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.