Rebuttal letter accompanying the revised version of manuscript HESS-2016-168

Gerrit H. de Rooij, Raneem Madi, Henrike Mielenz, and Juliane Mai

Below we address the reviewer's comments (in blue italics) point by point. In many cases, our response to these comments led to changes in the paper. These are highlighted in the submitted revision. All figures except Fig. 1-3 were also adapted.

Reviewer 1.

1.) No information is given on the four soils. I checked in Schelle et al. (2013) and discovered that they investigated samples with these four textures taken from three different sites in Germany and also for disturbed and undisturbed soil. The authors must give specific information on the location of the sampling sites, land use at the sites, whether the samples were disturbed (packed) or undisturbed, sample diameters and number of replicates.

Response: The requested information was added in section 3.1. All samples for the lower suctions were undisturbed, those for higher suction (where water content is determined by texture rather than structure) are disturbed. For this study, the land use at the sampling sites and their precise locations are not so important (we are not carrying out a site-specific study). The sample dimensions are more important.

2.) Looking at the figures, I am a little surprised by the apparent lack of structural pores that fill/drain in the tension range close to saturation, say < 10 cm (especially in the finer-textured soils). It does seem to me that throughout the paper the authors only consider the effects of textural pores and do not consider or acknowledge the existence of structural pores. Is this because you only looked at disturbed (packed) samples? Please discuss and clarify this point.

Response: We focus on unimodal soils only. We made this explicit in the Introduction and added a paragraph discussing structural pores and multimodal models, as requested by the reviewer. At this stage of the research the non-uniqueness problems that the large number of parameters of most multimodal models would interfere with the main objectives of the paper – we would chew off more than we could swallow in one paper.

3.) The authors comprehensively present the equations of the models, but they write nothing about their conceptual basis. A few introductory sentences are needed to explain the concepts and assumptions underlying these capillary bundle models, including the fact, for example, that they assume a mono-modal size distribution. Alternative approaches could also be mentioned (e.g. bimodal models, fractal models etc.).

Response: A theoretical background of the equations was added to the Introduction. There seems to be some confusion regarding terminology in this comment. The retention models never are capillary bundle models, only the conductivity models are. This is explained in the revised text. The revision also addresses multimodal models, as requested by the reviewer, but we had more trouble with the term 'fractal models'. The only fractal approach we found is that of Tyler and Wheatcraft (1990), which does not generate a water retention curve but rather underpins the Brooks-Corey model. We were left wondering how a fractal model would generate a water retention curve. In fractal models, the porosity itself is fractal and becomes dependent upon sample size. The expression for the retention curve would therefore have to be conditioned on the sample size, but we have not come across such a parameterization.

4.) A more extensive database than four soils would ideally be preferable to enable a reliable discrimination between alternative water retention models although I understand that few datasets include the very dry end of the range. The authors could discuss this.

Response: Reviewer 3 made a similar comment, but on the other hand kindly provided references to studies that already did just that. When we let go of the simulation tests and increase the number of soils for which we measure the performance of different parameterizations by comparing goodness-of-fit criteria we would be repeating earlier papers, notably Leij et al. (1997) and Khlosui et al. (2008). We argue instead for using simulations of unsaturated flow instead of merely goodness-of-fit to a static curve to evaluate the performance of any parameterization. This procedure is much more laborious than curve-fitting, and with the four soils we used we could already diagnose significant differences. Adding 4 (or 40) soils would not have made a massive difference but would have come at a considerable computational cost. We note that we also considered the conductivity curves associated with the retention curve, adding two or three cases for each parameterization that needed to be run for all soils under consideration. By selecting a limited number of soils that covered a wide range of textures we aimed to keep the work load and the computational burden manageable while at the same time being able to draw solid conclusions about the performance of the various parameterizations. In the revision we more explicitly contrasted our approach to those of earlier papers.

To our knowledge this is the first paper that compares existing parameterizations by including the effect on the soil hydraulic conductivity in the analysis, and we believe this is a significant contribution to the body of literature on the subject. We made that more explicit in the revised text.

5.) The simulation set-up does not appear to be optimal. My concern relates to the initial condition (hydrostatic equilibrium) in relation to both the bottom boundary condition (unit hydraulic gradient) and the length of the simulation, which was quite short (999 days). Judging from what the authors write (e.g. at lines 733-734), it appears that for this dry climate, this combination results in a simulated water balance that includes a non-negligible term for the change in profile water storage, which is not satisfactory. Water balances in the field should have a negligible change of storage in the long-term and scenario simulations with models should be set-up to mimic this as far as possible. As the authors note, the change of storage is different for the different models, which makes it difficult to compare them with respect to the important terms in the water balance (i.e. recharge, evaporation).

The best way to deal with this is to run a 'spin-up' ('warm-up') period first (separately for each model), then use the final state variables (water contents, potentials) at the end of this period as the initial condition for the actual simulation period (using the same driving data for both periods). The water balances for the second simulation period should then be checked to make sure that the change of storage is negligible. If it is not, the warm-up and simulation periods should be extended until it is. Only then can the water balances simulated by the different models be properly compared. If the authors do this, I suspect the differences between the model formulations will be smaller, though probably not negligible.

It would also be a good idea to summarize the simulated water balances for the different models in a simple table (precipitation, evaporation, recharge, change of storage).

Response: We adopted the suggestion of including a warm-up period and modified the text and the figures accordingly. This should lessen the valid concerns of the reviewer regarding the effect of the initial conditions.

We chose not to increase the simulation period to achieve a closed water balance. In semi-arid climates this might take decades or longer, as infiltration from heavy rainfall clusters lingers for decades in the deep unsaturated zone. The meteorological data for such a long-term study are lacking, and we would have had to set up the soil columns differently. We intend to eventually carry out such simulations, but as a follow-up to this work. We will have to rely on simulated rainfall for those studies since the 3-year record we use here is one of the longest we could find for this type of climate.

Incidentally, for our short columns, closure of the mass balance depends less on the duration of the monitoring period than on the strategic choice of the start and end time: if both the start and the end time are chosen after a prolonged dry period, all infiltrated water will have evaporated if all showers were small, and some of it will have contributed to deep drainage if some of the showers were heavy enough. Especially when heavy showers were lacking, not too much can be learned about the effect of the parameterizations on the partitioning between evaporation and deep drainage, even if the mass balance is nearly perfect. We therefore prefer our limited meteorological record, which luckily contained a wide range of rain showers, individually and in clusters.

We clarified the text in that we indicate that we want to study the effect on the fluxes of liquid water and vapor under widely different circumstances: large gradients and sharp contrasts in water content during infiltration after long dry spells, shallow infiltration and subsequent evaporation of small rain showers, and prolonged periods of combined liquid and vapor flow after deeper infiltration of heavier showers. The simulation period includes all these and is therefore well suited to compare the various parameterizations, even when the water balance does not close over the simulated period.

Reviewer 2.

The manuscript tries to address several issues, e.g., the deficiency of soil water retention (SWR) models near saturation, SWR models near the dry end, development of a general criterion for plausible hydraulic conductivity (K) curves, comparison of different SWR and K models, different methods for parameter optimization, numerical simulation to evaluate model selection on drainage and evapotranspiration, and model calibration/inverse model. Even the abstract contains multiple paragraphs, each of which addresses a different issue. As a result, neither of the issues is convincingly addressed.

In my opinion, the development of a general criterion (Eq. 4) for plausible K curves is interesting and can be the main issue of the manuscript. If so, the manuscript needs to provide convincingly theory and experiment results and the conditions a model can or cannot be used. However, to validate the correctness of the criterion, experiment errors need to be considered as well. For example, the SWRs were measured with several methods and the results differ more or less for a given soil. If the difference among different SWR models is less than the measurement error, the

SWR model should be fine. The manuscript needs to provide the implications to the readers how they can use the models correctly or appropriately. Section 2.2 is very long and can go to an appendix.

Response: The reviewer states that we did not address various issues comprehensively, yet later states that section 2.2 should become an Appendix on the grounds that it is long. This leaves us to believe that the reviewer failed to grasp crucial elements of the paper. The most obvious is the fact that nearly all of the parameterizations currently used in numerical Richards solvers are plainly physically wrong, including the one that has become the de facto standard worldwide. This is what section 2.2 sets out to prove, and does so convincingly according to the other reviewers.

This lack of understanding is also apparent from the choice of issues presented in the first paragraph. We pointed to some of the more glaring misconceptions in our on-line reply and will not address these here again. We make an exception for the erroneous thought expressed by the reviewer in the second paragraph that measurement errors somehow affect the validity of the mathematical analysis of the behavior of parametric conductivity models that generate infinite gradients near saturation that translate physically into the existence of a soil pore of infinite radius. This reasoning is so patently flawed that we cannot use it at all to improve the clarity of our text.

I don't think the numerical simulations using different SWR and K models can be used to validate or invalidate the models. First, modeling evaporation and drainage is challenging and different simulators can produce very different results because they may use different algorithms to solve the problem. Second, some models perform better for certain flow process (e.g., infiltration, redistribution) or soil types while other models perform better for different processes (e.g., evaporation, drainage) or soil types. Third, the assignment of initial and boundary conditions can lead to very different results. For example, for a soil that is never saturated for a simulation, the inaccuracy at the near saturation condition probably does not matter much.

Response: This paragraph has no relevance for the paper. Not only can model simulations very well be used to compare the performance of different parameterizations of the soil hydraulic properties, they are, in fact, the only viable way to do so considering the fact that goodness-of-fit type evaluations do not give an indication of the effect on water fluxes in soils that different parameterizations have. The sole purpose of the parameterizations of the soil hydraulic property curves is to allow numerical models to quantify fluxes in the unsaturated zone. We therefore fail to see how the reviewer can conclude that such calculations should not be used to test the performance of such parameterizations. The reviewer goes on by mentioning several different aspects of unsaturated flows that could be considered when evaluating the performance of the parameterizations but somehow missed that all of these were represented in the test scenario that we developed.

The dry-end issue may be left out because it was mentioned but not addressed. The parameter optimization should just be the methods to obtain parameters. It's better if the uncertainty in the optimized parameters be given.

For the reasons above, the manuscript is not publishable in the current form.

Response: Here too, the reviewer is off the mark: the fact that many of the parameterizations tested had their dry branches tailored to better represent observations in the dry range apparently escaped the reviewer. Also unnoticed went the fact that our scenario used boundary conditions representing a desert climate with very long dry spells under high evaporative demand. The explicit comparison between the fluxes generated by parameterizations with and without a specific dry branch also was also ignored. Thus, in this one comment the reviewer disregarded key elements of the theory section, the methodology, and of the results and discussion section.

<u>Reviewer 3</u>
Reviewer 3 starts with summarizing the main objectives and findings of the paper. In our view, this summary is correct, so we will not comment on it further. Below, we therefore only repeat the comments that follow this summary.

Based on my reading of the manuscript, I think it is in general significant even if the approach is not novel. The manuscript is fairly structured. The introduction of the paper illustrates quite clearly the rationale and the objectives of the work. However, it does not provide an exhaustive literature review about the approach used, with references missing important papers (since the 1990s) dealing with the same issue. Figures and Tables supports the findings, especially the part on the prediction of the hydrological processes selected for analysing the model performance in terms of functional properties.

Response: We gratefully acknowledge the literature provided by the reviewer and have included the suggested papers and others in the Introduction. We also better explained in the text that including the conductivity curves in the analysis and adding model simulations as a performance assessment tool are novel elements in our paper.

The strength of the work lies in the fact that the authors provide a systematic and comprehensive review of the WR and HC models available for hydrological analysis. Crucial in the manuscript is the effort to unify and generalize the analyses of the WR behavior provided by the different models, especially near saturation.

On the other side, I see some limits in the manuscript that can be summarized as follows: 1. The comparison among models is not novel and is based on a too limited WR and HC dataset . There are papers in the past dealing with the same issue of analyzing the performance of WR and HC functions, based on huge datasets, that the authors do not consider at all. I mainly refer for example to the work by Leij et al. (1997). The authors assembled different types of mathematical formulations and tested them on a large data set crossing practically the whole textural triangle. I would also add Cornelis et al. (2005).

Response: We included the papers mentioned by the reviewer in the Introduction. Given the availability of these and of Khlosi et al. (2008), we consider the comparison of soil water retention models based on goodness-of-fit to static water retention data points adequately covered in the literature.

We therefore included in the comparison the hydraulic conductivity curves associated with the retention curves according to three models, and evaluated the resulting combinations through numerical simulations. To our knowledge this is the first comparative study to do so, and we believe this contributes meaningfully to the existing body of literature. This approach is much more laborious than simply fitting curves to data and compare fits. Doing this for 40 instead of 4 soils is hardly feasible, and would not change our conclusions very much: we showed that the choice of parameterization makes a significant difference for four widely different soils, and we expect this to hold for a larger selection of soils as well.

We discussed the literature offered by the reviewer in the text and better explained how our approach differs from and adds to these earlier studies.

2. The approaches used are all unimodal. As noted by the anonymous Referee #1, the dataset misses most of the information on structural pores. It is not a case that most of the WR curves in figure 2 provide a similar flat behavior in the pF range 0-1. And yet, central in the manuscript is the behavior of the WR functions near saturation. It is well known that HC models based on the statistical capillary-bundle approach, which are based on the Hagen-Poiseuille law and which integrate the reciprocal of the pressure head to obtain the hydraulic conductivity (as in the case of the Mualem conductivity expression), are particularly sensitive to the slope of the water retention near saturation (Durner, 1994; Coppola, 2000). The effects of a wrong description of the WR close to saturation may have impressive effects on the hydraulic conductivity estimation, with an impact on the soil hydrological processes predictions which may well be larger than the effects observed by the authors in their unimodal analysis (see for example Coppola et al., 2012). Bimodality may also exist quite far from saturation. By looking at the figure 2d, the data trend may well suggest a bimodal behavior in the pF range 2-3 (more or less). It is thus not a case that for the silty loam all the WR functions give a poor description of the data in the drier region.

Response: In the revised text we discuss structural pores and multimodal models. We also explicitly acknowledge that we limit ourselves to unimodal models. We agree with the reviewer's statement regarding the sensitivity of the capillary bundle models near saturation, which was a major motivator for the mathematical analysis we carried out. This analysis is valid for multimodal models as well: each of the underlying unimodal equations should meet the criterion we formulated, so the generalization to this category of models is straight-forward.

We are not sure why the reviewer believes a combination of multiple retention curves would give a dramatically different hydraulic conductivity near saturation. Our analysis showed that models need a non-zero air-entry value to ensure physically realistic behavior of the conductivity. The reviewer's statement implies that she/he believes that the multimodal conductivity (comprised of the conductivities associated with the composing unimodal retention models) differs much more from the unimodal value at the air-entry value than the difference of either of them from the values from either the unimodal or the multimodal conductivity at zero matric potential derived from retention models without water-entry values. But those (incorrect) values will go to infinity (corresponding to the conductivity of a pore with infinite radius), so we think this statement cannot be mathematically correct.

N.B. The conductivity models are formulated as relative conductivities that are scaled by the value of the conductivity at saturation. This forces the relative conductivity at saturation to be equal to zero, even though its gradient there is infinite. The necessary consequence of this therefore is that both the relative and the absolute conductivity drop to zero as soon as the matric potential drops below zero. The physical analysis of the retention curve near saturation shows that there is a fraction of the pore space that only desaturates at zero matric potential, which corresponds to an infinite pore radius, which in turn leads to an infinite conductivity.

The problem with many of the multimodal models is the risk of non-uniqueness brought about by the large number of parameters, and we did not want that aspect to cloud our findings, which is why we chose not to extend the study to multimodal models at this time (see also our response to reviewer 1). The soils we selected had very little or no evidence of multimodality (we are less convinced than the reviewer that Fig. 2d provides evidence of this).

3. There is no effort for recommending which model combination is the most suitable to be used in a given water content range.

Response: We believe we should not give such a recommendation. Even in dry climates, infiltration into dry soils will result in a very high degree of saturation at the wetting front, so even then, the full range of water contents and matric potentials will be encountered. Conversely, even moderate climates have precipitation deficits in summer with crops reaching wilting point from time to time and top soils drying out strongly. Only under very humid climates, scenarios with year-long moderate matric potentials can be imagined, but not many simulation studies are carried out for such areas.

We also expect that the most frequently encountered matric potential range depends strongly on specific circumstances: soil type, weather, vegetation/crop, water management and irrigation regime, etc. A researcher with the competence to run a Richards solver is probably knowledgeable enough to make a sound judgement call on what parameterization to choose from the leads we provided in the paper and her/his own expertise and preliminary model runs.

4. One of the objectives of the paper is " . . . a robust fitting method applicable to various parameterizations and capable of handling data with different data errors" (see end of page 2). The authors introduce the reciprocal of the variance of the errors for weighting differently the single data in the various water content ranges. This is an extension of what one generally does when introducing different quantities in the objective function. And yet, they do not give any information on how they estimated these variances. Selection of data error variances seems to be crucial for determining the performance of the different WR models in describing the experimental data. In other words, the fitting results shown in the figure 2 may be partly an effect of the model parameterization and partly an effect of the error selection for single data.

Response: We actually give this information in section 3.1. We believe the reviewer overlooked this.

Some other remarks
By looking at the HC curves in the figure 3, it seems that the AS curves of the three soils are almost the same. The same may be said for the Mualem and Burdine curves for the silt and the silt-loam. It may only be a result of the axis extent used for the HC curves. Please, consider to show the curves for a smaller HC range of values. In general, the AS curve remains stably higher than the B and the M HC curves. As the authors are providing ". . .a critical evaluation of parametric expressions", the authors should even shortly explain the reason for this behavior compared to the Burdine and the Mualem models.

Response: The reviewer states that the AS curves in Fig. 3 are nearly the same for all thee soils. The vertical log scale still results in several orders of magnitude difference in the conductivity, for instance at pF 6. We agree that the curves according to Burdine and Mualem differ little between silt and silt loam, understimating the conductivity for silt and mostly overestimating it for silt loam.

Changing the axes does not bring much benefit in our view - we prefer to show the curves in their entirety and with the possibility to compare between the panels. The reviewer capitalized on by doing exactly that. Given those preferences, the scales are quite adequate. Also, the fact that HESS is an on-line journal allows the reader to zoom in on figures at will.

The reviewer would like us to explain the difference between the conductivity curves accordint to Burdine/Mualem on one hand and Alexander and Skaggs on the other, but we do not see what exactly the reviewer expects us to do. The fits of the retention curves are what they are and the corresponding parameter values translate directly into the three conductivity functions displayed in the figure. The three conductivity models differ only in the (fixed) values of their three additional parameters, resulting in small differences between Burdine and Mualem (also alluded to by van Genuchten (1980)), and a much larger difference between these two models and that of Alexander and Skaggs. The difference can necessarily only be caused by the difference in the values of the three parameters, but there no need to point that out for lack on an alternative explanation. The graphs do a nice job showcasing these differences, and are preferable over a verbal description of those differences.

In any case, in the section 4.2.1., what the authors consider as "physically implausible" behavior (when discussing about sustained, constant flux leaving the silt soil profile during prolonged dry periods in the AS case) strictly depends on the high values the AS HC curve keeps even for very dry conditions. I would not like it to be also the effect of numerical problems. For example, by looking at the graphs in the panel 4c, the silt-RNA_Mualem/Burdine cumulative drainage curves cross the silt-RNA_AS curve. The latter remains unexpectedly lower, given that the WR curve is the same and the AS curve is systematically higher than the Burdine and the Mualem curves). Maybe, the authors should give more details about the evolution of the pressure head at the bottom boundary conditions during the numerical simulations. They do this only for sand. The reader could do an effort for extending the discussion for the sand to the silt soil. However, the authors may agree that this may be quite laborious.

Response: We respectfully disagree with the reviewer and stand by our explanation. A constant bottom flux under conditions of zero influx requires the soil to dry out. For the flux to remain constant, the matric potential at the lower boundary must increase to compensate exactly for the non-linear drop in the conductivity brought about by the soil drying. By extension, something similar must happen throughout the drying profile to deliver the water to the lower boundary so it can flow out there.

The reviewer did not consider Fig. 14 in this comment. There we show clear evidence of infiltration during dry periods (with an upper boundary condition of a fixed, very low matric potential), which is direct evidence of numerical irregularities. We doublechecked the input files and found no errors in them. As we point out in the paper, there are various signals from the model output files can calculation times that indicate numerical difficulties.

In the figure 5, it seems that the evaporation fluxes are inversely related to the drainage fluxes. Higher drainage induces lower pressure head in the soil profile resulting in lower upward fluxes. Again, one should have a look at the pressure gradients at the soil surface. Nonetheless, in the dry region the AS curve may be even five or more orders of magnitude larger than the Mualem and Burdine HC curves. Thus, in the panel 4d (just as an example), I would not expect higher VGA_Mualem than VGA_AS evaporative fluxes, unless the hydraulic gradient at the soil surface in AS case be five or more order of magnitude lower than in the M/B cases.

Overall, I have no major problems with the manuscript and recommend publication after the authors have discussed these remarks.

Response: An inverse relationship between drainage and evaporation fluxes is to be expected: water that leaves the profile through the upper boundary cannot drain through the lower boundary and vice versa.

We believe the reviewer overstates the role of the drainage flux in driving the upward evaporative fluxes. The latter are driven by strong gradients very close to the soil surface. The gradient driving the drainage flux is generated by water 'escaping' to larger depths where the atmospheric boundary condition is not felt that strongly. The large gradients in the top soil dampen the influence of the atmospheric boundary condition at larger depths, and conversely strongly limit the influence of the lower boundary condition on the conditions in the top soil. The difference between AS and Mualem that the reviewer alludes to is largely an effect of the response to the initial condition. In the revision we corrected for that by treating the first 250 d as a burn-in period (as suggested by this reviewer), essentially eliminating the issue.