Reply to the comments of the anonymous referee #2

The authors wish to thank the anonymous referee for the time he has dedicated to provide his extensive comments. However, a misunderstanding regarding the scope of the paper seems to have occurred. The study presented in this paper is one of a series of studies, which are all related to the cal/val activities for SMOS in the Upper Danube Catchment. It is NOT the scope of this study to go into the details of all aspects of these activities (e.g., model validation with ground data, comparison with SMOS data). More detailed studies regarding model validation and comparisons with SMOS data have already been published (dall'Amico et al. and Schlenz et al., IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no.5, pp. 1507-1516 and pp. 1517-1529). The aim of the study presented here is to look specifically at the influence of the precipitation forcing data on the modelled soil moisture patterns. While the comparison with EMIRAD remotely sensed observational data suggests that either modelled precipitation patterns are plausible, any attempt to validate the modelled precipitation with either precipitation forcing data is not within the scope of this paper. A validation of modelled precipitation for the model in question with forcing data, e.g. from DWD, has been studied before (see Mauser and Bach 2009).

Please find detailed answers to the referee's remarks below.

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.

It is not clear to the authors why the paper is recommended for rejection given that the evaluation according to the objective criteria as suggested by HESS and presented by the same referee at the end of his comments, yields 2.4 points on average (on the scale from 1-excellent to 4-poor). The only question which is answered with 4 points by the referee regards the traceability of results – the traceability is necessarily limited since a commercial data product is used in this study. Thus, the referee rates the paper as good according to objective criteria while suggesting a "profoundly revised version" with several pages of detailed comments on what would extend the study in question to a study it was not meant to be from the authors' point of view. Apparently, this paper did not target the expectations and prime scientific interests of this only referee. This may be due to a lack of clarity in the introduction which has been indeed profoundly revised in order to make clear what the scope of the paper is. In this context, also many other parts of the paper have been clarified.

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.

It is not the objective of this study to compare simulated soil moisture patterns to SMOS data. The objective of this study is to gain an understanding of the uncertainties associated with the precipitation input or, using a different wording, to gain an understanding of the sensitivity of the simulated soil moisture on the SMOS scale to the precipitation input. In order to fulfil this objective, there is no need to compare simulated soil moisture with SMOS data. The introduction has been improved in order to emphasize the scope of this paper, and related studies on model validation and on comparisons with SMOS data are now cited in more detail.

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.

Simulated soil moisture is compared against ground-based observations in the same study area by Schlenz et al. (2011). In that study, measurements of several soil moisture stations were used for a comparison of a multi-year period which overlaps with the April to August 2010 period assessed in the study presented here. Additionally, hundreds of distributed soil moisture measurements stemming from a measurement campaign in the period from mid May to early July 2010 were used to validate simulated soil moisture patterns in the Schlenz et al. (2011) study. Again, the point made by the referee goes beyond the scope of the study presented here. Nevertheless, the results of Schlenz et al. (2011) are now summarised in the introduction.

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.

The comparison with EMIRAD data is the only way to extend the comparison of simulated soil moisture patterns to observational data (as done by Schlenz et al. 2011) to areas further away from the ground stations. It seems contradictory to demand further model validation against observational data and at the same time to recommend removing the parts related to the comparison to observed brightness temperatures.

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.

The authors agree that such a study would be very interesting, but disagree regarding the necessity to look into this in the context of this paper. For a validation of satellite data available every 2-3 days, near real time modelled soil moisture maps would have to be generated. For generating these soil moisture maps, the spatial distribution of the precipitation data forcing the model is very important. Clearly, it is difficult to spatially interpolate precipitation from point-like measurements due to its discontinuous nature (as opposed to e.g. air temperature), hence the inclusion in this study of the radar-derived precipitation product. The other variables listed by the referee are likely of a more continuous nature but even if this be not the case, those variables may capture the attention of some other publication and their inclusion here goes beyond the scope of this manuscript. This point has been made clearer in the manuscript.

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.

As stated in Section 2.5, PROMET calculations are performed with hourly resolution. No daily averages of soil moisture are used in this study. Our objective is to study the sensitivity to the precipitation input of simulated soil moisture data which are used in other studies for the validation of SMOS data. For this reason, we are particularly interested in the sensitivity of the simulated soil moisture to the precipitation input at 5 am UTC, as this is the approximate time of the SMOS overpass. PROMET soil moisture at 5 am UTC is taken as the approximate equivalent to the instantaneous measurement performed by SMOS. The question on which time scales the processes controlling the soil moisture in the upper five centimetres occur is certainly interesting, but not relevant to the objective of our study. This point has been made clearer in the manuscript.

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 km2 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.

The comparison of PROMET soil moisture with EMIRAD measurements provides evidence that PROMET soil moisture (averaged over the upper two PROMET soil layers) on the 1 km x 1 km grid is indeed useful in the desired context, as the achieved correlation corresponds to the correlation between EMIRAD measurements and in situ soil moisture measurements (see reference to dall'Amico et al. 2012). EMIRAD measures the microwave emission at the same wavelength as SMOS. The thickness of the soil layer contributing to this emission is indeed variable, but it is the same for EMIRAD and for SMOS. In a vegetated area with a temperate humid climate as in the UDC area, the thickness of the contributing soil layer will not exceed the 0-15 cm region (which is the depth of the upper two PROMET layers). The next PROMET layer lies fairly deep into the soil and is very unlikely to provide a substantial contribution to the remotely sensed brightness temperature in question. Regarding the usefulness of PROMET in this context, please refer to the answers to previous comments and the references cited in the text. This point has been made clearer in the manuscript.

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.

The fact that DWD data are not used in the context of the SMOS validation framework in the UDC derives from the need of near real time availability of hourly meteorological input data from a dense network which was provided by LfL but not by DWD at the time this study was conducted. An explanation has been added to the text. However, independently from the use of LfL stations or DWD stations, the model relies on the interpolation from a dense station network. Before deciding to use LfL rain gauge data, model interpolations from DWD stations have been compared in a different study (not published) to model interpolations from LfL stations and found to be very similar. In other words, there is no need and no advantage to use DWD data.

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.

As mentioned in other replies in the open discussion, and as also acknowledged by the referee, these details are not related to the scope of the presented study. However, a reference describing the methodology on which the production of the used data set is based has been added to the text.

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.

The scope of this study is to analyse simulated soil moisture, not stream flow. Biases in the precipitation products can be seen from the RMSEs in their intercomparison, as these RMSEs are not bias-corrected. Anyone interested in the implications of the different precipitation inputs on simulated stream flow is free to download LfL data and contact Meteomedia AG in order to conduct such a study.

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.

Access to the commercial data product was granted by Meteomedia AG for the time period from April to August 2010. To our knowledge, these are the first five months of operational production of this product. It appears contradictory that the referee asks for a study period covering several seasons and at the same time asks for the validation of simulated soil moisture for the exact time period and study area (see previous remarks). Anyone having experience with in situ soil moisture measurements should know that it is not feasible to acquire representative soil moisture measurements over several seasons in a 77000 km^2 area. In the period from May to August 2010 used for the two model runs, the distributed in situ soil moisture data with which Schlenz et al. 2011 validated PROMET soil moisture are available as well as EMIRAD data with which a comparison may be conducted over a larger area. Therefore, data availability is indeed limited.

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.

The concept of the PROMET model is that it is based on 1st order physical and physiological principles. It completely closes the water and energy balance. The model is thus not calibrated with measured discharge, but of course the discharge is used to validate model performance. The parameterisation of the land surface is done using remote sensing data source to a wide extent and standardized GIS tools, for example to derive routing parameters for each river section. That standard hydrological models need a calibration to realistically provide runoff data is considered as their weak point that PROMET targets to overcome. This information has been added to Section 2.5.

The authors see it as a general wrong way to "calibrate" errors in precipitation in the hydro model. Instead we need to improve the temporal and spatial precipitation fields that force the hydrological model. In this respect the paper is very relevant. Again a remote sensing data source (weather radar) is used to better interpolate station data. We see that the referee has a general different attitude to face the problem.

Of course the model was run before the investigation period started to define the initial conditions. In this case it was one year using interpolated station data as input.

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.

A bias could be seen in the RMSE since the given RMSE values are absolute values and not bias-corrected. More information on the agreement of the two precipitation input data (slopes and intercepts of the regressions) has been added to the text in Section 3.1.

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.

As mentioned in the reply to comment no. 2, the time scale of the processes influencing soil moisture are of minor relevance to the scope of this study, which is to gain an understanding of the sensitivity of modelled soil moisture maps for validation of SMOS data to the precipitation input (whereby this validation goes beyond the scope of this manuscript).

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.

When the analysis of hourly rainfall data is conducted using the proposed procedure (excluding all events with hourly rainfall sums less than 2 mm), the number of stations with correlation coefficients lower than 0.6 indeed increases from 4 to 47 (out of 133). However, the mean correlation coefficient is 0.66 and the mean RMSE (NOT bias-corrected) is 2.9 mm. This is still considered a good agreement for higher rainfall sums, keeping in mind that point-like precipitation measurements can easily have an error of the order of 10%. This point has been made clearer in the manuscript.

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.

The comparison of the novel precipitation data product with gauge station measurements is the pre-requisite to use this data set as forcing data, which then necessarily relates precipitation and simulated soil moisture. In order to enforce the connection between soil moisture and precipitation, precipitation data have been added to Figures 5, 6 and 8.

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.

This is correct (see also reply to comment no. 2). Comparisons with SMOS data are performed in a related study using the same model setup (dall'Amico et al. 2011). This point has been made clearer in the manuscript.

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.

Actually, the patterns of the two performance measures are quite similar with better agreement in the centre of the catchment and increased disagreement in the Alps and the western part of the catchment. Other, smaller differences are probably due to the choice of colour scale for the figure. A comparison of the spatial agreement of the two precipitation products on different temporal scales has in fact been conducted, but the results were not sufficiently different from the presented insights to justify their inclusion to this paper.

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.

The assumption in question is one of many that motivated the study in question, whose scope is mainly to assess the sensitivity of modelled soil moisture to the precipitation data set used to force the model. A systematic evidence of this particular assumption is more likely to lie within the scope of interest of the distributor of this commercial product.

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.

The first statement mentioned by the referee is supported by the results presented in Section 3. The other statements are clearly marked as speculations, which is legitimate in the conclusions section. This point has been made clearer in the manuscript.

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"?

Has been replaced by "upper soil layer".

2. The authors use the term assimilation several times. Forcing a hydrological model with a precipitation input must not be referred to as assimilation.

Agreed, has been replaced (see also the reply to referee #1).

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?

Reference has been added.

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!

Variability is indeed not quality. Nevertheless, the data need to represent the spatial variability in order to be of good quality as forcing data.

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?

"validation" has been replaced by "validation framework" in order to clarify.

6. The SMOS data should also be introduced shortly in the data section.

As no SMOS data are used in this study, this would not be very helpful. Nevertheless, more information on SMOS data has been provided in the introduction.

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.

done

8. On p. 3391, ll. 21 ff.: The meaning of brightness temperatures should be explained to the reader.

A certain degree of knowledge of terminology can be expected from the reader of a high-quality scientific journal. Besides, knowing the definition of brightness temperature is not crucial in order to understand the presented study.

9. On p. 3395, ll. 20-29: Here, the authors should also show the corresponding precipitation accumulations.

done

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.

SMOS data are delivered oversampled on the ISEA grid. This has been clarified in the text.

11. Fig. 2: Why not show one scatter plot including all stations?

done

12. Fig. 4 should be discussed in much more detail!

See reply to comment no. 14.

13. Fig. 5 should also show the two precipitation series so that the reader can understand the dynamics better.

done