# Response to the reviewers for hess-2019-478

High-resolution fully-coupled atmospheric–hydrological modeling: a cross-compartment regional water and energy cycle evaluation

Benjamin Fersch, Alfonso Senatore, Bianca Adler, Joël Arnault, Matthias Mauder, Katrin Schneider, Ingo Völksch, and Harald Kunstmann

The authors would like to thank the referees for their interest in the manuscript and their willingness to provide their valuable comments. A point by point discussion of the issues is enclosed below. Since several identical points were made by different referees, we decided to compile all answers into one document and to hyper-link among the related comments in the PDF file.

# **R**EFEREE # 1

Examining the ability of atmospheric-hydrological coupled modeling in the seasonal prediction is an important topic in hydroclimatic research. This study explored the ability of the WRF-hydro in capturing the regional hydrological cycle. Their results show considerable changes of latent and sensible fluxes between classic WRF and coupled WRF-Hydro though the calibrated parameters were same for both models. When compared to observations, the WRF-Hydro performed better than the classic WRF in the simulation for variety of variables including ET, sensible and ground heat flux, near surface mixing ratio and temperature. The value of this study lies on providing some information in the performance of WRF-Hydro that may be useful for weather/climate forecast service. I support it to be published in HESS if the following comments could be addressed.

### Comment # 1.1

For the Figure 2, since the parameter set was calibrated in WRF-Hydro (WRF-H\_SA), we can expect that when coupled with atmospheric model, the WRF-Hydro (WRF-H\_FC), which is the same model used in the calibration process, should have a better performance than the WRF standalone (WRF\_SA) which is a different model to that

used in the calibration process. Therefore, how could the authors demonstrate that the worse performance of WRF\_SA is not caused by that the parameters calibrated in WRF-Hydro did not work well with WRF\_SA?

#### Answer:

We would like to thank the referee for this comment, which offers the opportunity to clarify two important aspects related to fully-coupled modeling compared to uncoupled, which goes beyond the 'mere' performance comparison in a case study.

The first aspect is related to the fact that, in our experiments, WRF\_SA and WRF-H\_FC share exactly the same parameters: in agreement with previous literature, the results highlight that, with the same parameters, the fully-coupled model provides wetter soil conditions than the uncoupled. In our opinion, from a point of view of hydrological processes understanding, this result is quite interesting, because highlights that, discarding the (real) soil water redistribution processes with uncoupled model is a possible cause obliging to tune soil parameters to "unrealistic values for the sake of obtaining a good matching", as we wrote.

The second aspect concerns a methodological issue. The calibrated parameter set is achieved using a very peculiar variable, i.e. the discharge at the outlet, which summarizes the hydrological behavior of the whole upstream area. In other words, calibrating some LSM parameters using the hydrological model allows to take into account representative features of a wide (catchment) area, summarized in the discharge variable. On the other hand, the parameterization of the same parameters using only the WRF\_SA model would be less straightforward. Of course, they could be parameterized with respect to point observations (e.g., latent/heat fluxes or soil moisture measurements), which however do not guarantee the same accuracy in the remaining area. Also, a two-dimensional spatially distributed calibration could be attempted, e.g., with precipitation (whose spatial distribution is however always provided by other models, e.g., spatial interpolations, reflectivity-precipitation conversion formulas, etc.), which however would result much more difficult than the calibration based on the hydrological model. If this point is considered, the better results achieved with WRF-H\_FC in our experiments with flux and soil moisture observations are even more valuable. Nevertheless, it is clearly possible that other parameter combinations for WRF\_SA could lead to likewise or even better performance as with WRF-H\_FC.

In the updated manuscript, we will provide a detailed discussion about this issue.

While compared to WRF, the WRF-Hydro had a better performance, I am also curious about if compared to other classic hydrological models, such as SWAT, TOP-model, how the performance would be for the WRF-Hydro, especially on the stream flow modeling. Though this may be out of the scope of the manuscript, adding such information (maybe by citing other previous research) could much add the value of this manuscript.

# Answer:

We will add available information by collecting information from previous research at different locations, also including the region of our study.

# Comment # 1.3

P1, line 3: What are the traditional disciplines?

### Answer:

With the wording "traditional disciplines ", we wanted to express that studies are often restricted to the viewpoints of certain scientific disciplines such as meteorology or hydrology. We have changed the sentence to

Modeling instead is often narrowed to single compartments of the terrestrial system or bound to traditional viewpoints of definite scientific disciplines.

# Comment # 1.4

P8, line 13-14: Directly using soil moisture from WRF-H as an initial condition of WRF\_SA can not guarantee the land surface has been well spun-up for the WRF\_SA start.

# Answer:

The spinup of the land surface (canopy, and first soil layers) is usually a matter of hours to days. This is taken into account by having another 15 days of spinup (April 15 to May 01 2016). For the deeper soil, we agree that the system may not be in full equilibrium, however, we assume that the conditions in soil layers 3 and 4 are more realistic with the 6 month + 15 days spinup than with a 15 days only spinup. It is important that both models WRF-H\_SA and WRF-H\_FC start with identical conditions, so that the differences between standalone and coupled mode are commensurable. We will note that also in the manuscript.

P9, line 5: CML→CLM

### Answer:

Changed.

### COMMENT # 1.6

P9, line 11-13: Usually, these processes should be included in a land surface model. Does the WRF-Hydro use more advanced (or complicated) representations for those processes replacing the old ones in the Noah-MP, or there are no such processes in the Noah-MP at all?

### Answer:

In Noah MP it is possible to select a TOPMODEL based surface runoff parameterization and furthermore, a column based confined groundwater model can be activated. Whereas the groundwater model enlarges the vadoze zone storage body of the classic Noah model to the phreatic zone and thus provides an overall increased storage body, the TOPMODEL scheme does not offer an option for horizontal redistribution of infiltration excess water which is analogous to the Schake et al. model in Noah-MP that was used in this study. The surface and groundwater runoff produced by Noah MP can be used to drive separate runoff and channel routing tools. What WRF-Hydro uniquely provides is the overland routing of surface runoff that allows for reinfiltration and thus lateral redistribution of water on the land surface thus enabling feedback with the boundary layer. We slightly changed the paragraph in the manuscript to

Noah-MP provides different options for the computation of groundwater discharge and surface runoff (infiltration excess) but only WRF-Hydro enables the simulation of lateral hydrological processes such as the overland routing of surface runoff and channel (discharge) routing.

### Comment # 1.7

P10, line 18: Are there any reservoirs distributed in the basins? If so, why did you turn off the reservoir module? What uncertainties may come from this setting?

### Answer:

There is one single lake contained in the Ach subcatchment (Staffelsee, 7.66 km<sup>2</sup>).

The gauge Ach-OBN is located upstream and Ach-OBH is located downstream of the lake. From the observations we see that the lake dampens the hydrograph a bit with leads to reduced peaks and a slightly increased baseflow. When we look at the calibration and validation hydrographs, we see that upstream and downstream gauges are simulated reasonably. Therefore, we refrained from adding the reservoir module to the simulation as this would have further increased the number of parameters and thus the amount of calibration runs whereas the effect on the hydrographs are anticipated to be only minor. Since the lake evaporation is separately considered by the land surface model, we do not assume any impact on the land surface–boundary layer exchange. We updated the statement to expound our motivation as follows:

The reservoir module (retention of channel flow in lakes and reservoirs) is disabled for the simulations of this study because only one gauge (Ach-OBH) would be slightly affected by a 7.66 km<sup>2</sup> lake (Staffelsee) but the number of calibration parameters and thus calibration runs would considerably increase. Since the lake evaporation is explicitly considered by the Noah-MP land surface model we do not assume any impact on the land surface–boundary layer exchange by this simplification.

#### Comment # 1.8

P10, line 24-25: Please explain the meaning of "two and three dimensions of soil features representation".

### Answer:

In its standard version, Noah-MP supports only the domain wide assignment of soil layer thicknesses (typically set to 10, 30, 60, and 100 cm, for layers one to four), with 200 cm total depth. We have changed this domain-wide setting to a threedimensional array to support variable soil thickness at each grid point (please refer to Comment 1.15 for the discussion about the benefit of variable depth soil layers). The new implementation provides also the flexibility to use three dimensional information for other soil features, such as e.g., maximum soil moisture content or saturated hydraulic conductivity. However, we did not make use of this feature because it would have required three dimensional soil maps but these were not available. To make this more clear in the manuscript, we changed the text to:

Here, the model's general assumption of two meter soil thickness (or depth to bedrock) does not hold true, as it may lead to overestimated retention of infiltrating water. Therefore, the soil layer thickness definition was changed from a domain uniform

to a grid point based representation and soil layer depths were set to the Noah-MP standard (0.1, 0.3, 0.6, 1 m) distribution for hillslopes below 50 % and to more shallow values (0.05, 0.05, 0.1, 0.1 m), for all slopes >50 %. The 50 % threshold led to a realistic discriminability of valley bottoms and hillslopes.

Comment # 1.9

P12, line 12-13: What is this setting of REFKDT used for?

Related with Comment 3.2 and Comment 3.5.

# Answer:

The REFKDT parameter is erroneously mentioned here. Instead the parameter addressed in this sentence is REFDK which is the scaling parameter for the infiltration equation. We have set this in the simulations to the Noah-MP documented standard value which is 2e<sup>-06</sup> and the typo in the text was corrected.

# Comment # 1.10

P13, line 1-2: Does this simplification induce some errors in regarding to the landatmospheric interactions?

Related with Comment 1.11 and Comment 2.14.

# Answer:

The underestimation of baseflow is not relevant for the land surface–boundary layer exchange, since channel water does not interact with the LSM (channel loss is not considered). We added one sentence to the respective section in the manuscript:

The underestimation of long-term baseflow by the model has no influence on the land surface–atmosphere exchange as there is no interaction of the channel routing with the LSM.

# Comment # 1.11

P14, figure 4: Why there are no shifted hydrographs for Ach-OBH, Ach-OBN and Rt-RST?

# Related with Comment 2.14

# Answer:

Baseflow shifting is motivated by the hydrogeological conditions of the respective

subcatchments. The detailed answer to this comment is provided with the answer for Comment 2.14.

#### Comment # 1.12

P14, line 9: Could the authors show the net radiation for both observations and simulations?

#### Answer:

While the variables lw and sw allow to get a direct comparison, a net radiation comparison is less straightforward, since it could lead to "right results for the wrong reason" (e.g., compensation of under- and over-estimations in the variables of the radiation balance).

### Comment # 1.13

P15, figure 5: Same as P14, figure 4.

#### Answer:

Figure 4 depicts the calibration period whereas Figure 5 shows the validation period. We added the calibration period dates to the caption of Figure 4 to make this more clear.

Comment # 1.14

P29, line 12-13: How does the lateral water transport increase the soil moisture?

Related with Comment 3.9.

### Answer:

The infiltration excess computed by the LSM is transferred to a surface storage layer in WRF–Hydro. Thus, the surface runoff remains in place which allows for subsequent infiltration or becomes horizontally routed by the overland routing module once the retention depth of the surface is exceeded. Thus, the water remains on the surface until it infiltrates, evaporates or reaches a channel element. We will explain that in more detail in the model description section.

P29, line 16-18: It seems to me that the decrease of soil thicknesses made the results worse in the mountainous site. So why did you change the soil layers for the slope region?

# Related with Comment 2.25 and Comment 3.4.

# Answer:

When we had interpreted the simulated WRF-H\_SA hydrographs for the mountainous site (Am-OAG) we found that most of the observed peaks were missing whereas post-peak flow exceeded the measurements. We concluded that this is caused by too much retention of the effective precipitation on the mountain slopes. The 2 m soils on 50% sloped soils is not realistic, typically the bedrock layer starts between 10 and 50 cm depth (see also Ludwig et al., 2003; Hofmann et al., 2009, for a description of soils in the Ammer catchment region). Naturally, the assumption of a hard 22.5° threshold is not quite realistic. However, it gave the strongest gain in performance for our region among the range of values that we tested (between 20 and 60% slope). One could also think of a gradient dependent function that blends from the 2-m standard profile to the thin slope profile. However, this would be a study of its own. Figure 1 and 2 show the differences for standard and alternative soil profile distribution. Most affected are hydrographs for Am-OAG and Am-PEI, they also hold the largest proportion of steep slopes. Thus in terms of water balance, this quick runoff from the mountain slopes is an important process that needs to be considered in the model. We have added some lines to the calibration section and two figure to the supplementary material to explain this a bit more in detail:

The optimum slope gradient for the delineation of shallow and deep soil regions is found to be 50% (22.5°). Accounting for the shallow mountain slopes considerably improves the hydrographs for Am-OAG and Am-PEI, as it increases underestimated peaks and decreases overestimated retention. A comparison of the respective hydrographs and performance measures is shown in figures (SX1 and SX2, supplemental link).

### Comment # 1.16

P30, line 16: In terms of Earth System Modeling, how to calibrate the Earth System Model globally is quite challenging. Observation data are so limited on this scale.

Related with Comment 2.26. **Answer:** 



Figure 1: Simulated and observed discharge for calibration period (2015-05-01–2015-07-31. (a) Uniform soil layers throughout domain, (b) reduced soil layer thicknesses for slopes > 50%.



Figure 2: Simulated and observed discharge for calibration period (2015-05-01–2015-07-31. (a) Uniform soil layers throughout domain, (b) reduced soil layer thicknesses for slopes > 50%.

We agree with the referee that limited data and also computational constraints are a challenge for setting up a model like WRF-Hydro for global application. With our statement, we refer more to regional applications or up to continental extent which seems feasible. We will specify that more explicitly when we update the conclusion section based on the comments of referee #2.

# REFEREE # 2

Review of "Fersch et al. - High-resolution fully-coupled atmospheric–hydrological modeling: a cross-compartment regional water and energy cycle evaluation". The authors present a comprehensive effort in setting up, running and validating a WRF-hydro model run and comparing it to a non-hydro WRF setup. I have suggested major revisions and would like to see the following addressed before publication, but I would like to highlight, that with the changes made, the manuscript is indeed worthy of publication. I hope the comments can also be seen as improvements in readability.

#### Comment # 2.1

Introduction: P2L2-4: The introduction seems to have an abrupt start. Does this first paragraph really have the phrasing to initially frame the story? At least, add "... and magnitude..."after "spatial distribution".

#### Answer:

We will reconsider the introduction and look if we can find a more intuitive start for the story.

### Comment # 2.2

P3L1: The newest general HOBE publication is "Jensen, K.H., Refsgaard, J.C. (2018): HOBE – the Danish Hydrological Observatory, Vadose Zone Journal, 17:180059, doi:10.2136/vzj2018.03.0059". This could be changed or added.

#### Answer:

We have replaced the HOBE reference with the latest one.

The most prominent activities for Europe are HOAL (Hydrological Open Air Observatory, Blöschl et al., 2016), HOBE (The Danish Hydrological Observatory, Jensen and Refsgaard, 2018), LAFO (Land-atmosphere feedback observatory, Spath et al., 2018), and TERENO (TERestrial ENvironmental Observatories, Zacharias et al., 2011).

General intro: I believe an entire paragraph is missing here where the following is addressed: - What is the key relevance of this paper to the public and/or research domain? Why are you/we doing this? This is important but often overlooked. - What is new compared to previous Fersch/South-German studies as well as other studies? A lot of work is mentioned, but the relevance and new story here is overlooked. Please relate.

### Answer:

We agree with the reviewer that the study can be better related with the scientific context and that the motivation could be worked out more comprehensively. We will adapt the introduction section in that regard.

### Comment # 2.4

P7L8: Radolan reference?

#### Answer:

We have added the reference for RADOLAN to the modeling chain section, at the first occurrence of the acronym in the text.

Altogether, as outlined in Figure 2, the modeling chain encompasses the following four steps: 1) a classic standalone WRF run (2015-04-01–2016-10-31) with standard LSM parameters to derive driving variables for 2) the standalone WRF-Hydro simulations (WRF-H\_SA, 2015-04-01–2015-07-31 and 2016-04-15–2016-10-31) that ingest also gridded observed precipitation (RADOLAN, Bartels et al., 2004; Winterrath et al., 2012), 3) a fully coupled WRF-Hydro simulation (WRF-H\_FC 2016-04-15–2016-10-31) using calibrated parameters from WRF-H\_SA, and 4) a rerun of the classic standalone WRF (WRF\_SA, 2016-04-15–2016-10-31) with the same parameter set obtained from WRF-H\_SA.

### Comment # 2.5

P7L9-10: What is the uncertainty in doing this compared to using observations?

### Answer:

For the Ammer catchment, or the region in general, consistent station observations of all required variables are not available. For selected variables few stations are

spread throughout the domain but because of the complex terrain, geostatistical interpolation would very likely lead to unrealistic spatial patterns of the driving variables. Moreover, only for precipitation, gridded data products are available with the required resolution. As the focus with WRF-H\_SA is on discharge calibration we assume that the high resolution gridded RADOLAN precipitation product is most important to obtain good results. Deviations in any of the other driving variables would probably lead to variations in evapotranspiration and could probably affect the recession and baseflow components of the hydrographs.

#### COMMENT # 2.6

P7L3 to P8L6: I think this sections needs a paragraph that explains the key differences between WRF classic, WRF-H-SA and WRF-H-FC to be able to better understand the reasoning here.

Related with Comment 3.1.

### Answer:

We will add a paragraph to the section *Model setup and calibration* to make the differences between the different model setups clear.

## Comment # 2.7

P8L5: This "parameter set" has not yet been explained and confuses. What are these parameters? They belong to the atmospheric realm I believe and are therefore not hydrologic calibration parameters.

### Answer:

The new run of WRF-SA with the calibrated parameter set is needed to achieve our commensurable set of simulations for further comparison, since some of the calibrated parameters of the standalone hydrological model affect also the LSM (which is integrated in the WRF modelling system). We agree with the reviewer that this paragraph can be misleading if the parameter set is not known yet. However, for the sake of clarity and simplicity, we preferred to go into details of it later, when describing the Calibration procedure, rather than here, where the general structure of the experiment is presented. Therefore, we changed the text here to:

Altogether, as outlined in Figure 2, the modeling chain encompasses the following four steps: 1) a classic standalone WRF run (2015-04-01–2016-10-31) with standard LSM parameters to derive driving variables for 2) the standalone WRF-Hydro sim-

ulations (WRF-H\_SA, 2015-04-01–2015-07-31 and 2016-04-15–2016-10-31) that ingest also gridded observed precipitation (RADOLAN, Bartels et al., 2004; Winterrath et al., 2012). At this stage, a calibrated set of parameters influencing the overland flow routing, the baseflow model and the LSM is achieved; 3) a fully coupled WRF-Hydro simulation (WRF-H\_FC 2016-04-15–2016-10-31) using calibrated parameters from WRF-H\_SA, and 4) a rerun of the classic standalone WRF (WRF\_SA, 2016-04-15–2016-10-31) with the same calibrated LSM parameters obtained from WRF-H\_SA. Finally, this leads to a commensurable set of simulations, coupled versus uncoupled.

### Comment # 2.8

P9L2-11: Really hard to link between text and table (1). The schemes/physics as well as references do not match up and several domain levels are refereed to. Please elaborate/correct.

### Answer:

In table 1, we list the configuration of the physical schemes for the WRF model in the WRF\_SA and WRF-H\_FC simulation runs. We do not want to repeat all schemes in the text again. Only for cumulus and land surface model we added some important additional information. We have updated the paragraphs to give it a better structuring:

The WRF physics parameterization for the selected domains is listed in Tab. 1. Cumulus parameterization is only activated for the outermost domain (1), while explicit convection is chosen for the finer grids (domain 2 & 3), according to Skamarock et al. (2008).

For uncoupled and coupled simulations, Noah-MP (Niu et al., 2011) is used as the land surface model. For Noah-MP the selected configuration deviates from the default setup as follows: the Community Land Model (CLM) method is selected for stomatal resistance computation, the Schaake et al. (1996) method is used to determine infiltration and drainage (similar to classic Noah-LSM), the two stream radiation transfer applied to vegetated fraction (option 3) is chosen, and the dynamic vegetation option (Dickinson et al., 1998) is enabled.

Furthermore, we have updated the caption of Table 1 as follows:

Table1. Selected WRF\_SA and WRF-H\_FC model physical schemes.

P9L17: "only subsurface and surface overland flow routing is..."  $\rightarrow$  ok, but then state what is not taken into account in the atmosphere link – the non-WRF-hydro expert does not know this (me included), but would have to guess.

#### Answer:

The second paragraph about WRF-Hydro (starting with *Concerning the one-way...*) describes the processes that have no feedback with the atmospheric part of the fully coupled model. We changed the first sentence of the paragraphs so that it clearly refers as antagonist with the previous one:

Concerning the one-way coupled processes modeled in the WRF-Hydro system (i.e., no feedback with the atmosphere), in this study, the channel flow and baseflow modules are used.

### Comment # 2.10

Figure 1: Many site abbreviations, which are not really 'learned' when reading the manuscript. I don't know if a systematic naming approach could be thought of.

#### Answer:

The naming of the stations and locations follows the following scheme: for the TERENO Pre-Alpine sites they are the official site codes (also used in other flux databeses, e.g., ICOS network). The discharge gauges consist of an abbreviation for the channel (Rt=Rott, Ach=Ach, Am=Ammer) and the location of the gauges. So to our opinion, the naming scheme for the stations is already systematic and we would like to keep it this way. To make the figure slightly clearer, we propose to remove the ScaleX-tag from Figure 1b.

### Comment # 2.11

P10L26-27: "Furthermore..." – How was this done? Based on what? Which data?

### Answer:

In the original versions of WRF and Noah-MP the infiltration and percolation parameters (REFKDT and SLOPE) are read from a text file and are valid for the whole domain. In our version, we read the parameters from a 2-d netCDF fields, however, the values are uniform for each of the sub-catchments (lumped). We have rephrased the sentence in the manuscript to make this more clear: In addition, the infiltration (REFKDT) and percolation (SLOPE) parameter implementation was changed from domain-wide uniform values to subcatchment-wise distributed (lumped).

### Comment # 2.12

P11L28: You could also add: Larsen, M.A.D, Refsgaard. J.C., Jensen, K.H., Butts, M.B., Stisen, S., Mollerup, M. (2016): Calibration of a distributed hydrology and land surface model using energy flux measurements, Agricultural and Forest Meteorology, 217, 74-88, doi:10.1016/j.agrformet.2015.11.012. And; Stisen, S., McCabe, M.F., Refsgaard, J.C., Lerer, S., Butts, M.B. (2011): Model parameter analysis using remotely sensed pattern information in a multi-constraint framework. Journal of Hydrology, 409, 337-349. doi: 10.1016/j.jhydrol.2011.08.030.

# Answer:

The two papers were added to the list.

Several approaches are adopted to reduce or control this problem, particularly challenging for the emerging fully distributed paradigm in hydrology (Beven and Binley, 1992; Beven, 2001; Kelleher et al., 2017), either constraining the parameter set by means of various strategies (e.g., Cervarolo et al., 2010) and/or incorporating different observations than discharge in the calibration process (e.g., Thyer et al., 2004; Stisen et al., 2011; Graeff et al., 2012; Corbari and Mancini, 2014; Larsen et al., 2016b; Soltani et al., 2019).

### Comment # 2.13

P12L17: Computational resources  $\rightarrow$  should these not be mentioned in the manuscript?

# Answer:

We added a sentence to the modeling chain subsection to provide the numbers for computational demand

The computational demand was about 0.021 million core hrs for the WRF\_SA simulations, 0.32 million core hrs for the WRF-H\_SA calibration runs and 0.042 million core hrs for WRF-H\_FC on a 2.3 GHz Intel Haswell system.

and a further sentence in the calibration methods section:

The calibration period length of 3.5 months (2015-04-15–2015-07-31, including 14 days of spin-up) is selected as a compromise between the number of model runs (about 2000, which relates to about 0.32 MIO core hours for WRF-H\_SA), required during hypercube sampling and PEST optimization, and the available computational resources.

#### Comment # 2.14

P12L20-P13L1: Why/how did you add the baseflow. I see it was necessary, but please elaborate why/how etc. Also – please link the the "\_sh"runs in fig 4 + 5.

Related with Comment 1.11.

### Answer:

In the Ammer catchment, especially in the mountain valleys, glacial overdeepening created considerable aquifers with depths of up to several hundreds of meters, Seiler (1979); Frank (1979). These storage bodies with long residence time withhold precipitation input and act as low pass filters, sustaining the long-term baseflow dynamics. With the conceptual bucket in WRF-Hydro, it is currently not possible to cover these long-term dynamics. It would require either overlaying multiple buckets or a two- or three-dimensional description of groundwater flow, like, e.g., in TerrSysMP (Shrestha et al., 2014). Since we cannot reproduce these long-term baseflow dynamics in WRF-Hydro, we consider baseflow shifting of the simulated hydrographs to evaluate them as if the baseflow deficiencies would not exist. However, the shifting is only applied after the calibration of WRF-H\_SA, thus it is purely analytical. The shift amounts are manually derived so that the recession parts of the hydrographs are well resembled. For the subcatchments of the Rott and Ach (Rt-RST, Ach-OBN, Ach-OBH) the baseflow generation is governed by more shallow, quickly reacting aquifer systems and therefore adding constant baseflow would be inappropriate.

We changed the text as follows to explain the motivation and method of baseflow addition in detail. We also linked Figures 4 & 5 at the beginning of the paragraph.

The hydrographs for the calibration and validation periods of the WRF-H\_SA runs are presented in Figures 4 and 5. The final parameter sets and goodness of fit measures are listed in Table 3. For all subcatchments, reasonable configurations could be determined. However, for the three Ammer subcatchments (OAG, PEI, WM) it was required to add constant baseflow rates of 2, 4.71, and 5.13 m<sup>3</sup> s<sup>-1</sup> to the simulated hydrographs, respectively. Adding constant baseflow is justified by the fact that due to the glacial processes of the last ice-age, large storage bodies, that dip reversely with the surface elevation, were formed in the mountain valleys by overdeepening (Seiler,

1979; Frank, 1979). Further downstream, towards the opening of the valley, where the aquicludes reach towards the surface, springs are abundant. Channel inflow from such long-term storage cannot be realized by WRF-Hydro's conceptual bucket scheme. It would require the implementation of a more sophisticated groundwater model. The amounts for constant baseflow are derived manually after the PEST parameter optimization, so that the recession curves of the simulations agree well with the observations. For the Ach and Rott subcatchments channel baseflow is related with shallow aquifers with shorter residence times which should be solely captured by the model.

#### Comment # 2.15

Table 3: For Am-PEI "coeff" and "expon" the values are larger than the suggest span in table 2. Table 3: For the two parameters mentioned above as well as zmax = 1 (3 instances) and retdp ) = 5, the autocalibration seems to have hit a boundary limit, which could imply model deficiencies. Please mention/elaborate or re-run, using more sound limits.

#### Answer:

First, we would like to point out that in Table 2 the typical parameter ranges are either defined in the NOAH-MP user manual (the LSM parameters) or derived from the authors' experience. In the WRF-Hydro documentation, no ranges are defined for the calibration parameters. The ranges given in Table 2 are not aimed to represent tight limits, which have been exceeded also in other cases than those reported by the referee (e.g., refkdt < 0.1). In order to avoid any misunderstanding about the meaning of these 'typical ranges', we would like to skip this information from Table 2 in the revised manuscript.

Concerning the cases where the referee suggests that the autocalibration could have reached a boundary limit, it should be preliminarily recalled that the LH-OAT method was used to obtain a starting configuration for the automated parameter optimization. Then, the PEST method considers some upper and lower boundary limits for the parameters subject to calibration. Several times we experienced that the parameters hit the boundaries that we set, however in all cases we relaxed the constraints, allowing the parameters to exceed the previous limits. In the cases we found negligible (or even null) improvements, we preferred to come back to the previous borders (this typically occurred with lower zmax values, close to 1).

P13L10: First sentence: Not understood. Limited - how?

## Answer:

We have reworded the sentence to make the point more clear. Now it reads:

The commonly favored lumped calibration of WRF-Hydro seems to be rather limited with respect to the transferability of parameter sets among subcatchments and also concerning the numerical efficiency for automated calibration. Especially for complex terrain, e.g., as presented by this study, the distribution of discharge gauges does not agree with landscape units.

# Comment # 2.17

Table 4-9: I think you should add MAE.

#### Answer:

We agree with the referee and propose to replace RMSE by MAE throughout the manuscript.

### Comment # 2.18

P20L6-13: Isn't that just the latent heat results again? They look the same. If not, how are they different and why?

P20L6-9: Unclear which scenario is most realistic? Why not just use the best? Better word than scenario?

### Answer:

The simulated latent heat and evapotranspiration time series are indeed identical. However the observations for latent heat are obtained from a flux tower and those for evapotranspiration are from lysimeters. For the revised manuscript we propose to combine Figures 8 and 11 by showing the simulation results, the flux tower and the extensive lysimeter management data. On the other hand we would skip the intensive lysimeter management scenario (currently shown in Figure 11) as management type is not reflected by the implementation in the land surface model. Subsection 3.1.4 would then only contain soil moisture.

P22L3-4: Please see my previous comment on table 3 about the resulting parameters. "... often tuned to unrealistic values..."

#### Answer:

The statement here is not related to our model calibration and parameter ranges. It is a more general statment, based on Koster et al. (2009) who found that the soil representations in land surface models is typically far away from reality.

### Comment # 2.20

P22L14-15: Why? Discuss/relate! This is a discussions section.

### Answer:

The reason for the statement was actually given in the sentence before. Because of the accuracy of the HATPRO measurement, the conclusion for temperature is clear whereas for humidity we cannot be likewise sure about the underestimation by the model. The term "extremely likely" is then probably the wrong formulation. We therefore changed the statement to:

For temperature, the differences between HATPRO and the models are generally much larger than the HATPRO accuracy. For humidity, the mean deviation lies within the accuracy of the measurement. It can be followed that both models overestimate temperature and probably have a tendency to underestimate absolute humidity. As compared to WRF\_SA, WRF-H\_FC shows reduced deviations for both variables.

### Comment # 2.21

P26L19: Please write "(NSE)" after Nash Sutcliffe efficiency.

Related with Comment 3.6

### Answer:

Due to artifacts in the RADOLAN precipitation data, we have re-computed the 2016 WRF-H\_SA simulation and therefore updated the respective figure and passages. We have also added NSE as you suggested. The details about the changes are outlined in the answer for Comment 3.6.

P28L8-9: Why was this the case?

Related with Comment 1.11 and Comment 2.14. **Answer:** 

The detailed explanation for the baseflow shifting and underestimated baseflow is given in the calibration subsection (2.4.6). For the summary we have updated the sentence about baseflow shifting:

For the Ammer subcatchments (OAG, PEI, WM), due to long-term hydrogeological storage processes that cannot be reproduced by WRF-Hydro, it was required to correct the negative biases of the baseflow.

# Comment # 2.23

P29L2: "For the validation period"  $\rightarrow$  has this been addressed previously?

### Answer:

Yes, the results for the calibration were discussed in subsection (2.4.6). As we have re-run WRF-H\_SA for the validation period due to artifacts in the RADOLAN precipitation data, the context for this passage has changed. For the respective answer, please refer to Comment 3.6 (and Comment 2.21).

### Comment # 2.24

P29L9-10: Why is this a problem (the 1h time steps)? If it corresponds to "SA" mode.

### Answer:

In its fully coupled configuration (WRF-H\_FC), the Noah-LSM would call the hydrological subroutines (routing) at every WRF model time-step which would be about every four seconds. With these short time-steps and the high spatial resolution numerical effects lead to incorrect results. Therefore, to get commensurable simulations with all three model variants, it is required to upscale the hydrological simulation part to one hour resolution, similar as used for the WRF-H\_SA runs. We tried to explain this better by rewording the sentence to: For the fully-coupled WRF-Hydro run (WRF-H\_FC), to obtain commensurable quantities for the evaluation, instead of calling the hydrological subgrid functions at every WRF model timestep (4 sec), an hourly timestep identical to that of the WRF-H\_SA model was required. The strategy was selected to avoid numerical truncation effects in the overland routing routines that happen when timesteps are in the order of a few seconds and spatial resolution is about 100 m or below (see also section 2.4.4).

#### Comment # 2.25

P29L24: Reference? On alpine foothills soil texture.

Related with Comment 1.15 and Comment 3.4.

#### Answer:

There are only a few references about soils of the northern limestone Alpine region and most of them are in German. However, we found two application studies for our study region, Ludwig et al. (2003); Hofmann et al. (2009). Moreover, a web service of the Bavarian Environmental Agency (LfU) is available at https://www.umweltatlas. bayern.de/mapapps/resources/apps/lfu\_boden\_ftz/index.html?lang=de&layers=service\_ boden\_5%C2%A2er%3D4418466%2C5268428%2C31468&lod=6&stateId=b3ed6873-359d-46d7-ad68-73359d56d749. We added the references as follows:

This does not even rudimentarily reflect the complexity of the northern limestone Alps and foothills (compare Hofmann et al., 2009, for the Halbammer subcatchment of the Ammer), in particular with the high resolutions of 1 km and 100 m of the atmospheric and hydrological model sections. The incorporation of high-resolution continental or global soil maps like, e.g., recently made available by Hengl et al. (2017) and Tóth et al. (2017) could lead to further improvement here.

### Comment # 2.26

P30L15: I do not like the use of "hopefully". If you had framed the relevance of the study in the introduction, then this paragraph would be more easily written.

Related with Comment 1.16.

### Answer:

We will try to find an improved formulation for this paragraph, along with reworking the introduction section.

Minor: Abstract: "Nominal"–use better wording? P12L1: where "the" model. Sounds better to my taste. P26L25-26: "difficulty in reproducing" instead.

# Answer:

All changed. The sentence in the abstract reads now:

The comparison of classic WRF and fully coupled WRF-Hydro, both using the calibrated parameters from the offline model, shows only tiny alterations for radiation and precipitation but considerable changes for moisture- and heat fluxes.

# **R**EFEREE # 3

With the development of fully-coupled atmospheric-hydrological model especially the improvement in the hydrological part, whether the improvement in land surface process (e.g., hydrological process) can lead to a better simulation of regional water and energy cycle is an issue that is concerned by the whole literature. Fersch et al., presents a very comprehensive research on this issue by comparing the coupled model (WRF-Hydro) with the original one (WRF\_SA). Generally, their experiment is well designed, the results are convincing and the research does have added value to deepen our understanding on the performance of fully-coupled atmospheric–hydrological modeling. I think it can be published in HESS after addressing the following comments.

### Comment # 3.1

Introduction of WRF-Hydro. From the introduction of the WRF-Hydro, it seems that WRF-Hydro improves the physical description of surface flow and baseflow transportation. But do the WRF-Hydro and WRF\_SA have the same runoff scheme? Will and how the surface/subsurface lateral flow feed back to the land (e.g., through re-infiltration)? I suggest the author give some introduction on the difference of basic WRF and WRF-Hydro as not all readers are familiar with this model.

Related with Comment 2.6.

### Answer:

We will add a paragraph to the section *Model setup and calibration* to make the differences between the different model setups clear.

### Comment # 3.2

Meaning of different parameters. The current interpretation of physical meaning of parameters is distributed. For example, in P12,L12 "The LSM surface runoff scaling parameter REFKDT is globally set to 2e-06 as smaller values would have decreased infiltration to very small amounts."P28,L12 "Also the percolation parameter SLOPE was mainly reduced as compared to the standard value, meaning that a considerable portion of former infiltration excess water needed to be transferred to the bucket-storage to assure good performance for the simulated baseflow". They should be directly introduced in the calibration section. Moreover, I suggest a brief interpretation of the physical meaning of changes of parameter.

Related with Comment 3.5 and Comment 1.9. **Answer:** 

We will follow the suggestion of the reviewers and reformulate the calibration methods section to give an overview of all relevant parameters and also explain the physical meanings of the different parameters in more detail.

## Comment # 3.3

What is the uncertainty of observations especially for the evapotranspiration? Does the improvement exceed the uncertainty?

### Answer:

There are two types of uncertainty that come into play here. The first one is connected with the precision and accuracy of the instruments. The second one relates to the scale (spatial support) of measurement and simulations. However, given the accuracies of the sensors and the long term averaging for the diurnal cycles and the performance measures we can rate the improvements to outrange uncertainty. And we see the improvements also for the correlation measures that are not related with systematic over- or underestimation. With respect to evapotranspiration and latent heat flux, both observations state an uncertainty of about 10% (Mauder et al., 2006, 2018). Thus, the improvements seen in the diurnal cycles are largely exceeding the observation uncertainties.

### Comment # 3.4

P10, L25: What do the "two dimensional" and "three dimensional" mean? Does the "soil layer depths" mean the "soil layer thickness"?

Related with Comment 1.15 and Comment 2.25.

# Answer:

Yes, the soil layer depth refers to the soil layer thicknesses. We have reformulated the paragraph, and use the term thickness instead of depth. Also, we removed the terms "two dimensional" and "three dimensional":

Large parts of the modeling domain and the considered river catchments exhibit mountainous terrain with steep slopes covered by shallow soil layers. Here, the model's general assumption of two meter soil thickness (or depth to bedrock) does not hold true as it may lead to overestimated retention of infiltrating water. Therefore, in this study, the soil layer definition was changed from a domain uniform to a grid point based representation and soil layer thicknesses were set to the Noah-MP standard (0.1, 0.3, 0.6, 1 m) distribution for hillslopes below 50 % and to more shallow values (0.05, 0.05, 0.1, 0.1 m) for all slopes >50 %. The 50 % threshold led to a realistic discriminability of valley bottoms and hillslopes.

### Comment # 3.5

P12, L12: The LSM sets the REFKDT as 0.2e-6, this value is not within the range given in the Table 2.3. section 2.4.6: Please give the timescale of calibration and validation.

Related with Comment 3.2.

### Answer:

The REFKDT parameter is erroneously mentioned here. Instead the parameter addressed in this sentence is REFDK which is the scaling parameter for the infiltration equation. We have set this to the documented standard value which is 2e<sup>-06</sup>.

# Comment # 3.6

P13,L4-8: If you do not include the abnormally high value, which NSE efficiency you can get? Will it comparable to that in calibration period.

# Related with Comment 2.21.

### Answer:

We investigated the issue with the radar artifacts in the RADOLAN-RW dataset in more detail and identified two periods with erroneous signals. Especially the first one, end of June 2016 had a strong impact, in particular on the upper Ammer catchment. We cross-checked with gauge observations in the region, and also the suspicious cake-piece structure of the precipitation fields led to the decision to set the amounts for these less than 6 hour periods to zero rainfall. Figure 3 shows now the re-run of WRF-H\_SA for 2016, with updated performance measures. The NSE and VE values improved for all locations, in particular for the Ammer (shifted NSEs for Am-OAG:  $0.41 \rightarrow 0.71$ , Am-PEI:  $0.24 \rightarrow 0.54$ , Am-WM:  $0.30 \rightarrow 0.53$ ). For 2015, we did not find any issues with the RADOLAN precipitation data. We replaced the figure in the manuscript with the new version. Accordingly, the following passages were updated:

The performance statistics for the validation period are comparable to those of the calibration period and in the cases of Ach-OBN, Am-OAG, and Rt-RST even improved. For Ach-OBH, as expected due to the disabling of the reservoir option, the buffering



Figure 3: Re-simulated hydrographs for the validation period, 2016-05-01–2016-10-31 with artifact corrected RADOLAN precipitation data.

effect of the lake cannot be reproduced by the model, thus leading to an overestimation of most of the peak values. However, with respect to the overall discharge in the Ammer catchment, these cut-off peaks are rather small. The performance for Am-WM is lower for both 2015 and 2016, as it aggregates the mismatches of all the upstream subcatchments.

### and

For the validation period (2016-05-01–2016-20-31), the skill measures could largely outperform those of the calibration period. However, some structural deficiencies like the underestimated baseflow in the mountainous parts or negative bias remain. It is concluded that some of the processes in the catchment cannot be depicted because of the model physics or the lumped parameter estimation approach.

# Comment # 3.7

P26,L8: What do the 50% and 100% mean? Which reference they refer to?

# Answer:

The percent values are with respect to the WRF-H\_FC simulations. The range between 50 and 100% is for the different subcatchment. The sentence was updated to make this clear. Surface runoff (infiltration excess) with WRF\_SA is 50% (for some of the subcatchments more than 100%) higher than with WRF-H\_FC and WRF-H\_SA.

#### Comment # 3.8

P26,L17: Just a suggestion. Will the results improve if you use daily streamflow, as the hourly streamflow is difficult to simulate?

### Answer:

We computed also the daily discharge goodness measures and found that the results do not improve for WRF-H\_FC (the fully coupled simulation).

### COMMENT # 3.9

P29,L12: Again, I do not know how does the lateral water transport saturate the soil as you do not introduce this in the model introduction section.

Related with Comment 1.14.

# Answer:

There are two mechanisms in the overland routing of WRF-Hydro that affect infiltration, surface storage and surface routing. Surface storage keeps the infiltration excess in the model, whereas in WRF\_SA this quantity is removed from the model and not available for subsequent time steps. Thus, in WRF-H\_SA and WRF-H\_FC infiltration excess can infiltrate over multiple time steps. The surface gradient routing further increases infiltration capacity by transporting infiltration excess to adjacent grid cells that may remain without infiltration otherwise. We will update the model description section to make this also clear to the readers.

### REFERENCES

- Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., and Gerlach, N.: Projekt RADOLAN – Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), Deutscher Wetterdienst, Hydrometeorologie, 5, 2004.
- Dickinson, R. E., Shaikh, M., Bryant, R., and Graumlich, L.: Interactive Canopies for a Climate Model, J. Climate, 11, 2823–2836, doi:10.1175/1520-0442(1998)011(2823: icfacm)2.0.co;2, 1998.
- Frank, H.: Glazial übertiefte Täler im Bereich des Isar-Loisach-Gletschers: Neue Erkenntnisse über Aufbau und Mächtigkeit des Quartärs in den alpinen Tälern, im Gebiet des Murnauer Schotters und im Tölzer Lobus (erste Mitteilung), E&G – Quaternary Science Journal, 29, doi:10.23689/fidgeo-1329, 1979.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., BlagotiÄ, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, 1–40, doi:10.1371/journal.pone.0169748, 2017.
- Hofmann, M., Engelhardt, S., Huwe, B., and Stumpp, C.: Regionalizing soil properties in a catchment of the Bavarian Alps, European Journal of Forest Research, 128, 597–608, doi:10.1007/s10342-008-0242-6, 2009.
- Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., and Puma, M. J.: On the Nature of Soil Moisture in Land Surface Models, J. Climate, 22, 4322–4335, doi: 10.1175/2009jcli2832.1, 2009.
- Ludwig, R., Taschner, S., and Mauser, W.: Modelling floods in the Ammer catchment: limitations and challenges with a coupled meteo-hydrological model approach, Hydrol. Earth Syst. Sci., 7, 833–847, doi:10.5194/hess-7-833-2003, 2003.
- Mauder, M., Liebethal, C., Göckede, M., Leps, J.-P., Beyrich, F., and Foken, T.: Processing and quality control of flux data during LITFASS-2003, Boundary-Layer Meteorology, 121, 67–88, doi:10.1007/s10546-006-9094-0, 2006.
- Mauder, M., Genzel, S., Fu, J., Kiese, R., Soltani, M., Steinbrecher, R., Zeeman, M., Banerjee, T., De Roo, F., and Kunstmann, H.: Evaluation of energy balance closure adjustment methods by independent evapotranspiration estimates from lysimeters and hydrological simulations, Hydrological Processes, 32, 39–50, doi:10.1002/hyp. 11397, 2018.

- Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116, D12 109, doi:10.1029/2010JD015139, 2011.
- Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., and Chen, F.: Simple water balance model for estimating runoff at different spatial and temporal scales, J. Geophys. Res., 101, 7461–7475, doi:10.1029/95JD02892, 1996.
- Seiler, K.-P.: Glazial übertiefte Talabschnitte in den Bayerischen Alpen: Ergebnisse glazialgeologischer, hydrologischer und geophysikalischer Untersuchungen, E&G – Quaternary Science Journal, 29, doi:10.23689/fidgeo-915, 1979.
- Shrestha, P., Sulis, M., Masbou, M., Kollet, S., and Simmer, C.: A Scale-Consistent Terrestrial Systems Modeling Platform Based on COSMO, CLM, and ParFlow, Monthly Weather Review, 142, 3466–3483, doi:10.1175/MWR-D-14-00029.1, 2014.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, M., Duda, K. G., Huang, X. Y., Wang, W., and Powers, J. G.: A description of the Advanced Research WRF Version 3, Tech. rep., National Center for Atmospheric Research, 2008.
- Tóth, B., Weynants, M., Pásztor, L., and Hengl, T.: 3D soil hydraulic database of Europe at 250 m resolution, Hydrological Processes, 31, 2662–2666, doi:10.1002/hyp.11203, 2017.
- Winterrath, T., Rosenow, W., and Weigl, E.: On the DWD quantitative precipitation analysis and nowcasting system for real-time application in German flood risk management, Weather Radar and Hydrology, IAHS Publ, 351, 323–329, 2012.