Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-690-AC1, 2018

Hydrology and Earth System Sciences
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

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

NEED TO DEVELOP A NEW MODEL

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?

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 now state in the introduction:

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 based data streams including GRACE TWS along with estimates of snow water equivalent, evapotranspiration and runoff into a rather simple hydrological model. This model is designed as a combination of standard model formulations yet aims to maintain a low complexity in order to facilitate multi-criteria calibration and to focus on variables that can be constrained by observations. First, we explain the applied methods including the implemented model, the used data, [. . .]

Regarding the second question, a direct comparison of TWS and SWE from observations is an interesting suggestion that we considered thoroughly. There are two 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. 1-3). 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 larger effect of liquid water storages as compared to the modelled results (Fig. 3).

Minor Comments

USE OF SATELLITE DERIVED SOIL MOISTURE

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).

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.

DELINEATION OF THE STUDY AREA

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

We thank the reviewer for pointing out that our manuscript was missing this critical information, which will be included the revised manuscript:

We defined humid land surface based on an aridity index AI $\geq$ 0.65. Therefore we calculated AI as the ratio of precipitation and potential evapotranspiration (United Nations Environment, 1992), using the precipitation and potential evapotranspiration data that were also applied as model forcing (GPCP-1DD precipitation (Huffman et al., 2000) and potential evapotranspiration following the Priestley-Taylor formula based on CERES net radiation (Wielicki et al., 1996) and CRUNCEP v6.1 air temperature (Viovy, 2015)). 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° x 1° grid cell.

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. 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. 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. Viovy, N.: CRU-NCEPv6.1 Dataset, http://dods.extra.cea.fr/data/p529viov/cruncep/, 2015. Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B. I., Smith, L. G., and Cooper, J. E.: Clouds and the Earths Radiant Energy System (CERES): An Earth Observing System Experiment, Bulletin of the Amercian Meteorological Society, 77, 853-868, 1996.

ROUTING

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)?

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. (2009) 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 by adding the following:

Model Description: 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 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 simulations to measured river discharge of large basins in our model-data fusion approach. Limitations of the approach: [..], 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., 2009), and thus may propagate to the discrepancy between TWSobs and TWSmod in some regions.

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.

GLOBSNOW DATA

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

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 (Luojus et al., 2010). We state this in the revised manuscript.

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

MAPS OF TEMPORAL AVERAGE DATA UNCERTAINTIES

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.

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 (see belowFig. 4). As we apply a constant average uncertainty of 35 mm (as mentioned in line 26 page 9), an additional map is not included in the supplement.

PARAGRAPH OF SPATIAL COHERENCE

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.

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 revised the paragraph accordingly:

As this study intends to analyze 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 has to be considered. Some storage components, e.g., soil moisture, have much larger spatial variability than others. Due to this large small-scale heterogeneity, the effect on larger regional scale might actually be minimal, 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.

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.

OVERESTIMATION OF GPCP

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

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 include this statement in the revised manuscript.

Behrangi, A., Christensen, M., Richardson, M., Lebsock, M., Stephens, G., Huffman, G. J., Bolvin, D., Adler, R. F., Gardner, A., Lambrigtsen, 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.

VALUES OF THE 4 COST TERMS

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.

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 will not 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.

Table S1 (Fig. 9) 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 magnitude of the optimized values in general reflects the qualitative importance we assign to the individual data streams (as large cost values 'punish' the model during optimization), with the highest value for TWS, followed by SWE and Q and the smallest value for ET. These values represent a weighted Nash-Sutcliff efficiency of 0.37 (TWS), 0.44 (SWE), 0.57 (Q) and 0.80 (ET).

CORRELATION FOR SEASONAL VARIATIONS

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.

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 bias and RMSE metrics of ET and Q in the supplement of the revised manuscript.

REGIONS OF LARGE RMSE OF TWS

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?).

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:

[. . .] 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.

NEGATIVE TWS ANOMALY IN 2003

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

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. 5). 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 FigS4). In addition, there is a negative SWE anomaly of on average 5 mm (see Fig.4 a) 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. 6). 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. 7). We add a paragraph to the revised manuscript on the discrepancy regarding the 2003 anomaly.

AVERAGE VALUE OF CR

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

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

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

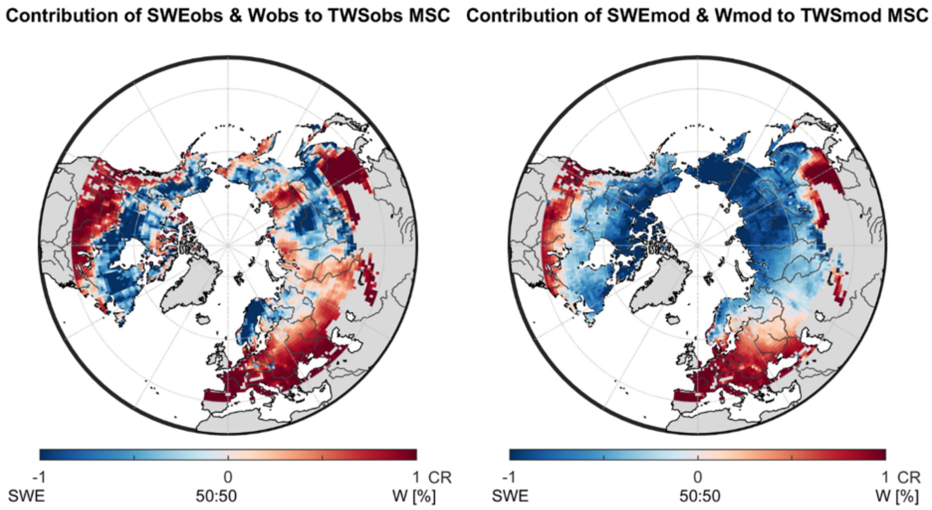

Fig. 1. Relative contribution based on CR of snow (SWE) and liquid water (W) storage anomalies to mean seasonal TWS anomalies based on observations (left) and based on the model (right)

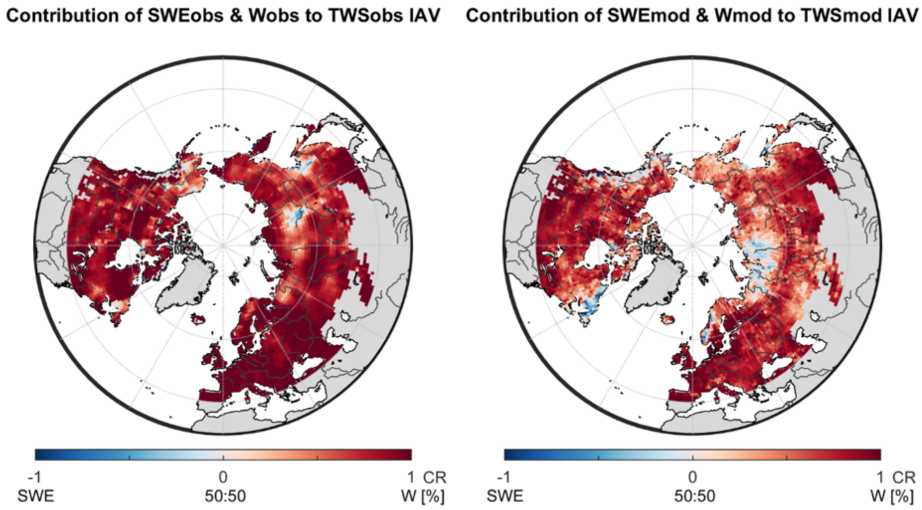

Fig. 2. Relative contribution based on CR of snow (SWE) and liquid water (W) storage anomalies to interannual TWS anomalies based on observations (left) and based on the model (right)

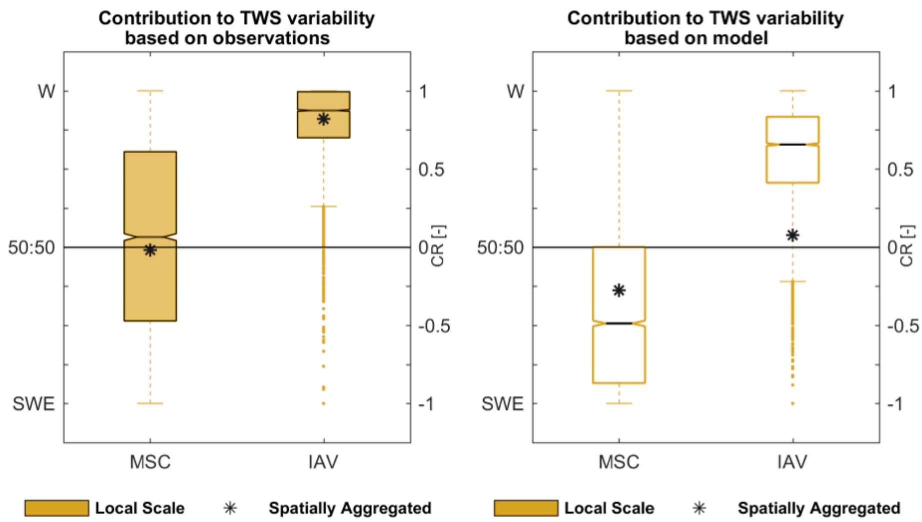

Fig. 3. Relative contribution of snow (SWE) and liquid water (W) to TWS variability on different spatial and temporal scales based on observations (left) and based on the model (right)

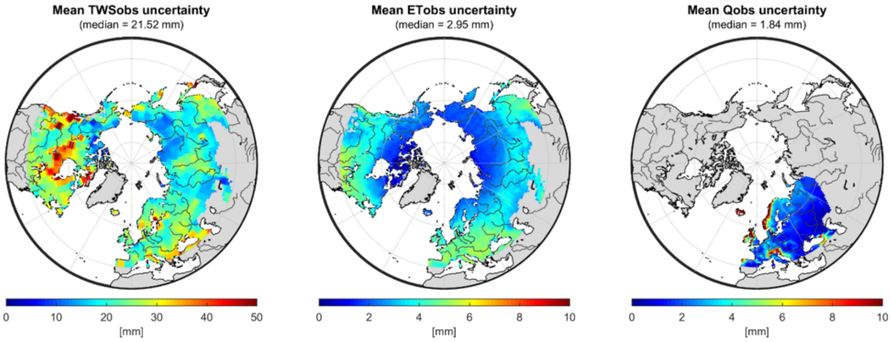

Fig. 4. Mean uncertainty of monthly TWSobs [mm], and of the mean seasonal cycle of ETobs [mm d-1] and Qobs [mm d-1] used for model calibration

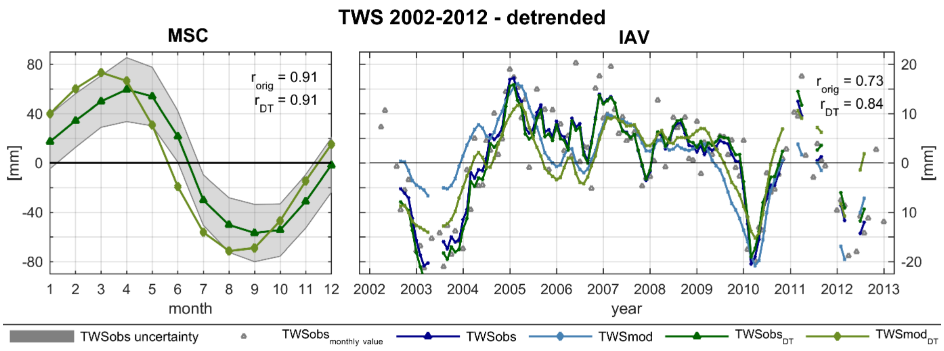

Fig. 5. Mean seasonal cycle and interannual variability of original (orig) and detrended (DT) TWSobs (GRACE) and TWSmod (forced with GPCP precipitation)

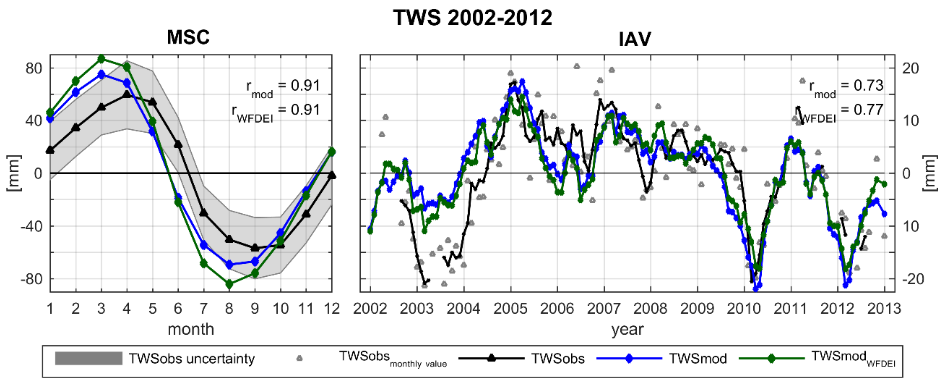

Fig. 6. Comparison of the mean seasonal cycle and interannual variability of TWSobs (GRACE), TWSmod (forced with GPCP precipitation) and TWSmodWFDEI (forced with WFDEI rain and snow fall)

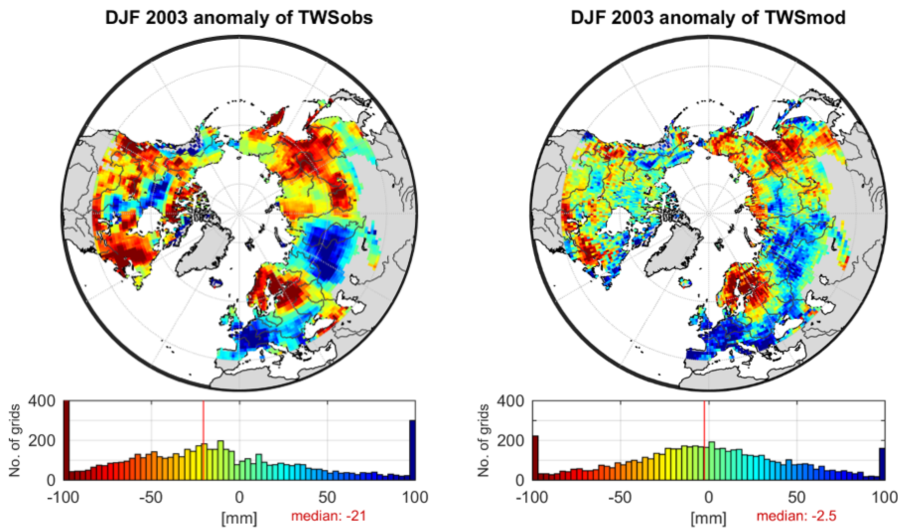

Fig. 7. Average IAV of winter 2002/2003 (December, January, February) of observed and modelled TWS

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

Fig. 8. Table S1: Cost values obtained with the default and the optimized model parameters using Eq. (1)