

Author Response to Reviews of:

Understanding terrestrial water storage variations in northern latitudes across scales

Tina Trautmann^{1,2}, Sujan Koirala¹, Nuno Carvalhais^{1,3}, Annette Eicker⁴, Manfred Fink⁵, Christoph Niemann^{1,5}, Martin Jung¹

Dear Prof. Jean-Christophe Calvet,

Dear Vincent Humphrey and Anonymous Referee #2,

we appreciate the positive feedback on our paper and thank the editor and the two reviewers for their constructive comments and suggestions. Please find our detailed response to all comments in the following sections. The corresponding changes in the revised manuscript and supplement are highlighted at the end of this document.

Apart from that, we ask for permission to include a few additional changes that are marked in the revised manuscript as well.

Please, don't hesitate to contact us, if further clarifications are needed.

Kind regards,

Tina Trautmann
on behalf of all co-authors

AUTHOR RESPONSE (RC – Referee Comment, AR – Author Response)

1. Referee #1 – Vincent Humphrey

RC: This study aims to evaluate the relative contribution of snow versus liquid water in (total) terrestrial water storage changes in northern latitudes. In order to investigate this question, the authors construct a simple bucket-type hydrological model with 10 free parameters which they calibrate against four different datasets (satellite observations of terrestrial water storage from GRACE satellites, snow water equivalent from the GlobSnow product which combines satellite and ground observations, evapotranspiration from FLUXCOM which is based on an ensemble of machine learning methods calibrated with in-situ observations, and E-RUN estimates of gridded runoff also based on a machine learning model calibrated with in-situ observations). Following a short evaluation of the performance of the presented model (and a comparison with the Earth2Observe ensemble of hydrological models), the main results are focused on distinguishing the respective contributions of snow and liquid storages to terrestrial water storage. Their analysis contrasts 1) local scale effects versus a spatially integrated average and 2) the mean seasonal cycle versus inter-annual variability. Consistent with previous studies, the authors find that the seasonal cycle is dominated by the snow component. The main finding of the paper is that liquid water storage clearly dominates inter-annual variability both at local scale and when considering a spatially integrated time series. They also find that the relative contribution of liquid water is weaker for the spatial integral compared to the local scale analysis. The authors argue that because snowpack evolution is primarily dependent on temperature (which has high spatial coherence (fig. 10)), this explains why the relative contribution of snowpack to large-scale inter-annual TWS variability is higher. In their conclusions, the authors comment on the usefulness of a simple hydrological modeling approach informed by multiple observational constraints. They suggest that long-term changes in water availability in northern latitudes might be driven by soil moisture rather than by snow dynamics.

This is a really good and well conducted paper. I find the results very interesting and worthy of publication in HESS. One can see that a lot of effort was invested in developing a custom hydrological model and this is reflected by the relatively important share of the methods and model evaluation sections in the paper. However, the authors manage to keep the results and discussion focused around the primary objective of quantifying the relative contribution of snow and non-snow storages to overall water storage variability.

I have four major comments/suggestions which I would like the authors to consider as well as some minor comments that are listed below.

AR: On behalf of all co-authors we very much thank Vincent Humphrey for very thoughtful and constructive comments. We appreciate the details and clarity in his remarks, and we have addressed all major and minor comments in the following. Further suggestions regarding terminology, clarity of formulations and figures were gratefully received and will be included in the revised manuscript.

Major Comments

1) PHASE LAG OF SEASONAL TWS

RC: In their results, the authors find that the modeled seasonal cycle of TWS has a systematic lag compared to observations (model TWS preceding observed TWS). This lag is also present in other models from the Earth2Observe ensemble. The analysis of the authors convincingly shows that their modeled snow storage seems to have a correct phase and is therefore not responsible for this lag between modeled and observed TWS. They also mention that adding delayed storage responses (as e.g. with a groundwater module) could not correct this effect either. I find this a major finding for the research community (which could be made more prominent in the conclusions) since it is often supposed that such model errors mainly stem from the lack of long memory water storages and poor representation of snow dynamics. Here the authors

conclude that neither of these seem responsible and that the origin of the phase lag in TWS must reside elsewhere, which brings me to my main suggestion below.

One important limitation that the authors fail to mention is that there is no consideration of permafrost and liquid/solid phase transitions of soil moisture content. In the proposed model, soil moisture does not have temperature neither does it store energy. In reality, it is well known that freeze/thaw dynamics are also a dominant factor for water and energy fluxes in high latitudes. Freeze/thaw is the on and off switch for evapotranspiration and vegetation growth. However, a phase lag between the availability of energy and the ET response cannot be modeled with the current model setup (the alpha parameter only conditions ET amplitude). Potentially, a lot of ground heat flux might be required before ET can actually take place.

In addition, from my understanding of the equations presented the supplementary material, actual ET is not reduced in the case of snow cover neither is it dependent on vegetation growth. This might introduce a too early response of ET to net radiation compared to reality, leading to a fast rise of soil moisture depletion already in early spring. Later, soil moisture would become limiting already in mid-summer and ET would peak in June and start to reduce already in July (Fig S1). The reference below suggests a peak of vegetation growth in August for a boreal forest (from one FLUXNET site).

The authors might consider exploring this direction and maybe check whether there is some evidence that FLUXCOM ET itself (the observational constraint) already contains such a phase lag. As this would require some additional work, it would also be fine if the authors prefer to simply mention this as a hypothesis to explore.

AR: We are grateful for the suggestions and a detailed explanation of the potential causes of the systematic lag in the modeled TWS.

- Biases in ET and its effect on TWS:

We have explored the relationship of potential biases in ET that may lead to a different timing in peaks of TWS. As already mentioned in the manuscript, we do not have ground heat flux and vegetation growth processes in the current model formulation, but we gratefully acknowledge it as an interesting opportunity for future investigations. In the current model, as the review correctly points out, actual ET is not reduced in the case of snow cover, which may lead to an early reduction of soil moisture and, consequently, TWS. To assess this effect, we scaled ET with the snow free fraction of a grid cell (1-FSC). Using the optimized parameter set presented in the manuscript together with the new scaling formulation of ET, there was a slight reduction of simulated ET in spring and a corresponding increase in July. This lead to a slight improvement of TWS timing (Fig. 1). On the other hand, when the model variant with snow cover scaling factor was optimized again, the marginal gain of performance was reduced (Fig. 2). This suggest that the observation data streams guide the model to optimal parameter values that would still result in the lag in the TWS. As a further check, we post-adjusted the simulated TWS with biases in ET simulation, representing the perfect ET simulation, but even that adjustment in the TWS was not enough to improve the lag in TWS.

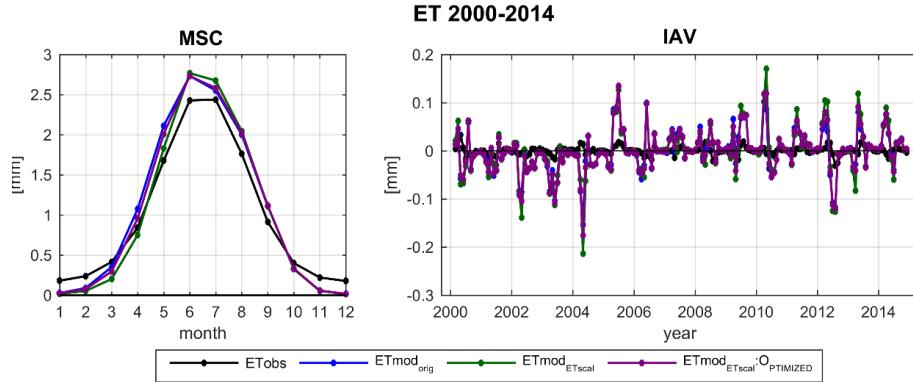


Figure 1. Comparison of the mean seasonal cycle and interannual variability of ETobs (FLUXCOM), ETmod_{orig} (as in the manuscript), ETmod_{ETscal} (parameter values as in the manuscript but scaling of actET with snow free fraction of each grid cell) and ETmod_{ETscalOPTIMIZED} (parameter values calibrated for scaling of actET with snow free fraction of each grid cell).

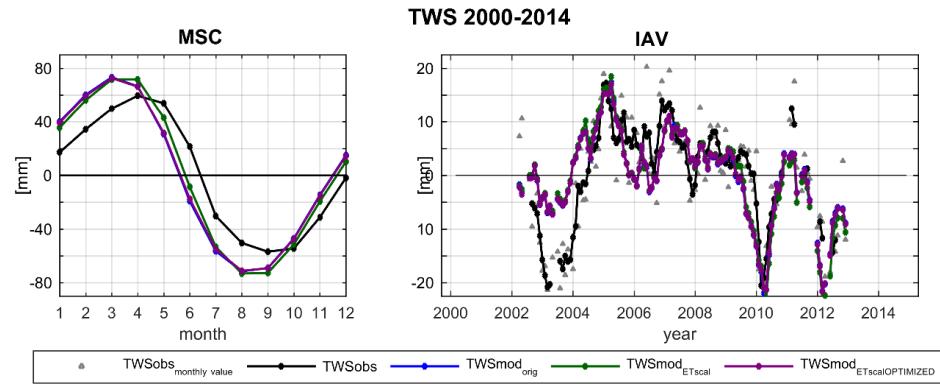


Figure 2. Comparison of the mean seasonal cycle and interannual variability of TWSobs (GRACE), TWSmod_{orig} (as in the manuscript), TWSmod_{ETscal} (parameter values as in the manuscript but scaling of actET with snow free fraction of each grid cell) and TWSmod_{ETscalOPTIMIZED} (parameter values calibrated for scaling of actET with snow free fraction of each grid cell).

- Permafrost and TWS variation:

As the reviewer points out, our model does not consider the permafrost dynamics. In order to identify the potential associations of the lag in TWS simulations against occurrences of permafrost, we compared the lag against permafrost fraction from the circum-Arctic map of permafrost and ground ice conditions (Brown et al., 1997). There is a tendency that the regions with the largest negative lag have a higher permafrost fraction (Fig. 3). This is especially visible in regions with sporadic permafrost (smf, slr, shr), as well as isolated patches of permafrost with high ground extent and thick overburden (ihf). One can expect that the sporadic permafrost is more active and may have larger influences in seasonal storage dynamics than more ‘permanent’ and larger permafrost. However, it should be noted that the ranges of permafrost fractions are large for both, short and long lags of TWS, suggesting a complex interaction between permafrost extent and its effect on lag in seasonal TWS dynamics as well as possible other factors related to the lag.

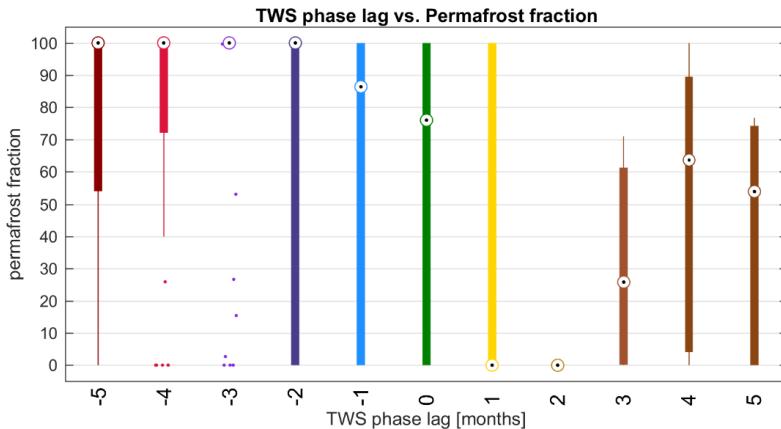


Figure 3. TWS phase lag compared to the permafrost fraction of the grid cell.

We have included the main findings of the above two analysis on potential relationships between the lag in TWS and biases in ET and the effects of permafrost and freeze/thaw dynamics in the revised manuscript. We have further highlighted the limitation that some potentially relevant processes are not yet accounted for in the current model setup.

2) MISLEADING TERMINOLOGY SNOW VS. LIQUID WATER

RC: *The separation of TWS into liquid and snow water seems a bit misleading since the liquid phase might implicitly include some frozen water as well (frozen soil moisture). As mentioned by the authors, there is a mismatch between explicitly represented processes and observed processes (TWS includes frozen water) that may be compensated by adjustments in model parameters. The expression “liquid phase” is hence misused in my opinion and might very well lead to confusion. It might be more accurate to refer to snow versus non-snow changes as done for example in page 27 line 30. I think this terminology should be extended to the rest of the manuscript.*

AR: The reviewer makes a valid point that the terminology might be misleading, especially with regards to observation. In reality, some part of TWS also includes solid or frozen water. However, in our study, the terminology of ‘snow’ vs ‘liquid water storages’ are used in the context of model simulation in which we do not account for frozen water storages. In order to avoid misunderstanding, we have elucidated that liquid water storages might implicitly include frozen water especially in the observation.

3) EFFECT OF PRECIPITATION FORCING

RC: *Figure S7 is quite pre-occupying because it suggests a dependency of your results on the forcing dataset. For instance, the difference might be related to your partitioning between snowfall and rainfall (which was not applied when using WFDEI). One possibility to check if this comes from uncertainty in the precipitation data might be to compare the regional mean time series of the two products and look for large differences in 2005 and 2010. This would also indicate whether GPCP-1DD appears superior to WFDEI. In relation to this -> Line 11-12 page 26: this is a rather unsubstantiated statement. Please give it more weight, for instance by replicating key figures (e.g. Fig 9) in the supplementary material.*

AR: We thank the reviewer for pointing to Fig. S7 (original supplement). Doing the analysis that he suggested, we found that in Fig. S7 of the submitted manuscript the shown time periods of TWS forced by WFDEI were shifted relative the observations and the modelled TWS based on GPCP forcing. The updated results (Fig. 4) are included in the revised

manuscript. The use of different precipitation forcing results in marginal difference in TWS simulations. TWSmod_{WFDEI} shows a larger seasonal amplitude because the amount of wintertime precipitation (snowfall and rain fall) is higher, while summertime precipitation is lower than estimated by GPCP (Fig. 5). However, the key findings of the dominant storage component remain the same (Fig. 6). We have included Fig. 6 in the supplement of the revised manuscript.

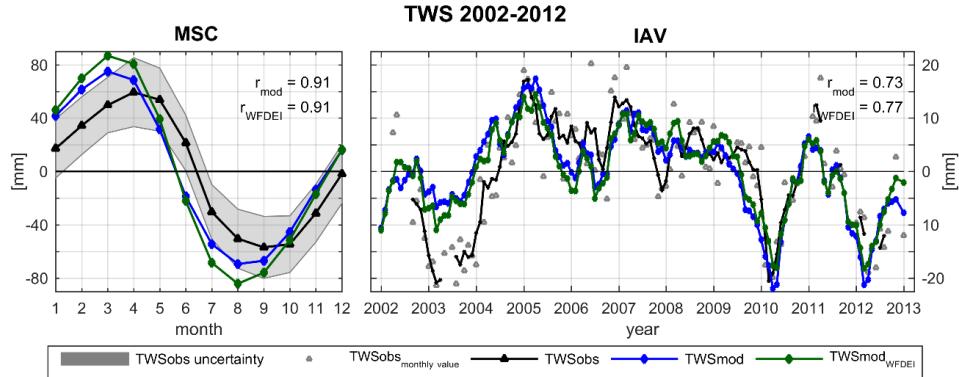


Figure 4. Comparison of the mean seasonal cycle and interannual variability of TWSobs (GRACE), TWSmod (forced with GPCP precipitation) and TWSmod_{WFDEI} (forced with WFDEI rain and snow fall).

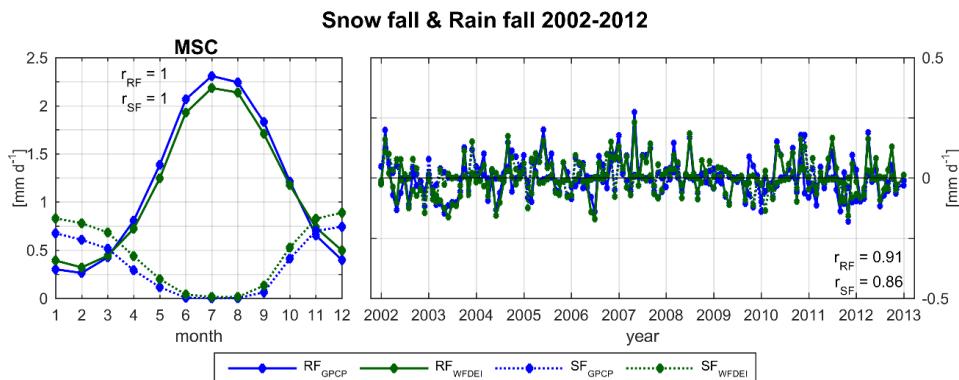


Figure 5. Comparison of the mean seasonal cycle and interannual variability of rain fall and snow fall from WFEI product and from GPCP (snow fall as in the optimized model: if $T < 0^\circ\text{C}$ and reduced to 67 % of original GPCP precipitation).

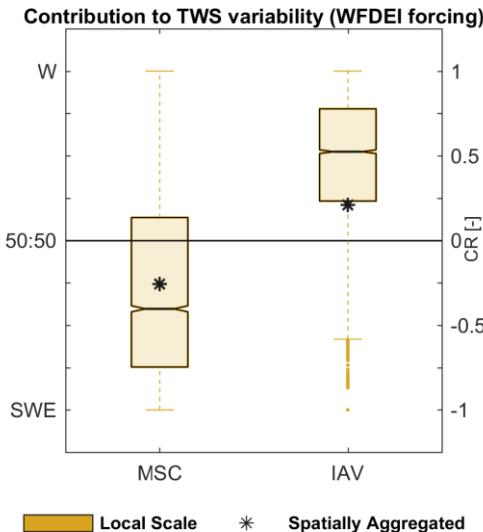


Figure 6. Relative contribution of snow (SWE) and liquid water (W) to TWS variability when forced with WFDEI snow and rainfall on different spatial (local grid scale, spatially integrated) and temporal (mean seasonal MSC, inter-annual IAV) scales based on CR (Eq.(3)). The boxplots represent the distribution of grid cell CR, with the dashed line marking the corresponding average. The star represents the CR calculated for the spatially integrated values.

4) CLARITY OF ABSTRACT

RC: You could make lines 24-29 of your abstract clearer. Upon first reading, I understood that snow dynamics dominate IAV on a large scale, which is not the case. It should be clearly said that “liquid water” dominates IAV at all spatial scales while snow dominates MSC at all spatial scales (Fig. 9). In addition, for IAV, the relative influence of snow increases with spatial aggregation due to the spatial coherence of T , a main driver of snowfall and snow melt. The wording “liquid water storages, comprising mainly of soil moisture” is also a bit misleading. It is not really clear what is implicitly incorporated in the soil moisture reservoir in order to fully reproduce TWS (as mentioned in the third comment).

Güntner et al. 2007 provides a similar analysis based on WaterGAP. This would be an interesting point of comparison since they indicate a contribution for IAV of 33% snow, 27% soil and 12% groundwater and 28% surface water (!) for cold climates (their table 5). I think this reference should be discussed and compared with your results.

AR: We thank the referee for highlighting this lack of clarity in the abstract. We have revised the abstract accordingly. Further, as the referee suggested, we have included a comparison with the results of Güntner et al. 2007 in the discussion of the revised manuscript.

Minor Comments

5) MODELLING FRAMEWORK

RC: Your work is very new and promising in the sense that multiple remote sensing or machine-learning observation-derived products are used simultaneously for calibrating a hydrological model. This is not easy to do and a research direction worth to explore. The overall modeling framework however still relies on a very standard land surface model structure. One missed opportunity may be to have used these observational datasets not only to calibrate model parameters, but also to identify functional relationships directly from the data (as opposed to fitting the parameters of a pre-defined equation to the data). Such research might be suggested as one possible future direction in the discussion. Finally, the paper does not emphasize on the added value of using remote sensing products to constrain the model (except for a lower RMSE

against observations, which is somewhat expected since other models were not calibrated with these observations). Could similar results have been obtained with the Earth2Observe ensemble? (especially on IAV?) If not, this would better show the merit and relevance of the presented approach.

AR: The referee is right in that our modelling framework still relies on a standard land surface model structure in terms of which processes are included. We yet want to highlight that this modelling framework still allows for more flexibility in the responses because we do not strictly constrain the model parameters that are often fixed in land surface models. We agree that identifying functional relationships directly from the observations represents new challenges in modelling the Earth system, especially when the modelling community shifts towards a hyper-resolution modeling, for which the classical formulation at coarse resolutions might not be valid. With the current availability and inconsistencies in the observational data, we could not address the challenge in the current study.

As pointed out, use of remote sensing data is advantageous in constraining the model over much larger spatial domains than using site-level or discharge measurements and thus improves the confidence in model results. The improved confidence, also reflected in the presented better performance metrics is also an important merit compared to the Earth2Observe ensemble. Despite, the lower performance metrics, the results of the Earth2Observe ensemble are in general similar to our study.

As the referee suggested, we have highlighted the merits of using remote sensing data, as well as potential future research in identifying functional relationships directly from observations.

6) CLARITY OF MODELLED LIQUID WATER AND MODELLED RUNOFF GENERATION

RC: *Liquid water is explicitly modeled as soil moisture + runoff routing but also likely includes river storage, lakes and wetlands implicitly (e.g. large water holding capacity mentioned on page 12, lines 7 and 19). This could also be made a bit clearer already in the model description in order to avoid some confusion later. Using a snow/non-snow terminology would also help resolving this.*

It could be made clearer that runoff is currently only generated from infiltration limitation (e.g. no baseflow in Eq. S10). Also mention that this is partially compensated by the recession time scale parameter that delays runoff generated on a specific day. Likely because the model is evaluated at monthly scale, this only has a limited impact on model performance and this model parameter is the least constrained by observations.

AR: Thanks for pointing this out, we have adjusted the model description in the revised manuscript.

7) DISTINCTION OF OBSERVATIONAL PRODUCTS

RC: *Methods: You could make a better distinction between purely observational products, and observation-based upscaled products such as Tramontana et al. or Gudmundson et al. which also rely on the quality of the underlying forcing data.*

AR: We see that such a distinction could help to underline the dependencies and uncertainties in the observational products. However, for each data product, information on its derivation is included in the original manuscript in the description of the input data. Besides, the line between purely observation based and upscaled isn't always that clear, as e.g. GlobSnow is based on a snow model, satellite data and site measurements, and the GRACE estimates rely on several models for data correction as well.

8) COVARIANCE OF SWE AND W

RC: *Line 29-30, page 10: this assumption seems a bit dangerous given figure 6. Could you please document the degree to which this assumption is correct and if this might affect the results qualitatively (possibly in supplementary information)?*

AR: We agree with the referee that the potential implication of this assumption should be discussed. We therefore have included a short discussion on the effect of the covariances between SWE and W on TWS variability in the supplement of the revised manuscript.

In-Text Comments

9) **RC:** *Line 29, page 2: and in addition, there can be no retrieval of SM in snow-covered or highly vegetated regions.*

AR: added.

10) **RC:** *Line 3, page 4: for clarity, maybe you could add an introductory sentence indicating that this whole section is meant to give an overview of the model setup.*

AR: added.

11) **RC:** *Line 10, page 4: E-RUN, based on E-OBS*

AR: changed.

12) **FLUXCOM ET**

RC: *Line 10-14, page 7: I thought FLUXCOM was based on an ensemble of machine learning algorithms (e.g. not only random forest). Could you also briefly comment on the performance of FLUXCOM in snow regions and high latitudes? Any idea if FLUXCOM is already accounting for sublimation?*

AR: The referee is right, FLUXCOM provides an ensemble of machine learning algorithms, but we only used the products from the random forest variant in this study. Even though FLUXCOM data have not been validated explicitly for snow-dominated regions, the cross validation of ET shows a good performance in most regions (Tramontana et al., 2016). In terms of sublimation processes, FLUXCOM conceptually includes sublimation processes as well, but the confidence in capturing such small fluxes is low due to lower signal to noise ratio in the underlying observations in FLUXNET sites. Therefore, we do not constrain modelled sublimation by FLUXCOM-based ET. We have clarified this in the revised manuscript.

Tramontana, G., Jung, M., Camps-Valls, G., Ichii, K., Raduly, B., Reichstein, M., Schwalm, C. R., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences Discussions, 1-33, 10.5194/bg-2015-661, 2016.

13) **RC:** *Page 9: Is there any reference for this cost function?*

AR: Unfortunately, there is no other reference for the cost function as it is applied in this paper.

14) **RC:** *Lines 17-20 page 9: I think this is indeed a very good idea!*

AR: Thanks.

15) **RC:** *Line 22, page 9: can you indicate where these commonly reported values can be found? (It's also fine if you decide to assume 10%).*

AR: We have assumed the 10 % and indicate this in the revised manuscript.

16) **RC:** Line 8, page 10: typo

AR: changed.

17) **REDUCTION OF SNOW FALL**

RC: Line 22-23 page 11: interestingly however, this also contradicts Behrangi *et al.* 2017 for mountainous regions...

Behrangi, A., Gardner, A. S., Reager, J. T., & Fisher, J. B. (2017). Using GRACE to constrain precipitation amount over cold mountainous basins. *Geophysical Research Letters*, 44(1), 219-227.

AR: Yes, Behrangi *et al.* 2017 found winter time precipitation is likely underestimated by popular precipitation products, including GPCP, compared to GRACE. However, their study focusses on two basins of the Tibetan Plateau, a region of which only the most northern parts with relatively less elevation and heterogenous topography are covered in our study domain. On contrary, Behrangi *et al.* (2016) and Swenson (2010), both showed an overestimation of precipitation by global precipitation products compared to purely observation based estimates in high-latitudes. Since high latitudes cover a larger fraction of the study domain, the reduction of snow fall by the calibrated global p_{sf} parameter rather reflects these regions instead of characteristics valid for the Tibetan Plateau.

Behrangi, A., Christensen, M., Richardson, M., Lebsack, M., Stephens, G., Huffman, G. J., Bolvin, D., Adler, R. F., Gardner, A., Lambrightsen, B., and Fetzer, E.: Status of high-latitude precipitation estimates from observations and reanalyses, *Journal of Geophysical Research: Atmospheres*, 121, 4468-4486, 10.1002/2015jd024546, 2016.

Swenson, S.: Assessing High-Latitude Winter Precipitation from Global Precipitation Analyses Using GRACE, *Journal of Hydrometeorology*, 11, 405-420, 10.1175/2009jhm1194.1, 2010

18) **RC:** Line 15 page 12: maybe this rather small value is in relation with the relatively large soil water holding capacity.

AR: The reviewer is correct, the parameter value of the soil water holding capacity influences the parameter for evapotranspiration and runoff generation.

19) **DOMINANCE OF IAV IN NA, E AURASIA**

RC: Line 12-13 page 15: For instance, Humphrey *et al.* 2016 figure 6 shows that the central North America and Eastern Eurasia is rather dominated by IAV (which appears more difficult to model according to your figure 5).

Humphrey, V., Gudmundsson, L., & Seneviratne, S. I. (2016). Assessing global water storage variability from GRACE: Trends, seasonal cycle, subseasonal anomalies and extremes. *Surveys in Geophysics*, 37(2), 357-395.

AR: We thank the reviewer for pointing to this reference. We have included the it in the revised manuscript.

20) **INACCURATE SENTENCE**

RC: Line 16 page 12: the sentence is inaccurate: a recession time scale of x days does not mean that only runoff of the preceeding x days contributes to “total runoff” (check Orth *et al.* 2013).

AR: Thank you very much for pointing this out! We have changed the sentence in the revised manuscript accordingly.

21) **GLOBAL UNIFORM PARAMETER VALUES**

RC: Line 13, page 13: maybe not necessary to say that these approaches are not commonly accepted as this might be a subjective statement in my opinion. The arguments you give just before (on overfitting) and the continental-scale focus of your study might be sufficient arguments. Another argument you could mention is that allowing locally varying parameters

would contaminate your conclusions: with locally dependent parameters, the differences in local-scale / large-scale contribution to IAV might due to the spatial dependency of parameters. But with your current setting, they can only be attributed to climate forcing. This is also why it makes a very clean experiment. This last point also calls for one caveat in the conclusion: your picture of the partitioning and scale-dependency of liquid versus snow might also change once you introduce spatial variability of the model parameters (e.g. snow melt factor might be very dependent on the vegetation cover, contrasting the responses of tundra versus boreal forests).

AR: The referee is right, this statement seems to transport a quite subjective opinion, yet it's based on Beck et al. (2016) 'Due to the lack of a commonly accepted approach for parameter regionalization, hydrologic models typically applied at continental to global scales (hereafter called macroscale) rarely use regionalized parameters [...]' . Therefore, we reformulated the sentence in the revised manuscript.

With his last comment, that the conclusions may change if we introduce spatial variability, the referee made a good point that is missing in our discussion. We have included this possible caveat when discussing the limitations of the approach.

Beck, H. E., Dijk, A. I. J. M. v., Roo, A. d., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resources Research, 52, 3599-3622, 10.1002/2015WR018247, 2016.

22) **RC:** Page 14, lines 3-6: essentially repeats page 13 line 10.

AR: We have removed the repeated lines.

23) **TRUNCATION OF VALUES IN FIGURES**

RC: Figure 3. If values were truncated (e.g. Fig3d) this should be indicated in the legend and in labels.

AR: In the legend of figures in the revised manuscript, we indicate if values were truncated.

24) **RC:** Line 1 page 19: TWSmod?

AR: clarified.

25) **RC:** Line 7 page 19: typo in earth2observe

AR: changed.

26) **RC:** Line 5 page 10: replace "grids" with "grid cells" idem on lines 5-6 page 20

AR: changed.

27) **RC:** Line 5 page 20: was the use of a subset mentioned also in the methods?

AR: Thanks for pointing to this. The use of a subset of grid cells for model calibration is mentioned in the methods of the revised manuscript now.

28) **RC:** Line 6 page 21: coincides

AR: changed.

29) **RC:** QUANTITAVE LABELS IN FIGURES SHOWING CR

Figure 7: It would be nice to add units to the colorbars (in addition to qualitative labels), same in Figure 8.

AR: We have added quantitative labels of CR in the revised manuscript.

30) **RC:** Line 18 page 22: is “received” the adequate word?

AR: changed.

31) **RC:** Line 7 page 23: frozen soil is not modelled

AR: removed.

32) MISUNDERSTANDING OF SPATIAL VARIABILITY

RC: Line 11 page 25: *On first read I could not follow since you cannot invoke geographic characteristics when you have spatially constant model parameters. The only source of spatial variability is in the model forcing. This is mentioned but only later on page 26 line 5, maybe you could reformulate this in a way that avoids such a misunderstanding.*

AR: Thanks for pointing to this possible misunderstanding. We intended the sentence to refer to general relations, not to the modelled variables. In order to avoid such a misunderstanding, we have clarified the sentence in the revised manuscript.

33) SPATIAL COHERENCE OF R_{net}

RC: Line 7-9 page 26: *note that ET is also influenced by R_{net} which might also be less spatially coherent.*

AR: The reviewer is right, R_n influences ET as well. However, as **Figure 7** suggests, inter-annual variations of net radiation in the study domain have a spatial coherence similar to temperature anomalies.

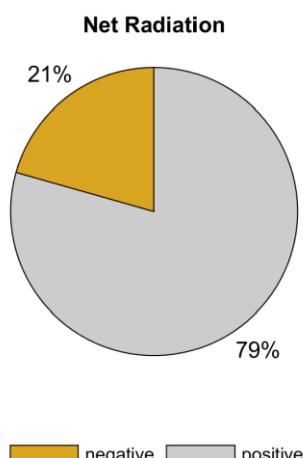


Figure 7. Proportion of total positive (grey) and negative (orange) covariances among grid cells for inter-annual variations of net radiation.

34) **RC:** Line 11-12 page 27: *and there can be no SM retrieval in snow-covered regions.*

AR: Thanks, we have added this fact in the revised manuscript.

35) **RC:** Line 15-18 page 27: solid/liquid phase transitions in soil moisture layers are another type of neglected effect relevant in the study domain as mentioned in the main comment.

AR: Thanks, we have added this fact in the revised manuscript.

36) **RC:** Line 10 page 28: the fact that snowpack anomalies are “erased” each summer and partially transferred to soil moisture through snow melt also largely explains this pattern. Hence, soil moisture also by construction allows for a longer memory than snowpack. This could be made more prominent in the discussion as well.

AR: Thanks for highlighting this, we have added it to the revised manuscript.

37) **RC:** Line 11-12 page 28: I would not qualify this as “diverging” since the sign is still the same (Figure 9). The non-snow storage only becomes less dominant when spatially integrated.

AR: changed.

38) **RC:** Line 23-25 page 28: adding this to the abstract would really explain better what you mean with the cryptic ending on line 32-34 page 1.

AR: The reviewer made a good point. We have added the essence of this lines to the abstract.

2. Anonymous Referee #2

RC: The main objective of this study is to analyse the spatial and temporal variability of snow pack and liquid water (mainly soil moisture) components at mid to high Northern latitudes and their respective contributions to total water storage (TWS) variations. To do so, a parsimonious hydrological model adapted to this purpose was first developed and calibrated using Earth Observations datasets, including TWS, snow water equivalent (SWE), evapotranspiration and gridded runoff. A comprehensive description of the model is provided in the Supplementary Material and a rather deep analysis of the calibration procedure is proposed. The authors also made a great effort in analyzing the performances of the model at different time scales (seasonal and interannual) and spatial scales (grid pixel and whole domain). Then the model is used over the 2000-2014 time period to evaluate the contribution of solid and liquid water components to TWS variations at these different spatio-temporal scales. Main conclusions are that TWS variations are mainly driven by snow dynamics at seasonal scales, while liquid water dominates TWS variations at interannual scales. Before concluding, the authors discuss some limitations of the method.

The paper is well written and organized. The conclusions are consistent with results presented all along the manuscript. The analysis of the calibration results (in terms of parameter values) is appreciable. Also appreciable is the comparison of the model to state-of-the-art global hydrological models from the earthH2Observe project, showing that despite its simplified structure, the current model performs reasonably well. I have only one major comment and some minor remarks and suggestions developed in the following.

AR: We very much thank the anonymous referee #2 for the helpful comments and suggestions on our paper. Please find the author's response in the following.

Major Comment

1) NEED TO DEVELOP A NEW MODEL

RC: *The need to develop a new model is not clearly stated, which is of prior importance since a large part of the paper is devoted to the presentation/validation of this model and the model outputs are used to draw the conclusions. Namely: - why not using existing models that show comparable performances and include more processes? -why not directly compare TWS and SWE from observations used here to calibrate the model? In that case, do the conclusions remain unchanged?*

AR: The reviewer made a good point that developing a new model should be better justified. As the reviewer suggested, existing models could have been used for our study in principle. We chose to implement our own version of a parsimonious model of water cycle processes which shares the common conceptualization of existing models and represents a recombination of established process formulations for conceptual and methodological reasons. Conceptually, simulations of simple models are advantageous with regards to interpretation and understanding of the responses. Furthermore, we think it is also useful to confront results of a simple model informed by observations with more complex and more ‘physically-based’ ones to elucidate the added value of increased model complexity or possibly to understand where the model requires more comprehensiveness. From a methodological point of view, the model-data fusion approach requires that the underlying model is parsimonious with respect to a) identifiability of model parameters, and b) computational tractability as thousands of simulations need to be performed during the model optimization. Unfortunately, to our understanding, both considerations are hardly achievable using most of the existing models. Additionally, the design is tailored by the globally available data and kept simple possible to provide the opportunity to identify the effect of the inclusion of the different data sets.

To address this comment in the manuscript we have adjusted the introduction in the revised manuscript.

Regarding the second question, a direct comparison of TWS and SWE from observations is an interesting suggestion that we considered thoroughly. There are three obstacles with respect to the suggested analysis: 1) the SWE data from GlobSnow suffer from a saturation effect above SWE values of about 100mm. This causes that these systematic errors in the snow data directly propagate to and corrupts the inferred ‘liquid’ storage component if the difference to GRACE-TWS is calculated. 2) There are frequent gaps in the SWE data which can be either due to the absence of snow or missing data. The suggested analysis would therefore be biased to respective grid cells and times without gaps and could not yield a representative picture. 3) Besides errors in GlobSnow SWE that propagate to the inferred ‘liquid’ water storages, errors and uncertainties of the GRACE TWS transfer to the ‘liquid’ water storage as well. We concluded that a joint interpretation of GRACE-TWS and GlobSnow SWE within an appropriate model-data fusion approach as done in this manuscript is preferred.

Nevertheless, we performed our analysis using GlobSnow SWE and GRACE TWS, as the referee suggested. We calculated liquid water as the difference between GRACE TWS and GlobSnow SWE and then compared the model results using the same data points as available from the observations (Fig. 8-10). On the interannual scale, we obtain similar conclusions when directly using the observations and when using the model. For the mean seasonal cycle, the main pattern persists as well, yet conclusions differ in some regions that likely suffer from saturation in GlobSnow SWE (e.g. Kamchatka) or in regions where permafrost and wetlands play a role (e.g. East Siberia). As the latter are not observed by GlobSnow, their contribution to observed TWS is included in calculated W. Additionally, the magnitude in GRACE TWS anomalies is in general much larger than the magnitude of GlobSnow SWE, and thus the magnitude in W based on these observations is larger than the magnitude in modelled W. Therefore, the relative contribution to TWS variability based on observations is shifted towards a larger effect of liquid water storages as compared to the modelled results (Fig. 8).

Contribution of SWEobs & Wobs to TWSobs MSC Contribution of SWEmod & Wmod to TWSmod MSC

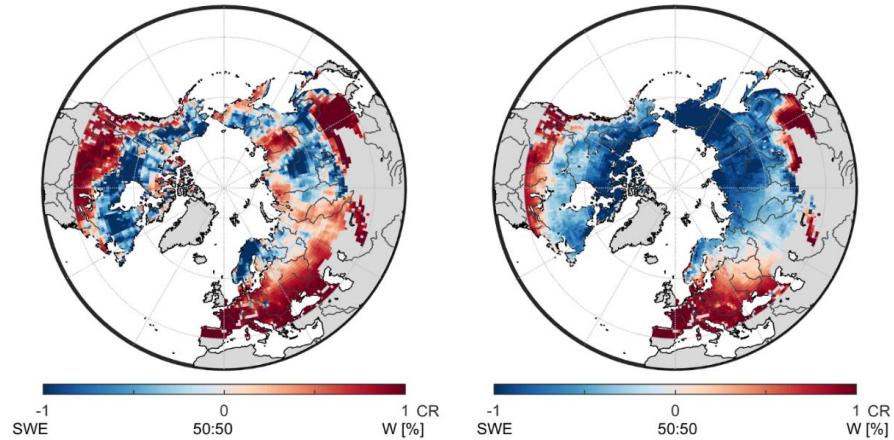


Figure 8. Relative contribution based on CR of snow (SWE) and liquid water (W) storage anomalies to mean seasonal TWS anomalies based on observations (left) and based on the model considering the same data points (right).

Contribution of SWEobs & Wobs to TWSobs IAV Contribution of SWEmod & Wmod to TWSmod IAV

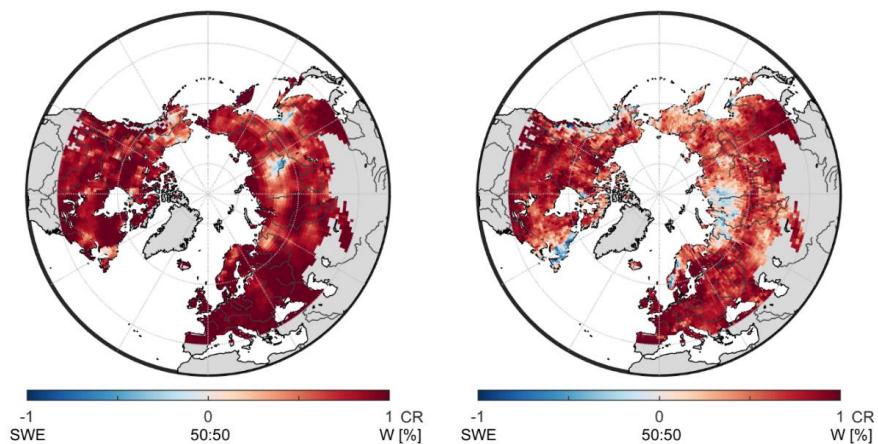


Figure 9. Relative contribution based on CR of snow (SWE) and liquid water (W) storage anomalies to interannual TWS anomalies based on observations (left) and based on the model considering the same data points (right).

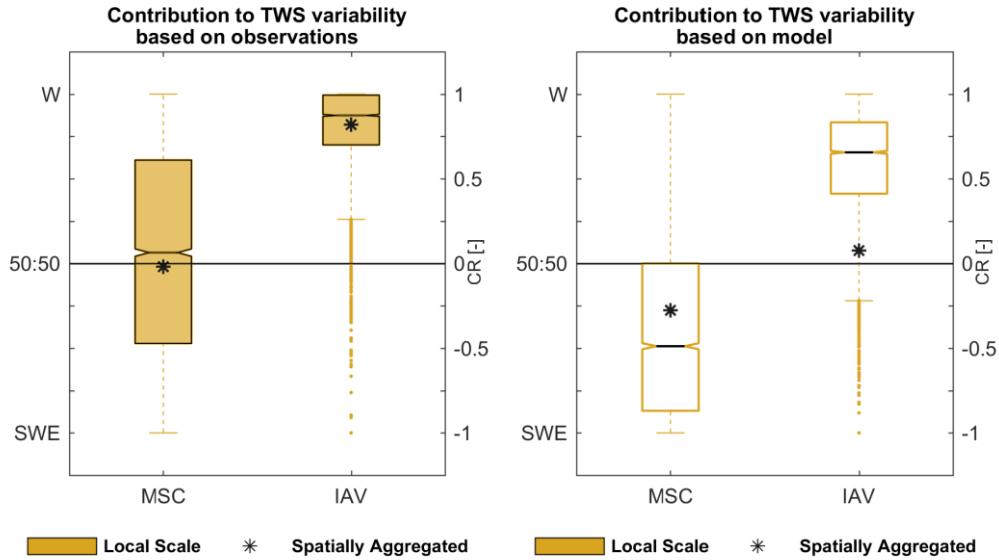


Figure 10. Relative contribution of snow (SWE) and liquid water (W) to TWS variability on different spatial and temporal scales based on observations (left) and based on the model considering the same data points (right).

Minor Comments

2) **RC: P1L20:** "...observed hydrological spatio-temporal patterns..."

AR: changed.

3) **USE OF SATELLITE DERIVED SOIL MOISTURE**

RC: P2L32: Some models explicitly simulate the upper soil layer using a multi-layer scheme (e.g., the ISBA land surface model, Decharme et al., 2011). In that case, satellite derived soil moisture can be compared to model outputs, and even assimilated with positive impacts on the model performances (Albergel et al., 2017).

AR: The authors are thankful for the reviewer's suggestion. We are considering the use of a multi-layer soil scheme with potential to assimilate satellite-derived soil moisture for future efforts, especially at the global scale, and mention this as outlook and potential improvement in the discussion of the revised manuscript.

4) **DELINEATION OF THE STUDY DOMAIN**

RC: P4L5: Which datasets are used to mask out such pixels?

AR: We thank the reviewer for pointing out that our manuscript was missing this critical information, which has been included the revised manuscript.

5) **ROUTING**

RC: 2.2 Model description: if I understand correctly, incoming water from upstream grid cells are not accounted for. At the monthly time scale, I agree that this would be negligible at the pixel scale, but is it still true at the basin scale (e.g., the Ob river basin)?

AR: As the reviewer pointed out, routing effects can be significant for large basins, especially in humid regions, which manifests in differences between surface runoff and river discharge at a given location. To address this, we do not use measured river discharge of large basins in our model-data fusion approach, but rather monthly runoff estimates for the European region at grid scale. Regarding the effect of river routing on TWS, e.g. Kim et al. (2009a) showed that the contribution of river storage to total TWS anomalies can be significant in the downstream regions of large basins. In northern high latitude catchments, this contribution is relatively smaller compared to continental tropical basins with large floodplains. Thus, we assume that although river storage is not explicitly represented in our model, the associated delay in surface runoff is sufficiently implicitly lumped into the delayed response of land runoff. Therefore, as the reviewer suggested, the effects of routing on the findings of this study can be expected to be small. We have clarified this in the revised manuscript.

Kim, H., Yeh, P. J. F., Oki, T., and Kanae, S.: Role of rivers in the seasonal variations of terrestrial water storage over global basins, Geophysical Research Letters, 36, doi:10.1029/2009GL039006, 2009.

6) **RC:** *P6L8: "...daily cumulated gridded precipitation..."*

AR: changed.

7) **RC:** *P6L11: "...that combines remotely-sensed precipitation..."*

AR: changed.

8) **GLOBSNOW DATA**

RC: *P7L2-4: Are these data [observed snow depth and radar data] assimilated into a snow model?*

AR: Yes, to our understanding the GlobSnow SWE processing applies a semi-empirical snow emission model and an assimilation scheme to produce maps of SWE estimates based on observations from passive microwave remote sensing and weather station observations (Luoju et al., 2010). We state this in the revised manuscript.

Luoju, K., Pulliainen, J., Takala, M., Lemmettyinen, J., Derksen, C., and Wang, L.: GlobSnow Snow Water Equivalent (SWE) Product Guide, ESA, 2010.

9) **MAPS OF TEMPORAL AVERAGE DATA UNCERTAINTIES**

RC: *2.3 Input Data: Since EO uncertainties are an important aspect of the calibration process, I suggest the authors to add a figure showing maps of temporal averages of each uncertainty for each dataset. This could help interpreting the model performances as shown in Figures 3, S1 and S2.*

AR: This is a helpful comment for making the manuscript comprehensive. We include the maps of the temporal averages of the uncertainty of observed TWS, ET and Q that are used for model calibration in the supplement of the revised manuscript. As we apply a constant average uncertainty of 35 mm (as mentioned in line 26 page 9, original manuscript), an additional map is not included in the supplement.

10) **RC:** *P10L8: "...Therefore..."*

AR: changed.

11) PARAGRAPH OF SPATIAL COHERENCE

RC: P11L4-9: This paragraph is unclear. Is it related to the smoothness of GRACE spatial patterns? In that sense, I think that for a better comparison with GRACE, modelled TWS should be first processed to remove high frequency spatial variability that is not observed by GRACE.

AR: We see that the methodological paragraph on compensatory effects and spatial coherence was not clear enough. It is not related to the smoothness of GRACE spatial patterns but is meant to provide background information for the analyses to explain the different importance of TWS components to the total TWS across different spatial scales (local grid scale vs. spatially aggregated). We have clarified the paragraph in the revised manuscript.

12) OVERESTIMATION OF GPCP

RC: P11L22-24: Is the overestimation found by Behrangi et al. (2016) and Swenson (2010) quantitatively comparable to this study?

AR: The reviewer pointed to an interesting comparison that was missing in the original manuscript. Due to the mismatch in the spatial and temporal domain, it yet is difficult to quantitatively compare the results reported in Behrangi et al. (2016) and in Swenson (2010) with this study in a precise manner. However, Behrangi et al. (2016) showed that average high-latitude annual precipitation of GPCP is 20 % higher compared to other precipitation products, and Swenson (2010) state that the GPCP undercatch correction is too large, resulting in too much cold season accumulation. This suggests that reducing GPCP snow fall in our study by 33 % seems quantitatively comparable. We have included this statement in the revised manuscript.

Behrangi, A., Christensen, M., Richardson, M., Lebsack, M., Stephens, G., Huffman, G. J., Bolvin, D., Adler, R. F., Gardner, A., Lambrightsen, B., and Fetzer, E.: Status of high-latitude precipitation estimates from observations and reanalyses, *Journal of Geophysical Research: Atmospheres*, 121, 4468-4486, 10.1002/2015jd024546, 2016.

Swenson, S.: Assessing High-Latitude Winter Precipitation from Global Precipitation Analyses Using GRACE, *Journal of Hydrometeorology*, 11, 405-420, 10.1175/2009jhm1194.1, 2010.

13) **RC:** P11L26: "...if SWE > 80 mm (parameter snc) ..."

AR: changed.

14) VALUES OF THE 4 COST TERMS

RC: 3.1 Model optimization: It would be interesting to discuss the values of the four cost terms in Eq. (2) obtained with the optimized parameters.

AR: We agree with the reviewer that the individual contributions to the total cost are a relevant methodological aspect. We have included a respective table in the supplement of the revised manuscript for completeness. To keep the manuscript concise, we don't add excessive discussion to this methodological detail, in particular since we present and discuss the evaluation of the model simulations against the individual data streams quite extensively in the manuscript.

15) CORRELATION FOR SEASONAL VARIATIONS

RC: P14L12: Are "seasonal variations" equal to the "mean seasonal cycle"? We understand after (from the figures) that yes. In this case, very high correlation values are not really surprising. Bias and RMSE would be more suited.

AR: The reviewer is correct, ‘seasonal variations’ are used synonymously to ‘mean seasonal cycle’. This has been clarified in the revised manuscript. We will also follow the suggestion of the reviewer and include RMSE metrics of ET and Q in the supplement of the revised manuscript.

16) REGIONS OF LARGE RMSE OF TWS

RC: *Figure 3: It seems from figure 3(d) that large RMSEs are found in regions affected by the Postglacial Rebound (Eastern Canada and Scandinavia) and near coastlines (ocean signal contamination?).*

AR: Yes, the reviewer is right, large RMSEs tend to occur in regions affected by Postglacial Rebound and near coastlines where the signal potentially is contaminated by the ocean. These limitations and errors of the GRACE TWS estimates are referred to in line 6-8 page 15 and are stated in relation to high RMSE in line 10-11 page 15 (original manuscript):

“[...] Second, although GRACE TWS passed through various pre-processing steps, the models to account e.g. for postglacial rebound or leakage between neighbouring grid cells introduce their own uncertainties and do not remove the effects completely. [...] This together is reflected in higher RMSE in arctic regions (e.g. surrounding the Hudson Bay), as well as in heterogeneous coastal and mountainous regions.”

17) NEGATIVE TWS ANOMALY IN 2003

RC: *P19L5: Do the authors have any possible explanation of the large negative anomaly in 2003 and why it is not captured by the model?*

AR: This is a very interesting point, which we investigated further. If we isolate interannual variations by removing the trends in GRACE and modelled TWS the agreement with respect to 2003 gets substantially better, as indicated by higher correlation scores (Fig. 11). This suggests that the trend in GRACE TWS is to some extent either subject to observational issues or represents a process that is not captured by our and the earthH2Observe models, which don’t reproduce the 2003 anomaly adequately either (see Fig. S6, revised supplement). In addition, there is a negative SWE anomaly of on average 5 mm (see Fig. 4a, manuscript) indicated in the GlobSnow data, that is not captured by our model, suggesting an issue with the precipitation forcing data. This not captured SWE anomaly appears to explain the remaining difference to the GRACE TWS anomaly in 2003 after detrending. The reason why this snow anomaly is not captured by the forcing remains unclear at this point – it persists when using the WFDEI forcing data set (Fig. 12). While the model reproduces the spatial pattern of the 2003 interannual TWS variability, the magnitude of observed TWS, especially in North America, is not captured by the forcing and thus by the model, either (Fig. 13).

We have added a paragraph to the revised manuscript on the discrepancy regarding the 2003 anomaly.

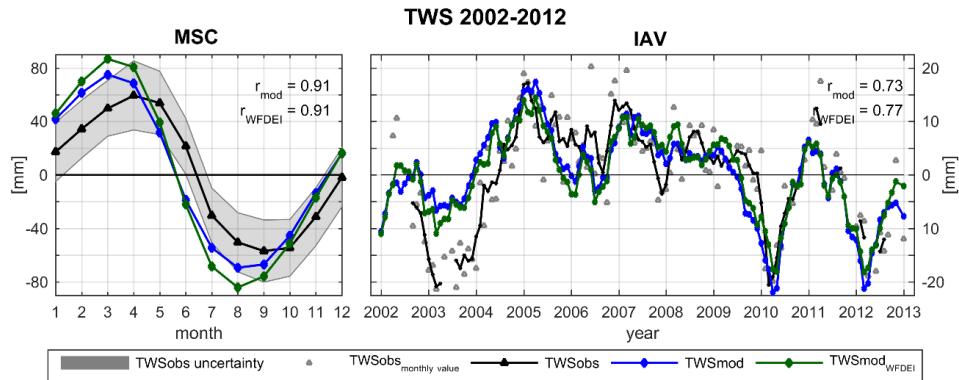


Figure 11. Comparison of the mean seasonal cycle and interannual variability of TWSobs (GRACE), TWSmod (forced with GPCP precipitation) and TWSmodWFDEI (forced with WFDEI rain and snow fall).

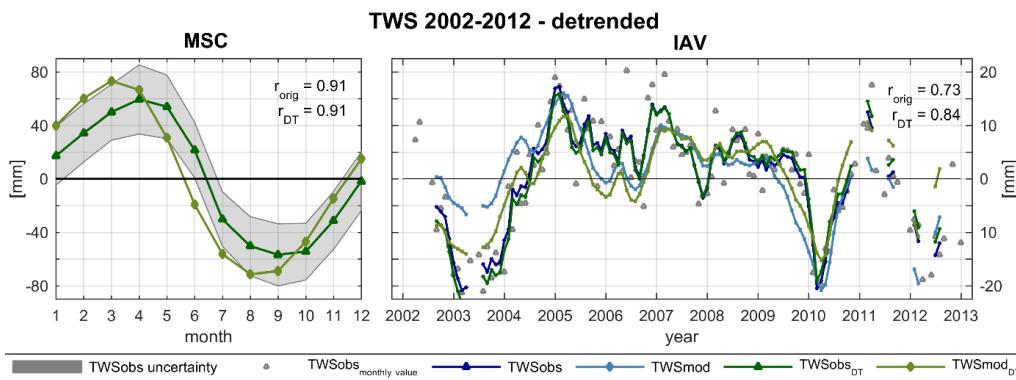


Figure 12. Mean seasonal cycle and interannual variability of original (orig) and detrended (DT) TWSobs (GRACE) and TWSmod (forced with GPCP precipitation).

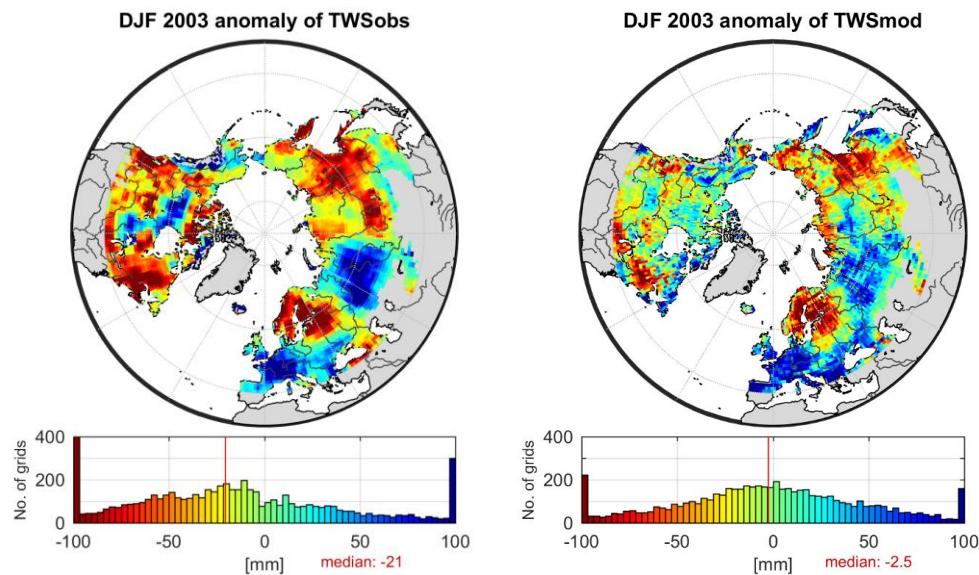


Figure 13. Average IAV of winter 2002/2003 (December, January, February) of observed and modelled TWS.

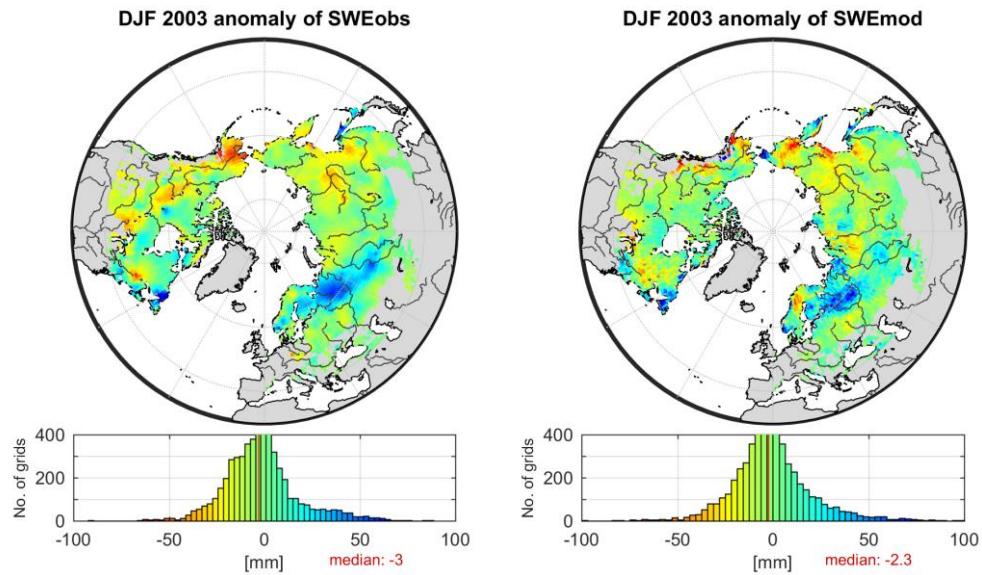


Figure 14. Average IAV of winter 2002/2003 (December, January, February) of observed and modelled SWE.

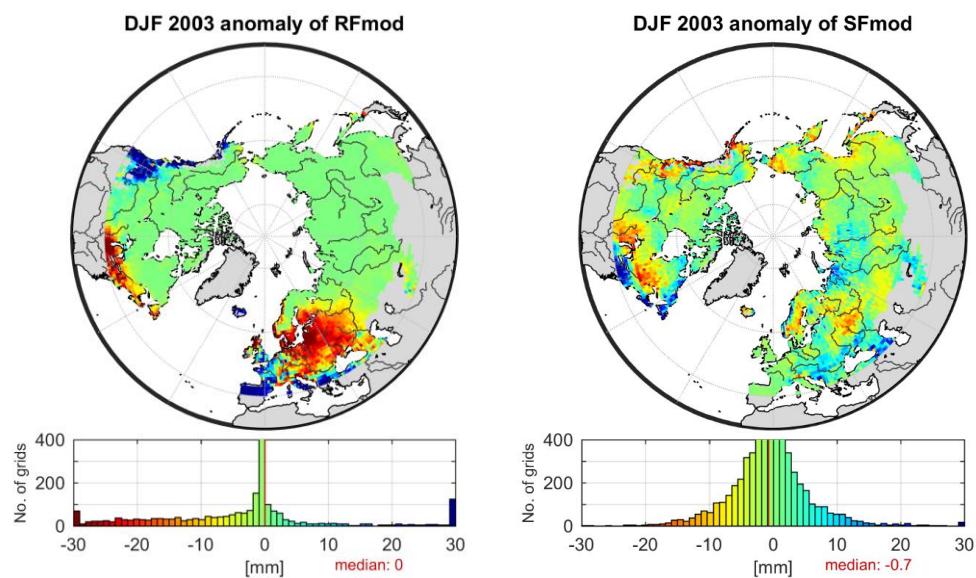


Figure 15. Average IAV of winter 2002/2003 (December, January, February) of modelled rain fall and snow fall (based on GPCP forcing).

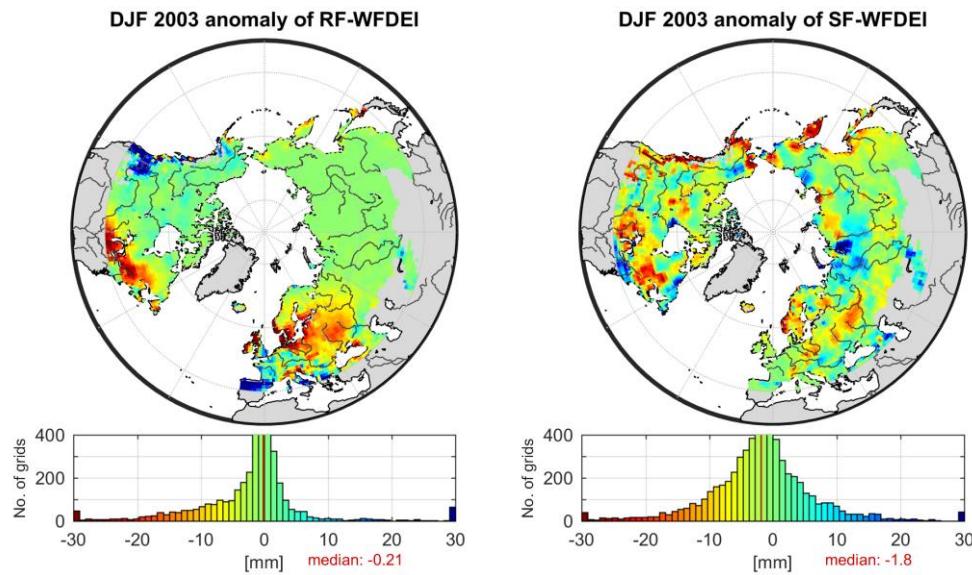


Figure 16. Average IAV of winter 2002/2003 (December, January, February) of WFDEI rain fall and snow fall data.

18) **RC: P23L4:** *The average value does not show that CR is positive over the entire domain.*

AR: We agree with the reviewer here and add quantitative labels of CR to the respective figures in the revised manuscript.

19) **RC: P23L14:** *“...less variable at interannual time scale...”*

AR: changed

Understanding terrestrial water storage variations in northern latitudes across scales

Tina Trautmann^{1,2}, Sujan Koirala¹, Nuno Carvalhais^{1,3}, Annette Eicker⁴, Manfred Fink⁵, Christoph Niemann^{1,5}, Martin Jung¹

¹Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, 07745, Germany

²International Max Planck Research School for Global Biogeochemical Cycles, Jena, 07745, Germany

³CENSE, Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, 2829-516, Portugal

⁴HafenCity University, Hamburg, 20457, Germany

⁵Department of Geography, Friedrich-Schiller University, Jena, 07743, Germany

Correspondence to: Tina Trautmann (ttraut@bgc-jena.mpg.de)

15 **Abstract.** The GRACE satellites provide signals of total terrestrial water storage (TWS) variations over large spatial domains at seasonal to inter-annual time scales. While the GRACE data have been extensively and successfully used to assess spatio-temporal changes in TWS, little effort has been made to quantify the relative contributions of snow pack, soil moisture and other components to the integrated TWS signal across northern latitudes, which is essential to gain a better insight into the underlying hydrological processes. Therefore, this study aims to assess which storage component dominates

20 the spatio-temporal patterns of TWS variations in the humid regions of northern mid-to-high latitudes.

To do so, we constrained a rather parsimonious hydrological model with multiple state-of-the-art Earth observation products including GRACE TWS anomalies, estimates of snow water equivalent, evapotranspiration fluxes, and gridded runoff estimates. The optimized model demonstrates good agreement with observed hydrological **spatio-temporal** patterns, and was

Kommentiert [TT1]: referee #2
/ comment 2

25 to total TWS variations. In particular, we analysed whether the same storage component dominates TWS variations at seasonal and inter-annual temporal scales, and whether the dominating component is consistent across small to large spatial scales.

Consistent with previous studies, we show that snow dynamics control seasonal TWS variations across **all** spatial scales in the northern mid-to-high latitudes. In contrast, we find that inter-annual variations of TWS are dominated by liquid water

30 **storages at all spatial scales. The relative contribution of snow to inter-annual TWS variations, though, increases when the spatial domain over which the storages are averaged becomes larger. This is due to a stronger spatial coherence of snow dynamics, that are mainly driven by temperature, as opposed to spatially more heterogeneous liquid water anomalies, that cancel out when averaged over a larger spatial domain.** comprising mainly of soil moisture. However, as the spatial domain over which the storages are averaged becomes larger, the relative contribution of snow to inter-annual TWS variations increases. This is due to a stronger spatial coherence of snow anomalies as opposed to spatially more heterogeneous liquid water anomalies that cancel out over large spatial domains.

Kommentiert [TT2]: referee #1
/ comment 4

35 The findings first highlight the effectiveness of our model-data fusion approach that jointly interprets multiple Earth observation data streams with a simple model. Secondly, they reveal that the determinants of TWS variations in snow-affected northern latitudes are scale dependent. **In particular, they seem to be not merely driven by snow variability, but rather are determined by liquid water storages on inter-annual time scales.** We conclude that inferred driving mechanisms of

Kommentiert [TT3]: referee #1
/ comment 38

TWS cannot simply be transferred from one scale to another, which is of particular relevance for understanding the short and long-term variability of water resources.

1 Introduction

Since the start of the mission in 2002, measurements from the Gravity Recovery and Climate Experiment (GRACE) provide 5 unprecedent estimates of changes in the terrestrial water storage (TWS) across large spatial domains (Tapley et al., 2004;Wahr et al., 2004). Due to its global coverage and independence from surface conditions, the data represents a unique opportunity to quantify spatio-temporal variations of the Earth's water resources (Alkama et al., 2010;Werth et al., 2009). Therefore, GRACE data has widely been used to diagnose patterns of hydrological variability (Seo et al., 2010a;Rodell et al., 2009;Ramillien et al., 2006;Feng et al., 2013), to validate and improve model simulations (Doll et al., 2014;Güntner, 10 2008;Werth and Güntner, 2010;Chen et al., 2017;Eicker et al., 2014;Giroto et al., 2016;Schellekens et al., 2017), and to enhance our understanding of the water cycle on regional to global scales (Syed et al., 2009;Felfelani et al., 2017).

Despite the high potential of GRACE data for hydrological applications (Döll et al., 2015;Werth et al., 2009), the measured signal vertically integrates over all water storages on and within the land surface, which challenges the interpretation of the driving mechanism behind TWS variations. To facilitate insight into the underlying processes, hydrological models are 15 frequently used to separate the measured TWS into its different components such as groundwater, soil moisture, and snow pack (Felfelani et al., 2017). However, as a consequence of uncertain model structure, forcing and parametrization, model-based partitioning is ambiguous (Güntner, 2008), and may lead to diverging conclusions especially on regional scale (Long et al., 2015;Schellekens et al., 2017).

While the uncertainties of catchment-scale hydrological models are commonly reduced by calibrating the model parameters 20 against discharge measurements, the majority of macro-scale models relies on a priori parametrization. So far, only few models used to assess hydrological processes on continental to global scales are constrained by observations, and if so, they are mainly calibrated against observed discharge of large river basins (Long et al., 2015;Döll et al., 2015). Recently, several studies showed the benefits of additionally including GRACE TWS data in model calibration (Werth and Güntner, 2010;Xie et al., 2012;Chen et al., 2017) or by means of data assimilation (Eicker et al., 2014;Forman et al., 2012;Kumar et al., 2016). 25 However, although these approaches improve model simulations, they do not reduce the uncertainty in partitioning of TWS due to the parameter equifinality problem (Güntner, 2008). Therefore, it is desirable to include multiple observations, ideally of several hydrological storages and fluxes, to constrain model results (Syed et al., 2009).

Nowadays, the increasing number and quality of Earth Observation based products provides valuable information on a variety of hydrological variables over large scales, and thus facilitates constraining model simulations with multiple data 30 streams simultaneously. While this can provide a more robust understanding of how variations of water storages translate into the observed TWS (Werth and Güntner, 2010), it is very challenging in practice and has rarely been implemented.

On the one hand, this is due to the limitations and inherent uncertainties of each Earth Observation based product that need to be considered when comparing simulations and observations. For example, satellite-based soil moisture retrievals only capture the upper 5 cm of soil [under snow-free conditions](#) and therefore are difficult to compare to modelled soil water 35 (Lettenmaier et al., 2015), while large scale observations of snow mass based on passive microwave sensors are known to suffer from uncertainties in deep and wet snow conditions (Niu et al., 2007) and multispectral sensors solely provide estimates of snow cover in the absence of clouds (Lettenmaier et al., 2015).

Besides, the application of multi-criteria calibration approaches is limited by the increasing complexity of most macro-scale hydrological models over time (Döll et al., 2015). This high model complexity is not only associated with conceptual issues 40 related to overparameterization (Jakeman and Hornberger, 1993) and large computational demand, but also has shown to not

Kommentiert [TT4]: referee #1
/ comment 9

necessarily improve model performance (Orth et al., 2015). Therefore, it is desirable to implement a rather parsimonious model structure (Sorooshian et al., 1993), especially in multi-criteria model-data fusion approaches.

Applying multiple observational constraints is in particular beneficial in regions, where hydrological dynamics are poorly understood and thus their representation in models varies widely. This is the case for snow-dominated regions as the 5 northern high-latitudes (Schellekens et al., 2017), which are among the areas most prone to the impacts of climate change (Tallaksen et al., 2015). These regions have been experiencing the strongest surface warming over the last century globally (IPCC, 2014), a trend which is expected to exacerbate in the future and to significantly change hydrological patterns (AMAP, 2017). Therefore, solid understanding of present hydrological processes and variations is crucial, yet the effect of complex snow dynamics on other storages and water resources is relatively unknown (van den Hurk et al., 2016;Kug et al., 10 2015). While it has been shown that snow mass is the primary component of seasonal variations of TWS in large northern basins (Niu et al., 2007;Rangelova et al., 2007), it is not known what drives the TWS variations on inter-annual or longer time scales in these regions. Moreover, most analysis so far focus on individual river basins and do not provide a comprehensive picture over large spatial scale.

15 In this study, we therefore aim to investigate the contributions of snow compared to other (liquid) water reservoirs to spatio-temporal variations of TWS in the northern mid-to-high latitudes. To do so, we establish a model-data-fusion approach that integrates multiple Earth Observation-observation based data streams including GRACE TWS along with estimates of snow water equivalent (SWE), evapotranspiration and runoff into a rather simple hydrological model. This model is designed as a combination of standard model formulations yet aims to maintain low complexity in order to facilitate multi-criteria calibration and to focus on variables that can be constrained by observations.

20 First, we explain the applied methods including the implemented model, the used data, and the multi-criteria calibration approach. The following section presents and discusses the results obtained with the optimized model. In the results, we describe the calibrated model parameters and evaluate the model performance with respect to observed patterns of TWS and SWE. Subsequently, the relative contributions of snow and liquid water storages to TWS variations are assessed on seasonal 25 and inter-annual scales. Thereby we first focus on spatially integrated values across the study domain, and secondly on the composition on local grid scale. Finally, we summarize our findings and draw the conclusions.

Kommentiert [TT5]: referee #2
/ comment 1

2 Data and methods

The following section provides an overview on the experimental set up, followed by a more detailed description of the model, the input data as well as the methods for model calibration and analysis.

Kommentiert [TT6]: referee #1
/ comment 10

30 2.1 Experiment design

To assess the composition of TWS variations in northern mid-to-high latitudes, we optimized a simple hydrological model on daily time steps at a $1^\circ \times 1^\circ$ latitude/longitude resolution. We defined the area of interest as humid land surface north of 40° N, excluding Greenland as well as grids with > 90 % permanent snow cover, and > 50 % water fraction. Humid areas are derived based on an aridity index AI ≥ 0.65 , which was calculated as the ratio of precipitation and potential evapotranspiration (United Nations Environment, 1992). Therefore, we used the same precipitation and potential evapotranspiration data as for model forcing (see 2.3). To mask out grids with > 90 % permanent snow cover and > 50 % water fraction, we applied the SYNMAP land cover classification (Jung et al., 2006). This dataset has an original resolution of 1 km and was used to determine the fraction of land cover classes within each $1^\circ \times 1^\circ$ grid cell.

35 Forcings with global observation-based climate data, the model parameters were constrained for a subset of the study domain 40 by multiple Earth observation data products using a multi-criteria calibration approach. These products include terrestrial

Kommentiert [TT7]: referee #2
/ comment 4

Kommentiert [TT8]: referee #1 / comment 27

water storage anomalies as seen by the GRACE satellites (Watkins et al., 2015; Wiese, 2015), measurements of snow water equivalent obtained in the GlobSnow project (Luoju et al., 2014), evapotranspiration fluxes based on FLUXCOM (Tramontana et al., 2016) and runoff estimates for Europe from E-RUN based on E-OBSU Grid (Gudmundsson and Seneviratne, 2016). Once the model parameters were calibrated, we evaluated the model against the same data, taking into account the entire study domain, and finally we applied the calibrated model to quantify the contributions of snow and liquid water storages to the integrated TWS. Thereby we considered different spatial domains (local grid cell and spatially aggregated) and temporal scales (mean seasonal and inter-annual variations).

Due to the differences in the temporal coverage of the observational data streams, model calibration and evaluation were conducted for the period 2002–2012, while analysis of TWS components cover the whole period of 2000–2014.

10 An overview on the experiment design and the selected time periods is provided by Fig. 1, while the following sections give a detailed description of the individual steps.

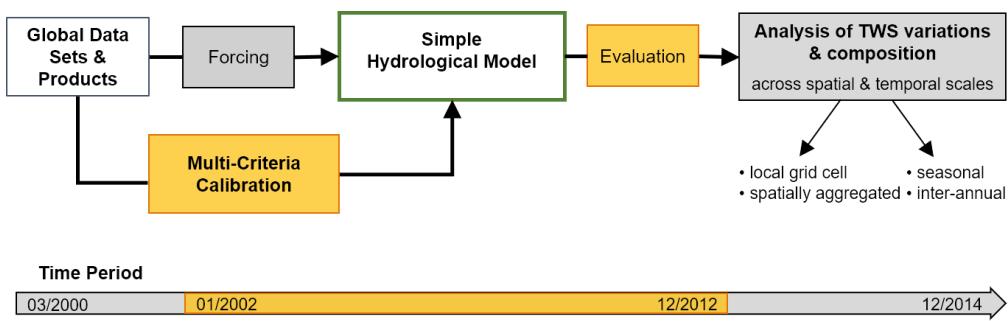


Figure 1. Experiment design and considered time periods for forcing/analysis (grey) and model calibration/evaluation (orange).

15

2.2 Model description

We designed a conceptual hydrological model with low complexity and a total number of 10 adjustable parameters. The model considers major hydrological fluxes as snow melt, sublimation, infiltration, evapotranspiration, and (delayed) runoff, and includes water storages in the snow pack, the soil, and due to delay in runoff (Fig. 2). It is forced by precipitation (P), air 20 temperature (T) and net radiation (Rn) and calculates all hydrological processes on daily time steps for individual grid cells. A simple schematic diagram of the model is shown in Fig. 2, while a detailed description of modelled processes is provided in S1.

25 In the first step, precipitation P is partitioned into liquid precipitation (rain fall) and snow fall based on a temperature threshold of 0 °C. Accumulating snow fall increases the snow pack represented by the snow water equivalent SWE [mm], which depletes by sublimation and melt if T exceeds 0 °C. We calculate sublimation based on the GLEAM model (Miralles et al., 2011c), and apply an extended day-degree approach to estimate snow melt (Kustas et al., 1994). Since the presence of snow can be highly variable in one grid cell, we model the fractional snow cover [-] following Balsamo et al. (2009) which is used to scale snow melt and sublimation.

30 Similar to the WaterGAP model (Döll et al., 2002), incoming water from rain and snow melt is allocated to soil moisture (SM) and land runoff (Q_s) depending on soil moisture conditions (Bergström, 1991). SM is represented by a one-layer bucket storage that depletes by evapotranspiration (ET). We calculate ET as the minimum of demand-limited potential ET following the Priestley-Taylor formula (Priestley and Taylor, 1972b) and supply-limited ET following Teuling et al. (2006).

Kommentiert [TT9]: referee #1 / comment 11

Kommentiert [TT10]: referee #1 / comment 27

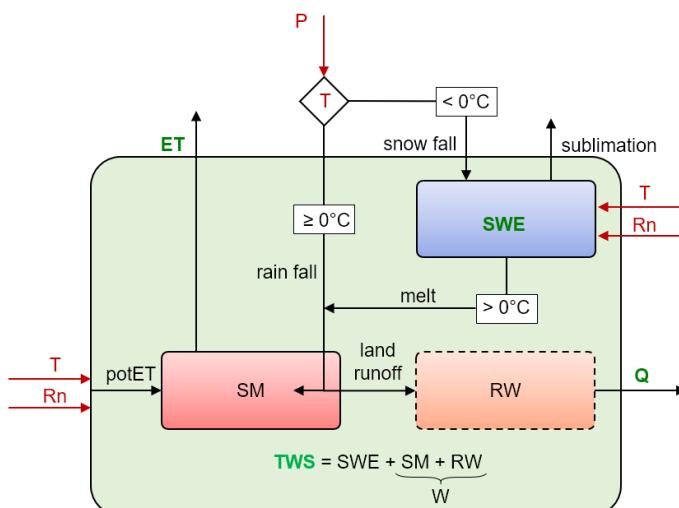
As land runoff results from an effective soil water recharge formulation, the calculated runoff is essentially all the water that cannot be stored in the soil. Thus, it implicitly contains both, surface and subsurface runoff as well as the percolation to deeper water storages such as groundwater, as well as contributions from surface water bodies. To account for runoff contributions from slow-varying storages, we calculate runoff from each grid cell (Q) by applying an exponential delay

5 function on Q_s (Orth et al., 2013). Based on mass balance, we derive the amount of retained land runoff (RW), which implicitly accounts for the effects of several water pools that are not explicitly represented in the model (groundwater, lakes, wetlands and the river storage). The sum of RW and SM is then taken as the total liquid water storage (W). To mimic runoff contributions from slowly varying reservoirs that are not explicitly represented in the model such as ground water and river storage, we use an exponential delay function (Orth et al., 2013). Frozen soil water is not explicitly included in the model.

10 The amount of retained land runoff (RW) together with SM represents the total liquid water storage (W). Further, the model does not account for lateral flow of water among grid cells and does not consider river routing explicitly. While the effect of the routing can be significant in large river basins of humid regions (Kim et al., 2009), it is negligible on the spatial scale of a grid cell (as also shown by small influence of the delayed storage component), and at the temporal scale of monthly aggregated values. To ensure that the model calibration is not affected by river routing, we do not compare simulated runoff

15 to measured river discharge of large basins in our model-data fusion approach.

Finally, the sum of liquid water storage and snow is taken as the modelled terrestrial water storage (TWSmod) of a grid cell for the given time step. Since the delayed runoff contribution is minor at the monthly time scale, we, for simplicity, only focus on the contributions of SWE and total W to TWS in this study.



20 **Figure 2.** Schematic structure of the model with calculation of TWS. Boxes denote the water storages [mm]: snow water equivalent SWE, soil moisture SM, retained water RW, liquid water W and total terrestrial water storage TWS. Fluxes are represented by arrows. Red colour identifies forcing data: precipitation P [mm d^{-1}], air temperature T [$^{\circ}\text{C}$] and net radiation Rn [$\text{MJ m}^{-2}\text{d}^{-1}$]; while green colour indicates variables constrained by observations: evapotranspiration ET [mm d^{-1}], runoff Q [mm d^{-1}], SWE [mm] and TWS [mm].

25

2.3 Input Data

As meteorological forcing we used globally available, daily cumulated gridded precipitation sums [mm d^{-1}], average air temperature [$^{\circ}\text{C}$] and net radiation [MJ m^{-2}] from March 2000 to December 2014.

Precipitation values originate from the 1° daily precipitation product version 1.2 of the Global Precipitation Climatology

30 Project (GPCP-1DD) (Huffman et al., 2000; Huffman and Bolvin, 2013), that combines remotely-sensed precipitation remote

Kommentiert [TT11]: referee #1 / comment 6

Kommentiert [TT12]: referee #1 / comment 2

Kommentiert [TT13]: referee #2 / comment 5

~~sensing~~ and observations from gauges. Temperature was obtained from the CRUNCEP version 6.1 dataset (Viovy, 2015), which is a merged product of Climate Research Unit (CRU) TS.3.23 observation-based monthly climatology (1901-2013) (New et al., 2000) and the National Center for Environmental Prediction (NCEP) 6-hourly reanalysis data (1948-2014) (Kalnay et al., 1996). Net radiation is based on radiation fluxes of the SYN1deg Ed3A data product of the Clouds and the 5 Earth's Radiant Energy Systems (CERES) program of the United States' National Aeronautics and Space Administration (NASA) (Wielicki et al., 1996).

Rather than using a single data stream, e.g. discharge measurements at the outlet of large continental catchments as used in traditional large-scale hydrological studies, we calibrated the model against multiple observation-based data streams on the 10 grid scale. The integrated datasets include terrestrial water storage anomalies (TWSobs) [mm], snow water equivalent (SWEobs) [mm], evapotranspiration (ETobs) [mm d⁻¹], and gridded runoff estimates for Europe (Qobs) [mm d⁻¹]. TWSobs is derived from the GRACE Tellus Mascon product version 2 based on the GRACE gravity fields Release 05 processed at NASA's Jet Propulsion Laboratory (JPL) (Watkins et al., 2015; Wiese, 2015). The GRACE solutions were corrected for geocentric motion coefficients, according to Swenson et al. (2008) and for variations in Earth's oblateness (C20 15 coefficient) obtained from Satellite Laser Ranging (Cheng et al., 2013). The Glacial isostatic adjustment has been accounted for using the model by A et al. (2013). The dataset provides monthly anomalies of equivalent water thickness relative to the January 2004–December 2009 time-mean baseline for the period 2002–2016. Unlike previous GRACE products based on spherical harmonic coefficients, the JPL RL05M dataset uses equal area 3° x 3° spherical cap mass concentration blocks (mascons) to solve for monthly gravity field variation. To ensure a clean separation along coastlines within land/ocean 20 mascons, a Coastline Resolution Improvement (CRI) filter has been applied (Watkins et al., 2015). For each mascon, uncertainties were estimated by scaling the formal covariance matrix. To enable hydrological analysis at sub-mascon resolution, we used the provided gain factors to scale the original TWSobs values.

To gain confidence in partitioning of the integrated TWS, we additionally used SWE estimates from the European Space Agency's (ESA) GlobSnow SWE v2.0 product (Luoju et al., 2014). The dataset provides daily SWE values [mm] ~~for the non-alpine Northern Hemisphere~~ based on assimilating passive microwave satellite data and observed snow depth from 25 weather stations ~~by applying a semi-empirical snow emission model~~ ~~for the non-alpine Northern Hemisphere~~. Compared to data from stand-alone remote sensing approaches, GlobSnow SWE shows superior performance, even though validation against ground based measurements still reveals a systematic underestimation of SWE under deep snow conditions due to a change in the microwave behaviour of the snow pack (Derksen et al., 2014; Takala et al., 2011; Luoju et al., 2014).

30 The ET product is based on FLUXCOM (www.fluxcom.org), i.e. upscaled estimates of latent energy that were derived by integrating local eddy covariance measurements of FLUXNET sites, remote sensing, and meteorological data using the Random Forest (Breiman, 2001) machine learning algorithm (Tramontana et al., 2016). ~~In this study, we apply the Random Forest (Breiman, 2001) realization of FLUXCOM-RS+METEO (see Tramontana et al. 2016 for details). While the product captures seasonality and spatial patterns of mean annual fluxes well, predictions of inter-annual variations remain highly uncertain~~ (Tramontana et al., 2016). In addition, the performance of FLUXCOM ET was found to be lower in extreme environments that are not well represented by FLUXNET sites such as the arctic. An underestimation in the order of 10–20 % of ET can be expected owing to missing energy balance correction prior to upscaling for this respective FLUXCOM 35 ET realization. ~~Although FLUXCOM ET performs relatively better than other gridded ET products, an underestimation in the order of 10–20 % can be expected owing to missing energy balance correction prior to upscaling. While the product captures seasonality and spatial patterns of mean annual fluxes well, predictions of inter-annual variations remain highly uncertain~~

40 To calculate ETobs [mm d⁻¹], we assume a constant latent heat of vaporization of 2.45 MJ ~~mkg⁻¹~~. Similar to TWS that represents the vertically integrated water storage, observations of river discharge spatially integrate hydrological processes within a basin. Thus, they provide an invaluable tool for model validation at large scales. However, it

Kommentiert [TT16]: referee
#2 / comment 8

Kommentiert [TT17]: referee
#1 / comment 12

is desirable to apply gridded products to evaluate model performance at local (grid) scale. Therefore, we used the observation-based gridded runoff product E-RUN version 1.1 (Gudmundsson and Seneviratne, 2016) as constraint for runoff processes. This dataset is based on observed river flow from 2771 small European catchments that was spatially disaggregated to upstream grid cells using a machine learning approach. The data provides mean monthly runoff rates per unit area for each grid, so that river routing is not necessary to *directly* compare runoff estimates *directly* with modelled runoff. Similar to the ET data, gridded runoff estimates show high accuracy for the mean seasonal cycle across Europe, and poorer agreement regarding monthly time series and inter-annual variations (Gudmundsson and Seneviratne, 2016).

Table 1 summarizes the main features of the data used in this study. If required, the data streams were resampled from their original resolution to a consistent $1^\circ \times 1^\circ$ latitude/longitude grid and common daily (meteorological forcing) respectively monthly (calibration data) time steps. Data preparation further included extraction of the relevant, overlapping time period and area under consideration.

Table 1. Overview on data applied for meteorological forcing and multi-criteria calibration resp. model evaluation (NH: Northern Hemisphere).

Variable	Dataset	Coverage and resolution		Reference	
		Spatial	Temporal		
Meteorological forcing					
P	precipitation	GPCP 1dd v1.2	$1^\circ \times 1^\circ$ global	daily 1996–present	Huffman et al. (2000); Huffman and Bolvin (2013)
T	air temperature	CRUNCEP v6.1	$0.5^\circ \times 0.5^\circ$ global	daily 1901–2014	Viovy (2015)
Rn	net radiation	CERES SYN1deg Ed3A	$1^\circ \times 1^\circ$ global	3-hourly 03/2000–05/2015	Wielicki et al. (1996)
Calibration and evaluation					
TWS	terrestrial water storage anomalies	GRACE Tellus JPL-RL05M v2	$0.5^\circ \times 0.5^\circ$ global	monthly 2002–2016	Watkins et al. (2015); Wiese (2015)
SWE	snow water equivalent	GlobSnow v2.0	$0.25^\circ \times 0.25^\circ$ non-alpine NH	daily 1979–2012	Luoju et al. (2014)
ET	evapo-transpiration	FLUXCOM	$0.5^\circ \times 0.5^\circ$ global	daily 1982–2013	Tramontana et al. (2016)
Q	runoff	EU-RUN v1.1	$0.5^\circ \times 0.5^\circ$ Europe	monthly 1950–2015	Gudmundsson and Seneviratne (2016)

2.4 Multi-criteria calibration

In this study, calibration is intended to identify the set of 10 model parameters (Table 2) that achieves the best fit between simulations and observations for all grids cells and regarding all observational data simultaneously. Thereby, we aimed to exploit the strength of each data stream, while considering known uncertainties and biases. For this purpose, we defined a cost function that takes into account the weakness of each observed variable and evaluates the overall model fit with one value of total cost (see subsequent section). To minimize total costs and thus find the optimal parameter values, we applied the Covariance Matrix Evolution Strategy (CMAES) (Hansen and Kern, 2004) search algorithm. The CMAES, as an evolutionary algorithm, is a stochastic, derivative-free method for non-linear, non-convex optimization problems. Compared to gradient-based approaches, it performs superior on rough response surfaces with discontinuities, noise, local optima

and/or outliers, and is a reliable tool even for global optimization (Hansen, 2014). Additionally, CMAES' guided search in the parameter space makes the algorithm less computationally demanding than other global optimization approaches which enumerate a large number of possible solutions (e.g. Monte Carlo Markov Chain methods) (Bayer and Finkel, 2007).

5 In order to keep computational demands low and to avoid overfitting by a very small sample size, we perform calibration for a subset of 1000 randomly chosen grid cells. Within these iterative calibration process, the model simulations are carried out on daily time steps, while costs are calculated based on monthly values. Further, each model run includes an initialization based on 10 random years that were selected a priori.

Kommentiert [TT18]: referee #1 / comment 27

Cost function

To objectively describe the goodness of fit, we defined a cost function based on model efficiency (Nash and Sutcliffe, 1970),

10 but with explicit consideration of the uncertainty σ_i of the observed data stream as:

$$\text{cost} = \frac{\sum_{i=1}^n \frac{(x_{obs,i} - \bar{x}_{obs})^2}{\sigma_i}}{\sum_{i=1}^n \frac{(x_{obs,i} - \bar{x}_{obs})^2}{\sigma_i}} , \quad (1)$$

where $x_{obs,i}$ is the observed data, \bar{x}_{obs} the average of x_{obs} , and $x_{mod,i}$ the modelled data of each space-time point i , respectively. Similar to model efficiency, the criterion reflects the overall fit in terms of variances and biases, yet with an optimal value of 0 and a range from 0–∞. Costs are calculated for each variable separately, considering only grid cells and time steps with 15 available observations, which vary for the different data streams. Additionally, to overcome the sensitivity to outliers arising from data uncertainties or inconsistencies, we adopted a 5 percentile outlier removal criterion (Trischenko, 2002), i.e. the data points with the highest 5 % residuals $x_{obs} - x_{mod}$ were excluded in the cost function.

The costs of each observed variable and its modelled counterpart are then added equally to derive a single value of total cost (Eq. (2)). Since a perfect simulation would yield a total cost of 0, calibration aims to find the global minimum of $\text{cost}_{\text{total}}$.

20

$$\text{cost}_{\text{total}} = \text{cost}_{\text{TWS}} + \text{cost}_{\text{SWE}} + \text{cost}_{\text{ET}} + \text{cost}_{\text{Q}} , \quad (2)$$

As the uncertainty σ of observational data in Eq. (1) is adapted to best reflect the strength of the individual data stream, we preselected the strongest aspect of the data to be included in the cost function. Owing to the larger uncertainties of ETobs and Qobs on inter-annual scales, we only employed the grid's mean seasonal cycles, while the full monthly time series of gridded TWSobs and SWEobs were taken into account.

25 25 As ETobs and Qobs do not explicitly provide uncertainty estimates, we assumedadopted an uncertainty of 10 % and minimal 0.1 mm, respectively based on commonly reported values. In order to define σ of TWSobs we utilized the spatially and temporally varying uncertainty information provided with the GRACE data. Additionally, the monthly values of observed and modelled TWS datasets were translated as anomalies to a common time-mean baseline of their overlapping period 01.01.2002–31.12.2012 before calculating the cost for TWS.

Kommentiert [TT19]: referee #1 / comment 15

30 For SWE, we applied an absolute uncertainty of 35 mm based on reported differences to ground-measurements (Liu et al., 2014;Luoju et al., 2014). Since GlobSnow SWE saturates above approx. 100 mm (Luoju et al., 2014), we do not penalize model simulations when both, SWEobs and SWEmod, are larger than 100 mm in order to prevent the propagation of data biases to calibrated model parameters.

For maps of the temporal average uncertainties see S2.

2.5 Evaluation of model performance

Once the parameters were optimized, we applied the model for the entire study domain, and evaluated its performance regarding all grid cells (6050) in terms of Pearson correlation coefficient r and root mean square error RMSE for each variable with observational data, respectively. On the one hand, the overall performance at local scale was assessed by

Kommentiert [TT20]: referee
#1 / comment 26

5 calculating r and RMSE for the monthly time series of each grid individually. On the other hand, the model performance over the entire study domain was evaluated by comparing the seasonal and inter-annual dynamics of the regional average.

Therefore, we defined inter-annual variation (IAV) as the deviation of the monthly values from the mean seasonal cycle (MSC). As with the calibration, we focused on the common time period 2002–2012, and considered only the grid cells and time steps with available observations.

Kommentiert [TT21]: referee
#2 / comment 10

10 In order to benchmark our model against current state-of-the-art hydrological models, we compared its simulations with the multi-model ensemble of the global hydrological and land surface models of the eartH2Observe dataset (Schellekens et al. 2017). This ensemble includes HTESSEL-CaMa (Balsamo et al., 2009), JULES (Best et al., 2011;Clark et al., 2011), LISFLOOD (van der Knijff et al., 2010), ORCHIDEE (Krinner et al., 2005;Ngo-Duc et al., 2007;d'Orgeval et al., 2008), SURFEX-TRIP (Alkama et al., 2010;Decharme et al., 2013), W3RA (van Dijk and Warren, 2010;van Dijk et al., 2014),
15 WaterGAP3 (Flörke et al., 2013;Döll et al., 2009), PCR-GLOBWB (van Beek et al., 2011;Wada et al., 2014) and SWBM (Orth et al., 2013). For consistency, we processed the model estimates in the same manner as our model simulations to directly compare modelled SWE and TWS to observations from GlobSnow and GRACE, respectively. While each model provides simulated SWE, they vary in the representation of other storage components. We calculated modelled TWS for each model by summing up the available water storage components, respectively. Thus, the variables contributing to
20 modelled TWS vary between the models, which impedes detailed comparison. Additionally, we calculated the multi-model mean of SWE and TWS simulations.

2.6 Analysis of TWS variations and composition

Finally, the contribution of snow and liquid water to seasonal and inter-annual TWS variability was quantified across spatial scales. For this, we ran the model with optimized parameters for the entire study domain from 2000 to 2014, and translated

25 simulated storages as anomalies to the time-mean baseline. As in the model evaluation, the MSC and IAV of SWEmod, W and TWSmod anomalies were calculated at local scale for each grid individually and as spatial average over all grid cells. To assess storage variability, the variance in the MSC and the IAV of each storage component was computed. Assuming negligible covariance of snow and liquid water (see S8), their relative contribution to TWS variance was calculated as the contribution ratio CR:

Kommentiert [TT22]: referee
#1 / comment 8

30

$$CR = \frac{var(W)}{var(TWSmod)} - \frac{var(SWEmod)}{var(TWSmod)}, \quad (3)$$

While $CR = 0$ indicates equal contribution of snow and liquid water to TWS variability, positive (negative) values of CR imply that variations of TWSmod mainly result from variations in liquid water (snow pack), with $CR = +1$ meaning that all variation is explained by liquid water and $CR = -1$ suggests determination solely by snow.

35 From Eq. (3) and the assumption that $var(W) + var(SWE) = var(TWS)$, the percentage contribution of liquid water storages to the variability of TWS can be inferred as CW:

$$CW = \frac{var(W)}{var(TWSmod)} = \frac{CR+1}{2}$$

(4)

As this study intends to analyse the effects of storage components on TWS at different spatial scales (local grid scale and large (regional) spatial averages), the difference in spatial heterogeneities of these components is considered. Some storage components, e.g., soil moisture anomalies, have much larger spatial variability than others. Due to this high small-scale heterogeneity, the effect on larger regional scale might be smaller than expected, as different local scale heterogeneities compensate each other when the regional averages are calculated (Jung et al., 2017). Thus, we assessed the spatial coherence of simulated patterns of SWE and W by calculating the proportion of total positive and total negative covariances among grid cells (Eq. (4,5) in Jung et al. (2017)). If the sum of positive covariances outweighs the sum of negative covariances, it implies some degree of spatial coherence of the anomalies. Spatial coherence of anomalies then causes a larger variance of the averaged anomalies compared to the sum of the variances of individual grid cells. This assessment of spatial coherence of SWE and W anomalies allows for understanding different contributions of SWE and W to TWS variability at local scale compared to the regional scale.

When analysing fluxes and storages on different spatial scales, one has to take into account that some of them vary highly on small scales. The local scale heterogeneity, especially regarding water storage variations, can lead to compensatory effects when averaging the variables over large spatial domains (Jung et al., 2017). Thus, we assessed the spatial coherence of simulated patterns of TWS components by calculating the proportion of total positive and total negative covariances among grid cells (Eq.(4,5) in Jung et al. (2017)). Predominance of either one of them implies spatial coherence, whereas balance between both suggest spatially diverging pattern that compensate each other out when analysing large spatial domains.

Kommentiert [TT23]: referee #2 / comment 11

3 Results and discussion

The following sections present and discuss the results obtained with the calibrated model. First, we review the calibration approach and the optimized parameter values. Then the model is validated with respect to its overall performance at grid scale, as well as the reproduction of average seasonal (MSC) and inter-annual (IAV) dynamics. Subsequently, we assess the driving component of spatially integrated TWS variations and the relative contributions of snow and liquid water to TWS variability on local scale. Finally, we summarize the results across spatio-temporal scales.

3.1 Model optimization

Optimization of the model identifies the parameter values listed in Table 2 to be most suitable regarding all data constraints simultaneously. The CMAES search algorithm converged after 3272 function evaluations as no further improvement of $costs_{total}$ could be achieved, which suggests a reliable estimate of the global optimal parameter set. The individual cost terms obtained with default and optimized parameter values are contrasted in Table S1. Overall, this parameter set obtained for a subset of 1000 random grids is reasonable with respect to reported ‘plausible’ parameter ranges, with none of them reaching their physically and/or technically defined upper and lower calibration bounds.

In detail, snow fall is reduced by p_{sf} to 67 % of precipitation occurring at $T < 0$ °C. This reduction agrees with Behrangi et al. (2016), who found GPCP to overestimate snowfall over Eurasian high latitudes by about 20 % compared to other precipitation products. Similar, overestimation of precipitation undercatch correction in GPCP has also been reported by Swenson (2010). Taking into account the mismatch in temporal and spatial domain, as well as the experimental definitions, reducing GPCP snow fall in our study by 33 % is roughly consistent with to both studies. Therefore, p_{sf} allows to reduce inconsistencies between the precipitation forcing and the water storages as given by GlobSnow SWE and GRACE TWS.

Kommentiert [TT24]: referee #2 / comment 14

Kommentiert [TT25]: referee #2 / comment 12

Kommentiert [TT26]: referee
#2 / comment 13

Further, each grid is assumed to be completely covered by snow if SWE \geq 80 mm (sn_c). On the one hand, the snow pack can be reduced by sublimation, with $sn_a = 0.44$ indicating relatively high sublimation resistance, compared to a default of $sn_a = 0.95$ proposed by Miralles et al. (2011b). The divergence probably results from interaction with snow melt, as net radiation also contributes to melt with 0.9 mm MJ $^{-1}$ (m_r) if T exceeds 0 °C. Nevertheless, melt is mainly induced by temperature, as the estimated day degree factor (m_t) is 2.63 mm K $^{-1}$, which is close to typical values of 3 mm K $^{-1}$ (Müller Schmied et al., 2014; Stacke, 2011). These parameter interactions underline ~~that neither an equifinality issue between~~ modelled snow melt ~~nor and~~ sublimation ~~are well constraint by data due to missing data constraints~~, resulting in ~~larger parameter uncertainties~~ ~~for due to parameter equifinality between~~ sn_a , m_r and m_t . However, for the objective of this study it's not primarily relevant whether sublimation or radiation induced melt decreases the snow pack, as the total ~~melt-snow loss~~ amount remains relatively unchanged for different parameter combinations.

The maximum soil water holding capacity is set to 515 mm after calibration, a comparatively high value that is likely to include storages in surface water bodies such as lakes and wetlands within our study domain. The optimized value of s_{exp} is 1.46, which suggests a non-linear relationship between soil moisture storage and runoff generation. For the same amount of incoming water (rain fall and snow melt), the non-linear relationship produces a smaller runoff and larger infiltration than a linear relationship ($s_{exp} = 1$).

Regarding evapotranspiration, the alpha coefficient (et_a) in the Priestley-Taylor formula is generally taken as 1.26 ~~for well-watered crops~~ based on experimental observations (Priestley and Taylor, 1972b; Eichinger et al., 1996). Thus, the optimized value of 1.20 for et_a reflects a plausible value. Further, et_{sup} indicates that 2 % of the available soil moisture can evaporate per day (including transpiration), which lies within the range of site-specific ET sensitivities from 0.001 – 0.5 d $^{-1}$ ~~and is close to the median value (5 %)~~ (Teuling et al., 2006).

~~Finally, the calibrated recession time scale that delays land runoff is 13 days (q_r). Compared to much smaller alpine catchments for which Orth et al. (2013) reported q_r of 2 days, the longer delay coefficients are reasonable at a spatial resolution of 1° x 1° grids, because the elevation gradients are much smaller within a large spatial area. Finally, land runoff of the preceding 13 days contributes to total runoff for a given day (q_r). Compared to much smaller alpine catchments for which Orth et al. (2013) reported q_r of 2 days, this seems reasonable for 1° x 1° grids that rarely reach steep average slopes.~~ At first glance, 13 days appear to be quite a short effective time period, as the delay is supposed to comprise contributions from much slower depleting reservoirs, such as lakes and deep groundwater. However, implementing and calibrating a simple groundwater storage, that is recharged with some proportion of land runoff and linearly depletes over time, led to similar retardation times. ~~Further, as the calibrated value is far from the parameter bounds, q_r seems to represent the best compromise between various storage components that deplete at different rates. The parameter, despite the limited physical interpretation, mimics the average effect of slow runoff components in the most efficient way.~~

The uncertainty in the optimized parameter vector was estimated by quantifying each parameter's standard error as the square root of the product between the diagonal elements of the parameters' covariance matrix (calculated from the Jacobian matrix) and the sum of residual squares according to Omlin and Reichert (1999) and Draper and Smith (1981). ~~The resulting relative parameter uncertainty is particularly instructive for comparing how well individual parameters could be constrained. Most of the parameters have uncertainties smaller than 10 % were well constrained (Table 2) suggesting . This suggests that our model-data fusion method, fed by multiple observation streams, succeeded in reducing is able to reduce the initial theoretical parameter ranges (up to 500 %) to much narrower ranges. Nevertheless, some parameters have a larger uncertainty range than others (e.g. q_r , sn_c , m_t), which may highlight a limitation in suitable observations to constrain them, as well as a lower sensitivity of the model results and the cost function used. Further, given that the model only considers the spatial variability of climate, the uncertainty in global parameters obtained from inversion may reflect the natural variations in these parameters that arise from differences in local land surface characteristics such as topography or land cover.~~

Kommentiert [TT27]: referee
#1 / comment 20

Table 2. Adjustable model parameters, their meaning, calibration range (theoretical range in brackets), optimized value including estimated uncertainty, and the corresponding equation in S1.

Parameter	Description	Unit	Range (theoretical)	value	Optimized ± uncertainty (%)	Eq.
Snow						
p_{sf}	scaling factor for snow fall	-	0–3 (∞)	0.67	$\pm 1e^{-3}$ ($\textcolor{red}{<10\%}$)	(S2)
sn_c	minimum SWE that ensures complete snow cover of the grid	mm	0–500 (∞)	80	± 19 (24 %)	(S3)
m_t	snow melt factor for T	mm K $^{-1}$ d $^{-1}$	0–10	2.63	± 0.26 (10 %)	(S4)
m_r	snow melt factor for Rn	mm MJ $^{-1}$ d $^{-1}$	0–3	0.90	± 0.05 (6 %)	(S4)
sn_a	sublimation resistance	-	0–3	0.44	± 0.01 (3 %)	(S5)
Soil						
s_{exp}	shape parameter of runoff-infiltration curve	-	0.1–5	1.46	± 0.02 (2 %)	(S12)
s_{max}	maximum soil water holding capacity	mm	10–1000 (0– ∞)	515	± 9 (2 %)	(S12)
et_a	alpha coefficient in Priestley-Taylor formula	-	0–3	1.20	± 0.01 (1 %)	(S14)
et_{sup}	ET sensitivity / SM fraction available for ET	d $^{-1}$	0–1	0.02	$\pm 6e^{-5}$ ($\textcolor{red}{0\textcolor{red}{-}1\%}$)	(S18)
Runoff						
q_t	recession time scale for land runoff	d	0.5 (0)–100	13	± 4 (31 %)	(S20)

5 We adopted the calibrated parameter values as ~~the~~ global constants for model simulations over the entire study domain. Even though the globally uniform parameters may not provide perfect simulation for all grids over a large study domain, this approach represents a compromise between a priori parametrization of the model and its calibration at local or regional (e.g. basin) scale. While local and regional model calibration enables good adaption to geographic characteristics, it easily leads to overfitting of the model and thus propagates the constraints' inherent errors and uncertainties to the modelling result. As

10 these uncertainties often vary in space, globally uniform parameter values diminish overfitting uncertainties. In addition, calibration for several independent grids is computationally demanding and subsequently requires a parameter regionalization approach (He et al., 2011). Since such approaches are not commonly accepted ~~and not considered practical~~ (Sood and Smakhtin, 2015; Bierkens et al., 2015), macro-scale models mostly apply a priori parameters based on empirical values or on expert knowledge, that yet may lead to suboptimal simulations (Beck et al., 2016; Sood and Smakhtin, 2015).

15 ~~Therefore, a global parameter set estimated for a subset of randomly chosen grid cells allows adaption to observational constraints, while on the other hand reduces potential propagation of data uncertainty due to overfitting of the model in a specific region. In this regard, a global uniform parameter set, that still allow good model performance, can be seen as an indicator for the robustness of the model and the modelling approach.~~

Kommentiert [TT28]: referee #1 / comment 21

3.2 Model performance

20 For model validation, we used the optimized parameter values to simulate hydrological fluxes and states of the 2002–2012 period over the entire study domain, and evaluated the model results against the observation-based data of TWS, SWE, ET and Q.

Kommentiert [TT29]: referee #1 / comment 22

In general, all observed patterns are reproduced very well, taking into account the specific data weaknesses. We achieve a 'near perfect' correlation of 0.99 and 0.94 for mean seasonal variations of ET and Q, respectively. The median RMSE of mean seasonal ET is 11 mm month⁻¹ and 9.5 mm month⁻¹ for Q, which represent 15 % resp. 17 % of the average observed annual amplitude. At the inter-annual scale, though, larger discrepancies exist, which at least partly arise from larger 5 uncertainties in ETobs resp. Qobs (S42). Thus, we assume high confidence in modelled ET and Q fluxes and subsequently focus on evaluation of the water storages TWS and SWE.

Kommentiert [TT30]: referee

#2 / comment 15

3.2.1 Performance on local grid scale

Overall, the model performs well compared to the observations of monthly time series of SWE and TWS (Fig. 3). More than half of the grid cells obtain correlation values higher than 0.74 between SWEobs and SWEmod. In general, the median 10 RMSE is 20 mm, which is smaller than the average uncertainty of 35 mm in SWEobs. The correlation reduces in lower latitudes where seasonal snow accumulation and thus variability is small. Further, the correlation is also relatively weaker in arctic North America and the Rocky Mountains, while larger deviations between observed and modelled snow quantities center around mountainous and coastal regions (e.g. Rocky Mountains, Kamchatka), and regions with the largest seasonal 15 snow accumulation (Labrador Peninsula, North Siberian Lowland and northern West Siberian Plain). There are several reasons for this relatively poorer performance. First, the GlobSnow measurements do not cover mountainous areas due to the sub-grid variability of snow depth and high uncertainties in the microwave measurements in complex alpine terrains (Takala et al., 2011). As the resampling and the coarse resolution of each grid in this study compound a distinct alpine/non-alpine classification, these uncertainties leak to the surrounding areas. Second, neither the input forcing data nor our model include 20 the sub-grid scale heterogeneity of climate (e.g., precipitation and temperature) and hydrological processes, that may be significant in mountain-near or coastal regions. Additionally, the accuracy of observed large snow accumulation is limited as the radar-retrieval methods tend to saturate at large SWEobs values, which then leads to large RMSE of the model simulation.

Similar to SWE, more than half of the grid cells show a strong correlation of 0.71 between TWSobs and TWSmod, which reflects a realistic temporal variation in the model simulation. Compared to SWE, the RMSE of TWS is somewhat higher, 25 yet the median of 43 mm still reflects the range of ± 22 mm average uncertainty in GRACE TWSobs of the study domain (Wiese, 2015). However, when comparing GRACE TWS with model simulations, several aspects have to be considered. First, TWSobs as an integrated signal comprises all water storages, not all of which are (sufficiently) represented in the model structure. Second, although GRACE TWS passed through various pre-processing steps, the models to account e.g. for postglacial rebound or leakage between neighbouring grid cells introduce their own uncertainties and do not remove the 30 effects completely. Further, with a native resolution of 3° , uncertainties remain for grids that comprise large variability at sub-grid scale and depend on the model used to estimate GRACE scaling factors (Wiese et al., 2016). This together is reflected in higher RMSE in arctic regions (e.g. surrounding the Hudson Bay), as well as in heterogeneous coastal and mountainous regions. Additionally, our model shows a weaker performance in subarctic and arctic wetlands, and in central 35 North America and Eastern Eurasia. The latter both are relatively dry regions that are rather dominated by inter-annual TWS variations (Humphrey et al., 2016). Discrepancies between TWSobs and TWSmod thus relate to a low signal-to-noise ratio in TWSobs due to small seasonal TWS variations. On the other hand, the anthropogenic influence for irrigational withdrawal is very large in these regions, yet such processes are not considered in our model. We also lack explicit surface water storages (including wetland dynamics), which may be the reason for poorer performance especially in North American wetland regions.

Kommentiert [TT31]: referee

#1 / comment 19

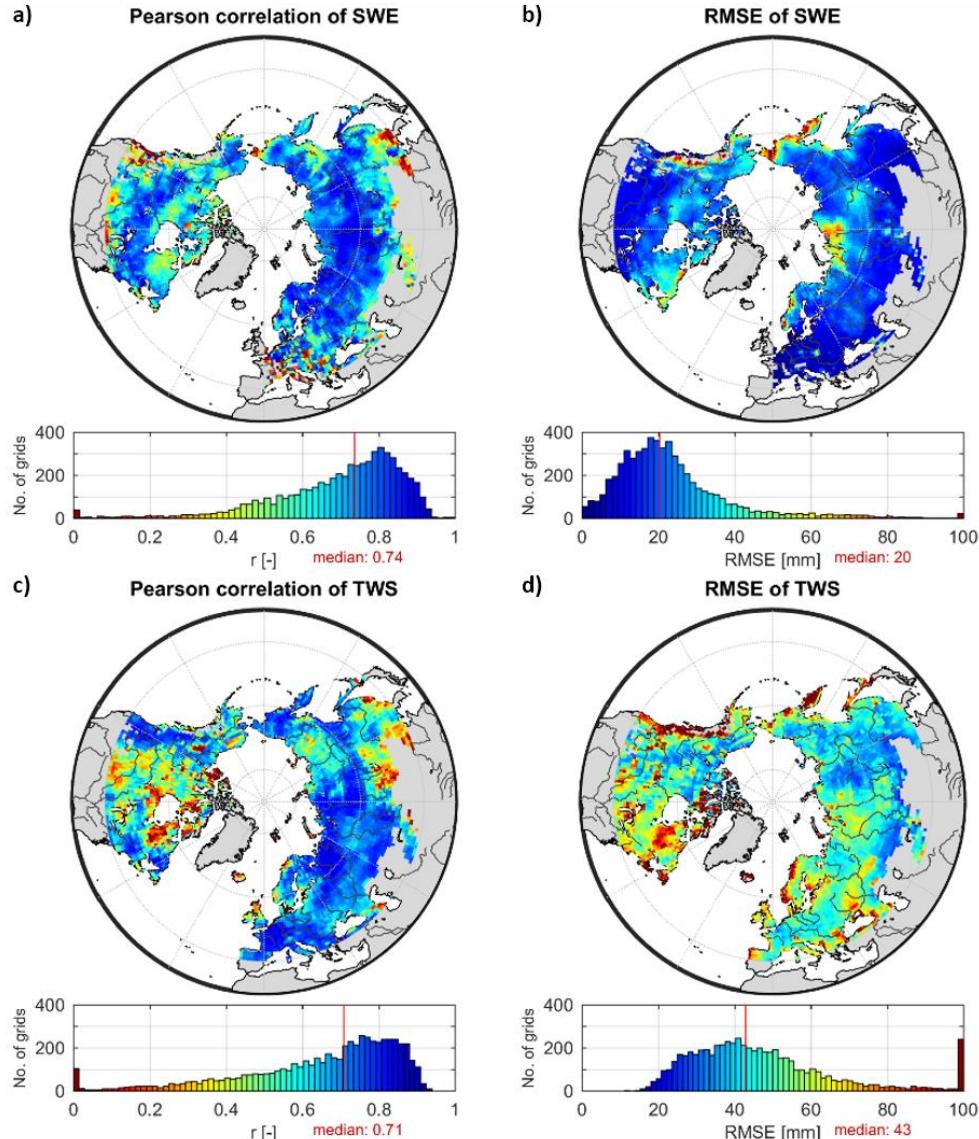


Figure 3. Pearson correlation coefficient r (a,c) and root mean square error RMSE (b,d) between monthly values of modelled SWE and GlobSnow SWE (a,b), as well as modelled TWS and GRACE TWS (c,d) for the period 2002–2012 and for each $1^\circ \times 1^\circ$ grid cell of the study domain. Values of r are truncated to the range 0–1 (a,c), and values of RMSE to the range 0–100 (b,d).

Kommentiert [TT32]: referee
#1 / comment 23

5

3.1.2 Performance of the spatially integrated simulations

Since the aim of this study is to analyse the composition of TWS across temporal scales, we additionally evaluated average (spatially integrated) MSC and IAV of SWE and TWS (Figure 4). While the mean seasonal variations of both observational data streams are relatively robust and have been used for model evaluation before (Alkama et al., 2010; Döll et al., 10 Döll et al., 2014; Schellekens et al., 2017; Zhang et al., 2017), their inter-annual variations are more uncertain and contain considerable noise. This clearly reduces the information content in the observational data, so that we evaluate the IAV in more qualitative terms.

As with the comparison at grid scale, the spatially averaged SWEmod compares well to SWEobs, with a correlation of 0.95 suggesting a good reproduction of seasonal snow accumulation and ablation processes (Fig. 4a). Owing to the high uncertainty of SWEobs peaks due to signal saturation, the higher amplitude of SWEmod seems reasonable. Although inter-annual variations are not as well represented as the MSC, general tendencies, e.g. increasing/decreasing positive/negative 5 anomalies, coincide.

Similar to SWE, the spatial average TWS shows high correlation of 0.91 for seasonal variations, with positive anomalies from December to May/June and negative anomalies during summer and autumn months (Fig. 4b). Even though the modelled amplitude is slightly larger than the observed one, it stays within the uncertainty range of TWSobs for most months, suggesting reliable simulations. However, TWSmod precedes TWSobs on average by one month, reaching the 10 maximum in March instead of April, and the minimum in August instead of September. A similar phase shift of one month between GRACE TWS and modelled TWS has been reported by several state-of-the-art global models (Döll et al., 2014; Schellekens et al., 2017). It should be noted that some areas such as East North America, Kamchatka, Scandinavia and Western Europe do not show phase differences, while the lag in South East Eurasia is even larger, as already suggested by 15 lower overall correlation (Fig. S53). In general, the disagreement in timing is attributed to the lack of sufficient water storages and delay mechanism within the model, so that the modelled system reacts too fast (Schellekens et al., 2017; Döll et al., 2014; Schmidt et al., 2008). Thus, we implemented model variants with an explicit groundwater storage to delay 20 depletion of TWS, with spatially varying soil properties to better represent heterogeneous infiltration and runoff rates, as well as a variant that applied a more sophisticated approach to calculate snow dynamics based on energy balance. Despite the efforts, we achieved no improvement in terms of reducing the phase shift. Therefore, the question arose, whether it is not 25 primarily the model formulation that prevents correction of the temporal delay, but rather the combination of forcing data respectively observational constraints. To further preclude possible errors due to such data inconsistencies e.g. between GRACE TWS and GlobSnow SWE, we excluded GlobSnow SWE data from calibration. Although this could slightly improve the agreement of TWS' MSC, it led to unrealistic behaviour of snow dynamics, and thus did not offer any advantages. Besides, we found no major differences in the magnitude or spatial distribution of the phase shift resulting from 30 25 the precipitation forcing (GPCP vs. WFDEI) or compared to other GRACE solutions (S65). Further, the lag in TWS simulation can occur due to several mechanisms and processes that are not yet considered in the current model structure such as lateral flow and surface storages (wetland and lakes), vegetation processes, glacier melt, and human influence with dams and reservoirs. However, we don't observe a general or systematic relationship with either elevation, land cover type, soil properties, and the occurrence of lakes and wetlands. There is a tendency that larger negative lags occur more frequently in regions with sporadic permafrost, but the ranges of permafrost fractions are large for both, short and long lags in TWS, suggesting a complex interaction between permafrost extent and its effect on lag in seasonal TWS dynamics. Finally, potential biases in timing of ET due to snow cover and/or vegetation processes may also affect the timing of depletion of SM and TWS. Since we obtained no general correlation to either elevation, land cover, soil properties, the presence of lakes and wetlands, or the distribution of permafrost, we attribute the difference in TWS phase to a combination of missing, yet 35 spatially varying processes and storage components. As suggested in previous studies, those processes likely relate to human impacts, glacier melt, groundwater reservoirs, wetlands, river storage and/or river routing. Additionally, high uncertainties of the precipitation forcing and GlobSnow SWE product in mountain (near) regions, as well as leakage errors in the GRACE signal influence the accuracy of both, TWSobs and TWSmod. Although these shortcomings should be kept in mind, we assumed that they do not affect our results regarding to the relative contributions of snow and liquid water to TWS 40 significantly.

Kommentiert [TT33]: referee
#1 / comment 1

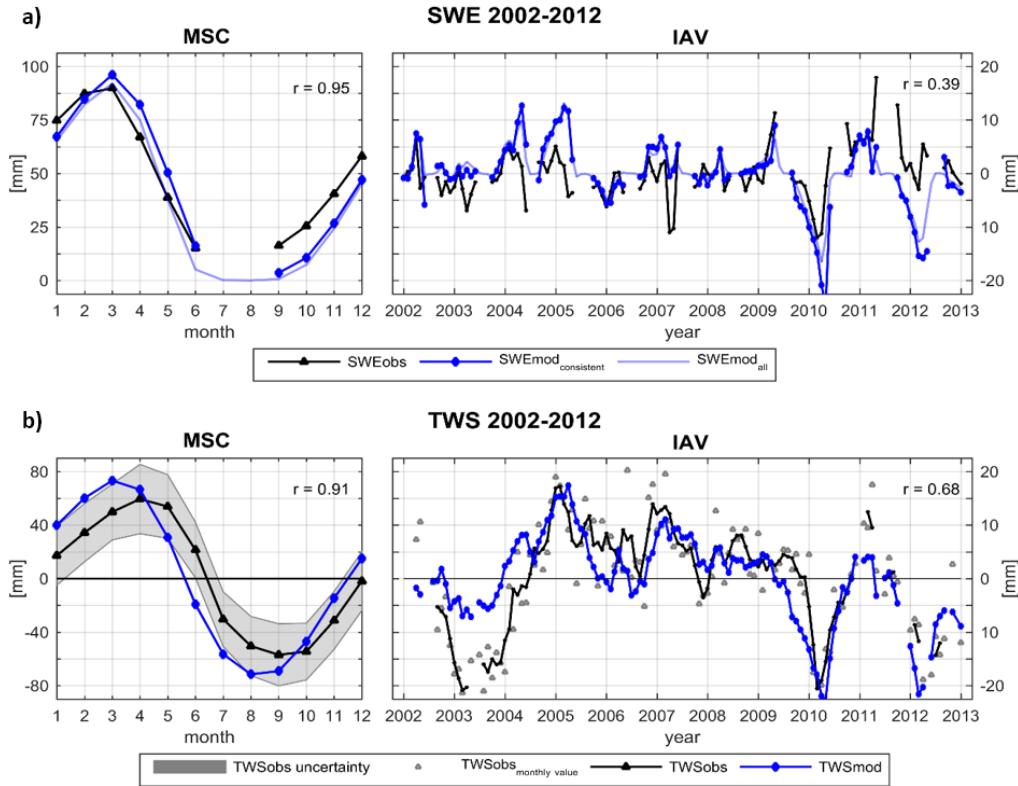


Figure 4. Spatially averaged mean seasonal cycle (MSC) of the period 2002–2012 as well as inter-annual variability (IAV, difference between monthly values and the MSC) for **a)** SWE and **b)** TWS. In a), SWEmod_{consistent} refers to modelled SWE considering only data points with available SWEobs, while SWEmod_{all} incorporates all time steps for all grids of the study domain. Correlation r is calculated only for consistent data point, respectively. In b), IAV, TWSobs_{monthly} value shows the original IAV of individual TWSobs months, while TWSobs and TWSmod are smoothed using a 3-month average moving window filter. Correlation r refers to the smoothed values. For the MSC in b) no smoothing is applied.

In terms of inter-annual variations, the variance in monthly TWSobs values is highly underestimated [by modelled TWS](#), which on the one hand relates to noise within the GRACE signal, but on the other hand may again reflect missing process representation in the model. To reduce the noise, we applied a three-month-moving-average filter on the monthly time series. The smoothed time series then shows fairly good agreement of inter-annual dynamics, with correlation $r = 0.68$ (Figure 4b, solid lines). Solely the amplitude of the large negative anomaly in 2003 is not captured by the model. [While the spatial pattern of this negative TWS anomaly can be simulated, the forcing data doesn't show large anomalies in 2003, so that the model fails to reproduce the magnitude of observed TWS, especially in North America. Issues with the precipitation forcing are further suggested by a negative SWE anomaly of on average 5 mm \(see Fig.4 a\) indicated in the GlobSnow data, that is not captured by the model, either. The reason why this snow anomaly is not captured by the forcing remains unclear at this point – it persists when using the WFDEI forcing data set. Besides, the agreement between GRACE and modelled TWS IAV gets substantially better when isolating inter-annual variations by removing the trends in both TWS time series \(increase in correlation \$r\$ from 0.77\). This suggests that the trend in GRACE TWS is to some extent either subject to observational issues or represents a process that is not captured by the model.](#)

Kommentiert [TT34]: referee #1 / comment 24

Kommentiert [TT35]: referee #2 / comment 17

3.1.3 Comparison with the eartH2Observe model ensemble

Compared to the model ensemble of the [eartH2Observe](#) dataset, we achieve equally good or better performance for the spatially integrated SWE and TWS on both, MSC and IAV, scales (Fig. 5, S6). Besides, the majority of the model ensemble obtains similar spatial patterns of performance criteria for SWE as well as for TWS (not shown).

Kommentiert [TT36]: referee
#1 / comment 25

5 The average observed MSC of SWE is captured with a correlation in the range of 0.79 (PCR-GLOBWB) to 0.99 (ORCHIDEE), whereby only ORCHIDEE ~~outperforms shows a better correlation than~~ our model ($r = 0.95$). However, modelled snow accumulation exceeds that of SWEobs for the majority of the models, which also reflects in higher RMSE (Fig. [S6](#), [Fig. S7](#), [Fig. S8](#)). On IAV scales, the correlation in general is lower, yet again we obtain a better fit ($r = 0.39$) than the model ensemble ($r = 0.12$ (ORCHIDEE) to 0.28 (LISFLOOD)). However, it remains uncertain, whether the 10 discrepancies between SWEobs and SWEmod represent model deficiencies or evolve from issues related to the GlobSnow SWE retrieval (Schellekens et al., 2017).

Regarding average seasonal TWS variations, our model performs as well as the model ensemble ($r = 0.91$), with the range of the eartH2Observe ensemble spanning from $r = 0.83$ (ORCHIDEE) to $r = 1.00$ (PCR-GLOBWB). The amplitudes in the MSC of TWSmod (95 to 156 mm) are comparable to the observed amplitude of 118 mm, except for SWBM, whose 15 amplitude is twice as large as that of TWSobs. This discrepancy is reflected in relatively high RMSE values for SWBM (Fig. [S8](#)). The model ensemble precedes observed seasonal TWS variations by 1 to 1.4 months, similar to our estimates of TWSmod (-1.1 month). Only PCR-GLOBWB, with a higher correlation than other models, shows a smaller average lag of less than 1 month (-0.3 months). This minor difference results from balancing out of preceding and succeeding in different regions over the study domain. Additionally, Schellekens et al. (2017) found that PCR-GLOBWB shows unrealistic snow 20 accumulation over time in Europe and boreal North America. Except for PCR-GLOBWB, the majority of the models obtains comparable spatial pattern of preceding TWS, with small differences at regional scales. Even though the difference in the MSC is commonly attributed to the lack or inadequate size ~~and number~~ of water storages (Kim et al., 2009b), a relationship between model performance and model complexity is not obvious. Relatively complex models, such as HTESSEL, SURFEX, and JULES, show similar phase differences as simpler models, such as SWBM and our model (-1.0 resp. -1.1 25 months). Since Schellekens et al. (2017) found the largest phase differences in cold regions, they postulate the implementation of processes associated with snow as important factor for this phase lag. In this context, constraining the model with snow observations as done in our study should increase confidence in the representation of snow processes. Nevertheless, we obtain a similar phase difference, which points to the importance of other hydrological processes and 30 storages even in snow-affected regions.

30 ~~Although~~In summary, compared to the model simulations in the EartH2Observe ensemble, our modelling framework assimilates information from more data streams, e.g. GRACE and GlobSnow data, we only used a subset of 1000 random grid cells to constrain the model parameters. Despite, our model performs better than EartH2Observe ensemble over the whole domain (6050 grids). This improvement in model performance is also consistent among several modelled variables and not limited to storage components only. This suggests that remote sensing data, with larger spatial coverage than site
35 ~~measurements, have a large potential in improving hydrological simulations over a large domain. In addition, remote sensing data also hold potentials beyond the use as an observational constraint and can provide information on identifying and formulating functional relationships across several spatial and temporal scales, that should be addressed in future efforts.~~none of the models, regardless of its complexity, outperforms others in terms of all variables and at all spatial and temporal scales. Compared to the eartH2Observe model ensemble, our model performs equally well or better regarding all variables and performance criteria. However, we note that the comparison is not completely fair as the eartH2Observe models unlike our model were not informed by GRACE and/or GlobSnow data beforehand. Nevertheless, we only used a subset of 1000 random grids to calibrate the model parameters, which suggests that inclusion of multiple observations better constrains the model over a large domain (6050 grids).

Kommentiert [TT37]: referee
#1 / comment 5

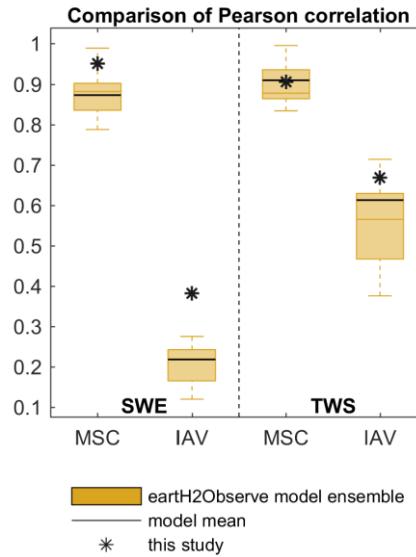


Figure 5. Pearson correlation for the spatially integrated SWE (left) and TWS (right) achieved by this study compared to the model ensemble of earth2Observe dataset across temporal scales. In each box, the central orange line represents the median and the edges the 25 % and 75 % percentiles of the model ensemble, while the solid black line marks the performance of the ensemble mean.

5

All in all, we conclude that our simple model with a global uniform parameter set achieves considerably good performance regarding observed patterns, especially compared to well-established, more complex models, and with respect to its simplicity and given uncertainties of forcing and calibration data. Thus, we found the model results to be suitable to analyse the composition of TWS across spatial and temporal scales.

10 3.3 TWS variation and composition

3.3.1 Spatially integrated

To assess the average composition of seasonal and inter-annual TWS variations, we spatially integrated modelled TWS anomalies as well as modelled anomalies of snow (SWE) and liquid water storages (W) across all grids of the study domain (Fig. 6).

15 Regarding the MSC, all water storages show a clear seasonal pattern. The maximum TWSmod in March coincides with the maximum in SWEmod. On contrary, W remains nearly constant throughout the winter, as related to the lack of evapotranspiration losses and missing infiltration due to prevailing solid precipitation. Starting from March, snow melt decreases SWEmod, and thus TWSmod, progressively. Thereby TWSmod declines with some delay, as positive W anomalies in April and May suggest buffering of melt water in the soil and on the surface before being transferred to runoff

20 or evapotranspiration. During the summer months, snow is absent, while W decreases due to higher summertime evapotranspiration, and preceding runoff of temporarily stored water. With W and SWEmod being at their minimum in August/September, TWSmod reaches its minimum, too, before starting to increase again in September/October. This rise relates to dropping evapotranspiration rates in combination with more precipitation input (increasing W) and beginning snow accumulation (increasing SWEmod). Despite the interplay of SWEmod and W on seasonal variations of the integrated

25 TWSmod, the amplitude of W (62 mm) is considerably lower than the one of SWEmod (92 mm) and TWSmod (144 mm). Thus, the seasonal accumulation of snow largely determines the magnitude of TWSmod. Additionally, W anomalies at least partly result from snow melt, whereas liquid water does not influence the snow pack. Thus, we conclude that average seasonal TWS variations in northern mid-to-high latitudes are mainly driven by annual snow accumulation and ablation

Kommentiert [TT38]: referee #1 / comment 28

processes. The Contribution Ratio CR (Eq.(3)) based on the spatially averaged MSC underlines this, as CR = -0.26 indicates that variations in SWEmod explain 63 % of seasonal TWSmod variability (CW = 37 %). This agrees with previous findings of Güntner et al. (2007), who found that SWE contributes to 68 % of seasonal TWS variations in cold regions using the WaterGAP model.

Kommentiert [TT39]: referee #1 / comment 4

5 On IAV scales, the pattern seems less homogeneous (Figure 6). In contrast to the MSC, CR = 0.25 suggests larger influence of liquid water anomalies (CW = 62.5 %) than snow anomalies to inter-annual TWS variations. Thereby, we found the main contributor to TWSmod anomalies being dependent on the phase of previous precipitation anomalies, in that they define whether snow fall anomalies lead to anomalies in the SWEmod, or whether rain anomalies directly influence W. Additionally, precipitation input shows larger inter-annual variability than evapotranspiration or runoff losses, and thus
10 dominates the change in water storages on IAV scales (not shown). Large TWSmod anomalies, such as in 2005, 2010 and 2012, follow anomalies in wintertime precipitation and go along with anomalies in SWEmod (Figure 6). On contrary, summertime anomalies related to W are usually less pronounced in their magnitude (e.g. 2003, 2006). We attribute this to accumulating effects of snow storage anomalies over the cold period, as they integrate all anomalies of previous cold months while the impact of evapotranspiration and runoff is reduced. Accordingly, largest TWSmod anomalies are obtained in early
15 spring before snow melt starts and when snow accumulation is highest. Nevertheless, since W is influenced by the quantity of snow melt, anomalies in SWEmod implicate subsequent changes in W. Additionally, snowpack anomalies are eliminated each summer, while W anomalies dominate the summer. As a result, W anomalies in any case affect TWSmod variability on
IAV scales when analysing the spatial average composition.

Besides, Güntner et al. (2007) demonstrated a shift from short-term storages with high seasonality such as SWE on MSC
20 scales towards storages with longer delay times on IAV scales. (Güntner et al., 2007) Although modelled W mainly represents soil moisture, it implicitly includes surface and ground water storages. Thus, its contribution of CW = 62.5 % to inter-annual TWS variations is roughly comparable to 55 % contribution from soil moisture (27 %) and surface water (28 %) in cold regions found by Güntner et al. (2007). Despite, the relatively large influence of surface water bodies shown by
25 Güntner et al. (2007) suggests that the lack of explicit surface water storages in this study may contribute to remaining discrepancies with GRACE and the lower magnitude of modelled inter-annual TWS variability compared to GRACE estimates.

Kommentiert [TT40]: referee #1 / comment 36

Kommentiert [TT41]: referee #1 / comment 4

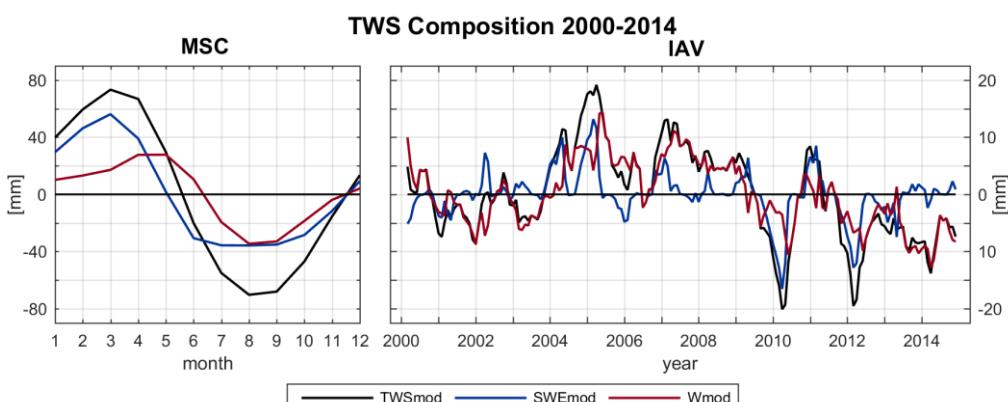


Figure 6. Spatially averaged mean seasonal cycle (MSC) of the period 2000–2014 as well as inter-annual variability (IAV, difference
30 between monthly values and the MSC) for modelled TWS, SWE and W anomalies to the time-mean of 2000–2014.

3.3.2 Local grid scale

Based on CR (Eq.(3)), Fig. 7 shows the relative contribution of SWEmod and W variances to total TWSmod variability on MSC and IAV time scales for each grid. Thereby, blue colours represent prevailing SWEmod variations as indicated by $CR < 0$, while red colours show dominance of variations in W ($CR > 0$).

5 Accordingly, variations in the MSC of TWSmod are mainly influenced by snow in northern regions, with the mean CR = -0.30 indicating that on average 65 % of seasonal TWSmod variability can be explained by SWEmod (CW = 35 %) (Fig. 7Figure 7a). The contribution of variation in liquid water in general increases southwards and prevails seasonal TWSmod variability south of approximately 50° latitude. An exception is Europe, where the influence of W reaches up to 60° latitude, and where the transition to snow dominated regions is more gradual. Since the calculated variations in RW are low, the
10 majority of modelled W represents variability in SM.

This obtained pattern confirms earlier studies, that ~~showed received~~ dominance of snow in northern latitudes in North America (Rangelova et al., 2007), and prevailing soil moisture dynamics further South e.g. in the Mississippi basin (Ngo-Duc et al., 2007;Güntner et al., 2007). Besides, the north extent of dominating W reflects the temperature gradient in North America and Eurasia. Comparison with average annual temperature suggests, that for $T > 10$ °C variability of W dominates,
15 while for $T < 0$ °C SWEmod dynamics prevail. This is plausible, as temperature determines annual snow accumulation, and the relative contribution of liquid water increases in the absence of snow. Yet, it further highlights the dependency on the used temperature data set, especially in a model that calculates snow fall and snow melt based on temperature thresholds as ours.

Opposed to the MSC, the variability of W dominates TWSmod variations on IAV scales in the entire study domain, as
20 clearly indicated by average CR = 0.63 (Fig. 7b). Inter-annual variations of SWEmod seem to be relevant only in regions that receive highest annual snow amounts, such as the Canadian Arctic Archipelago, the northern west coast of North America, North East Siberia and the northern West Siberian Plain. Due to a prolonged cold period there, the time span with rain ~~fall and~~ evapotranspiration ~~and unfrozen soil is~~ short, decreasing the occurrence of potential variability in W. However,
25 even in these regions the influence of SWEmod is low compared to the MSC. This reduced importance of snow on inter-annual scales again agrees with the findings of Güntner et al. (2007).

Apart from that, and sSince we already showed a link between average TWSmod IAV and previous precipitation anomalies, and as precipitation represents the main model forcing data, we investigated the relative contribution of rain and snow fall to inter-annual variability of total precipitation (Fig. 8). Similar to the composition of TWSmod on local scale, rain anomalies prevail for most of the grid cells (mean CR = 0.68). This is consistent when snow fall is not scaled by p_{sf} and thus suggests
30 that the greater contribution of W to inter-annual variations of TWSmod on local scale relate to highly variable (liquid) summertime precipitation events. On contrary, monthly snow fall seems less variable between years, resulting in less pronounced variations in SWEmod compared to W. Exceptions are regions of high maximum SWEmod, that accordingly show a considerable relative contribution of snow to the inter-annual TWSmod variability.

Kommentiert [TT42]: referee
#1 / comment 30

Kommentiert [TT43]: referee
#1 / comment 31

Kommentiert [TT44]: referee
#2 / comment 19

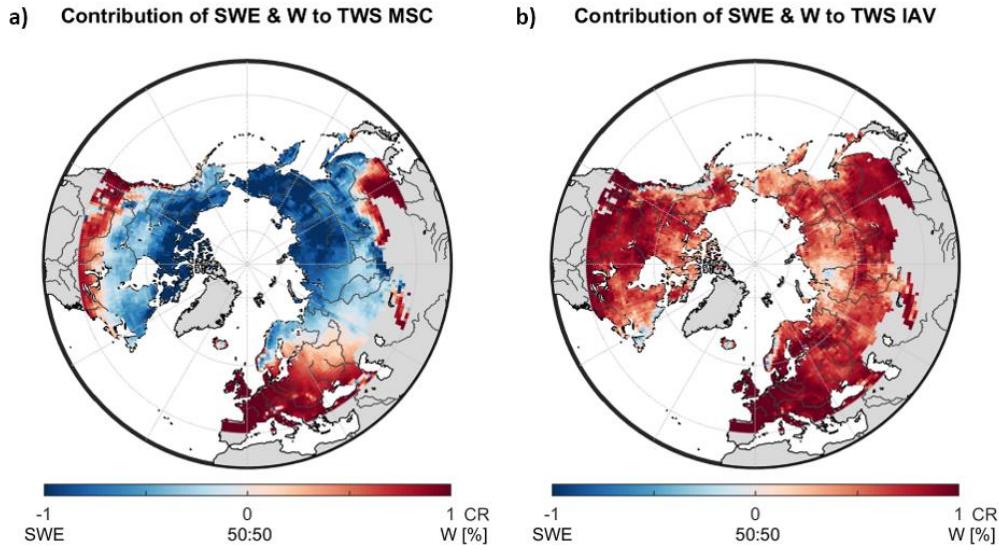
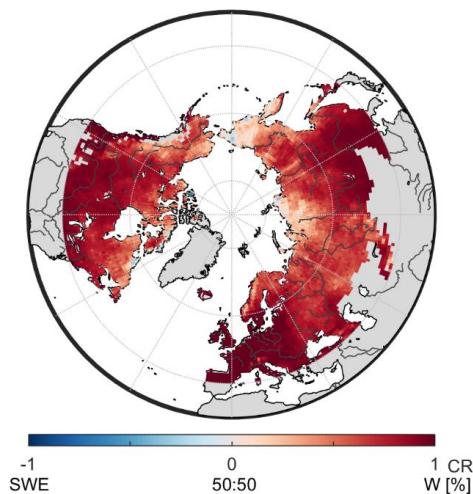


Figure 7. Relative contribution based on CR (Eq.(3)) of modelled snow (SWE) and liquid water (W) storage anomalies to **a**) mean seasonal variations from 2000–2014 of modelled TWS anomalies, and **b**) inter-annual variations of modelled TWS anomalies for each grid cell of the study domain, respectively.

Kommentiert [TT45]: referee #2 / comment 18
referee #1 / comment 29

Contribution of snow fall & rain fall to P IAV



5

Figure 8. Relative contribution based on CR (Eq.(3)) of modelled snow fall and rain fall to total precipitation (P) anomalies on inter-annual (IAV) scales for each grid of the study domain.

3.3.3 Comparison of different scales

10 Figure 9 summarizes the above presented contributions to TWSmod variability across spatial and temporal scales. As explained in the previous sections, we obtained two scale dependent differences in the relative contribution to TWSmod variability: (1) in general between temporal scales, and (2) for inter-annual variability between spatial scales. Regarding (1), Fig. 9 emphasizes again that seasonal variations of TWSmod are mostly determined by seasonal snow dynamics, while on inter-annual scales TWSmod variability mainly originates from variations in liquid water. As previously

stated, determination by SWEmod dynamics on MSC scales relates to the pronounced magnitude of seasonal snow variations in northern mid-to-high latitudes. In comparison, average monthly changes in W were found to be minor and additionally influenced by snow ablation. Thereby, the spatially integrated CR (black star) roughly agrees with the average of local contributions (dashed line).

5

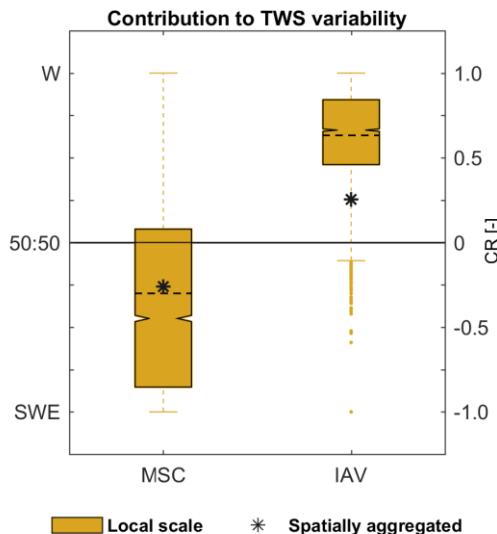


Figure 9. Relative contribution of snow (SWE) and liquid water (W) to TWS variability on different spatial (local grid scale, spatially integrated) and temporal (mean seasonal MSC, inter-annual IAV) scales based on CR (Eq.(3)). The boxplots represent the distribution of grid cell CR, with the dashed line marking the corresponding average. The star represents the CR calculated for the spatially integrated values.

Kommentiert [TT46]: referee
#2 / comment 18
referee #1 / comment 29

Concerning IAV scales, we found that the determination of TWSmod variability by W relates to larger inter-annual variations in liquid precipitation compared to snow fall. However, considerable differences between spatial scales exist (Figure 9). Opposed to the MSC, the spatially integrated CR (black star) for IAV is not within the interquartile range of local contributions CR. This indicates a relatively larger effect of SWEmod variations when looking on the spatially integrated values compared to local values. Since liquid Liquid water storages are determined by various geographic characteristics (e.g. topography, soil properties, land cover) and interacting processes (precipitation, evapotranspiration, runoff), their variations are highly heterogeneous in space, mainly due to heterogeneous rainfall anomalies. On contrary, snow variability is affected by fewer factors, and mainly regulated by a range of temperature values that control freezing and melting. Temperature per se se anomalies in turn shows apart from small scale variability (e.g. related to topography) sizeable spatial coherence across large areas. To assess the spatial coherence of W compared to SWEmod, we calculated the proportion of total positive and total negative covariances among grid cells (Fig. 10).

For inter-annual variations of SWEmod, the sum of positive covariances prevails (Fig. 10a), whereas positive and negative values are more in balance for W (Fig. 10b). This suggests SWEmod anomalies to be more spatially coherent than anomalies of W. Thus, when spatially averaging, the more homogeneous snow anomaly patterns maintaining importance. On contrary, opposed anomalies of W compensate each other more strongly. This leads to a relatively larger influence of SWEmod to the spatially integrated inter-annual TWSmod variability compared to when analysing the local grid scale. Since positive covariation clearly dominates for temperature anomalies, the spatial coherence of SWEmod relates to their homogeneous patterns (Fig. 10c). Similar to W, positive covariances only slightly outweigh for year-to-year variations in rain fall (Fig. 10d). The same is true for snow fall (not shown). Therefore, the spatial coherence of SWEmod anomalies is less pronounced

Kommentiert [TT47]: referee
#1 / comment 32

than for temperature, as snow is additionally influenced by snow fall anomalies. Regarding W anomalies, this indicates that the spatial heterogeneity in our model, which misses explicit information on soils, topography, etc., mainly results from inhomogeneous patterns in rain fall anomalies. Thereby, the ~~greater slightly more~~ balanced ~~between~~ positive and negative covariations for W compared to rain fall can be ascribed to the additional impact of ~~primarily radiation-driven~~ evapotranspiration to SM. ~~Since evapotranspiration anomalies, as influenced by temperature, do not necessarily show the same spatial pattern as rain fall anomalies, the combined effect introduces more spatial variability in W than rain fall or evapotranspiration obtain on their own.~~

In order to ensure that these results are not artificially caused by the forcing data, we did the same analysis running the model with rain and snow fall estimates of the WFDEI product (Weedon et al., 2014). Since we observed the same patterns, 10 we assume our findings to be robust (S7.1).

~~Apart from the relatively larger contribution of SWEmod to the spatially integrated TWSmod on IAV scales, the spatial coherence of snow dynamics also explains the agreement between the average of local CRs and the CR of the spatially integrated TWSmod on MSC scales.~~

15

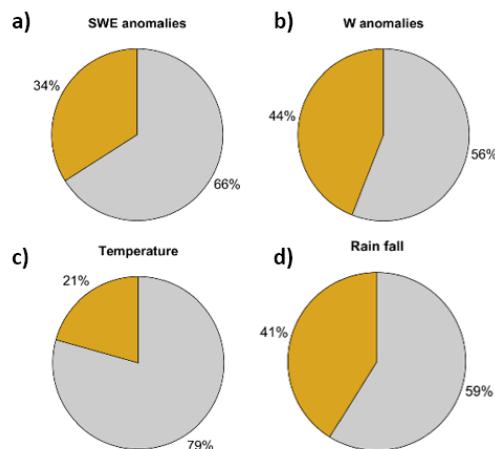


Figure 1017. Proportion of total positive (grey) and negative (orange) covariances among grid cells for inter-annual variations of **a)** snow (SWE), **b)** liquid water storages (W), **c)** temperature, and **d)** rain fall.

20 3.4 Limitations of the approach

Although the model of this study reproduces observed hydrological patterns well and achieves comparable results to state-of-the-art models, its low complexity and the applied calibration approach ~~imply some are associated with~~ limitations in terms of process understanding and predictive power.

First of all, the simple structure only allows inferences on represented processes, that likely include effects of fluxes and 25 storages not considered explicitly. For example, ~~the model does not resolve individual a more distinct partitioning of~~ liquid water storages ~~into its components such as soil moisture, such as~~ deep groundwater and surface water ~~explicitly, is not feasible with the current structure.~~ As discussed previously, ~~our~~ delayed land runoff comprises various (intermediate) storages and delay times, and thus cannot be associated with one distinct storage component. Even though soil moisture is distinguished from these slowly varying reservoirs, its quantity and pattern have not been directly validated. Future research 30 is required to increase confidence by including remote sensing based data of soil moisture in calibration and/or validation. However, these satellite data still have limited value as the microwave signals can only capture moisture in the upper 5 cm of

soil and do not provide estimates under snow-cover and dense vegetation (Döll et al., 2015; Lettenmaier et al., 2015). Therefore, a multi-layer soil scheme is needed to compare model outputs to satellite derived soil moisture estimates, as for example successfully demonstrated by Albergel et al. (2017).

Kommentiert [TT48]: referee
#1 / comment 34

Further, the model does not include any human-induced changes of water storages, which yet contribute to observed TWS variability in many regions (Döll et al., 2015; Rodell et al., 2015). Other simplified or ignored hydrological processes include the coincident occurrence of rain and snow fall, liquid water capacity of snow, interception, freeze/thaw dynamics within the soil, capillary rise and other surface-groundwater interactions, the effect of vegetation or other surface properties, as well as lateral flow from one grid cell to another. Especially in the downstream areas of large basins the latter represents a potential input that may significantly affect total TWS (Kim et al., 2009a), and thus may contribute to the discrepancy between TWSobs and TWSmod in some regions. Besides, the model does not account for spatial variability sub-grid heterogeneity of topography and land surface characteristics, ~~except for the fractional snow cover used to estimate snow melt and sublimation~~.

Kommentiert [TT49]: referee
#2 / comment 3

With regards to model parameter, we apply a global uniform parameter set and do not regionalize the parameters according to spatially distributed physio-geographical characteristics. In contrast, most macro-scale hydrological models include spatially distributed soil properties to define parameters related to infiltration, soil water holding capacity and percolation, as well as vegetation types to assess the effects of different plant functional types on evapotranspiration and canopy storage (Sood and Smakhtin, 2015). In contrast, our Our model only implicitly considers the effects of vegetation for example on ET, but not its spatial variability, as the associated impacts are included in the observational constraint. Spatial variability of model parameters might affect the relative contributions of different storage components to TWS variability at different spatial scales. However, the comparison with eartH2Observe models, which generally involve spatial heterogeneity in model parameters, suggests that the main conclusions remain unchanged. Additionally, However, we want to highlight that spatial distribution of model parameters depend on assumptions and some degree of simplification as well, and thus does not necessarily represent reality better than improve model performance compared to the a global uniform parameter set obtained from multiple observational data. Further, as we encountered issues with parameter equifinality, especially between modelled snow melt and sublimation, future efforts should include a stronger utilization of runoff data in the calibration and validation process. This would help to better constrain between water fluxes to the atmosphere and liquid water fluxes, that can contribute to the runoff.

Kommentiert [TT50]: referee
#1 / comment 1
referee #1 / comment 35

Finally, though the implemented cost function explicitly accounts for the uncertainty of the calibration data, additional uncertainties of other input data, their processing and characteristics remain partly unaddressed.

Kommentiert [TT51]: referee
#2 / comment 5

30 Conclusion

In this study, we assessed the relative contributions of snow pack versus soil and retained water variations to the variability of total terrestrial water storage (TWS) for northern mid-to-high latitudes. To do so, we constrained a parsimonious hydrological model with multi-criteria calibration against multiple Earth observation data streams including TWS from GRACE satellites and snow pack estimates from GlobSnow. The optimized model showed considerably good agreement with observed patterns of hydrological fluxes and states, and was found to perform comparable or better than simulations from state-of-the-art macro-scale hydrological models. This underlines the potential of simple hydrological models tied to observational data streams as powerful tools to diagnose and understand large scale water cycle patterns. Further, it highlights the benefits of considering multiple, complementary data constraints to overcome their individual shortcomings. Consistent with previous studies, we found that seasonal TWS variations are dominated by the development of snow pack during winter months in most places of the mid-to-high northern latitudes. In contrast to this seasonal pattern, our study reveals that not snow but anomalies in liquid water storages, mainly comprising soil moisture, drive inter-annual TWS

variations in almost the entire spatial domain. This counter-intuitive pattern was found to relate to larger rainfall anomalies as compared to snowfall anomalies.

Apart from the time-scale dependent dominant controls on TWS variations, we additionally observed ~~different verging~~ behaviour across spatial scales. In terms of seasonal variations, the spatially integrated contribution reflects the average of the spatial domain. However, and more interestingly, the relative contribution of snow pack variations to total TWS inter-annual anomalies appears to be larger when spatially integrated than at local scale. We found this pattern results from stronger spatial coherence of snow pack anomalies compared to anomalies in other storages, such that the latter cancel out more strongly than the former when calculating an average across large spatial domains. The stronger spatial coherence of snow pack anomalies seems to be driven by the nature of spatially coherent temperature anomalies that determine snow accumulation and melt. These findings imply that patterns from large scale integrated signals should not be associated with locally operating processes, since spatial covariations of climatic variables can confound the picture.

Overall, our study underlines the benefits of GRACE TWS as a model constraint, and moreover, stresses the importance of temporal and spatial scale when assessing the determinants of TWS variability. Clearly, insights obtained at one scale cannot be transferred to another, as is often (unintentionally) done. Hence, TWS variations in northern latitudes seem to be not merely subject to snow variability, but rather are driven by soil moisture on inter-annual scales - which may be of considerable importance when assessing long-term water availability in the context of climate changes.

Kommentiert [TT52]: referee #1/ comment 37

Authors contribution

TT, SK and MJ designed the research in extensive collaboration with NC, AE and MF. CN processed and integrated runoff estimates from EU-OBS in model calibration. NC contributed to parameter estimation and uncertainty analysis. TT conducted the analysis. All co-authors contributed to the preparation of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research was carried out within the initiative for the development of SINDBAD, the Strategies to Integrate Data and Biogeochemical models, framework.

References

A, G., Wahr, J., and Zhong, S.: Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada, *Geophysical Journal International*, 192, 557-572, 10.1093/gji/ggs030, 2013.

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., and Meurey, C.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8. 0: LDAS-Monde assessment over the Euro-Mediterranean area, *Geoscientific Model Development*, 10, 3889, 2017.

Alkama, R., Decharme, B., Douville, H., Becker, M., Cazenave, A., Sheffield, J., Voldoire, A., Tyteca, S., and Le Moigne, P.: Global Evaluation of the ISBA-TRIP Continental Hydrological System. Part I: Comparison to GRACE Terrestrial Water Storage Estimates and In Situ River Discharges, *Journal of Hydrometeorology*, 11, 583-600, 10.1175/2010jhm1211.1, 2010.

10 AMAP: Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017, Arctic Monitoring and Assessment Programme (AMAP), Oslo, Norway, xiv + 269, 2017.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, *Journal of Hydrometeorology*, 10, 623-643, 10.1175/2008jhm1068.1, 2009.

15 Bayer, P., and Finkel, M.: Optimization of concentration control by evolution strategies: Formulation, application, and assessment of remedial solutions, *Water Resources Research*, 43, 10.1029/2005WR004753, 2007.

Beck, H. E., Dijk, A. I. J. M. v., Roo, A. d., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, *Water Resources Research*, 52, 3599-3622, 10.1002/2015WR018247, 2016.

Behrangi, A., Christensen, M., Richardson, M., Lebsack, M., Stephens, G., Huffman, G. J., Bolvin, D., Adler, R. F., Gardner, A., 20 Lambrightsen, B., and Fetzer, E.: Status of high-latitude precipitation estimates from observations and reanalyses, *Journal of Geophysical Research: Atmospheres*, 121, 4468-4486, 10.1002/2015jd024546, 2016.

Bergström, S.: Principles and Confidence in Hydrological Modelling, *Nordic Hydrology*, 22, 123-136, 1991.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment 25 Simulator (JULES), model description – Part 1: Energy and water fluxes, *Geosci. Model Dev.*, 4, 677-699, 10.5194/gmd-4-677-2011, 2011.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, *Hydrological Processes*, 29, 310-320, 10.1002/hyp.10391, 2015.

Breiman, L.: Random Forests, *Machine Learning*, 45, 5-32, 10.1023/a:1010933404324, 2001.

Brown, J., Ferrians Jr, O., Heginbottom, J., and Melnikov, E.: Circum-Arctic map of permafrost and ground-ice conditions, US Geological Survey Reston, 1997.

Chen, X., Long, D., Hong, Y., Zeng, C., and Yan, D.: Improved modeling of snow and glacier melting by a progressive two-stage calibration 35 strategy with GRACE and multisource data: How snow and glacier meltwater contributes to the runoff of the Upper Brahmaputra River basin?, *Water Resources Research*, 53, 2431-2466, 10.1002/2016wr019656, 2017.

Cheng, M., Tapley, B. D., and Ries, J. C.: Deceleration in the Earth's oblateness, *Journal of Geophysical Research: Solid Earth*, 118, 740-747, 10.1002/jgrb.50058, 2013.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, 40 O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, *Geosci. Model Dev.*, 4, 701-722, 10.5194/gmd-4-701-2011, 2011.

d'Orgeval, T., Polcher, J., and de Rosnay, P.: Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes, *Hydrol. Earth Syst. Sci.*, 12, 1387-1401, 10.5194/hess-12-1387-2008, 2008.

Decharme, B., Martin, E., and Faroux, S.: Reconciling soil thermal and hydrological lower boundary conditions in land surface models, 45 *Journal of Geophysical Research: Atmospheres*, 118, 7819-7834, 10.1002/jgrd.50631, 2013.

Derkzen, C., Lemmettyinen, J., Toose, P., Silis, A., Pullainen, J., and Sturm, M.: Physical properties of Arctic versus subarctic snow: Implications for high latitude passive microwave snow water equivalent retrievals, *J Geophys Res-Atmos.*, 119, 7254-7270, 10.1002/2013jd021264, 2014.

Döll, P., Mueller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related 50 groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, *Water Resources Research*, 50, 5698-5720, 10.1002/2014wr015595, 2014.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, *Journal of Hydrology*, 270, 105-134, 2002.

Döll, P., Fiedler, K., and Zhang, J.: Global-scale analysis of river flow alterations due to water withdrawals and reservoirs, *Hydrol. Earth 55 Syst. Sci.*, 13, 2413-2432, 10.5194/hess-13-2413-2009, 2009.

Döll, P., Fritzsche, M., Eicker, A., and Müller Schmied, H.: Seasonal Water Storage Variations as Impacted by Water Abstractions: Comparing the Output of a Global Hydrological Model with GRACE and GPS Observations, *Surveys in Geophysics*, 35, 1311-1331, 10.1007/s10712-014-9282-2, 2014.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., and Wada, Y.: Modelling Freshwater Resources at the Global Scale: Challenges and 60 Prospects, *Surveys in Geophysics*, 37, 195-221, 10.1007/s10712-015-9343-1, 2015.

Draper, N., and Smith, H.: *Applied Regression Analysis*, John Wiley, New York, 1981.

Eichinger, W. E., Parlange, M. B., and Stricker, H.: On the concept of equilibrium evaporation and the value of the Priestely-Taylor coefficient, *Water Resources Research*, 32, 161-164, 1996.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., and Müller-Schmied, H.: Calibration/data assimilation approach for integrating GRACE data into the WaterGAP global hydrology model (WGHM) using an ensemble Kalman filter: First results, *Surveys in Geophysics*, 35, 1285-1309, 10.1007/s10712-014-9309-8, 2014.

Felfelani, F., Wada, Y., Longuevergne, L., and Pokhrel, Y. N.: Natural and human-induced terrestrial water storage change: A global analysis using hydrological models and GRACE, *Journal of Hydrology*, 553, 105-118, <https://doi.org/10.1016/j.jhydrol.2017.07.048>, 2017.

Feng, W., Zhong, M., Lemoine, J.-M., Biancale, R., Hsu, H.-T., and Xia, J.: Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements, *Water Resources Research*, 49, 2110-2118, 10.1002/wrcr.20192, 2013.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., and Alcamo, J.: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, *Global Environmental Change*, 23, 144-156, 10.1016/j.gloenvcha.2012.10.018, 2013.

Forman, B. A., Reichle, R. H., and Rodell, M.: Assimilation of terrestrial water storage from GRACE in a snow-dominated basin, *Water Resources Research*, 48, 10.1029/2011wr011239, 2012.

Girotto, M., De Lannoy, G. J. M., Reichle, R. H., and Rodell, M.: Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model, *Water Resources Research*, 52, 4164-4183, 10.1002/2015wr018417, 2016.

Gudmundsson, L., and Seneviratne, S. I.: Observation-based gridded runoff estimates for Europe (E-RUN version 1.1), *Earth System Science Data*, 8, 279-295, 10.5194/essd-8-279-2016, 2016.

Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K., and Merz, B.: A global analysis of temporal and spatial variations in continental water storage, *Water Resources Research*, 43, 10.1029/2006WR005247, 2007.

Güntner, A.: Improvement of Global Hydrological Models Using GRACE Data, *Surveys in Geophysics*, 29, 375-397, 10.1007/s10712-008-9038-y, 2008.

Hansen, N., and Kern, S.: Evaluating the CMA Evolution Strategy on Multimodal Test Functions, in: *Parallel Problem Solving from Nature - PPSN VIII*, edited by: Yao, X., Burke, E., Lozano, J. A., Smith, J., Merelo-Guervós, J. J., Bullinaria, J. A., Rowe, J., Tino, P., Kabán, A., and Schwefel, H.-P., Springer, Berlin, 2004.

The CMA Evolution Strategy: <https://www.lri.fr/~hansen/cmaesintro.html>, access: 20-03-2017, 2014.

He, Y., Bárdossy, A., and Zehe, E.: A review of regionalisation for continuous streamflow simulation, *Hydrol. Earth Syst. Sci.*, 15, 3539-3553, 10.5194/hess-15-3539-2011, 2011.

Huffman, G. J., Adler, R., Morrissey, M. M., Bolvin, D., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global Precipitation at One-Degree Resolution from Multisatellite Observations, *Journal of Hydrometeorology*, 2, 36-50, 2000.

Huffman, G. J., and Bolvin, D.: Version 1.2 GPCP One-Degree Daily Precipitation Data Set Documentation, NASA, 2013.

Humphrey, V., Gudmundsson, L., and Seneviratne, S. I.: Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes, *Surveys in Geophysics*, 37, 357-395, 10.1007/s10712-016-9367-1, 2016.

IPCC: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Geneva, Switzerland, 3-87, 2014.

Jakeman, A. J., and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, *Water Resources Research*, 29, 2637-2649, 1993.

Jung, M., Henkel, K., Herold, M., and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, *Remote Sensing of Environment*, 101, 534-553, <https://doi.org/10.1016/j.rse.2006.01.020>, 2006.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlstrom, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rodenbeck, C., Tramontana, G., Viovy, N., Wang, Y. P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO₂ sink changes to temperature, *Nature*, 541, 516-520, 10.1038/nature20780, 2017.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., and Woollen, J.: The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, 77, 437-471, 1996.

Kim, H., Yeh, P. J. F., Oki, T., and Kanae, S.: Role of rivers in the seasonal variations of terrestrial water storage over global basins, *Geophysical Research Letters*, 36, doi:10.1029/2009GL039006, 2009a.

Kim, H., Yeh, P. J. F., Oki, T., and Kanae, S.: Role of rivers in the seasonal variations of terrestrial water storage over global basins, *Geophysical Research Letters*, 36, n/a-n/a, 10.1029/2009GL039006, 2009b.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochemical Cycles*, 19, n/a-n/a, 10.1029/2003GB002199, 2005.

Kug, J.-S., Jeong, J.-H., Jang, Y.-S., Kim, B.-M., Folland, C. K., Min, S.-K., and Son, S.-W.: Two distinct influences of Arctic warming on cold winters over North America and East Asia, *Nature Geoscience*, 8, 759, 10.1038/ngeo2517, 2015.

Kumar, S. V., Zaichik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., Lannoy, G. D., Cosh, M. H., Hain, C. R., Anderson, M., Arsenault, K. R., Xia, Y., and Ek, M.: Assimilation of Gridded GRACE Terrestrial Water Storage Estimates in the North American Land Data Assimilation System, *Journal of Hydrometeorology*, 17, 1951-1972, 10.1175/jhm-d-15-0157.1, 2016.

Kustas, W. P., Rango, A., and Uijlenhoet, R.: A simple energy budget algorithm for the snowmelt runoff model, *Water Resources Research*, 30, 1515-1527, 1994.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F.: Inroads of remote sensing into hydrologic science during the WRR era, *Water Resources Research*, 51, 7309-7342, 10.1002/2015wr017616, 2015.

Liu, J., Li, Z., Huang, L., and Tian, B.: Hemispheric-scale comparison of monthly passive microwave snow water equivalent products, *Journal of Applied Remote Sensing*, 8, 10.1117/1.JRS.8.084688, 2014.

Long, D., Longuevergne, L., and Scanlon, B. R.: Global analysis of approaches for deriving total water storage changes from GRACE satellites, *Water Resources Research*, 51, 2574-2594, 10.1002/2014wr016853, 2015.

Luoju, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Derksen, C., and Wang, L.: *GlobSnow Snow Water Equivalent (SWE) Product Guide*, ESA, 2010.

5 Luoju, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Kangwa, M., Eskelinen, M., Metsämäki, S., Solberg, R., Salberg, A.-B., Bippus, G., Ripper, E., Nagler, T., Derksen, C., Wiesmann, A., Wunderle, S., Hüsler, F., Fontana, F., and Foppa, N.: *GlobSnow2 - Final Report*, 2014.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, *Hydrology and Earth System Sciences*, 15, 967-981, 10.5194/hess-15-967-2011, 2011a.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation 10 estimated from satellite-based observations, *Hydrology and Earth System Sciences*, 15, 453-469, 10.5194/hess-15-453-2011, 2011b.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, *Hydrology and Earth System Sciences*, 18, 3511-3538, 10.5194/hess-18-3511-2014, 2014.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models Part I - A discussion of principles, *Journal of Hydrology*, 15 10, 282-290, 1970.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate, *Journal of Climate*, 13, 2217-2238, 2000.

20 Ngo-Duc, T., Laval, K., Ramillien, G., Polcher, J., and Cazenave, A.: Validation of the land water storage simulated by Organising Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) with Gravity Recovery and Climate Experiment (GRACE) data, *Water Resources Research*, 43, 10.1029/2006WR004941, 2007.

Niu, G.-Y., Seo, K.-W., Yang, Z.-L., Wilson, C., Su, H., Chen, J., and Rodell, M.: Retrieving snow mass from GRACE terrestrial water storage change with a land surface model, *Geophysical Research Letters*, 34, 10.1029/2007gl030413, 2007.

Omlin, M., and Reichert, P.: A comparison of techniques for the estimation of model prediction uncertainty, *Ecological Modelling*, 115, 45-59, 10.1016/S0304-3800(98)00174-4, 1999.

25 Orth, R., Koster, R. D., and Seneviratne, S. I.: Inferring Soil Moisture Memory from Streamflow Observations Using a Simple Water Balance Model, *Journal of Hydrometeorology*, 14, 1773-1790, 10.1175/jhm-d-12-099.1, 2013.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, *Journal of Hydrology*, 523, 147-159, 10.1016/j.jhydrol.2015.01.044, 2015.

Priestley, C. H. B., and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, *Monthly Weather 30 Review*, 100, 81-92, 1972.

Ramillien, G., Lombard, A., Cazenave, A., Ivins, E. R., Llubes, M., Remy, F., and Biancale, R.: Interannual variations of the mass balance of the Antarctica and Greenland ice sheets from GRACE, *Global and Planetary Change*, 53, 198-208, 10.1016/j.gloplacha.2006.06.003, 2006.

Rangelova, E., van der Wal, W., Braun, A., Sideris, M. G., and Wu, P.: Analysis of Gravity Recovery and Climate Experiment time-variable 35 mass redistribution signals over North America by means of principal component analysis, *Journal of Geophysical Research: Earth Surface*, 112, n/a-n/a, 10.1029/2006JF000615, 2007.

Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, *Nature*, 460, 999-1002, 10.1038/nature08238, 2009.

Rodell, M., Beaudoin, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayton, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., 40 Schlosser, C. A., Sheffield, J., and Wood, E. F.: The Observed State of the Water Cycle in the Early Twenty-First Century, *Journal of Climate*, 28, 8289-8318, 10.1175/jcli-d-14-00555.1, 2015.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Spera Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the earth2Observe Tier-1 dataset, *Earth System Science Data*, 45 9, 389-413, 10.5194/essd-9-389-2017, 2017.

Schmidt, R., Petrovic, S., Guntner, A., Barthelmes, F., Wünsch, J., and Kusche, J.: Periodic components of water storage changes from GRACE and global hydrology models, *Journal of Geophysical Research: Solid Earth (1978-2012)*, 113, 10.1029/2007JB005363, 2008.

Seo, K.-W., Ryu, D., Kim, B.-M., Waliser, D. E., Tian, B., and Eom, J.: GRACE and AMSR-E-based estimates of winter season solid precipitation accumulation in the Arctic drainage region, *Journal of Geophysical Research*, 115, 10.1029/2009jd013504, 2010.

50 Sood, A., and Smakhtin, V.: Global hydrological models: a review, *Hydrological Sciences Journal*, 60, 549-565, 10.1080/02626667.2014.950580, 2015.

Sorooshian, S., Duan, Q., and Gupta, V. K.: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, *Water resources research*, 29, 1185-1194, 1993.

Stacke, T.: Development of a dynamical wetlands hydrology scheme and its application under different climate conditions, Max-Planck-55 Institute for Meteorology, Hamburg, 2011.

Swenson, S., Chambers, D., and Wahr, J.: Estimating geocenter variations from a combination of GRACE and ocean model output, *Journal of Geophysical Research*, 113, 10.1029/2007jb005338, 2008.

Swenson, S.: Assessing High-Latitude Winter Precipitation from Global Precipitation Analyses Using GRACE, *Journal of Hydrometeorology*, 11, 405-420, 10.1175/2009jhm1194.1, 2010.

60 Syed, T. H., Famiglietti, J. S., and Chambers, D. P.: GRACE-Based Estimates of Terrestrial Freshwater Discharge from Basin to Continental Scales, *Journal of Hydrometeorology*, 10, 22-40, 10.1175/2008jhm993.1, 2009.

Takala, M., Luojus, K., Pullainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, *Remote Sensing of Environment*, 115, 3517-3529, 10.1016/j.rse.2011.08.014, 2011.

Tallaksen, L. M., Burkhart, J. F., Stordal, F., Berntsen, T. K., Bryn, A., Etzelmüller, B., Hagen, J. O. M., Hamran, S.-E., Halvorsen, R., Kääb, A., Kristjánsson, J. E., Krüger, K., Lande, T. S., Schuler, T. V., Westermann, S., Wisland, D., and Xu, C.-y.: Land Atmosphere Interactions in Cold Environments (LATICe): The role of Atmosphere-Biosphere-Cryosphere-Hydrosphere interactions in a changing climate, EGU General Assembly, Vienna, Austria, 2015.

Tapley, B. D., Bettadpur, S., Watkins, M., and Reigber, C.: The gravity recovery and climate experiment: Mission overview and early results, *Geophysical Research Letters*, 31, n/a-n/a, 10.1029/2004gl019920, 2004.

10 Teuling, A. J., Seneviratne, S. I., Williams, C., and Troch, P. A.: Observed timescales of evapotranspiration response to soil moisture, *Geophysical Research Letters*, 33, 10.1029/2006gl028178, 2006.

Tramontana, G., Jung, M., Camps-Valls, G., Ichii, K., Raduly, B., Reichstein, M., Schwalm, C. R., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, *Biogeosciences Discussions*, 1-33, 10.5194/bg-2015-661, 2016.

15 Trischenko, A. P.: Removing Unwanted Fluctuations in the AVHRR Thermal Calibration Data Using Robust Techniques, *Journal of Atmospheric and Oceanic Technology*, 19, 1939-1953, 2002.

United Nations Environment, P.: World atlas of desertification / UNEP, United Nations Environment Programme, Accessed from <http://nla.gov.au/nla.cat-vn624121>, Edward Arnold, London ; Baltimore, 1992.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, *Water Resources Research*, 47, W07517, 10.1029/2010WR009791, 2011.

20 van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M., Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Materia, S., Law, R. M., and Sheffield, J.: LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome, *Geosci. Model Dev.*, 9, 2809-2832, 10.5194/gmd-9-2809-2016, 2016.

25 van der Knijff, J. M., Younis, J., and De Roo, A. P. J.: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation, *International Journal of Geographical Information Science*, 24, 189-212, 10.1080/13658810802549154, 2010.

van Dijk, A. I. J. M., and Warren, G.: The Australian Water Resources Assessment System. Technical Report 4. Landscape Model (version 0.5) Evaluation against observations, CSIRO, 2010.

30 van Dijk, A. I. J. M., Renzullo, L. J., Wada, Y., and Tregoning, P.: A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, *Hydrology and Earth System Sciences*, 18, 2955-2973, 10.5194/hess-18-2955-2014, 2014.

Viovy, N.: CRU-NCEPv6.1 Dataset, <http://dods.extra.cea.fr/data/p529viov/cruncep/>, 2015.

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, *Earth Syst. Dynam.*, 5, 15-40, 10.5194/esd-5-15-2014, 2014.

35 Wahr, J., Swenson, S., Zlotnicki, V., and Velicogna, I.: Time-variable gravity from GRACE: First results, *Geophysical Research Letters*, 31, n/a-n/a, 10.1029/2004gl019779, 2004.

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., and Landerer, F. W.: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons, *Journal of Geophysical Research*, 120, 2648-2671, 2015.

40 Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resources Research*, 50, 7505-7514, 10.1002/2014wr015638, 2014.

Werth, S., Güntner, A., Petrovic, S., and Schmidt, R.: Integration of GRACE mass variations into a global hydrological model, *Earth and Planetary Science Letters*, 277, 166-173, 10.1016/j.epsl.2008.10.021, 2009.

45 Werth, S., and Güntner, A.: Calibration analysis for water storage variability of the global hydrological model WGHM, *Hydrol. Earth Syst. Sci.*, 14, 59-78, 10.5194/hess-14-59-2010, 2010.

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B. I., Smith, L. G., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment, *Bulletin of the American Meteorological Society*, 77, 853-868, 1996.

Wiese, D. N.: GRACE monthly global water mass grids NETCDF RELEASE 5.0 Ver. 5.0 Mascon Ver. 2, PO.DAAC, CA, USA, 2015.

50 Wiese, D. N., Landerer, F. W., and Watkins, M. M.: Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution, *Water Resources Research*, 52, 7490-7502, 10.1002/2016wr019344, 2016.

Xie, H., Longuevergne, L., Ringler, C., and Scanlon, B. R.: Calibration and evaluation of a semi-distributed watershed model of Sub-Saharan Africa using GRACE data, *Hydrology and Earth System Sciences*, 16, 3083-3099, 10.5194/hess-16-3083-2012, 2012.

Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R., and Thomas, M.: Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations, *Hydrology and Earth System Sciences*, 21, 821-837, 10.5194/hess-55-21-821-2017, 2017.

REVISED SUPPLEMENT including marked changes

Supplement

5 S1 Detailed model description and formulas

The model consists of three components: (1) a snow component that simulates accumulation and ablation of snow, (2) a soil water component to calculate soil moisture, evapotranspiration and land runoff, and (3) a runoff component that derives total runoff. All modelled fluxes and states correspond to the spatio-temporal resolution of the forcing data, which in this study is a $1^\circ \times 1^\circ$ latitude/longitude grid and daily time steps.

10 The following describes all implemented processes and equations in detail.

S1.1 Snow Component

Snow storage is implemented as a simple accumulation and melt approach, which further is extended by consideration of sublimation and fractional snow cover. The snow storage as described by the snow water equivalent SWE [mm] at time t [d] is calculated as mass balance:

15

$$SWE_t = SWE_{t-1} + SF_t - ETSub_t - M_t \quad (S1)$$

where SWE_{t-1} [mm] is the snow water equivalent of the preceding time step which is increased by snowfall SF_t [mm d^{-1}] and reduced by the amount of sublimation $ETSub_t$ [mm d^{-1}] and snow melt M_t [mm d^{-1}].

All precipitation P [mm d^{-1}] is assumed to fall as snow at temperatures below 0°C . Since precipitation estimates, especially 20 during the cold season, are known for biases due to substantial under-catch (Rudolf and Rubel, 2005; Seo et al., 2010), P is scaled using the parameter p_{sf} to derive SF at time t :

$$SF_t = p_{sf} \cdot P_t \quad | \quad T < 0^\circ\text{C} \quad (S2)$$

In order to incorporate sub-grid variability, the fraction of the grid cell covered by snow is computed following the H-25 TESSEL approach (Balsamo et al., 2009; ECMWF, 2014):

$$FSC_t = \min\left(\frac{SWE_{t-1}}{sn_c}, 1\right) \quad (S3)$$

with fractional snow cover FSC [-] at time t being linearly dependent on SWE_{t-1} of the preceding time step and the parameter sn_c [mm] being the minimum SWE that ensures complete snow coverage of the grid cell.

30 Further, snow melt M and sublimation $ETSub$ are assumed to only emerge from snow covered area by using FSC as scaling factor in the calculation of these fluxes.

Snow melt M occurs when snow storage is present and temperature exceeds melting temperature. Based on the restricted degree-day radiation balance approach described by Kustas et al. (1994), melt M [mm d^{-1}] at time t depends on temperature 35 T_t [$^\circ\text{C}$] and net radiation Rn_t [$\text{MJ m}^{-2} \text{d}^{-1}$]:

$$M_t = (m_t \cdot T_t + m_r \cdot Rn_t) \cdot FSC_t \quad | \quad T > 0^\circ C \quad (S4)$$

where the degree-day factor m_t [mm $^\circ C^{-1}$] and the radiation factor m_r [mm MJ^{-1}] control the melt rate.

The derivation of snow sublimation ETSub is adapted from the approach implemented in the GLEAM model. This technique
5 is based on the Priestley and Taylor (1972) formula, which calculates evaporation rate as latent heat flux LE [$MJ m^{-2} d^{-1}$] based on the available energy Rn [$MJ m^{-2} d^{-1}$], ground heat flux G [$MJ m^{-2} d^{-1}$] and a dimensionless coefficient sn_a that parameterizes evaporation-resistance. LE at time t is derived by

$$LE_t = \left(sn_a \cdot \frac{\Delta sn_t}{\Delta sn_t + \gamma sn_t} \cdot (Rn_t - G) \right) \cdot FSC_t \quad (S5)$$

10 with Δsn_t being the slope of the temperature/saturated vapor pressure curve [$kPa K^{-1}$] and γsn_t representing the psychrometric constant [$kPa K^{-1}$]. Both, Δsn and γsn , are modified for snow covered areas according to Murphy and Koop (2005).

They calculate Δsn_t as a function of T_t [K] (Eq. (S6)), and γsn_t as a function of atmospheric pressure Pair [kPa], specific heat of air at constant pressure c_p [$MJ kg^{-1} K^{-1}$], the ratio molecular weight of water vapor/dry air MW and latent heat of sublimation of ice λsn [$MJ kg^{-1}$] (Eq. (S7)).

15

$$\Delta sn_t = \left(\frac{5723.265}{T_t^2} + \frac{3.53069}{T_t - 0.00728332} \right) \cdot e^{9.550426 - \frac{5723.265}{T_t} + 3.53068 \cdot \ln(T_t) - 0.00728332 \cdot T_t} \quad (S6)$$

$$\gamma sn_t = \frac{Pair \cdot c_p}{MA \cdot \lambda sn_t} \quad (S7)$$

In Eq.(S7), Pair is assumed to be time- and space-invariant with a uniform value of 101.3 kPa and $c_p = 0.001$ $MJ kg^{-1} K^{-1}$. MA is a constant of 0.622 and λsn is defined by Murphy and Koop (2005) as a function of T_t [K]. With a molecular mass of
20 water of 18.01528 g mol^{-1} , λsn can be calculated as:

$$\lambda sn_t = \left(46782.5 + 35.8925 \cdot T_t - 0.07414 \cdot T_t^2 + 541.5 \cdot e^{-\left(\frac{T_t}{123.75}\right)^2} \right) \cdot \frac{0.001}{18.01528} \quad (S8)$$

Since snow-covered ecosystems can be assumed to be unstressed due to the sufficient availability of water, LE corresponds to actual sublimation ETSub (Miralles et al., 2011). And ETSub [mm d^{-1}] can be converted from LE through division by λsn :

25

$$ETSub_t = \frac{LE_t}{\lambda sn_t} \quad (S9)$$

Altogether, the model calculates ETSub as a function of T_t , Rn_t , Pair, G , sn_a and FSC_t . While T_t , Rn_t and FSC_t are variable in time and space and depend on input data, the approach postulates constant Pair = 101.3 kPa and $G = 0$ $MJ m^{-2} d^{-1}$.

S1.2 Soil component

30 The central part of the model is the soil water component, which distributes input from rain fall and snow melt to soil water storage SM [mm], actual evapotranspiration ET [mm d^{-1}] and land runoff Qs [mm d^{-1}].

Like snow, the calculation of soil water storage as represented by soil moisture SM [mm] at time t follows the mass balance

$$SM_t = SM_{t-1} + In_t - ET_t \quad (S10)$$

with SM_{t-1} [mm] being the soil moisture of the preceding time step which is increased by infiltration In_t [mm d⁻¹] and reduced by actual evapotranspiration ET_t [mm d⁻¹].

5

On the one hand, the amount of infiltration In [mm d⁻¹] depends on the possible inflow IW [mm d⁻¹], which is the sum of rain fall RF (precipitation P if $T \geq 0^\circ\text{C}$) and snow melt M at time t :

$$IW_t = RF_t + M_t \quad (S11)$$

10 On the other hand, a part of IW may not infiltrate due to current soil moisture conditions but contribute to (direct) land runoff Qs [mm d⁻¹]. To estimate the partitioning of IW into SM and Qs , Qs at time t is calculated after Bergström (1995) as:

$$Qs_t = IW_t \cdot \left(\frac{SM_{t-1}}{s_{max}} \right)^{s_{exp}} \quad (S12)$$

15 In Eq. (S12) Qs_t depends on the inflow IW_t , the runoff coefficient s_{exp} and the actual soil moisture SM_{t-1} compared to its maximum water holding capacity s_{max} . Thus, no land runoff occurs if the soil water storage is empty and all IW is allocated to land runoff if the soil is completely saturated. Between these points, s_{exp} determines the amount of inflow that converts to Qs . While low values of s_{exp} lead to a high amount of Qs even if the soil moisture deficit is low (e.g. low SM/s_{max} ratio), higher values of s_{exp} increase the proportion of IW that infiltrates.

Infiltration In at time t is derived in accordance to the law of conservation of mass as:

20

$$In_t = IW_t - Qs_t \quad (S13)$$

Potential evapotranspiration $potET$ [mm d⁻¹] at time t is derived from net radiation Rn [MJ m⁻² d⁻¹] and air temperature T [°C] according to the Priestley-Taylor formula (Priestley and Taylor, 1972), where et_a is the alpha coefficient:

$$potET_t = et_a \cdot \left(\frac{\Delta_t}{\Delta_t + \gamma_t} \cdot \frac{Rn_t}{\lambda_t} \right) \quad (S14)$$

25

where Δ_t is the slope of the temperature/saturated vapor pressure curve [kPa K⁻¹], λ_t the latent heat of vaporization [MJ kg⁻¹] and γ_t the psychrometric constant [kPa K⁻¹].

The slope of the saturated vapor pressure curve Δ_t , as well as the latent heat of vaporization λ_t are functions of T at time t :

$$\Delta_t = \frac{4098 \cdot 0.611 \cdot e^{\frac{17.27 - T_t}{T_t + 237.3}}}{(T_t \cdot 237.3)^2} \quad (S15)$$

30

$$\lambda_t = 2.501 - (2.361 \cdot 10^{-3}) \cdot T_t \quad (S16)$$

Analogue to Eq. (S7), γ_t depends on a constant atmospheric pressure P of 101.3 kPa, the specific heat of air at constant pressure c_p [MJ kg⁻¹ K⁻¹], the constant MA and the latent heat of vaporization λ_t :

$$Y_t = \frac{Pair \cdot c_p}{MA \cdot \lambda_t} \quad (S17)$$

In order to avoid complete depletion of the soil water storage and to account for cohesion of water in the soil matrix, only a fraction of soil moisture after infiltration is assumed to be available for evapotranspiration. We express the sensitivity of 5 evapotranspiration to available water similar to Teuling et al. (2006) by the parameter et_{sup} . Thus, et_{sup} determines the portion of the sum of infiltration In_t [mm d⁻¹] and soil moisture SM_{t-1} [mm], that represents evapotranspiration supply $supET$ [mm d⁻¹] at time t:

$$supET_t = et_{sup} \cdot (SM_{t-1} + In_t) \quad (S18)$$

10 Finally, actual evapotranspiration ET [mm d⁻¹] at time t is derived by comparing $potET_t$ [mm d⁻¹] and $supET_t$ [mm d⁻¹]:

$$actET_t = \min(potET_t, supET_t) \quad (S19)$$

S1.3 Runoff component

As total runoff comprises fast direct runoff as well as delayed interflow and base flow, it's appropriate to consider retardation (Orth et al., 2013). Accordingly, total runoff Q [mm d⁻¹] at time t results from the accumulated effects of all land 15 runoff Q_s [mm d⁻¹] generated during the preceding 60 time steps:

$$Q_t = \sum_{i=0}^{60} Q_s_{t-i} \cdot \underbrace{\left[e^{-\frac{1}{q_i t}} - e^{-\frac{1+i}{q_i t}} \right]}_{\text{delay component}} \quad (S20)$$

where the recession time scale q_i [d] determines how quickly land runoff is transformed into streamflow. In theory, an 20 infinite number of time steps would be necessary to ensure that all generated Q_s is transformed into Q . However, the arbitrary number of 60 days allows accounting for > 99 % of Q_s (Orth et al., 2013), as long as q_i is below 13 days. To allow longer recession times when calibrating the model and still account for > 99 % of Q_s within the 60 days-window, the delay component of Eq. (S20) is scaled with its sum.

Introducing temporal delay leads to retention of a portion of Q_s , and thus to an additional, temporal storage of retained water 25 RW [mm]. The change of retained water storage ΔRW [mm d⁻¹] at time t can be inferred using the water balance:

$$0 = P_t - actET_t - Q_t + \Delta TWS_t \quad (S21)$$

with the change of total water storage ΔTWS [mm d⁻¹] resulting from

$$\Delta TWS = (SWE_t - SWE_{t-1}) + (SM_t - SM_{t-1}) + W_t \quad (S22)$$

30

so that solving Eq. (S21) and Eq. (S22)

$$\Delta RW_t = actET_t + Q_t - P_t - (SWE_t - SWE_{t-1}) - (SM_t - SM_{t-1}) \quad (S23)$$

The amount of retained water RW_t [mm] at time t then results from

$$RW_t = RW_{t-1} + \Delta RW_t \quad (S24)$$

Finally, the integrated terrestrial water storage TWS [mm] at time t represents the sum of all storage components:

5

$$TWS_t = SWE_t + SM_t + RW_t \quad (S25)$$

S2 Uncertainty of the observational constraints

Kommentiert [TT53]: referee #2 / comment 9

Maps of the temporal average uncertainties of observed TWS , ET and Q that are used for model calibration are shown in Figure S1. For observed SWE a constant average uncertainty of 35 mm is applied.

10

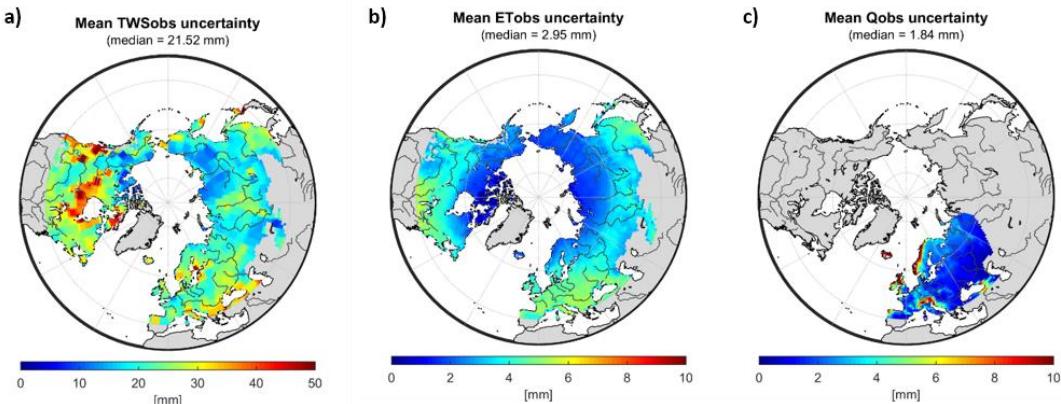


Figure S1. Mean uncertainty of monthly TWS_{obs} [mm], and of the mean seasonal cycle of ET_{obs} [$mm\ d^{-1}$] and Q_{obs} [$mm\ d^{-1}$] used for model calibration. [Values are truncated to 50 mm resp. 10 mm.]

Kommentiert [TT54]: referee #1 / comment 23

S3 Cost terms

Kommentiert [TT55]: referee #2 / comment 14

Table S1 shows the cost terms achieved with the default and the optimized parameter set. Compared to the default parameter values, total costs clearly improve after calibration. The shown optimized values represent a weighted Nash-Sutcliffe efficiency of 0.37 (TWS), 0.44 (SWE), 0.57 (Q) and 0.80 (ET) (weighted Nash-Sutcliffe = 1 – cost value).

20

Table S1. Cost values obtained with the default and the optimized model parameters using Eq. (1).

parameter values	TWS	SWE	ET	Q	total
default	0.84	0.54	0.15	1.00	2.55
optimized	0.63	0.56	0.20	0.43	1.82

S4 Model performance regarding evapotranspiration and runoff

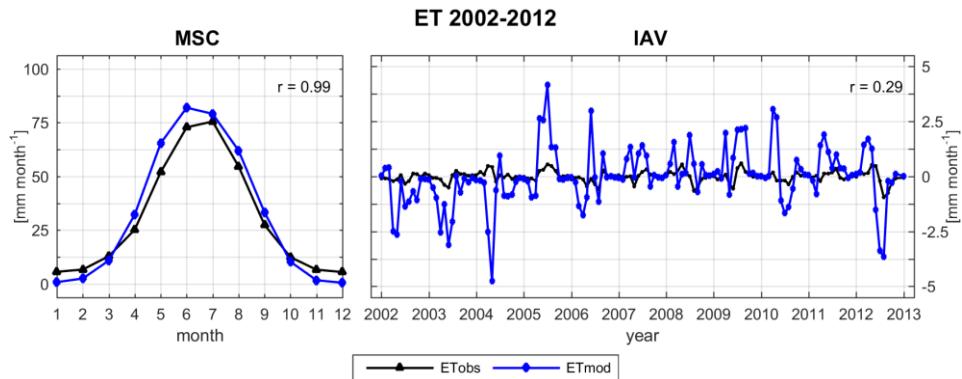


Figure S2. Spatially averaged mean seasonal cycle (MSC) of the period 2002–2012 and inter-annual variability (IAV, difference between monthly values and the MSC) for ETmod and FLUXCOM based ETobs.

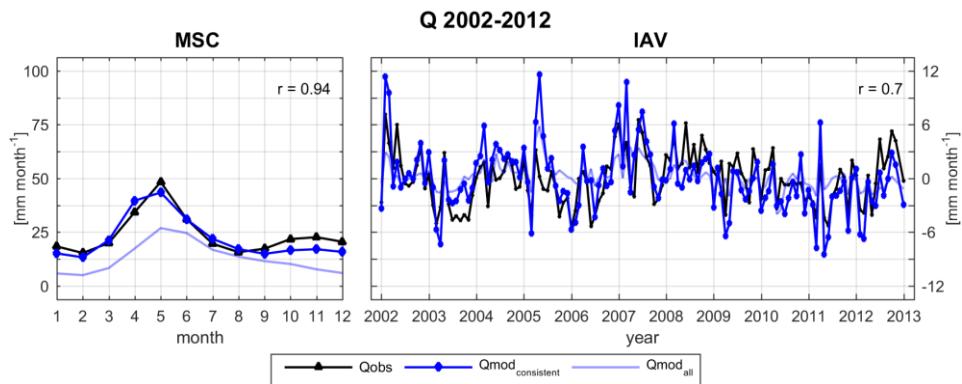


Figure S3. Spatially averaged mean seasonal cycle (MSC) of the period 2002–2012 and inter-annual variability (IAV, difference between monthly values and the MSC) for Qmod and EU-grid runoff Qobs. Qmod_{consistent} solely considers grid cells that coincide with Qobs, while Qmod_{all} is based on modelled runoff for all grids of the study domain.

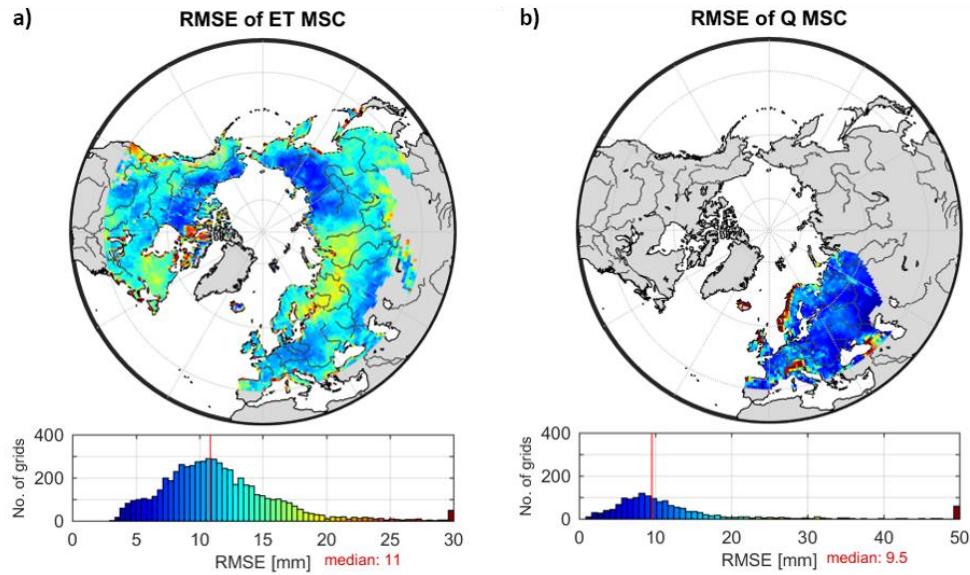


Figure S4. RMSE of the mean seasonal cycle of simulated and observed a) ET [mm month^{-1}] and b) Q [mm month^{-1}]. RMSE values have been truncated to the range 0–30 (a) resp. 0–50 (b).

Kommentiert [TT56]: referee #2 / comment 15
referee #1 / comment 23

5 S5 Phase shift in mean seasonal TWS

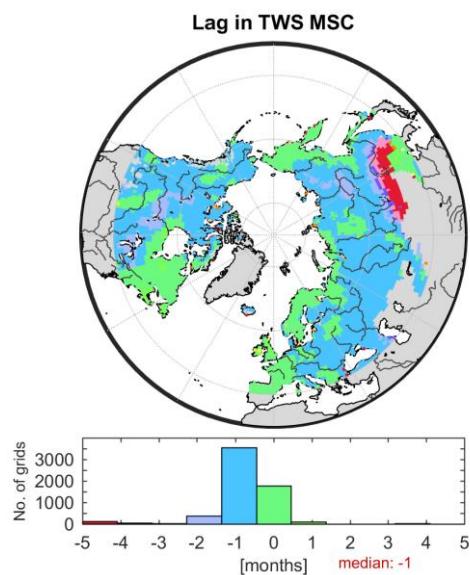


Figure S5. Grid wise phase lag [months] between mean seasonal TWSobs and TWSmod. Negative values indicate preceding of the model compared to GRACE TWS.

S6 Comparison with earthH2Observe models

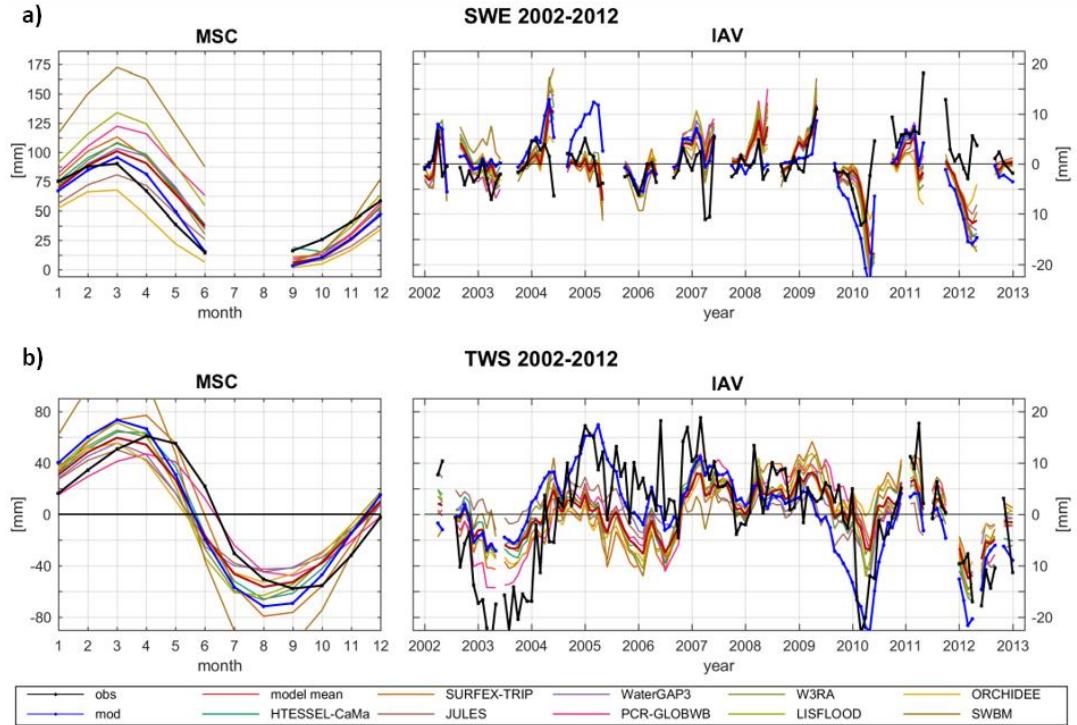


Figure S6. Comparison of spatially averaged observed (obs) **a**) SWE (GlobSnow) and **b**) TWS (GRACE) to simulations of this study (mod) and earthH2Observe models (incl. ensemble mean) in terms of average mean seasonal cycle (MSC) and inter-annual variability (IAV). MSC is calculated for the period 2002–2012, and IAV represents the difference of monthly values from the MSC. Only data points consistent between all models and the respective observational data are considered.

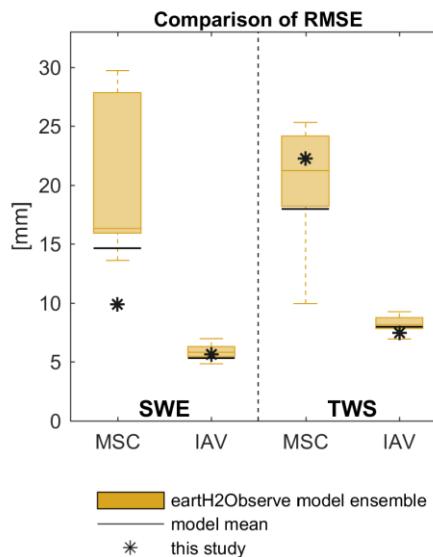


Figure S7. RMSE for the spatially averaged SWE (left) and TWS (right) achieved by our model compared to the model ensemble of earthH2Observe models and the ensemble mean across temporal scales.

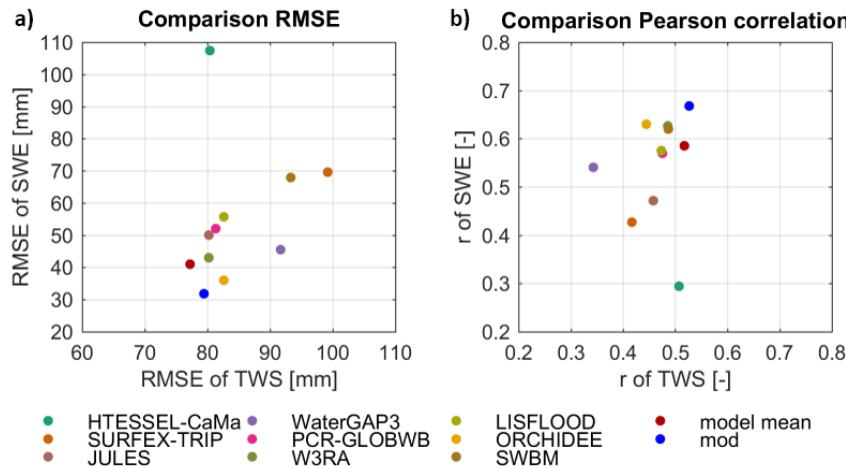


Figure S8. Comparison of a) RMSE and b) Pearson correlation r for monthly SWE and TWS time series simulated with the earth2Observe models, the model ensemble mean (model mean) and by our model (mod).

5 S7 Uncertainty due to forcing and calibration data

S7.1 Comparison to WFDEI precipitation forcing

To assess the uncertainty in TWSmod and SWEmod that emerges from the choice of precipitation forcing, we calibrated our model in the same manner as before, yet used rain fall and snow fall estimates from the reanalysis based WFDEI product (Weedon et al. 2014) instead of GPCP-1DD precipitation data. Since precipitation is likely the most uncertain input data

10 (Herold et al. 2015, Schellekens et al. 2017), we did not change the temperature and net radiation data sets. The global meteorological WFDEI data for land area is generated by applying the Water and Global Change (WATCH) forcing data methodology to ERA-Interim reanalysis data (Dee et al. 2011). The advantage of the WFDEI product is that it already provides separate values for snow and rain fall, as diagnosed by the reanalysis (Weedon et al. 2014). Therefore, it is not necessary to partition precipitation based on a temperature threshold within the model. We rather applied the provided rain

15 and snow fall estimates directly, and also desisted from scaling snow fall.

Regarding the MSC, we obtained similar model performance in terms of SWE and TWS for both, the spatially averaged dynamics (Figure S9, Figure S10) and the spatial pattern (not shown). Although the dynamics and thus the correlation coincide, we obtain a higher amplitude in TWSmod when using WFDEI as forcing compared to the original TWSmod

20 (and TWSobs). This higher amplitude relates to larger seasonal snow accumulation in SWEmod_{WFDEI}, because the scaling parameter for snow fall is not calibrated. In terms of IAV, the correlation between observation and WFDEI forced model is comparable for both, TWS and SWE. However, the key findings (Figure S11) remain the same as with GPCP precipitation forcing.

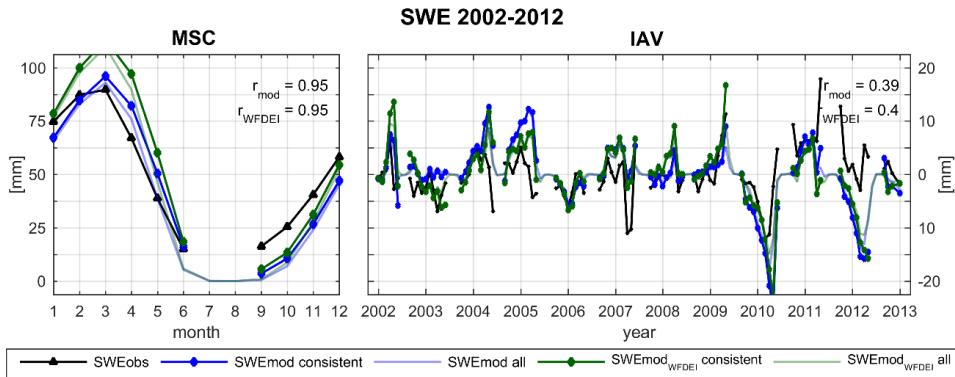


Figure S9. Comparison of the spatially averaged mean seasonal cycle (MSC) and inter-annual variability (IAV, difference between monthly values and the MSC) of observed SWE (SWEobs), modelled SWE (SWEmod), and modelled SWE based on WFDEI precipitation forcing (SWEmod_{WFDEI}). SWEmod consistent and SWEmod_{WFDEI} consistent refers to modelled SWE considering only data points with available SWEobs, while SWEmod all and SWEmod_{WFDEI} all incorporates all time steps for all grids of the study domain.

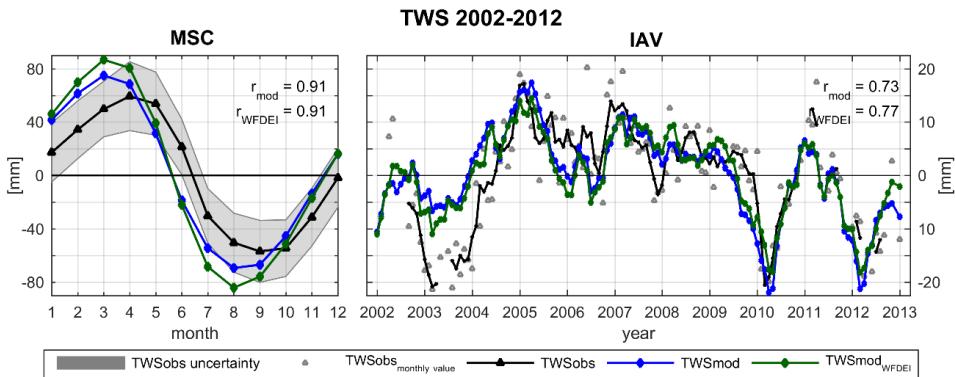


Figure S10. Comparison of the spatially averaged mean seasonal cycle (MSC) and inter-annual variability (IAV, difference between monthly values and the MSC) of observed TWS (TWSobs), modelled TWS (TWSmod), and modelled TWS based on WFDEI precipitation forcing (TWSmod_{WFDEI}). For IAV, TWSobs_{monthly value} shows the original IAV of individual TWSobs months, while TWSobs, TWSmod and TWSmod_{WFDEI} are smoothed using a 3-month average moving window filter. Pearson correlation r refers to the smoothed values. For the MSC no smoothing is applied.

Kommentiert [TT57]: referee
#1 / comment 3

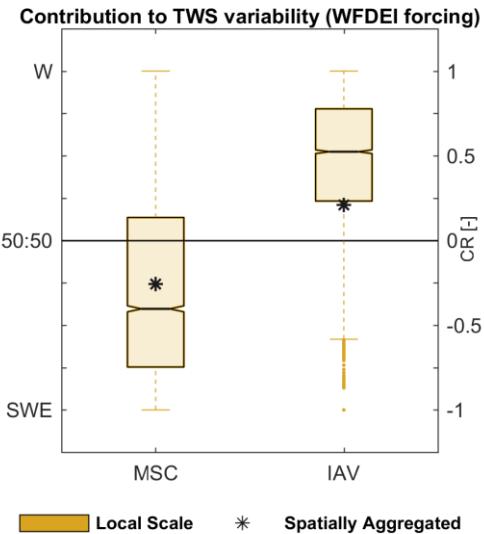


Figure S11. Relative contribution (based on CR (Eq.(3))) of snow (SWE) and liquid water (W) to TWS variability on different spatial (local grid scale, spatially integrated) and temporal (mean seasonal MSC, inter-annual IAV) scales when forced with WFDEI rain and snow fall. The boxplots represent the distribution of grid cell CR, with the dashed line marking the corresponding average. The star represents the CR calculated for the spatially integrated values.

Kommentiert [TT58]: referee
#1 / comment 3

S7.2 Comparison to other GRACE solutions

In this study we used TWS estimates from the JPL mascon RL05 product for model calibration and evaluation (Watkins et al., 2015; Wiese, 2015). However, various GRACE solutions for TWS from different institutions and using different processing approaches exist. To assess the potential uncertainty resulting from the choice of TWS solution, we compared modelled TWS (mod) and the JPL mascon solution (JPL_{masc}) with other solutions based on different processing approaches. They include the mascon product from the Center of Space Research (CSR at the University of Texas) (CSR_{masc}) (Save et al., 2016), as well as three RL05 solutions based on spherical harmonics provided by JPL, CSR and GeoforschungsZentrum (GFZ) (Swenson and Wahr, 2006; Landerer and Swenson, 2012; Swenson, 2012). As recommended, we also considered the average of the latter three ($Avg_{JPL/CSR/GFZ}$). All TWS estimates were taken as anomalies to the respective time-mean of 2002–2012, and scaled with the provided gain factors (except for CSR_{masc} that does not require scaling (Save et al., 2016). For comparison, we calculated the spatial average mean seasonal cycle (MSC) and inter-annual variability (IAV) across all grid cells of the study domain (Figure S12).

Thereby, we find that the spatial average MSC of all GRACE TWS estimates agrees in its dynamics, albeit minor differences in the solutions' amplitudes exist (by ± 15 mm). This results in comparable correlation and RMSE with modelled TWS. As the signal itself is noisier on IAV scales, the GRACE solutions show broader variability for IAV than at MSC scales as well. However, the qualitative pattern between the solutions remains, and modelled TWS is not closer to one specific solution or another during the entire time period. Therefore, the uncertainty evolving from the choice of GRACE solution used for model calibration can be assumed to be minor.

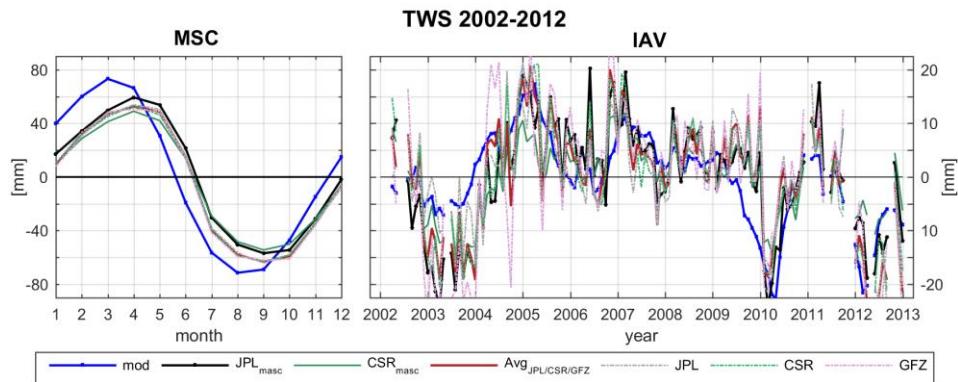


Figure S12. Comparison of the spatially averaged mean seasonal cycle (MSC) and inter-annual variability (IAV, difference between monthly values and the MSC) of modelled TWS (mod) and observed TWS of different GRACE solutions.

S8 Covariances between SWE and W

Kommentiert [TT59]: referee #1 / comment 8

Figure S13 and Figure S14 compare the contribution of the combined SWE and W variances and the covariance of both storages to the total variance of the spatially aggregated TWSmod. On the interannual and spatially aggregated scale, 81 % of TWS variability is explained by the variances in SWE and W, suggesting that the covariance between SWE and W only has minor effect. This is underlined by high percentage of SWE and W variance on total TWSmod variance for all grids of the study domain (Figure S14). On mean seasonal scales, the majority of spatially aggregated TWS variability is still explained by variances in SWE and W, but the contribution of the covariance increases. This can be expected, as the seasonal variation of snow storage affects the subsequent availability of liquid water storages through the snowmelt process. At the local scale, though, the percentage of SWE and W variance on total TWSmod variance remains high in regions where the dominance of either snow or liquid water components are clear (Fig. 7 of the manuscript). In regions where covariances of two storage components is larger, the contribution of two storage components to TWS variability are similar resulting in a CR value of around 0. Therefore, we conclude that while the covariances of snow and liquid water can be remarkable on the seasonal scale over a large spatial domain, it does not affect or change the dominant components on the TWS.

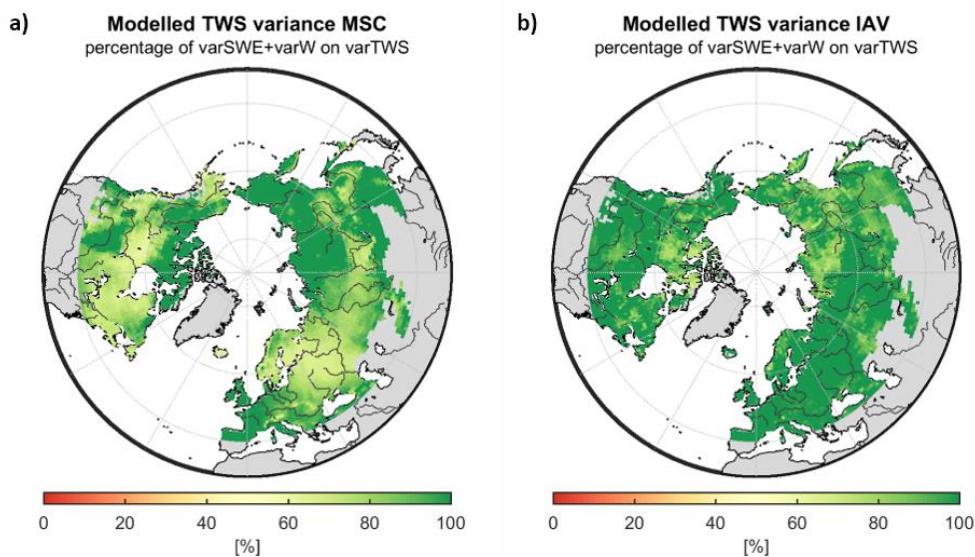


Figure S13. Percentage of SWE and W variance on total TWSmod variance on mean seasonal (MSC) and interannual (IAV) scales.

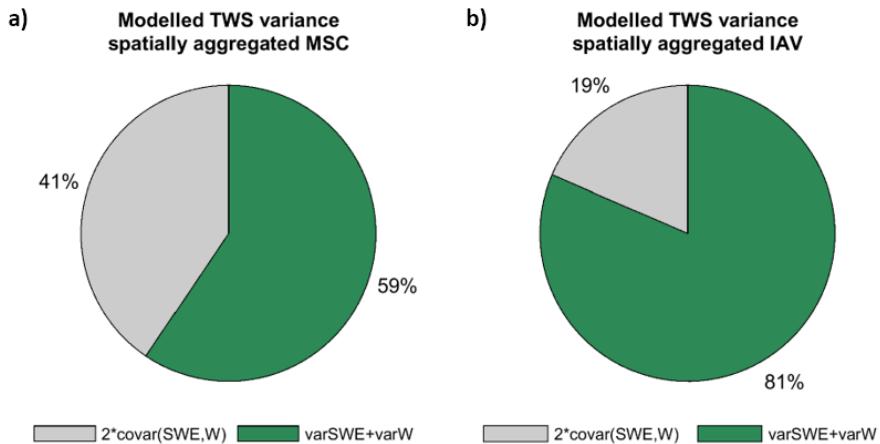


Figure S14. Percentage composition of spatially aggregated TWSmod variance from the combined variances of SWE and W, and two times the covariance of SWE and W on mean seasonal (MSC) and interannual (IAV) scales.

References Supplement

5 Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, *Journal of Hydrometeorology*, 10, 623-643, 10.1175/2008jhm1068.1, 2009.

Bergström, S.: The HBV model, in: *Computer models of watershed hydrology*, edited by: Singh, V., 443-476, 1995.

ECMWF: IFS Documentation - Cy40r1: Part IV: Physical Processes, in, European Centre for Medium-Range Weather Forecasts, Reading, 10 England, 2014.

Kustas, W. P., Rango, A., and Uijlenhoet, R.: A simple energy budget algorithm for the snowmelt runoff model, *Water Resources Research*, 30, 1515-1527, 1994.

Landerer, F. W., and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resources Research*, 48, 10.1029/2011wr011453, 2012.

15 Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, *Hydrology and Earth System Sciences*, 15, 453-469, 2011.

Murphy, D., and Koop, T.: Review of the vapour pressures of ice and supercooled water for atmospheric applications, *Quarterly Journal of the Royal Meteorological Society*, 131, 1539-1565, 2005.

Orth, R., Koster, R. D., and Seneviratne, S. I.: Inferring soil moisture memory from streamflow observations using a simple water balance 20 model, *Journal of Hydrometeorology*, 14, 1773-1790, 2013.

Priestley, C., and Taylor, R.: On the assessment of surface heat flux and evaporation using large-scale parameters, *Monthly Weather Review*, 100, 81-92, 1972.

Rudolf, B., and Rubel, F.: Global Precipitation, in: *Observed Global Climate* edited by: Hantel, M., Geophysics Springer, Berlin, 567, 2005.

25 Save, H., Bettadpur, S., and Tapley, B. D.: High-resolution CSR GRACE RL05 mascons, *Journal of Geophysical Research: Solid Earth*, 121, 7547-7569, 10.1002/2016jb013007, 2016.

Seo, K. W., Ryu, D., Kim, B. M., Waliser, D. E., Tian, B., and Eom, J.: GRACE and AMSR-E-based estimates of winter season solid precipitation accumulation in the Arctic drainage region, *Journal of Geophysical Research: Atmospheres*, 115, 2010.

30 Swenson, S., and Wahr, J.: Post-processing removal of correlated errors in GRACE data, *Geophysical Research Letters*, 33, 10.1029/2005gl025285, 2006.

Swenson, S.: GRACE monthly land water mass grids NETCDF RELEASE 5.0 CA, USA, <http://dx.doi.org/10.5067/TELND-NC005>, 2012.

Teuling, A. J., Seneviratne, S. I., Williams, C., and Troch, P. A.: Observed timescales of evapotranspiration response to soil moisture, *Geophysical Research Letters*, 33, 10.1029/2006gl028178, 2006.

35 Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., and Landerer, F. W.: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons, *Journal of Geophysical Research*, 120, 2648-2671, 2015.

Wiese, D. N.: GRACE monthly global water mass grids NETCDF RELEASE 5.0 Ver. 5.0 Mascon Ver. 2, PO.DAAC, CA, USA, 2015.