second step, extrema in the GPR signal are detected (Sect. 2.3.2). 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).

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.

Accounting for energy dissipation in 3D requires additional manipulation of the amplitude. Assuming a direct ray path and horizontal reflector with the reflector distance d and mean square root of dielectric permittivity $\overline{\sqrt{\varepsilon}}$ along the ray path, this is done via

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

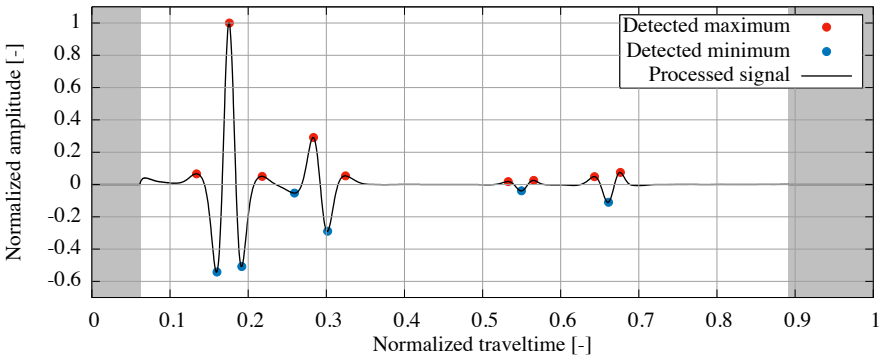

Figure 4. The detected events of the first trace of the synthetic radargram analyzed in Sect. 3.2. The amplitude of a trace is searched for extrema with a neighborhood search algorithm. For the subsequent evaluation, the amplitude of the detected events is normalized to the maximal absolute amplitude 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 preprocessing step (Sect. 2.3.1) and possible events close to these signals are ignored.

2.3.2 Event detection

To facilitate the identification of relevant events at large signal travel times, the normalized amplitude (original amplitude) preprocessed amplitude is amplified quadratically with travel time using an arbitrary gain function that showed to work well. Subsequently, the extrema extrema of the amplified amplitude are detected with a local neighborhood search. Then, we We keep a predefined number of events (15) with the largest amplified absolute amplitude. If the original non-amplified preprocessed amplitude of an detected extremum is below a predefined amplitude threshold (0.006)it , 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 original non-amplified preprocessed amplitude data. The resulting amplitude and travel time of the extremum are is directly used for the following evaluation.

2.3.3 Event selection

After the event detection, the measured signal is inspected manually together with and the detected events are inspected manually. In this one-time preprocessing step events can either be deleted or added manually. This ensures 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.

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. Therefore, 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 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. We use Finally, the combination with the minimal summed absolute travel time difference is used. Afterwards, the events 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). We iterate over all according combinationsto find After iterating over all allowed combinations, the association with the maximal number of associated events and the minimal summed absolute travel time difference . It is used. Note that it is critical to also consider combinations where some intermediate events (e.g., $(t_{\mathrm{s},2}, A_{\mathrm{s},2})$ in Fig. 5) can not be associated.

After the association of the events, outliers are detected by calculating the mean and standard deviation of the travel time differences. All associations are discarded which 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 in of each trace.

2.4 Parameter estimation

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 the Latin hypercube algorithm (implemented by the pyDOE package, https://github.com/tisimst/pyDOE).

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). Hence, as Since time-lapse GPR data are highly correlated in experiment time (e.g., Fig. 14), we equidistantly subsample the number of traces of the time-lapse GPR radargram and generate a data set with lower temporal resolution. We use those data Those data are used to improve the 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 , fast, and easy to implemented and fast parameter update. Subsequently, the resulting parameters The resulting parameters then serve as initial parameters for the Levenberg–Marquardt algorithm (Sect. 2.4.3) concluding which concludes the preconditioning step. The preconditioned parameter sets 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.

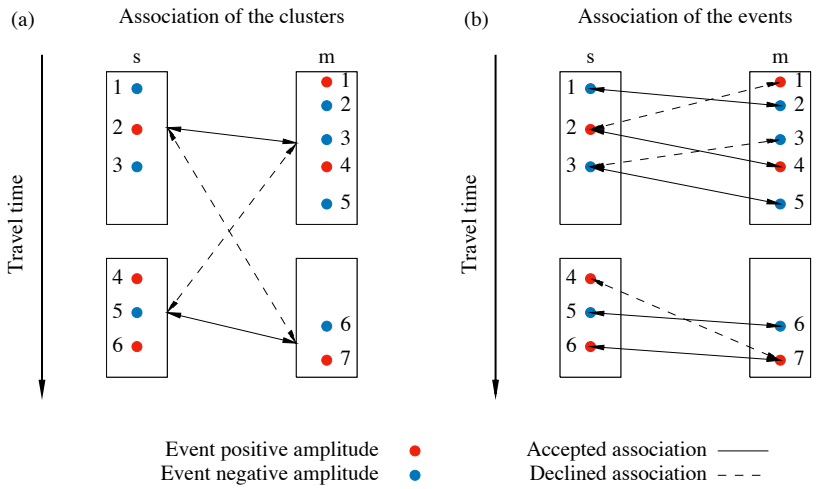

Figure 5. Exemplary association of simulated (s) and measured (m) events with indices 1–6 and 1–7, respectively. The color of the dots indicates the sign of amplitude of the events. (a) The detected events (Fig. 4) are aggregated in clusters to minimize the number of possible event combinations. The 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 $(t_{\mathrm{s},1}, A_{\mathrm{s},1})$ is associated with event $(t_{\mathrm{m},2}, A_{\mathrm{m},2})$, event $(t_{\mathrm{s},2}, A_{\mathrm{s},2})$ can not be associated with event $(t_{\mathrm{m},1}, A_{\mathrm{m},1})$, if $t_{\mathrm{m},1} < t_{\mathrm{m},2}$ or $\mathrm{sign}(A_{\mathrm{s},2}) \neq \mathrm{sign}(A_{\mathrm{m},1})$. 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.

2.4.1 Objective Cost function

Assuming P parameters p_π and M associations 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 χ^2 objective function cost function $S(\boldsymbol{p})$ is given by

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

with the constant standard deviation of the measured normalized travel times $\sigma_{t,\mu} = \sigma_t$ and of the measured normalized amplitudes $\sigma_{A,\mu} = \sigma_A$leading to the . 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 analysis applied GPR data evaluation (Sect. 2.3), the χ^2 cost function is not necessarily convex and may even be discontinuous at some points, in particular as the number of associated events M is not necessarily constant may change during the minimization process. To compensate for Hence, adding and removing simulated or measured events associations of events requires a compensation to prevent the cost function from becoming discontinuous. To this end, Buchner et al. (2012) introduced *tagging*. If the number of measured events is smaller

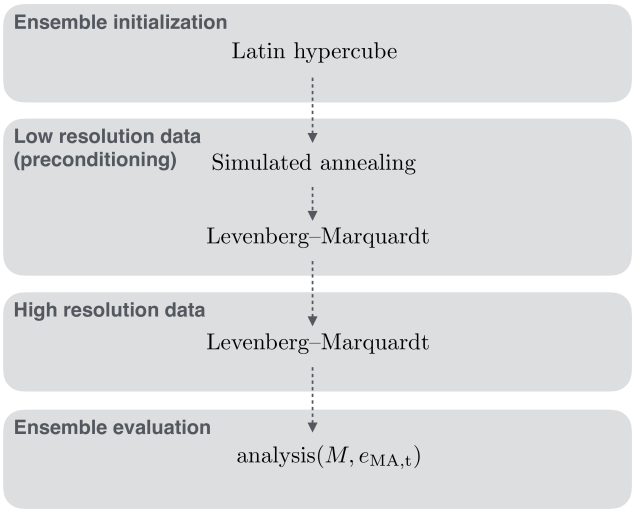

Figure 6. We choose an iterative a sequentially coupled parameter estimation procedure which (i) allows to minimize the computational cost and (ii) facilitates the implementation of tagging (Sect. **??**2.4.1). Therefore, we precondition the Latin hypercube sampled parameter sets are preconditioned with a data set with reduced number of traces (low resolution data) by sequentially coupling the simulated annealing algorithm and use optimization algorithms which compare the parameter sets sequentiallyLevenberg–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).

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 objective cost function is added to the previous objective function valuecost. 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 objective function valuecost.

2.4.2 Simulated annealing

We choose the simulated annealing algorithm (Press, 2007) to start the minimization of the objective cost function (Eq. 11), because this algorithm is gradient-free and updates the parameters statistically. Additionally, randomly. Besides the implementation of tagging (Sect. **??**)2.4.1), which can be implemented easily and it also allows uphill steps, which can be favorable if the events are not yet associated to their appropriate reflection.

If the parameter update is drawn from the whole parameter space, the algorithm is globally convergent. However, this approach is typically inefficient. We mainly use the simulated annealing algorithm to find a parameter set that associates the events to their appropriate reflection such that the more efficient gradient-based algorithm can take over. Hence, we search the neighborhood for better parameters starting from Latin hypercube sampled initial parameters $\boldsymbol{p}_{\pi,0}$. 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 $\boldsymbol{p}_{\pi,\max}$ and $\boldsymbol{p}_{\pi,\min} p_{\pi,\max}$ 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 costfunction value. 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). However, in order to successfully apply The application of this gradient-based algorithm to GPR data , the optimization has to be regularized. Therefore, we focus with this algorithm on GPR data requires the implementation of tagging (Sect. 2.4.1) and 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, we tag 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 abs $(r_\mu(\boldsymbol{p}_\mathrm{perturbed}) - r_\mu(\boldsymbol{p})) > 50$ $|r_\mu(\boldsymbol{p}_\mathrm{perturbed}) - r_\mu(\boldsymbol{p})| > 50$ are also set to zero together with entries of the Jacobi matrix that are larger than 10^4.

We choose $\lambda_\mathrm{initial} = 5$

We choose $\lambda_\mathrm{LM,initial} = 5$ as initial value for λ λ_LM and apply the delayed gratification method by decreasing (increasing) λ 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, we apply the presented methods to the acquired are applied to GPR data. Therefore, we first explain To this end, the setup of the case studyand the , its implementation and the detailed setup of the parameter estimation procedure

are explained first (Sect. 3.1.1). Then, we test the method with synthetic data to understand the phenomenology of the data and capabilities of the method (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 3.2). Finally, we apply the method to the measured the methods are applied to measurement data (Sect. 3.3)and analyze the accuracy of the resulting parameters.

3.1 Setup of the case study

3.1.1 Implementation

For the simulation of the GPR signal of antenna 3, 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). An absorbing layer is used as boundary condition.

The numerical solution of the

The Richards equation (Eq. 1) is based on 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-dimension) on a structured grid with a resolution of ≈ 0.005 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.

The simulated water content is converted to relative permittivity via the CRIM using the mean soil temperature $T_\mathrm{s} = 8.5\ ^\circ\mathrm{C}$ (Sect. 2.2.4).

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-dimension. 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 a Ricker an excitation function (first derivative of a Gaussian–shaped function) that leads 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 initial electromagnetic field in the domain is zero.

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

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

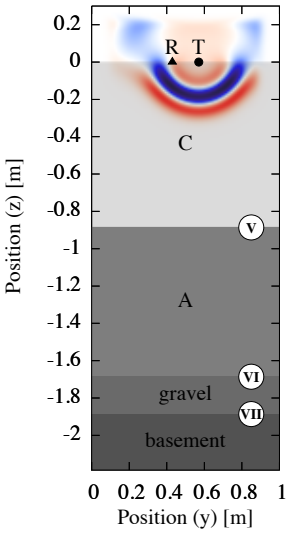

Figure 7. For the simulation of the GPR signal, we assume a layered subsurface architecture (Fig. 1). The transmitter of the antenna is represented with an infinitesimal dipole (T) and the electric field is read at the position of the receiver antenna (R). An absorbing layer is used as boundary condition. The x-component resulting electric field after 5 ns is shown. The markers for the material interfaces are used consistently in this work. This figure only shows the electromagnetic simulation. The hydraulic simulation is not shown here, such that the subsurface architecture can be defined clearly. 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.

with the speed of light in vacuum c_0, maximal frequency $f_{\max} = 2{\cdot}400$ MHz, and $\varepsilon_{\mathrm{r,max}} = 31.25$ corresponding to $\theta_{\mathrm{s,max}} = 0.5$. Hence, we choose the numerical resolution $\Delta z = 0.005$ m for the 2D isotropic, structured, and rectangular grid. Therefore, the simulated one dimensional relative permittivity distribution is extruded in the y dimension. 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, as no evaluation of air wave or ground wave is done 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 direct current electrical conductivity of the subsurface σ_{dc} 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).

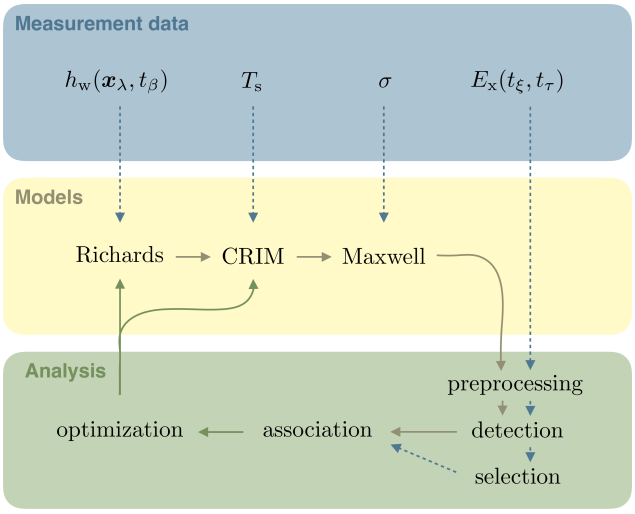

Figure 8. The available hydraulic potential h_{wt} is measured at the position of the groundwater well $\boldsymbol{x}_\lambda$ times t_β. These measurements are used as a boundary condition for the Richards equation (Sect. 2.2.1). Estimates for the soil temperature T_s and the direct current electrical conductivity σ_{dc} σ are derived from TDR related measurements. The actual signal of the GPR system is proportional to the $x$component -component of the electric field E_x and measured discretely at experiment time t_ξ and signal travel time t_τ. This signal is preprocessed (Sect. 2.3.1) and used for event detection and (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. 2.32.3.3). 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 preprocessing 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 objective function value during the cost in the optimization step (Sect. 2.4). Dashed arrows indicate initial preprocessing steps, whereas solid arrows indicate iterative steps required for the optimization.

3.1.2 Setup of the parameter estimation

General setup of the optimization is explained with Fig. 8. This setup is used in an iterative a sequential approach (Fig. 6) , where we selected ever fewer 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. We draw 40 initial parameter sets with the Latin hypercube algorithm within Hence, both data sets subsample the actually measured number of traces (one trace per $\approx$ 30s) equidistantly in time. Within the sample range given in Table 1and use , 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 . Therefore, we run 200 iterations with running 200 iterations of the simulated annealing algorithm(Sect. 2.4.2). Notice that the parameter fit range given in Table 1 determines the parameter update via $\boldsymbol{p}_{\pi,\max}$ and $\boldsymbol{p}_{\pi,\max}$ ($p_{\pi,\max}$ and $p_{\pi,\max}$ according to Eq. 12). After the application of the simulated annealing algorithm, we run maximally 15 iterations of the Levenberg–Marquardt algorithm (Sect. 2.4.3) which

completes the precondition 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, we use the mean absolute error in normalized travel time $e_{\mathrm{MA,t}}$, because is used since this statistical measure is independent of the number of associated events. These are The number of associated events is accounted for by choosing those evaluating those those 10 members with minimal $e_{\mathrm{MA,t}}$ where that associated at least 85% of the measured eventsare associated. Each of these members has locally optimal parameters. However, the exact position of these local minima typically depends on (i) the settings of the optimization algorithms, (ii) the choice of the events to be evaluated, and (iii) 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 χ^2 surfacecost 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. Notice that the mean parameter set is not necessarily optimal. However, if uncertainty on the of the resulting parameters is small, this result the mean parameter set is typically more reliable results than the 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 amplitudesis used as the standard deviation of the residuals in the objective function (Sect. **??**).

3.2 Synthetic data

3.2.1 Phenomenology

The phenomenology of the transition zone reflection for characteristic times during imbibition, relaxationequilibration, 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. Therefore, we simulated To this end, the water content distribution in of the 1D profile located at 17.05 m of ASSESS using parameters typical the ASSESS site (Fig. 1) was simulated using typical parameters for coarse-textured sandy soils(Table 2). The results are visualized over time . These parameters are given together with the estimated parameters in Table 2. The groundwater table measured in the well (Fig. 9a) and over the water content (Fig. 9b). Initialized with

Initialized with static hydraulic equilibrium, the simulation starts with the an initial drainage step (Sect. 2.2.3) 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 the subsequent equilibration step, the groundwater table is raised during the subsequent imbibition step. Generally, the The Brooks–Corey parameterization (Eq. 2) features a sharp kink where air enters the material at the upper end of the capillary fringe. AdditionallyFurthermore, the imbibition causes another introduces an additional kink in the water content distribution (at marker (2) in Fig. 9b), because the relaxation time from the hydraulic non-equilibrium is much shorter

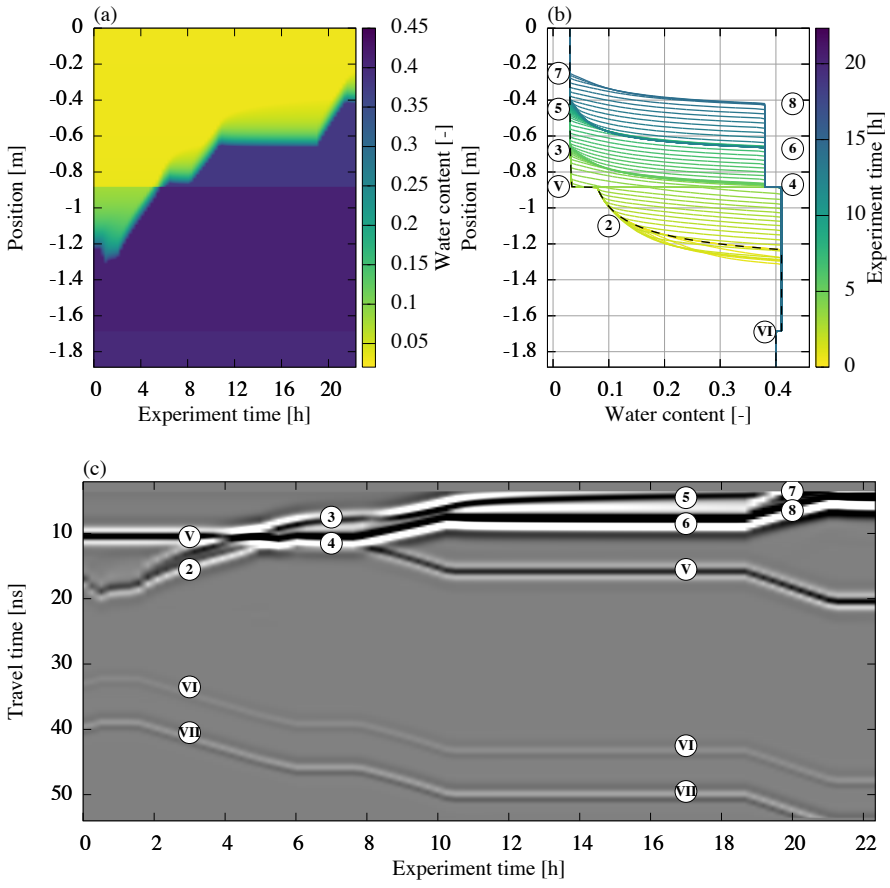

Figure 9. We used The true synthetic data are simulated with hydraulic parameters representing that represent coarse-textured sandy soils (Table 2)to generate the synthetic data for parameter estimation according to Sect. 3.1.1. Subfigure (a) shows the simulated water content in color code over experiment time, whereas subfigure (b) shows the same data in a line plot emphasizing the temporal development of the water content distribution. 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 (2). The temporal evolution of this reflection is sensitive on to the initial water content distribution, the soil water characteristic and the hydraulic conductivity function. Except for the normalization, the The data shown are processed according to Sect. 2.3 , including a dewow filter and 2D to 3D conversionexcept for the normalization. In contrast to the quantitative evaluationdata that are evaluated, the shown radargram is normalized to the maximal its maximum absolute amplitude, facilitating the visual comparison of the traces. The markers are used consistently in this paper study and are further explained in the text.

at high water contents compared to the relaxation time at low water contents. This is due to the strong 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 sharpened decreased during the imbibition .

During the equilibration step after the first imbibition, the transition zone additional kink smoothes. Thus, the water content increases in the material with low water content (3) and decreases in the material with high water content at the upper end of the capillary fringe (4). This smoothing process strongly 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. The Note that since the water content changes continuously, the signal in between these wavelets is a superposition of infinitesimal reflections which contain detailed also contain information about the form of the transition zone. Notice that 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

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 correlation coefficients for the mean parameter set are analyzed in detail in the text. Notice that the porosity of the gravel ($\theta_{s,G}$) as well as the position of the material layers (h_1 and h_2) can be reliably estimated from single channel GPR when evaluating both the signal travel time and amplitude.

We calculated (a) the resulting water content distribution of the 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 over all 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 evaluation the synthetic GPR data is separated in three parts: Subfigure (a) shows the detected (Sect. 2.3.2) and selected (Sect. 2.3.3) events which are used as synthetic measurement data. Except for the normalization, the data are processed according to Sect. 2.3, including a dewow filter and 2D to 3D conversion. The radargram is normalized to the maximal absolute amplitude, facilitating visual comparison of the traces. Subfigure (b) shows resulting differences in travel time and amplitude of the mean parameter set. The differences of the amplitude are given in arbitrary units which are consistent for the whole work. Subfigure (c) shows standardized residuals of the differences, essentially zooming into the small differences given in subfigure (b). Note that outliers are set onto the boundary. The markers for the reflections are used consistently in this paper.

The 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 on 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 $h_{0,\mathrm{C}}$ and λ_C 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.

As

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 Cto $\theta_{s,\mathrm{A}}$. 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.

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). As 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 on $K_{\mathrm{s,C}}$. Yetto 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 (Sect. 3.2.1markers 3 and 5 in Fig. 9) is sensitive to the hydraulic conductivity. The hydraulic conductivity function (Eq. 3) is not unique if K_s K_s is not fixed. This leads to a strong correlation of the parameters $K_{\mathrm{s,C}}$ and τ_C 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 in the soil water characteristic of material C A (Fig. 10c) is largest for low water contents, as because there are only few data points available. In particular, this increases the uncertainty of λ_A λ^A (± 0.7, Table 2). The material properties

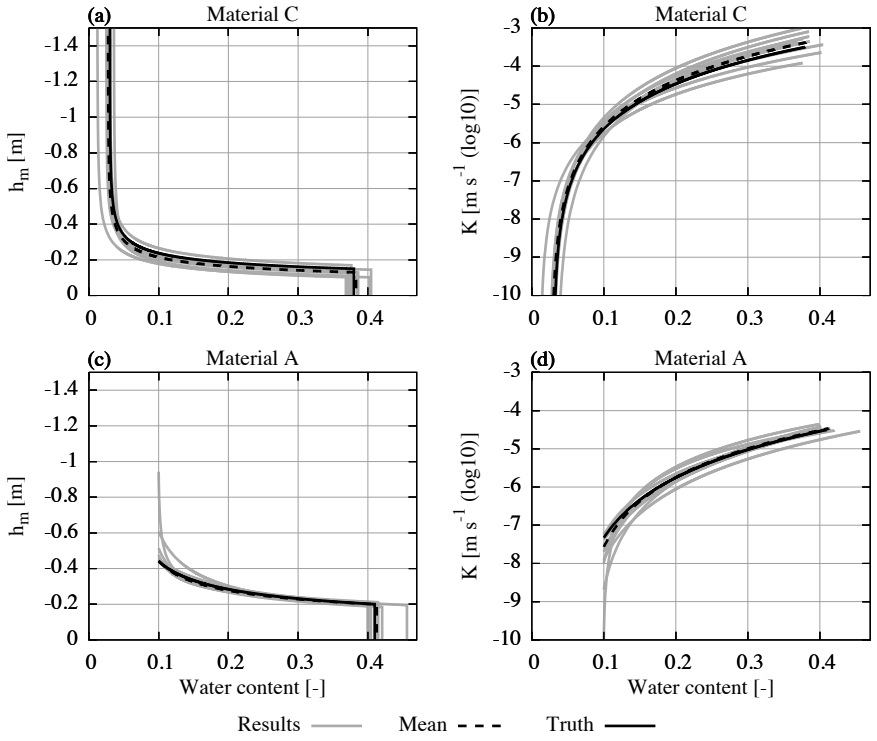

Figure 10. The resulting material parameters estimated from synthetic data are shown for the 10 best ensemble members (Sect. 3.1.2) together with the mean of these parameter sets and the true parameter set (Table 2). The plot range of the parameters is adjusted to the water content range of the data.

of 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 $h_{0,\mathrm{A}}$ and λ_A h_0^A and λ^A are strongly correlated (-0.6). Yet, the uncertainty in $h_{0,\mathrm{A}}$ h_0^A is relatively small (± 0.008, Table 2) mainly because it is essentially uncorrelated to other parameters. In contrast, the parameter $\theta_{s,\mathrm{A}}$ θ_s^A is correlated to the parameters $h_{0,\mathrm{C}}$, λ_C, $\theta_{s,\mathrm{C}}$, and $\theta_{r,\mathrm{C}}$ as h_0^C, λ^C, θ_s^C, and θ_r^C, because wrong parameters for material C introduce changes in the total water content which can be partially balanced out by adjusting

The uncertainty of the saturated hydraulic conductivity of material A (Fig. 10d) is comparably small as the largest fraction of the data are influenced by this parameter. Hence, the parameters τ_A and $K_{\mathrm{s,A}}$ τ^A and K_s^A are only very weakly correlated.

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 channelsingle-channel time-lapse radargram. Evaluation methods that merely exploit the signal travel time (e.g., Gerhards et al., 2008), require additional measurements a multi-channel approach to achieve this goal.

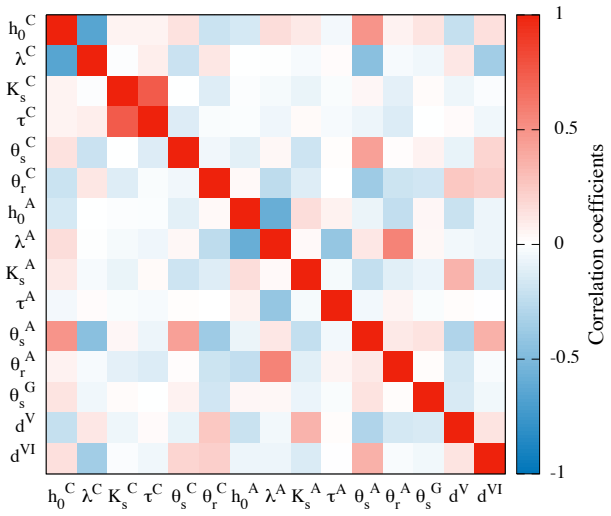

Figure 11. The correlation coefficients for the mean parameter set show in particular that the porosity of the gravel layer (θ_s^G) as well as the position of the material layers (d^V and d^{VI}) can be reliably estimated from single-channel GPR when evaluating both the signal travel time and amplitude.

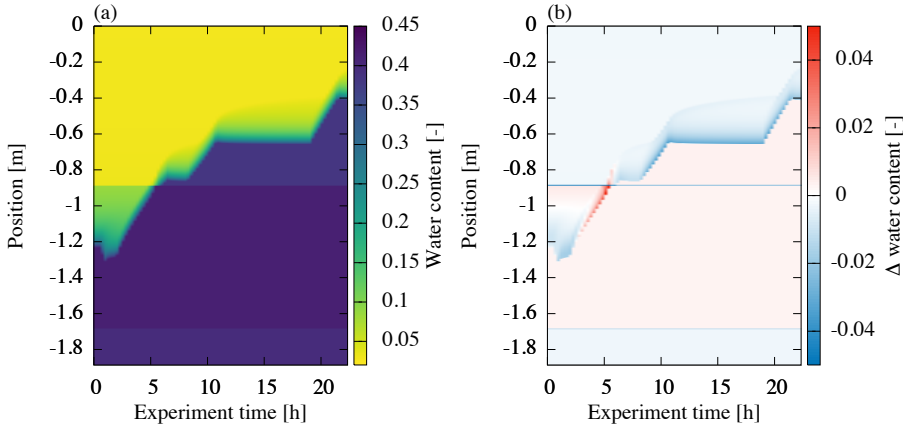

Figure 12. This figure shows (a) the water content distribution simulated with the resulting mean parameter set and (b) the difference to the true water content distribution (Fig. 9). The mean absolute deviation of the volumetric water content is 0.004. The over all balance of the volumetric water content 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.

In order to further investigate the quality of the mean parameter set, we simulated the resulting water content distribution (Fig. 12a) and calculated the difference to the true water content distribution (Fig. 12b). Due to the narrow pore size distribution

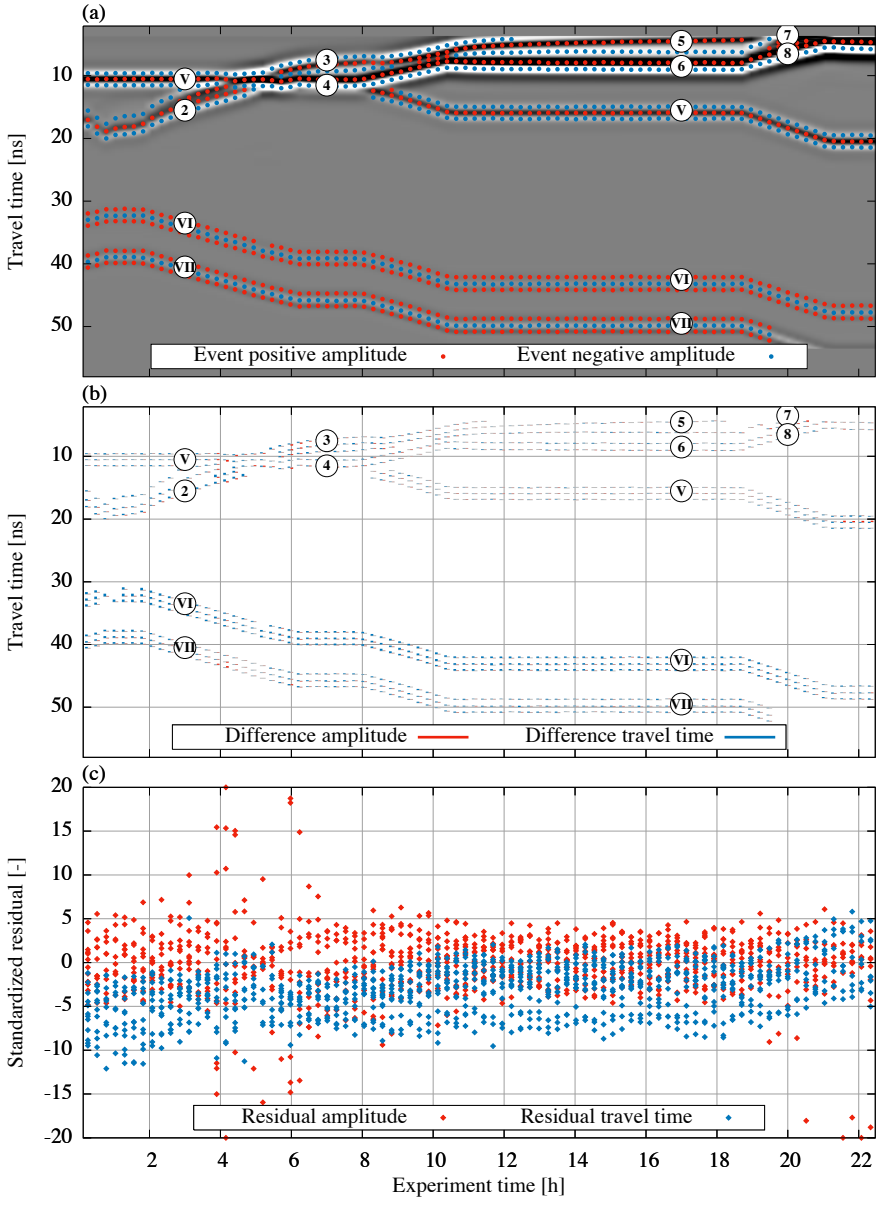

Figure 13. The evaluation the synthetic GPR data is separated into three parts: Subfigure (a) shows the selected events (Sect. 2.3.3) which are evaluated with the optimization. The data shown are processed according to Sect. 2.3 except for the normalization. In contrast to the data that are evaluated, the shown radargram is normalized to its maximum absolute amplitude, facilitating the visual comparison of the traces. The markers are used consistently in this study and are further explained in the text. Subfigure (b) shows resulting differences in travel time and amplitude of the mean parameter set. The differences of the amplitude are given in arbitrary units which are consistently used in this study. 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.

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.

These deviations also

The remaining deviations in soil water content after the parameter estimation cause residuals in the GPR signal (Fig. 13), which . 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 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

Similar to the analysis of the deviation in water content (Fig. 12), the largest residuals in unsaturated hydraulics 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.

Apart from that, the deviations in the material properties of unsaturated material C do not lead to significant residuals in the GPR signal, although the deviations in water content are considerable.

3.3 Measured data

3.3.1 Phenomenology

The common offset GPR measurement (Fig. 14a) reveals Before starting the experiment a common offset measurement was acquired that allows to analyze the initial state of ASSESS . (Fig. 14). The reflections of the material interfaces are marked with uppercase roman Roman numbers (I, II, III, IV, V, VI, and VII) which were introduced in Fig. 1. Compaction interfaces were generated during During the construction process of ASSESS compaction interfaces were generated. The most pronounced of them are marked with lowercase roman Roman numbers (i, ii, iii, iv, and v). In particular, the reflection of compaction interface (iv) close to the reflection of the initial position of the groundwater table (1) are is difficult to separate from reflections from material interfaces. Reflections from confining walls are most visible around 1 m (W) but influence the signal for more than 2 m. The reflection of the edge of the L–element (L) is particularly prominent. As ASSESS is confined by walls at all four sidesand approximately 4 m wide, the . The walls parallel to

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)and its amplitude increases. After passing the material interface, the reflection of the groundwater table (2) splits in two separate reflections (3)

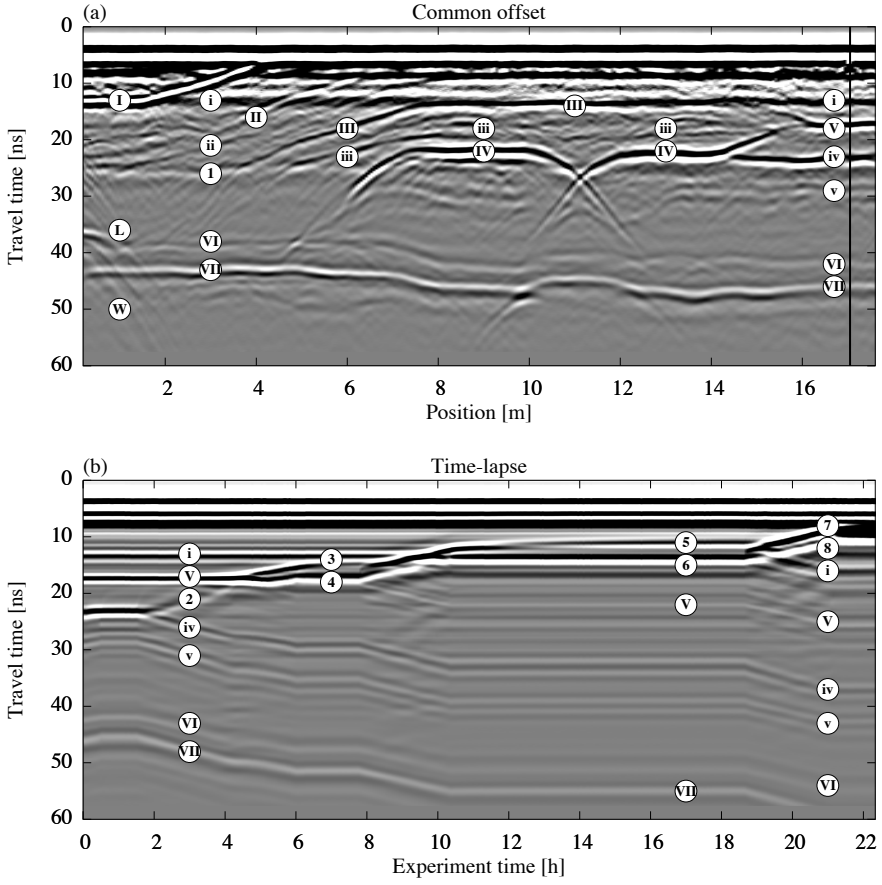

Figure 14. (a) A common offset measurement of the hydraulic state of ASSESS at the beginning of the experiment is shown. For this measurement, the antenna was moved over the site at one point in time. The temporal evolution of the subsequent hydraulic dynamics was monitored with a stationary antenna at the position indicated with the vertical black line. The resulting time-lapse measurement data are shown in subfigure (b). Note that the common offset and the time-lapse data were measured with different but structurally identical antennas. Thus, in particular the measured GPR signal of the direct wave and the ground wave is slightly different. 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, lowercase Roman markers for compaction interfaces, 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.

and (4). This is due to the strong dependency of the hydraulic conductivity on the water content (Sect. 3.2.1). As 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

According

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 relaxation 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 exacerbates 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).

We show a common offset measurement of the hydraulic state of ASSESS at the beginning of the experiment (a). The vertical line indicates the position of the time-lapse measurement shown in subfigure (b). The common offset (time-lapse) data was measured with antenna 2 (3). Hence, in particular the measured GPR signal of the direct wave and the ground wave is slightly different. Both radargrams are measured with internal channels with an antenna separation of $a = 0.14$ m. Except for the normalization, the data are processed according to Sect. 2.3. In contrast to the quantitative evaluation, the radargram is normalized to the maximal absolute amplitude, facilitating the visual comparison of the traces. The markers are used consistently in this paper and are further explained in the text.

3.3.2 Results and discussion

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 plot range of the parameters is adjusted to the water content range of the corresponding data.

The evaluation the measured GPR data is separated in three parts: Subfigure (a) shows the detected (Sect. 2.3.2) and selected (Sect. 2.3.3) events which are used as synthetic measurement data. Except for the normalization, the data are processed according to Sect. 2.3. The radargram is normalized to the maximal absolute amplitude, facilitating visual comparison of the traces. Subfigure (b) shows resulting differences in travel time and amplitude of the mean parameter set. The differences of the amplitude are given in arbitrary units which are consistent for the whole work. Subfigure (c) shows standardized residuals of the differences given in subfigure (b). Note that outliers are set onto the boundary. The markers for the reflections are used consistently in this paper. Since the hydraulic 1D model cannot represent lateral flow present in initial drainage, we neglect the measured events of the first 2 h. Additionally, outlying events which have a different sign amplitude are not considered for event association (Sect. 2.3.4).

As Since the GPR measurements cover only a small portion of the subsurface architecture, we restricted the hydraulic representation is restricted to 1D (Sect. 3.1.1). Hence, we neglect 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, we merely focus the evaluation on the focus of this study is on evaluation of the imbibition phase of the experimentas , 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).

The identification of relevant events is more difficult for the measured data than for synthetic data due to additional reflections of the GPR signal which are not represented. The most important representation errors are presumed to be the (i) reduced dimensionality of the representation of the inherently three dimensional GPR antennas and ASSESS test site using a 1D hydraulic and a 2.5D electrodynamic model, (ii) neglected small-scale heterogeneity in particular associated with compaction interfaces, (iii) neglected reflections from confining walls, (iv) neglected roughness of material interfaces, (v) influences of the antenna characteristics on the GPR signal in particular including the direct wave, the ground wave, temporal fluctuation of time-zero, and the source wavelet, (vi)assumption of a constant direct current conductivity, and (vii) assumption of a constant soil temperature for the calculation of the relative permittivity of water.The

Investigating the resulting material properties of the inversion (Fig. 15), the main findings concerning the mean parameters for the synthetic data (Sect. 3.2.2) can also be identified for the measured data, i. e. . These comprise (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

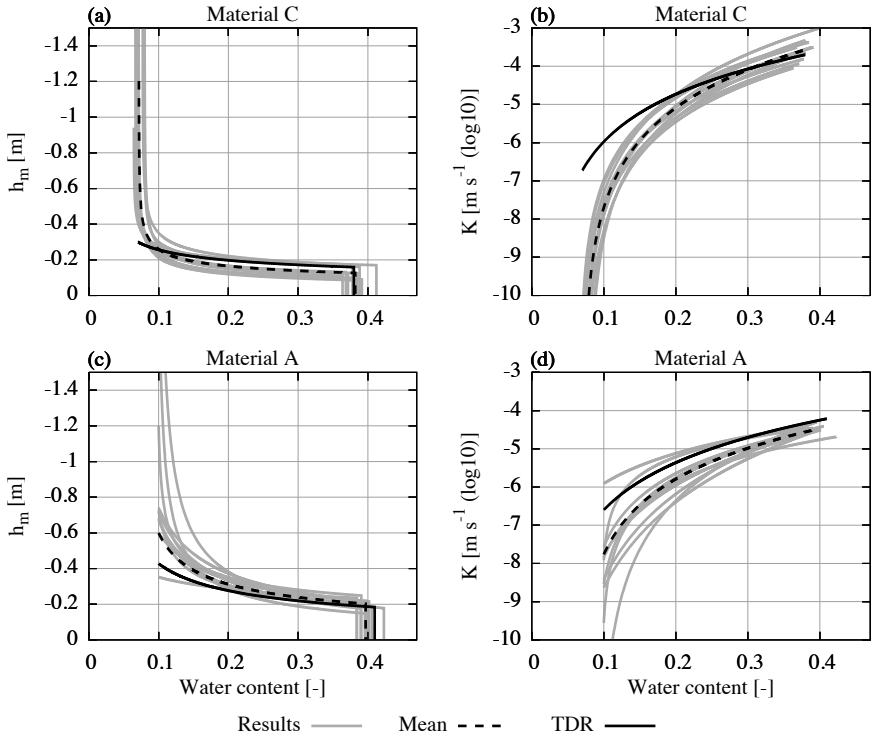

Figure 15. The resulting material parameters estimated from measured data are shown for the 10 best ensemble members (Sect. 3.1.2) together with the mean of these parameter sets (Table 3) and a reference parameter set determined from TDR data acquired during the same experiment (Jaumann and Roth, 2017). The plot range of the parameters is adjusted to the water content range of the corresponding data.

Compared to the uncertainties based on synthetic data (Table 2), the uncertainty of the resulting mean parameters (Table 3) mostly increaseddue to representation errors. Yet, except for four parameters, the parameters estimated from TDR measurements (Jaumann and Roth, 2017) are within the one standard deviation of the mean parameter setexcept for four parameters (Table 3). The deviations of these parameters are analyzed in the following.

The parameter $\theta_{r,\mathrm{C}}$

The parameter θ_r^{C} estimated from the GPR data is approximately a factor 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), we use the integrated water content for is included in the inversion. This also includes 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 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 .

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, as this compaction interface is not represented in the model, the resulting $\theta_{r,\mathrm{C}}$ 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 partially adapted by changing $\theta_{r,\mathrm{C}}\theta_r^{\mathrm{C}}$. 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, evaluating

Third, analyzing the TDR data, we find that an underestimation of $\theta_{r,\mathrm{C}}$ θ_r^{C} is likely due to the lack of TDR measurements at low hydraulic potential. Hence, the

Compared to the evaluation of the TDR data, the resulting value for parameter τ_{C} τ^{C} is a factor 2 larger for the GPR evaluation. 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 low water contents. Yet, 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.

Similar to parameter $\theta_{r,\mathrm{C}}\theta_r^{\mathrm{C}}$, the parameter $\theta_{r,\mathrm{A}}$ θ_r^{A} yielded from the GPR evaluation is approximately a factor of 3 smaller than the result from the TDR evaluation. However, this parameter can only be approximated evaluating the GPR data as they lack events that are influenced by low water content.

The resulting value for parameter $K_{\mathrm{s,A}}$ K_s^{A} is factor 2 smaller for the GPR evaluation compared to the result from the TDR evaluation. This parameter limits the flux through the lower boundaryas , 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 a 3D flux (Jaumann and Roth, 2017). Since this 3D flux is not represented, the hydraulic potential and hence at the bottom and thus the water flux is overestimatedtoo large. This is compensated

The estimated interface position of the material A and C (d^{V}) corresponds well the to ground truth measurements acquired during the construction of ASSESS 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 are still within the uncertainty of the ground truth measurements when compared to GPR measurements.

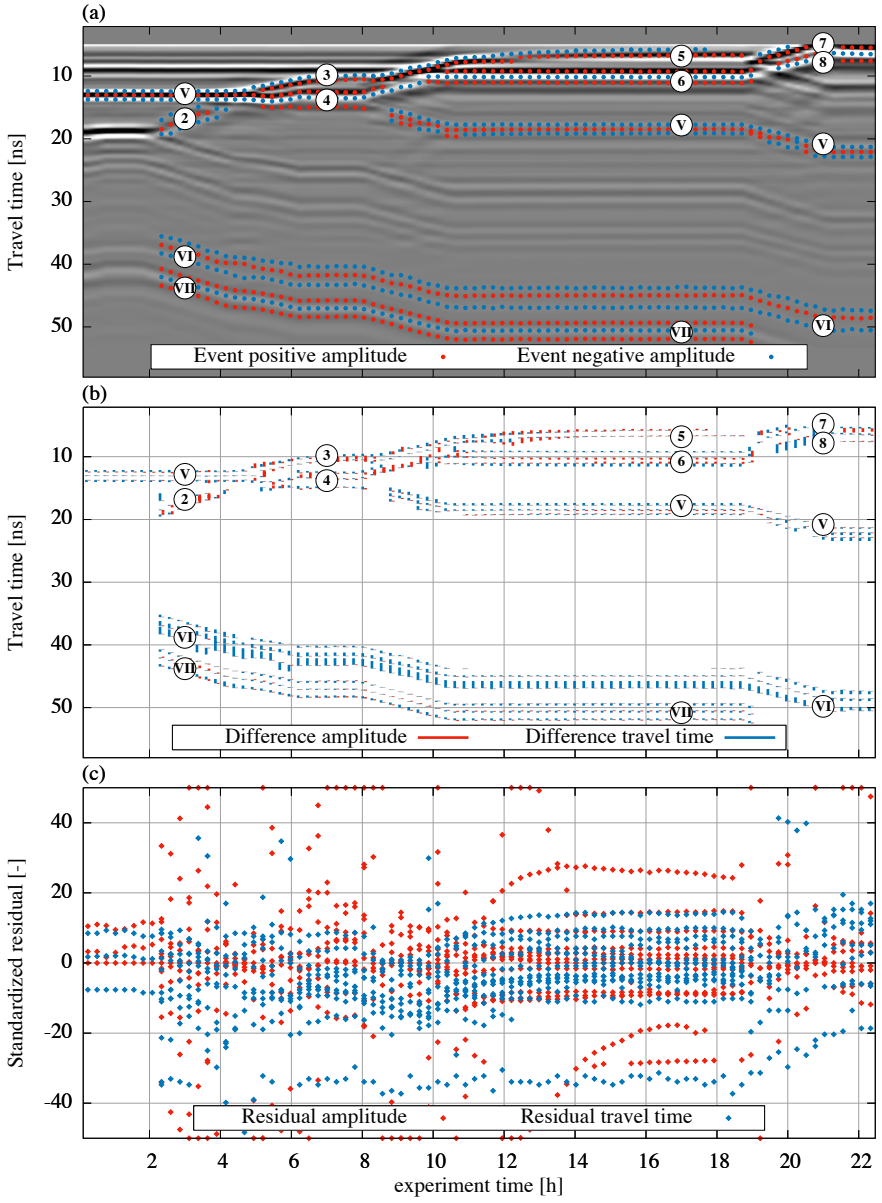

Figure 16. The evaluation the measured GPR data is separated in three parts: Subfigure (a) shows the selected events (Sect. 2.3.3) which are evaluated with the optimization. The data shown are processed according to Sect. 2.3 except for the normalization. In contrast to the data that are evaluated, the shown radargram is normalized to its maximum absolute amplitude, facilitating the visual comparison of the traces. The markers are used consistently in this study and are further explained in the text. Subfigure (b) shows resulting differences in travel time and amplitude of the mean parameter set. The differences of the amplitude are given in arbitrary units which are consistently used in this study. 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. Since the measurement data acquired with a stationary antenna can not spatially resolve the lateral flow present in the initial drainage phase, the measured events of the first 2 h are neglected.

The evaluation of the GPR measurement data remaining residuals after the optimization (Fig. 16b) shows that deviations from the shape of the reflected wavelet contributes wavelet contribute to the residuum significantly. These deviations have three main origins: (i) unknown shape of the source wavelet of the GPR antenna when coupled to the subsurface, (ii) the assumption of a constant direct current conductivity for the whole subsurface, and (iii) neglected roughness of the material interfaces which influences the shape of the reflected wavelet.Evaluating the In particular evaluating the different reflections (V) and (VII), the residuals in travel time suggest an incorrect width of the wavelet. At the beginning of the experiment, the simulated wavelet is too broad for reflection (V) whereas it is too narrow for the reflection (VII). This indicates that effect can be explained with (i) the assumption of a constant direct electric conductivity in the whole architecture is suboptimal and that a possible frequency dependency of the electrical conductivity and (ii) the direct electric conductivity can be assessed with GPR measurements. The roughness of the material interfaces (Dagenbach et al., 2013). The latter is in particular noticeable for the residuals for reflection (VI) which are the major contribution to the cost functionas different representation errors are combined. Of all the events in the wavelet, the events with the largest travel time exhibit the highest residuals. Due to the large grain size of the gravel, the real material interface is rougher than its representation. This leads to a non–symmetric non-symmetric broadening of the measured compared to the simulated wavelet (Dagenbach et al., 2013).

The residual of the other events in the wavelet can be attributed to a combination of representation errors associated with the boundary condition, suboptimal parameters, incorrect direct current conductivity, and also to lateral water flow which not represented in 1D model.

A large part of the residuals associated to the reflections (3), (4), (5), and (6) originates from interferences with the compaction interface (i) which is not represented. Interferences typically are not quantitatively evaluable if not all contributions are correctly represented. This seems also to be the case for the signal during the forcing which is highly sensitive to the shape of the wavelet.

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

Notice that 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. The TDR sensors are distributed over a 2D slice of ASSESS measuring in all available materials (Fig. 1). Yet, the measurement volume is limited to the position of the sensors yielding the average permittivity along the TDR rods. Hence, these measurements are subject to representation errors such as small-scale heterogeneity or uncertainty in the sensor position (Jaumann and Roth, 2017). The In contrast, the GPR measurement data 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 the electrical conductivity. Hence,

these measurement data are subject to representation errors such as neglected (i) compaction interfaces, (ii) spatial variation of the direct current electrical conductivity, and (iii) roughness of certain material interfaces. Hence, the previous analysis illustrates how GPR-determined parameters can differ from TDR-determined ones making joint evaluation procedures challenging and promising since they open a window to the soils multi-scale nature.

We demonstrated that effective

4 Summary and conclusions

TDR measurements are a standard method that provides measurement data for the estimation of soil hydraulic material propertiesand the subsurface architecture can be estimated accurately from single channel time-lapse GPR measurement data. The GPR-determined subsurface architecture corresponds well to the ground truth and the resulting material properties compare favorably to material properties determined from an independent analysis of TDR measurements acquired during the same experiment.These results are based on (i) a fluctuating groundwater table experiment at ASSESS, (ii) a new heuristic

In this study, we propose a new heuristic semi-automatic approach to automatically extract and associate relevant parts

In order to identify and extract the relevant information given in GPR data, a heuristic evaluation method was developed. First, this method 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 radargrams, and (iii) an elaborate optimizationprocedure coupling different optimization algorithms which employ a subsampled data set to precondition the initial parameter proposals. We

Synthetic and measured single-channel common offset time-lapse GPR data are first analyzed qualitatively. It was confirmed that a fluctuating groundwater table experiment introduces characteristic transition zone reflections . A detailed qualitative analysis of its phenomenology indicates that this type of reflection is sensitive on that are likely to provide valuable information for the parameter estimation. Subsequently, the subsurface architecture and soil hydraulic material properties .

Employing the presented approaches on synthetic data shows that the true parameters are within the standard deviation corresponding to 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 the TDR measurement data. We discussed the deviations of the parameters and basically associated them with representation errors or the lack of available measurement data. Critical representation errors for the GPR data comprise the neglected (i) compaction interfaces, (ii) spatial variation of the direct current electrical conductivity, and (iii) roughness of certain material interfaces.

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 the codes to process the ASSESS architecture raw data and the GPR data. 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 two anonymous reviewers, who all 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.

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.

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.

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

Roth, K., Schulin, R., Flühler, H., and Attinger, W.: Calibration of Time Domain Reflectometry for Water Content Measurement Using a Composite Dielectric Approach, Water Resources Research, 26, 2267–2273, , 1990.

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 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: 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 the 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_0 (m s^{-1}) 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_0 (m s^{-1}) 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	h_1 (m) d^{V} (m)	0.90	1.10	0.95	1.05
	h_2 (m) 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 ten 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_0 (m s^{-1}) 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_0 (m s^{-1}) 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	h_1 (m) d^{V} (m)	1.00	0.99 ± 0.02
	h_2 (m) 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 given shown in Fig. 15. The reference parameters for the materials A and C are determined from TDR data acquired during the same experiment investigated in this work (Jaumann and Roth, 2017). Note that the according standard deviations for these reference parameters are determined formally from a single Levenberg–Marquardt run and hence 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. Notice that 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_0 (m s^{-1}) 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_0 (m s^{-1}) 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	h_1 (m) d^{V} (m)	0.99 ± 0.05	0.97 ± 0.02
	h_2 (m) d^{VI} (m)	0.13 ± 0.05	0.17 ± 0.02