

Gabriel and co-workers present a very comprehensive study on the effect of barley cover-cropping and minimum tillage on the hydraulic soil properties based on a 10-year experiment and subsequent inverse modelling for parameter identification. The study and the manuscript are well thought and worked out. It definitely fits in the scope of HESS and will be a fruitful contribution to the field.

Thank you very much for this comment.

That said and with all respect to this work, I however see some fundamental issues with the presented study which I would suggest the authors to reconsider:

1. Why is there no direct soil data from some time within the 10-year experiment used for validation?

We agree that validation is always an important issue in modelling studies. To strengthen the validation status, we used independent datasets for the calibration/validation of the crop module, in previous soil hydrological studies. Such approaches were followed for instance in Gabriel et al. 2012 or in Alonso-Ayuso et al. 2018, where the WAVE model was used as a prediction and scenario simulation tool. However, in this study, we do not want to deploy the model in a scenario analysis. Rather we want to use the Richard's equation based model as a tool to interpret observed data and to improve the modelling concept, demonstrating in this case that soil hydraulic properties should be considered dynamic when modelling mid-term soil hydrodynamics. In just case, we are just proposing a new methodology for measuring the evolution of the hydraulic parameters based on observed data and Richard's equation.

2. How can one account for the effect of changing soil conditions on the initially calibrated sensor readings?

We totally agree that this is an important issue that was not addressed in this version of the manuscript. However, as we presented in Gabriel et al. 2010 (the manuscript dealing with the sensor calibration in this field experiment), the sensor results were validated not only during two irrigation events (trying to identify the fast response of the sensors to fast wetting-drying processes) but also with field observed soil water content measurements sampled in seven different dates distributed along almost the 3 first years of experiment. The most important changes of the hydraulic properties were identified during these 3 years. The rather reliable validation of the sensor readings as reported in Gabriel et al. 2010 could be due to the fact that we are analysing volumetric soil water content inferred from dielectric data at the macroscopic scale. Even when soil particles could be redistributed in many ways and modify soil structure, it is suggested that this occurs at scales smaller than the capacitance probes footprint and hence does not dominate the sensor calibration. For sure, a larger data base would be needed to explore this more in detail, but such a data base could not be developed in the present study.

3. How do the authors account for equifinality and parameter interaction in their inverse modelling?

Equifinality is indeed an important issue when inverse calibration (with so many parameters) is used. We avoid equifinality in two ways: 1) We limit the parameter range to the real range directly observed in the soil and 2) We use the Markov chain parameter sets that start searching independently from several points in order to explore better the n dimension matrix of possible results and increasing the probability of reaching the global fit.

4. Is the WAVE-WOFOST model system evaluated for its general necessity/suitability in the study and potential parameter interaction and identifiability of the required soil parameters?

The crop model (based on WOFOST, but not WOFOST) only interact with WAVE providing a potential crop water demand at each soil layer (supplied or not depending on previous WAVE available soil water simulation) and a soil cover in order to estimate potential transpiration rate.

**(1 & 2)** The authors base their study on continuous measurements of soil moisture in 6 depth levels (0.1 to 1.1 m, 0.2 m increments) on six treatment plots. In addition meteorological state variables are recorded on site. As very good practice, the soil moisture sensors have been calibrated under field and lab conditions to the specific soil condition previous to the experiment period. At this time a pedohydrological analysis was done, too. Since the study is about the effect of cover crop and tilling routines on soil hydraulic properties, I suspect that the authors have considered to directly sample the soil repeatedly to answer their research question.

We collected indeed data for this, but only for the top soil layer. The plots are only 12x12 m<sup>2</sup> in size and measurements deeper in the soil are very destructive. Therefore depth measurements were avoided as much as possible. We have depth measurements at the end of the experiment. These measurements were published by García-González et al 2018. This reference and these results have been included in the discussion section on the page 11 lines 12 and 27.

What I find especially challenging is that a calibration of the capacitive sensors to a specific soil condition might be in question once the soil matrix properties change. As such, there might be an issue that the changing soil properties under study may have indiscernible effect on the actual measured dynamics. With regard to interaggregate and macropore structure formation, the effect should be more clear to be identified as the resulting dynamics do not rely on precisely measure the absolute soil moisture but the relative dynamics during events. However, the authors do not show direct pedophysical measurements from during the experiments, nor do they discern different (event-scale) soil moisture dynamics (diffusive redistribution vs. initially advective percolation in soil macro structures). I find that an unnecessary flaw in the well-thought study. With regard to the presented results, I suspect that a more detailed view to different temporal integrals of the hydrological responses might allow for more specific and less equifinal parameter identification.

This remark has already been addressed above. We agree that this is a potential flaw in the set-up; i.e. the temporal stability of the petrophysical and pedo-dielectric relationships was not further tested after the identified rather stable sensor calibration reported in Gabriel et al., (2010). It would be interesting to take up this point in a separate experiment in future similar field studies.

**(3)** The authors present a multi-step parameterisation scheme, which first addresses the crop parameters based on observed properties like biomass production, ground cover and root development. In a second step they used the Shuffled Complex Evolution Metropolis algorithm for optimisation (and uncertainty assessment) of the pedohydrological parameters. While this appears as very reasonable choice to derive the parameters for the system, it still cannot resolve the issue of potential parameter interaction and equifinality.

We partially address equifinality issues by using the Ceff as model evaluation statistic (Ritter and Carpena, 2013). The Ceff evaluates the Goodness of Fit in a more efficient way than the classical RMSE, and is less prone to equifinality. We could have included the graphical comparison between simulated/observed data for each layer (x4), treatment (x2) and year (x10) but this would lead to 80 figures (with a temporal line of around 190 dots). We prefer therefore the use of statistical performance indicators, and preferably those that are robust towards equifinality.

The separation of the crop parameter calibration/validation from the hydraulic calibration/validation in this study also aimed to reduce equifinality risks. The potential crop parameter remain constant along the 10 years (as should be because the barley variety was always constant). So if the drying / wetting dynamics of the soil could not be simulated for a given season, we calibrated only the soil properties (and not the crop parameters).

Especially so, when the objective function evaluates the full observation time series and does not include the crop parameters. Given the criterions (NSE and RMSE), I would expect the parameters to trend towards acceptable timing and fit of the mean soil moisture. This to my understanding would not be sensitive to changes in bulk density but might only weakly consider the creation of macro-structure (which the authors point out to be one important process). A Richards-type single domain model might still be capable to assess the research questions, but I suspect that some event responses and respective percolation properties, should be studied in more detail. For this, a more data-focussed approach might be the first step.

This convergence of the model to the mean would be true if we only NSE or RMSE were used. Yet, as stated above, we also use Ceff that penalizes greatly the local differences. By definition, positive Ceff values are values that fit better the observations than just the observed mean. Moreover, we are getting values always higher than 0.57 and many times over the 0.80, which are very high Ceff values in this kind of studies. (Ritter and Carpena, 2013)

(4) Assumingly in order to dynamically determine evaporation, the authors coupled the Richards-type soil water model WAVE with the crop model WOFOST. While this appears to be a very reasonable step, it might not be central for the event-scale soil water dynamics. Moreover, it introduces a quite large number of potentially interacting parameters. Even if one reduces the about 120 crop parameters to the most sensitive 10, their interaction with the soil parameters is still in question. From the biomass plot in fig. 2 I get the impression that the model actually only simulates 3 to 4 different “classes” of production, which are not really coherent with the observations (maybe except for the very lowest ones). This actually is in line with my expectations that WOFOST (unfortunately) is not really capable to be used in a dynamic eco-hydrological setup. For the study at hand, erroneous crop simulations may be susceptible to blur the actual soil parameter effects without providing the desired advances of more correct evaporation simulations. Since apparently ground cover of the cover crops has been observed occasionally and since WOFOST suggests only 3-4 crop development scenarios, maybe a more direct implementation of an ET estimate might reduce ambiguity in the analysis?

We agree that the crop module (it is not WOFOST, but a module based on WOFOST) impacts simulated evapotranspiration. However, the sensitivity of the crop parameters for the soil water dynamics in this case is low, since the crop parameters only determines the potential fluxes which will be reduced to actual fluxes. This reduction, and therefore also the actual fluxes, are mainly controlled by the soil hydraulic properties. Hence, there is a direct implementation of the ET calculation in the model. Also, as previously said, the crop parameters were fixed for the 10 years on the calibrated values from just one year. With that assumption, crop simulation during the 10 years fit quite well to the observed data. So the potential fluxes (potential evapotranspiration) are expected to be well simulated.

It should also be noted that the variability of the field data is much higher than the simulated. It should also be noted that, when the crop production is low (below 2000 kg/ha), it was not always water that was the limiting factor. We were also confronted with bad germinations, pest, nutrient deficiencies... These additional stress factors

were not considered in the model, but partially explains the variability of observations in the field.

Despite my concerns and suggestions for the methods used in the study, with regard to the current setup I find it necessary that the authors give more insight into the observed and modelled soil water dynamics, the parameters apart from fig. 3, the used time stepping, and at best some more details about the actual model realisations (especially since the authors use their own Matlab derivatives). This could partly also be given as supplement.

As such, I suggest the manuscript to be considered for major revisions.

**Minor comments:**

**P2L4:** soil size pore distribution » pore size distribution?

It has been changed

**P2L31f:** I am not sure, if inverse modelling is specifically useful “to overcome a parameter limitation problem” as it faces the issue of parameter interaction and equifinality. I would expect more specific explanation and citations here.

It has been modified

**P3L2:** Although I agree to the general attitude that “multi-sensor” probes have advantages, I do not see that the study could not be done with more standard soil moisture probes. Especially with regard to the nature of capacitive sensors being potentially more effected by changes in the soil properties, one could also think of alternative setups - eg. using TDR probes.

At this point we did not try to state capacitance sensors over other sensors. We are just saying that to have continuous measurements at different depths (multi-sensors) allow this approach, but of course, other water content measurement sensors could be used. Anyway we have rewritten the sentence for clarification.

**P3L3ff:** I would expect that the uncertainties are not directly depending on the identification strategy of the parameters. Thus a consistent measurement over depth might allow for the assumption that uncertainties between the individual records might be reduced. However, under natural conditions there might always be air and gravel entrapments altering the control volume. Maybe I misunderstood the statement?

I think you misunderstood. We are trying to say that only with a good parametrization the uncertainties in the predicted results could be reduced (but not deleted, because intrinsic natural soil uncertainty will be always there). We have rewritten it in order to clarify.

**P3L18:** I would expect the weather station’s sensors to be more important than the logger...

It has been changed.

**P3L24f:** I can grasp the study layout from the description. However, I would suggest a small plot, clarifying on the locations of the random plots, the respective treatments and the locations of the observation stations.

All this information is already available in previous published manuscripts. We preferred to make figures for the new results.

**P4L2:** I do not quite understand: There are 8 randomly chosen plots, but only in 6 soil moisture was monitored?

This is correct. We had 4 plots per treatment for biomass and soil sampling but only in three per treatment we could afford to have moisture sensors.

**P4L4f:** This calibration is very good practice. However, your experiment might raise the question if such a calibration remains valid for changing soil conditions: : I expect this also holds true for the soil hydraulic properties in general. I would suggest to include a paragraph on this in the discussion. Section 2.3: The model system appears very parameter-rich and finally rather complex. Although I can follow your description having once coupled WOFOST with the hydrological model SWAP, I am not convinced that this description suffices to be able to understand the coupled model system and to reproduce your results. Moreover at this stage of reading, I slightly doubt that the model system is actually required to answer the research question.

As commented before, this interaction has been clarified.

**P4L28f:** Was WOFOST used to determine the crop development and soil water use? How has it been integrated? From my experiences, coupling WOFOST and any hydrological model may result in even worse identifiable parameter sets since the crop parameters compensate for soil definitions and vice versa.

As said before, the crop model only interact with WAVE providing a water potential demand at each soil layer (supplied or not depending on previous WAVE available soil water simulation) and a soil cover in order to estimate potential soil direct evaporation rate. But as the crop parameters are fixed in a first step and not changed along the 10 years, there are no interaction of these parameters with the soil parametrization each year.

**P6L15:** I would not consider 612 mm as “very humid” » relatively humid?

It has been changed

For the rest of the paper, I refer to my general remarks above.

Ritter and Carpena, 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. Journal of Hydrology 480, 33-45. <https://doi.org/10.1016/j.jhydrol.2012.12.004>