Review in support of the editor's decision for

Mangel et al: Reflection tomography of time-lapse GPR data for studying dynamic unsaturated flow phenomena

The paper under review for publication in HESS by Mangel et al aims at employing a reflection tomography based inversion algorithm, which is well-established for calculating subsurface velocity distributions from CMP GPR measurements in stationary conditions for deriving – by proxy – subsurface water content – distributions. In contrast to previous publications, here the focus is on dynamically changing conditions during infiltration experiments.

First of all, I would like to specifically laud the authors for their dedicated experimental approach and congratulate them for their laboratory setup and the undoubtedly involved data set which may yet hold the key to studying the infiltration experiments they monitored by GPR in so much detail.

However, the key question for whether the currently submitted work warrants a dedicated publication is whether the authors found a novel and robust way to extract meaningful and relevant information from this great dataset. Unfortunately, I am convinced that the inversion approach chosen for this publication in its current form falls short of achieving that aim (i.e., as the title states: Usage of this algorithm for "studying dynamic unsaturated flow phenomena") and does not give justice to the information potentially contained in their elaborated dataset.

The inversion algorithm's trouble is quite clearly shown already by the simulation based results the authors present in Figure 3: Here, the authors first calculate water content distributions from HYDRUS-2D (figure 3, left column), then derive GPR profiles from these distributions (examples shown in figure 2) and feed these into their tomography algorithm to retrieve the respective water content distributions (figure 3, center column). In the first case, as the authors admit themselves, their algorithm fails completely to capture the velocity profile, since there is simply not enough information for this approach to work with. OK. However, this remains true for the second case – the algorithm basically does not resolve the infiltration plume at all (3d-f). In the third case (3g-i), the algorithm actually outputs an infiltration plume - which could be expected since the input in this case is to first order approaching a two-layered system and no longer includes a water table below. However, and this is in my opinion crucial if such an approach is supposed to be used for studying infiltration experiments, the algorithm misplaces the position of the infiltration front by about a factor of two (the "true depth" of the plume is about 0.23 m judging from figure 3g, the calculated position clearly surpasses 0.4 m). If the results are aimed at "informing models of hydrologic processes" (L210), adding this information on top of the rather large water content deviations will certainly not be beneficial to the output of any model.

From the examples in figure 3, only the very last case (Figure 3 j-l) might be deemed an acceptable result, although the shape of the infiltration front and lateral expansion is still not captured (which would be important information for the hydrological model!). As stated above, this is most likely due to the fact that as the infiltration plume advances into the medium and increases in size, it resembles a much more simple two-layered medium case – again without the presence of a water table. To give a better indication of whether this algorithm could - at least based on a numerical study - provide an output, which would be useful for studying the actual hydrologic infiltration process it would be necessary to present a detailed time-lapse assessment of how a progressing infiltration plume can be resolved in the first place. At minimum this could start from a time-lapse representation (e.g., a movie) of results with a good enough temporal resolution: E.g., of the "true" water content calculated by HYDRUS-2D on the left and the tomography result on the right – depicting the temporal evolution of both the infiltration event and the corresponding tomography result for each timestep. This could in principle then be used both for a rigorous error assessment, which is missing so far, and for discriminating periods in time during the infiltration process in which the situation is just too complex for the current tomography approach and where it deliver at least useful information. From the examples shown in the paper, I take it that first, the imaging fails completely, then the infiltration is resolved as being much faster than in reality while in the end a simpler situation is reached in which an acceptable result may be achieved: Hence this looks like there is a point where the inversion actually somehow converges towards reality which should be clearly identified and discussed. Without such an assessment, which does not only encompass comparing average water contents, I do not see much reason for trusting the results of the measurement inversions shown later. In 2019, for studying infiltration processes with GPR, a quantitative "average error of 5-10%" in water content is not enough if not at least the dynamics can be qualitatively resolved much better. In fact, it would be truly a pity if matching average water contents to within 10% would really be all that can be done with your elaborated dataset.

From the work presented here it seems clear that for studying dynamic unsaturated flow phenomena the authors should attempt to leverage much more of the information actually contained in the dataset. Information is already scarce for tomography algorithms based on surface data in stationary conditions. In such a dynamic infiltration experiment context, any viable approach will therefore have to give credit to the specific strengths of such a dataset. Getting more acceptable results may, e.g., include concurrently considering information from the air/groundwave and the wetting front reflection – which would likely not be directly possible in the framework of the present version of the inversion algorithm. I would also encourage the authors to take another look at the dynamics of the wetting front reflection for a source of additional information.

For getting better results by adapting the currently employed algorithm, an approach could be to constrain the inversion based on CMPs acquired at a specific time by the results from previous and subsequent time steps. Basically: If timelapse movies helped in visual interpretation of the dataset – there is no reason to expect that this will not also be the case for an automated evaluation… In my opinion, the fundamental limitation in the case presented here is not so much in the information content of the data set in itself (as stated in L.205), but in the limitations of the algorithm which would have to be discussed in a lot more detail in this paper to warrant a publication. The author's claim that "automated high-speed GPR data acquisition coupled with reflection tomography algorithms can provide a new approach to hydrologic monitoring" – will only hold if these algorithms actually leverage the additional information contained in the temporal domain. As far as I understood the author's approach, for each example shown, the pertaining spatially distributed series of CMPs is inverted without taking into account the information obtained at different times. Maybe each inversion is actually starting from a different starting model – but to what extent this is actually the case is not clear to me from the paper and would warrant a whole discussion of its own, e.g.: how does the starting model evolve over the time series? How much does the final inverted velocity model differ from the respective starting model? Could the starting model be in some clever way constraint by results from a previous – or in an iterative approach even a subsequent – timestep? How dense would the temporal resolution have to be for such an approach to work (btw. – the inversion seems to be quite computationally intensive, which should also be discussed in terms of potential limitations: how dense could such a temporal sampling from a computational point of view actually be?

In conclusion, so far I do not see enough evidence in the paper presented here to sustain the author's main claim that "reflection tomography in the post-migrated domain is a viable method for resolving transient soil moisture content in 2D".

Hence, which way forward? I do see two possible roads to follow:

- Since the main claim can so far not be sustained, the only reason for publishing this paper would be to provide a much more thorough performance assessment of the employed algorithm under such dynamic infiltration conditions. Hence, radically refocus the publication to concentrate on assessing the true capabilities of the present algorithm under dynamic conditions based on (potentially a series of additional) numerical simulations – including some sort of time-lapse analysis /movies etc. as hinted at above. Improve on constraining the starting model and discuss in the framework of a rigorous error assessment. As stated above: Deriving average water content error is just a small part of the task if this is to be useful for studying dynamic cases. Correctly resolving the position of infiltration-induced interfaces over time is another. Water balance would be yet another – e.g., to what extent is the total amount of infiltrated water actually retrieved?
- Otherwise I would advice to keep this publication as is in the status of a discussion paper and focus the efforts on a larger inversion framework in which the present results can be one source of information, to be augmented by evaluating different aspects of the dataset. Please leverage much more of the information contained in the temporal nature of this great dataset.

In light of my rather substantial objections to publishing the current manuscript, I will not continue adding additional minor comments at this point.