Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-690-AC2, 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Understanding terrestrial water storage variations in northern latitudes across scales" *by* Tina Trautmann et al.

Tina Trautmann et al.

ttraut@bgc-jena.mpg.de

Received and published: 9 April 2018

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

PHASE LAG OF SEASONAL TWS

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.

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.

*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.

We include 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. We further highlight the limitation that some potentially relevant processes are not yet accounted for in the current model setup and add the following paragraph in the discussion of the revised manuscript.

Performance of the spatially integrated simulations: [...] 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 and a 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. Additionally, high uncertainties of the precipitation forcing and GlobSnow SWE [...]

Limitations of the approach: [...] 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 growth, as well as lateral flow from one grid cell to another.

MISLEADING TERMINOLOGY SNOW VS. LIQUID WATER

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.

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 elucidate that liquid water storages might implicitly include frozen water especially in the observation.

[. . .] The amount of water storages in retained land runoff (RW) and SM represents the liquid water storage (W). Frozen water, e.g. in soil, is not explicitly included in the model, yet might implicitly be accounted for in W after model calibration.

EFFECT OF PRECIPITATION FORCING

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.

We thank the reviewer for pointing to Figure S7. Doing the analysis that he suggested, we found that in Figure 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. TWSmodWFDEI 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 include Figure 6 in the supplement of the revised manuscript.

CLARITY OF ABSTRACT

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.

We thank the referee for highlighting this lack of clarity in the abstract. We revise the abstract accordingly:

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

Further, as the referee suggested, we include a comparison with the results of Güntner et al. 2007 in the discussion of the revised manuscript.

Minor Comments

MODELLING FRAMEWORK

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?) I f not, this would better show the merit and relevance of the presented approach.

The referee is right in that our modelling framework still relies on a standard land sur-

face 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 highlight the merits of using remote sensing data, as well as potential future research in identifying functional relationships directly from observations.

Comparison with the eartH2Observe model ensemble: [. . .] Compared to the model simulations in the EartH2Observe ensemble, our modeling framework assimilates information from more data streams, e.g. GRACE and GlobSnow data. Even though we only used a subset of 1000 random grid cells to constrain the model parameters, 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 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.

CLARITY OF MODELLED LIQUID WATER AND MODELLED RUNOFF GENERATION

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.

Thanks for pointing this out, we adjust the model description in the revised manuscript as following:

As land runoff is generated with an effective infiltration excess formulation, this excess runoff is essentially all the water that cannot be stored in soil water storage, and thus implicitly contains both, surface runoff as well as the percolation to deeper water storages such as groundwater. Therefore, we use an exponential delay function (Orth et al., 2013) to mimic runoff contributions from slow-varying storages, such as groundwater and surface water bodies. After model calibration, this retained land runoff (RW) is supposed to implicitly include 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).

DISTINCTION OF OBSERVATIONAL PRODUCTS

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.

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.

COVARIANCE OF SWE AND W

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)?

We agree with the referee that the potential implication of this assumption should be discussed. We therefore include a short discussion (see below) on the effect of the covariances between SWE and W on TWS variability in the supplement of the revised manuscript.

Figure S9 compares 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 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 (Fig. S9). 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.

In-Text Comments

FLUXCOM ET

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?

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 clarify this in the revised manuscript.

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 machine learning algorithms (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 ET realization. To calculate ETobs [mm d-1], we assume a constant latent heat of vaporization of 2.45 MJ m-2.

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.

INACCURATE SENTENCE

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).

Thank you very much for pointing this out! We change the sentence accordingly to: Finally, the calibrated recession time scale that delays land runoff is 13 days (qt). Compared to much smaller alpine catchments for which Orth et al. (2013) reported qt of 2 days, the longer delay coefficients are reasonable at a large spatial resolution of 1° x 1° grids, because the elevation gradients are much smaller within a large spatial area.

GLOBAL UNIFORM PARAMETER VALUES

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).

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 reformulate the sentence accordingly:

[. . .] Since such approaches are not commonly accepted, macro-scale models mostly apply a priori parameter values based on empirical relationship or on expert knowledge that may lead to suboptimal model simulations (Beck et al., 2016;Sood and Smakhtin, 2015).

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 include this possible caveat when discussing the limitations of the approach on page 27 line 24:

[. . .] Considering the 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 partly involve spatial heterogeneity in model parameters, suggests that the main conclusions should remain unchanged. Additionally, we want to highlight [. . .]

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. Sood, A., and Smakhtin, V.: Global hydrological models: a review, Hydrological Sciences

Journal, 60, 549-565, 10.1080/02626667.2014.950580, 2015.

TRUNCATION OF VALUES IN FIGURES

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

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

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.

We add quantitative labels of CR in the revised manuscript.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690, 2017.

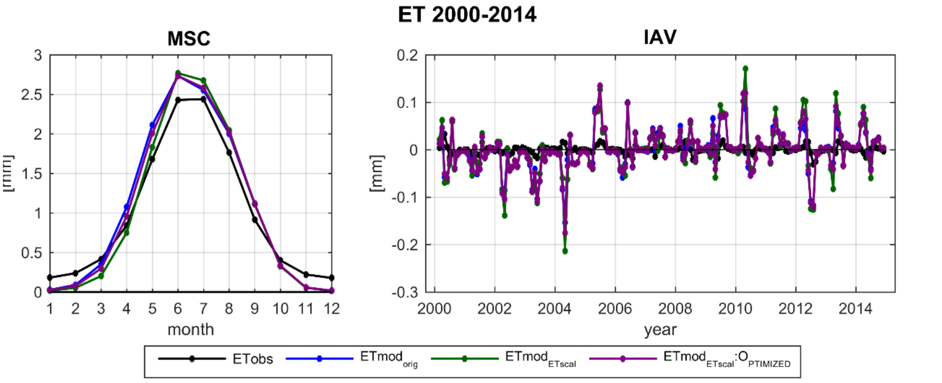

Fig. 1. Comparison of MSC and IAV of ETobs, ETmodorig (as in manuscript), ETmodETscal (parameter as in manuscript but actETscaled) and ETmodETscalOPTIMIZED (actET scaled and optimized)

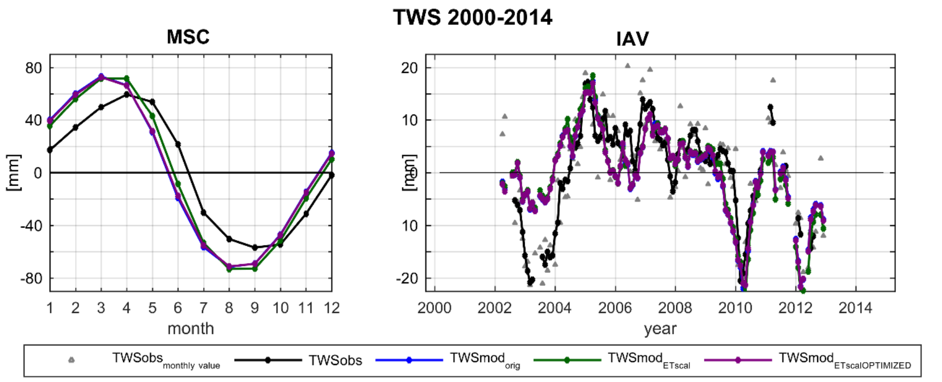

Fig. 2. Comparison of MSC and IAV of TWSobs, TWSmodorig (as in manuscript), TWSmod-ETscal (parameter as in manuscript but actETscaled) and TWSmodETscalOPTIMIZED (actET scaled and optimized)

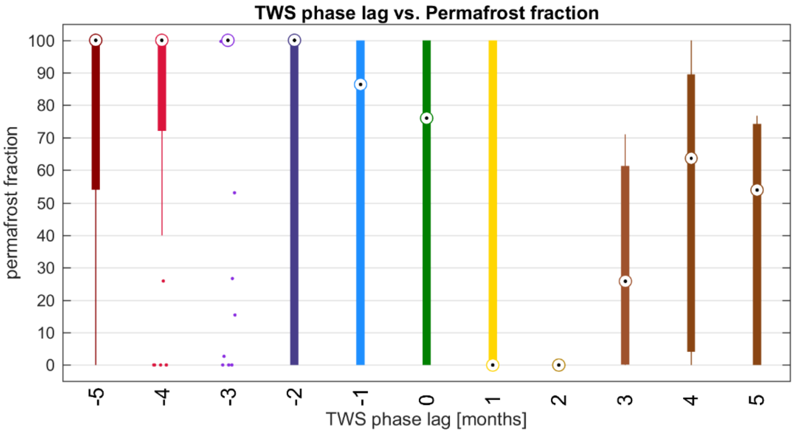

Fig. 3. TWS phase lag compared to the permafrost fraction of the grid cell (colors relate to the TWS lag class)

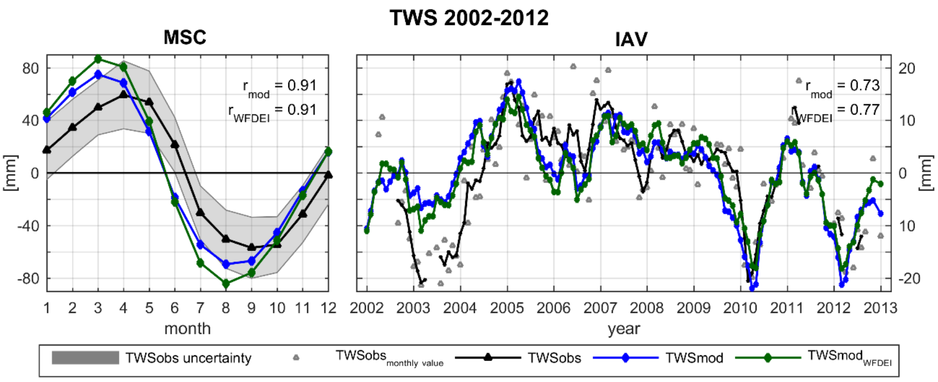

Fig. 4. Comparison of the mean seasonal cycle and interannual variability of TWSobs (GRACE), TWSmod (forced with GPCP precipitation) and TWSmodWFDEI (forced with WFDEI rain and snow fall)

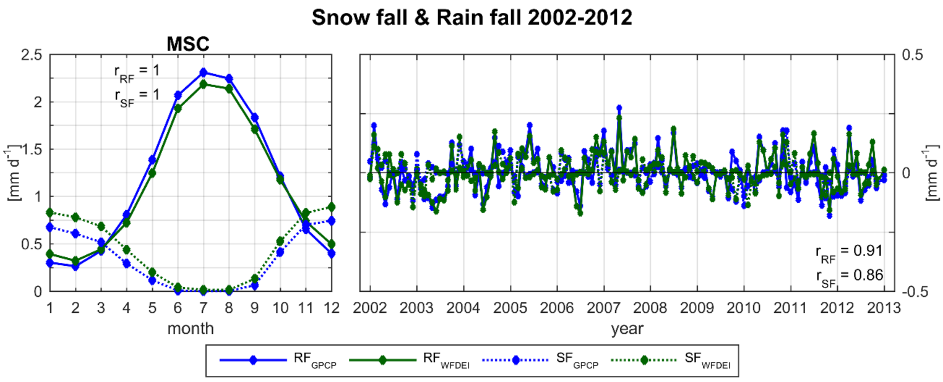

Fig. 5. Comparison of the mean seasonal cycle and interannual variability of rain fall and snow fall from WFEI product and from GPCP (snow fall as in the optimized model)

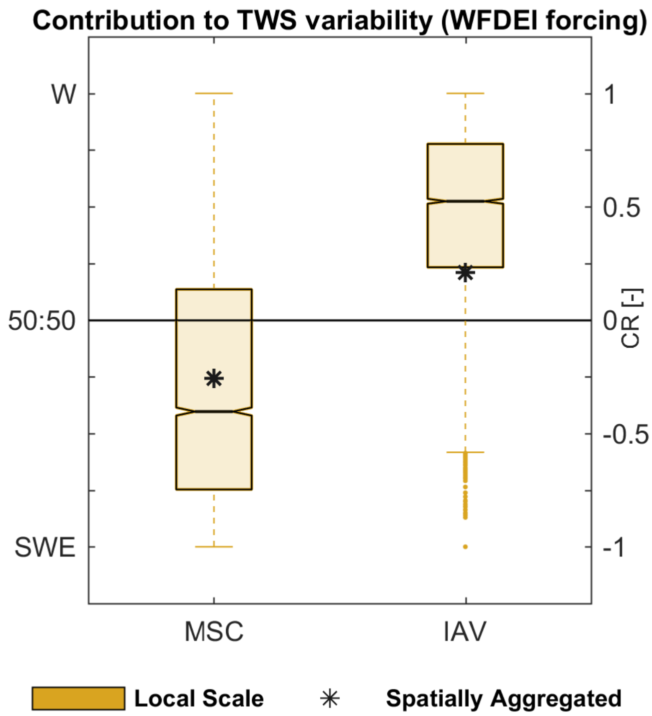

Fig. 6. Relative contribution of snow (SWE) and liquid water (W) to TWS variability when forced with WFDEI snow and rainfall on different spatial and temporal scales

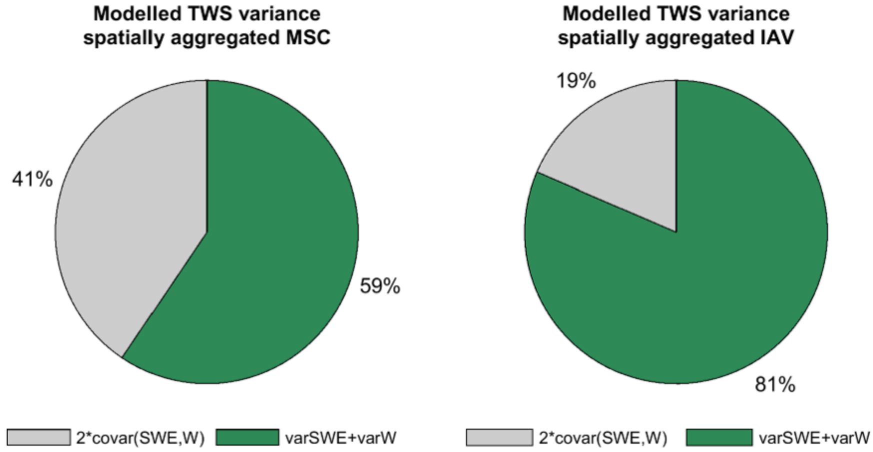

Fig. 7. Percentage composition of spatially aggregated TWSmod variance from the combined variances of SWE and W, and two times the covariance of SWE and W on mean seasonal and interannual scales

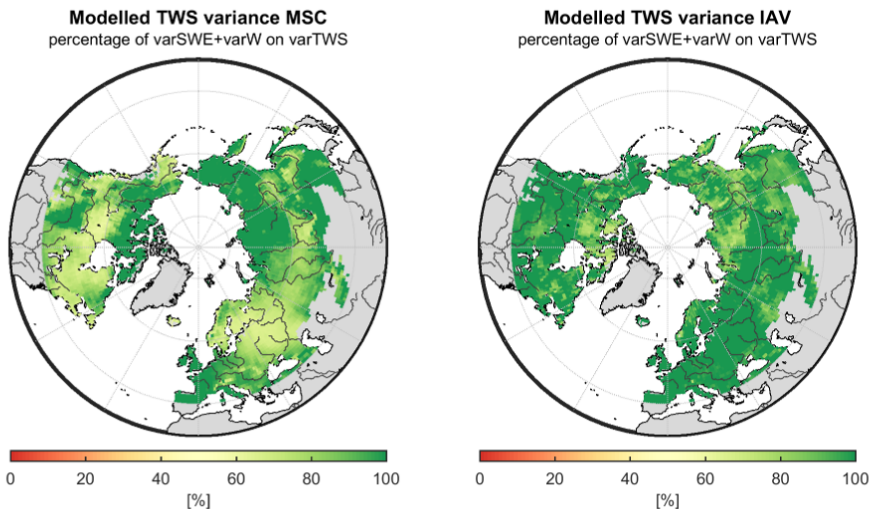

Fig. 8. Percentage of SWE and W variance on total TWSmod variance on mean seasonal (MSC) and interannual (IAV) scales