General comments

In their manuscript *"Precipitation fields interpolated from gauge stations versus a merged radar-gauge precipitation product: influence on modelled soil moisture at local scale and at SMOS scale"*, J.T. dall'Amico and colleagues present their results on the effects of different precipitation inputs on simulated soil moisture at different scales.

The topic of remote sensing of soil moisture and its comparison to soil moisture patterns simulated by hydrological models is relevant to the scope of HESS: Remote sensing of soil moisture has a vast potential to contribute to hydrological assessments, particularly in data scarce areas. The current uncertainties in remote sensing of soil moisture still call for additional joint research efforts of both hydrologists and remote sensing scientists. The paper in general is well written and well structured and in most parts well comprehensible.

However, the manuscript does not reach the scientific depth and level of innovation required for publication in HESS. In my opinion, the study is not yet complete – neither in terms of methods, nor results or conclusions. Therefore, **I recommend rejecting the paper** in its current form. A profoundly revised version might though be suitable for resubmission; however, this would require a substantial amount of additional work.

In the following, I would like to explain this recommendation.

The authors state that *"this study aims at analysing the potential of using a merged radar-gauge precipitation input dataset for the modelling of soil moisture fields in the Upper Danube Catchment (UDC) […] an understanding can be gained of the uncertainties in the SMOS cal/val activities in the UDC area that are associated with the precipitation input"* (page 3388). In general, this is a valid objective. The main motivation of the author's objective is to provide simulated soil moisture patterns as a reference for the evaluation of remotely sensed soil moisture (exemplified by SMOS). To my mind, the authors do not fulfill their objective: In fact, the study does not compare simulated soil moisture patterns to remotely sensed SMOS data or to any other soil moisture reference.

In a nutshell, the authors force a hydrological model (PROMET) with two different precipitation inputs and compare the simulated soil moisture patterns for different levels of spatial averaging. Their main conclusions are that the differences between the simulated soil moisture patterns are small and that these differences decrease with a higher level of spatial smoothing. In my opinion, these conclusions are quite weak. Beyond, these conclusions do not substantially relate to the authors' aim since the actual comparison to observed soil moisture patterns lacks. Such a comparison is prerequisite to evaluate the "potential" of a precipitation input for simulating valid soil moisture patterns. Differences between the soil moisture patterns simulated by using different QPEs do not imply that one of both is "wrong". In the same way, an agreement between the two simulated soil moisture patterns does not mean that both are "right". One of the authors' main conclusions (p. 3399, ll. 11-14) is based on this assumption, though. It is not sufficient to refer in general to previous successes in the simulation of soil moisture with PROMET (see pp. 3388, ll. 24 ff.). In order to actually evaluate the potential of a precipitation product to serve as input for simulating soil moisture patterns, simulated soil moisture patterns need to be compared against ground-based observations for the exact study period and study area.

The authors extend their analysis with a small comparison to brightness temperatures observed by an airborne EMIRAD sensor. However, due to a lack of substantial rainfall events in the corresponding period this analysis remains highly inconclusive. I would recommend removing the parts related to brightness temperatures from the manuscript.

Beyond these major shortcomings, there are a number of substantial issues which need to be addressed by the authors (I will not comment on technical details at this stage of the review process). These issues will be discussed in the following as related to scope, methods and conclusions.

Scope

1. On p. 3388, ll. 9-10, the authors refer to Juglea et al. (2010) in that *"[…] at SMOS scale, soil moisture variability is mostly driven by atmospheric forcing effects."* The authors leave it unaddressed whether they think that this atmospheric forcing only includes precipitation or also climate variables affecting evapotranspiration. It would be really interesting to add to Juglea et al. (2010) by analyzing which processes and variables in fact govern soil moisture variability at SMOS scale. I would expect that on a smaller time scale the precipitation is dominant while on larger time scales the interaction of soil, topography, vegetation, temperature and wind might become dominant. But this is just an assumption. Anyway, the authors need to look into this: If other processes than precipitation turn out to be dominant on a specific time scale, it might not very insightful to focus on differences in soil moisture patterns which were simulated by different precipitation inputs.
2. Beyond, the authors focus in their analysis mainly on the daily time scale, i.e. they compute measures of agreement on a daily time step. I think the authors need to provide evidence that this is actually the time scale of processes controlling the soil moisture in the upper five centimeters of the soil. This might be achieved via literature research or by own analyses. In my opinion, it is not sufficient to refer to the fact that the SMOS data is available for only one point in time per day – since this is no daily average but an instantaneous measurement. The basic question is how quickly the upper five centimeters react on rainfall pulses and how long a rainfall pulse affects the soil moisture in the upper five centimeters.
3. The authors need to address the issue of the penetration depth of the microwaves in relation to SMOS more explicitly. A penetration depth of 5 cm does certainly not hold for all conditions. Instead, the depth is influenced by vegetation, the soil itself and even its moisture. Thus, simply averaging over the top five centimeters of simulated soil moisture might not be enough. In addition, the authors should show (or refer to the literature more explicitly) that PROMET can actually simulate soil moisture in the top two or five centimeters of the soil on a resolution of 1 km^2 and is thus potentially useful in the desired context. This should also include a discussion of the meaningfulness of soil hydraulic parameters on this spatial scale.

Data

4. The authors use hourly rain gage data from the LfL. Why do they not use additional rain gage data from the German Weather Service (DWD) – at least for validation? In my opinion, there is no justification in ignoring the DWD data as many are freely accessible online and others against a very small fee.
5. The reference which the authors provided for the radar-based QPE from Meteomedia does not provide any information on the radar data or the merging procedure whatsoever. Neither does it contain information about the network of rain gages used for the merging. I understand that the actual QPE methodology is not in the focus of the authors nor the manuscript. However, in order to evaluate potential weaknesses of the radar-based QPE it is absolutely necessary to get an idea about the methodology of the merging, the level of radar data processing (clutter correction, attenuation correction, Z/R transformation etc.) and the underlying rain gage data.
6. Why do the authors not use observed stream flow data in order to get an idea about the hydrological implications of the different precipitation inputs? This would be a very obvious way e.g. to analyze potential biases in the QPE products.
7. Why do the authors restrict their analysis to the short period of five months? If they really want to highlight the effects of small scale heavy rainfall, the study period should cover several seasons and I do not see any basic limitations in data availability.

Methods

8. On p. 3395, ll. 3-4, the authors state that *"for the first run, LfL station precipitation data were used and interpolated […] for the second run, the Meteomedia data precipitation fields were used directly to force the model […]. All remaining model configurations, initializations and input*

data were the same for the two runs". I think the authors should thoroughly analyze the implications of this procedure: Based on which precipitation input was the model calibrated? And based on which precipitation input were the models initial conditions computed? How long might the initial conditions influence the soil moisture patterns? – the actual study period only comprises five months! Similarities in simulated soil moisture patterns might to some extent simply be due to the joint initial conditions. In a similar way, this applies to the calibrated model parameters. It has been shown that (systematic) errors in precipitation inputs can be substantially counterbalanced by model calibration (Heistermann and Kneis, 2011). What does that imply for the use of only one joint parameter set? At least, the author need to calibrate the model for each precipitation input and they need to compute the initial conditions of the study period based on each precipitation input. Otherwise, the results might be seriously flawed.

9. Another important aspect is the use of measures for the agreement of two time series. It is not sufficient to look only at the correlation and the RMSE since this does not give any idea about the level of systematic disagreement. This applies to both precipitation *and* simulated soil moisture. Particularly radar-based rainfall products tend to be biased and this bias is typically not spatially homogeneous. The impact of a biased precipitation estimate on simulated soil moistures is obvious. The authors certainly need to look into this.
10. The evaluation of the precipitation products is also incomplete in other respects: The authors urgently need to validate the gage-based, interpolated QPE in the same way as they did for the Meteomedia product. If they lack independent data, the authors could use cross-validation instead. Beyond, the evaluation of the precipitation products on hourly resolution should receive much more attention. The results already indicate that errors smooth out on larger temporal scales. However, infiltration and surface runoff rather happen on the hourly time scale and are of course major processes influencing soil moisture.
11. In the precipitation validation, the authors compute the quality measures by using all time steps. This procedure is known to produce good quality measures since "no rain" or "low rain" events are generally estimated very well. In their precipitation validation, the authors should focus on time steps with rainfall sums which significantly affect the soil moisture – e.g. hourly rainfall sums larger than 1 or 2 mm or even more. I expect that the overall error for such relevant events is much higher. The resulting quality measures would be more helpful for the comparison of the two precipitation estimates.
12. The authors do not really relate the results of the precipitation validation to the evaluation of the simulated soil moisture. Of course, this would require – as put above – the validation of the gage-based precipitation input and the use of more meaningful evaluation measures beyond correlation and RMSE.
13. If I understood correctly, the authors only compare simulated soil moistures at 05:00 UTC (p. 3395, ll. 4-6). Why is this? If they do not compare the modeled output to actual SMOS observations anyway, they should not exclude 23/24 of the time steps. But maybe, I misunderstood that. In that case, the authors need to clarify their procedure.
14. The authors should discuss in more detail the differences in the simulated soil moisture patterns – this particularly applies to Figure 4: There are strong differences between the patterns of correlation and RMSE – what are the reasons? Is it the systematic error? A comparison of the spatial agreement of the two precipitation products on different temporal scales might be helpful.

Conclusions

15. The authors suggest that a radar-based QPE enables a better simulation of soil moisture at small spatial scales in case of convective rainfall events. Of course, this is a valid hypothesis. However, they provide no systematic evidence to support this assumption - only one single point in time for one small sub-region. The authors should provide a more systematic analysis of convective events which not only takes into account the simulated soil moisture, but also the actual rainfall distribution.

16. On p. 3398, ll. 6-8, the authors state that *"uncertainties increase for hourly precipitation data, especially concerning the exact timing and amount of short but intense precipitation events."* This conclusion is speculative and not supported by evidence. The same holds true for p. 3398, ll. 11-12, ll. 22-24 and p. 3400, ll. 5-9. I strongly suggest that the authors draw their conclusions based on the actual results presented in section 3.

Some specific comments

1. *On p. 3387, ll. 4-5: "…while the water content of the soil surface is of great importance for applications in meteorology and hydrology."* - What is meant by *"soil surface"*?
2. The authors use the term assimilation several times. Forcing a hydrological model with a precipitation input must *not* be referred to as assimilation.
3. On p. 3387, ll. 27-28: *"This spatial resolution is rather low when compared to the available in situ measurements"* – to which in situ measurements are the authors referring?
4. On p. 3388, ll. 6-8: *"The quality of hydrological model output crucially depends on the quality of the input data, in particular on the spatial variability of rainfall"* - Variability is not quality!
5. On p. 3389, l. 4: *„Schlenz et al. (2011) studied explicitly the uncertainties of the SMOS validation…"* - what is meant by uncertainty of validation?
6. The SMOS data should also be introduced shortly in the data section.
7. On p. 3390, ll. 21-22: *"Hellmann rain gauges installed at 1m above ground […]"* - typically, a Hellmann is referred to as a totalizing precipitation measurement.
8. On p. 3391, ll. 21 ff.: The meaning of brightness temperatures should be explained to the reader.
9. On p. 3395, ll. 20-29: Here, the authors should also show the corresponding precipitation accumulations.
10. On p. 3396, ll. 1-7: In the introduction, the authors mention a SMOS resolution of 40 km, but here the SMOS resolution is referred to as 12.5 km. This should be clarified.
11. Fig. 2: Why not show one scatter plot including all stations?
12. Fig. 4 should be discussed in much more detail!
13. Fig. 5 should also show the two precipitation series so that the reader can understand the dynamics better.

Evaluation according to the criteria as given by HESS

Scientific Significance: Fair

Scientific Quality: Fair

Presentation Quality: Good

For the following criteria are evaluated with points between 1 (excellent) and 4 (poor):

1. Does the paper address relevant scientific questions within the scope of HESS? **2**
2. Does the paper present novel concepts, ideas, tools, or data? **3**
3. Are substantial conclusions reached? **3**
4. Are the scientific methods and assumptions valid and clearly outlined? **2**
5. Are the results sufficient to support the interpretations and conclusions? **3**
6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? **4**

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? **3**
8. Does the title clearly reflect the contents of the paper? **2**
9. Does the abstract provide a concise and complete summary? **2**
10. Is the overall presentation well structured and clear? **2**
11. Is the language fluent and precise? **2**
12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? **2**
13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? **Section 3.4 and related parts of the paper**
14. Are the number and quality of references appropriate? **3**