Author response to Referee #1 comments from August 30th, 2016

We are thankful to Referee #1 for his/her valuable comments and suggestions, which will certainly improve our manuscript. In the following, the response to the individual comments is given with some new figures at the end. The original review is quoted in italics, whereas the author response is given in normal font.

This manuscript presents results of comparing estimated terrestrial water storage (TWS) from four hydrological models with GRACE derived TWS in 31 hydrological basins. Four metrics were used in evaluating model performances. Components of TWS as well as actual and potential ET were examined in selected basins to show the impact of model physics on estimated TWS.
The results and discussions are generally well presented and justified. But I think the paper can be further improved in a few areas. For instance, the fact that three of the four models do not model groundwater, which may contribute significantly to TWS changes, is not explicitly mentioned and discussed in the paper. In addition, the four metrics used in evaluation may be good for summarizing the differences but they do not necessarily reflect the actual discrepancies between modeled and GRACE derived TWS. For instance, the amplitude and phase differences may not be important if TWS exhibits strong inter-annual variability.

-Besides annual amplitude and phase differences, we also show the explained variance with the seasonality removed to evaluate the agreement of models with GRACE in terms of inter-annual variability (Page 6, Line 1-2; Fig. 5 in the revised manuscript (RM)). We respond to the groundwater issue in the answer to the reviewer's next comment below.

Additional comments: Page 3, data set. Please emphasize the fact that three of the four models do not simulate groundwater and discuss its potential impacts on model estimated TWS in the result section. Also, do these models account for anthropogenic impacts such as groundwater abstraction? If not, how would this affect the comparison with TWS from GRACE which does detect changes associated with groundwater withdrawals?

-Thanks for the suggestion, and this is indeed an important information. In Line 12 on Page 4 of RM we propose to include the following sentences: Some of the main characteristics of the four numerical models are presented in Table 1, which provide more information on how models are different with each other. For instance, although soil moisture and snow water are included in all models, surface water and groundwater are simulated differently. JSBACH is the only model which does not include surface water. Groundwater is simulated by WGHM, where the anthropogenic impact such as groundwater abstraction is also considered. JSBACH does not include groundwater explicitly. However, soil moisture in deep layers below the root zone is simulated and buffers extreme soil moisture conditions in the layers above. Thus, some of the characteristics of real groundwater are considered. We use the term subsurface water for both soil moisture and groundwater. But the impact from consideration of groundwater to TWS variations in WGHM will be investigated in the following discussion.

A Fig. 1 is added in the manuscript showing the differences of explained variances from WGHM with and without groundwater. The positive values indicate that WGHM with groundwater exhibits better agreement with GRACE than the one without groundwater. The large impact mainly locates at basins such as Toscantins, Niger, Huang He, Mekong and Mississippi. Only in three basins (Lena, Indus and Yukon), the effect of groundwater consideration in the model is negative.

Page 4, Line 8: I understand why you removed the trend but the ability to predict trend is also an important part of the models. Can you provide a scatter plot comparing trends from the models and those from GRACE in the 31 basins?

-A scatter plot comparing trends from the models and those from GRACE in the 31 basins is shown in Fig 2. The TWS trends from various models do perform quite differently among each other and with GRACE.

Page 5 Line 10, should the second "GRACE" be TWS?

-Yes, and it has been changed accordingly.

Page 6. Line 20, I don't think it is appropriate to compare GRACE errors with the RMSE since the former represents instrument and post-processing errors and has nothing to do with how well models perform.

-The GRACE errors are calculated to indicate the TWS uncertainties from GRACE, which can be applied to indicate about where GRACE might be suitable as a validation tool for models and where not. Special attention should be paid to basins with large GRACE errors, as the large discrepancies could be related to large GRACE TWS uncertainties, but not to model differences. We would like to stress that we do not directly compare GRACE errors with RMSE, but that we use the GRACE errors only as indicator of observation uncertainty. The way to estimate GRACE errors is introduced in Zhang et al. (2016). The error estimation is also investigated through an end-to-end simulation performed by Flechtner et al. (2016). We thus believe that the errors we calculated are plausible.

In addition, basin-scale GRACE errors are smaller than the gridded errors which are spatially correlated. Did you consider spatial correlation of errors (in both modeled and GRACE TWS) when calculating basin-scale RMSEs between the model and GRACE? Either way, I think it only makes sense to compare RMSEs among the models.

-The correlation between the gridded errors from GRACE is much larger than the one from models and is considered by using the squared exponential covariance function to estimate the statistical covariance between two grids as proposed by Landerer and Swenson (2012). The error estimates from the gridded data set also show consistent results with the ones derived directly from Stokes coefficients.

The statistics in Table 2 shows the models generally did not performed well in the tropical climate. Why is that? Does it have something to do with runoff estimates as ET is energy limited in this type of climate? You don't necessarily need to collect in situ stream flow data, but some discussions and plots on runoff may be needed to explain this result.

-The large RMSE values in tropical regions are partly related to the fact that the TWS variability in this region is comparably large. Besides, the runoff comparison is shown for three basins affected by the tropical climate (Fig. 3). The runoff is calculated from the models following the equation: Runoff=Precipitation-Actual evapotranspiration-TWSC (Ramillien et al., 2006). It is seen that the performance of a certain model is connected with its differently simulated runoff. At Amazon basin, the comparably large runoff simulated from MPI-HM also leads to smaller variability in TWS, which is also shown at Orinoco basin. At Mekong basin, the larger amplitude in TWS from JSBACH compared with GRACE is related to the apparent small amplitude in its runoff.

Page 7. Line 5 to 12. As you pointed out that AET does not have a significant impact on TWS in humid areas, then what is the purpose of including three basins from that climate in Fig. 7? I think including more basins from drier climate is more useful here.

-This is true. We remove the basins in humid areas and focus on three basins (Niger, Chari and Indus) in the dry zone in Fig. 9 (RM).

Page 7. Line 13-22, I didn't learn anything from this paragraph and it can be removed. As you correctly pointed out that AET may be significantly different from PET which does not help much in explaining the result. Again, I think presenting runoff estimates is more useful.

-Following the reviewer's suggestion we shorten the discussion on PET in the revised manuscript and add a figure showing comparison of runoff data (Fig. 10 (RM)).

Figs. 2-5: It would be very helpful if you provide time series of TWS for a basin(s) with the largest deviation from GRACE in either of these metrics. For instance, it's hard to visualize how significant a 45 degree difference in phase is. In addition, two of these metrics measure differences in seasonality which may not mean much when the interannual variability of TWS is much stronger. So providing actual TWS time series along with some discussions will be helpful for readers to understand the usefulness and limitation of these metrics.

-Fig. 4 showing the time series of TWS for basins with the largest deviation from GRACE is added in the manuscript along with some discussions.

In Line 22 of Page 6 (RM), some sentences are added: As each metric usually focuses only on one specific property of statistical performance and has its own limitations, the time series of TWS are given for some basins with the largest deviation between GRACE and the model. We choose Yukon basin, where both WGHM and JSBACH exhibit the largest deviation of annual amplitudes from GRACE. Although the annual amplitude is simulated better by LSDM and MPI-HM, apparent negative phase differences are shown. Amur basin is also shown, as LSDM, WGHM and MPI-HM all have the largest negative phase differences with GRACE here. Models generally capture the inter-annual signals but perform quite differently among each other and with GRACE in terms of seasonality. Almost opposite phase differences are found for these models. The smallest explained variance for MPI-HM happens at St. Lawrence basin, where a much larger amplitude and a negative phase difference compared with GRACE are found. When the annual signal is removed, models perform differently in terms of the explained variance. In Nile basin, large inter-annual variations simulated by LSDM lead to even negative explained variance compared with the other models.

Fig. 7. I think including runoff instead of PET is more appropriate here. Also, please try to use the same y-axis range for all plots which makes it to compare the magnitude of TWS and ET.

-The y-axis range is changed and the runoff comparison is also shown (Fig. 3).

Fig. 8, "Subsurface water" should be soil moisture + groundwater storage for WGHM and soil moisture for all other models. Again, please use the same range for all y-axis if possible. In the caption, snow water content should be snow water equivalent.

-This has been changed accordingly.

References

Flechtner, F., Neumayer, K.-H., Dahle, C., Dobslaw, H., Fagiolini, E., Raimondo, J.-C., and Güntner, A.: What Can be Expected from the GRACE-FO Laser Ranging Interferometer for Earth Science Applications?, Surveys in Geophysics, 37, 453–470, doi:10.1007/s10712-015-9338-y, http://dx.doi.org/10.1007/s10712-015-9338-y, 2016.

Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resour. Res., 48, doi:10.1029/2011WR011453, W04531, 2012.

Ramillien, G., Frappart, F., Güntner, A., Ngo-Duc, T., Cazenave, A., and Laval, K.: Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry, Water Resources Research, 42, n/a–n/a, doi:10.1029/2005WR004331, http://dx.doi.org/10.1029/2005WR004331, w10403, 2006.

Zhang, L., Dobslaw, H., and Thomas, M.: Globally gridded terrestrial water storage variations from GRACE satellite gravimetry for hydrometeorological applications, Geophys J Int., 206, 368–378, doi:10.1093/gji/ggw153, 2016.

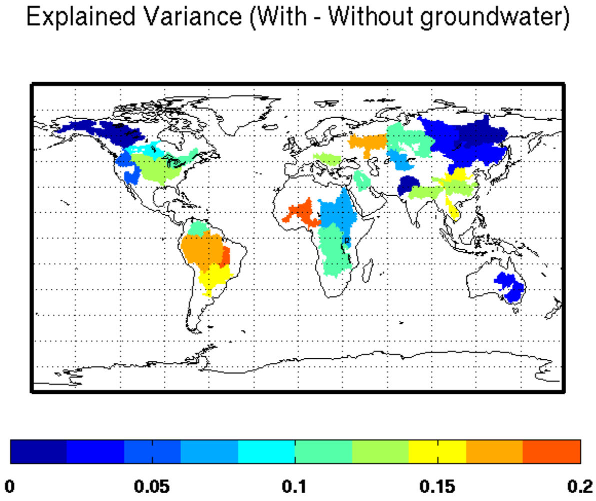

Figure 1. The differences between the explained variance values from WGHM with and without groundwater.

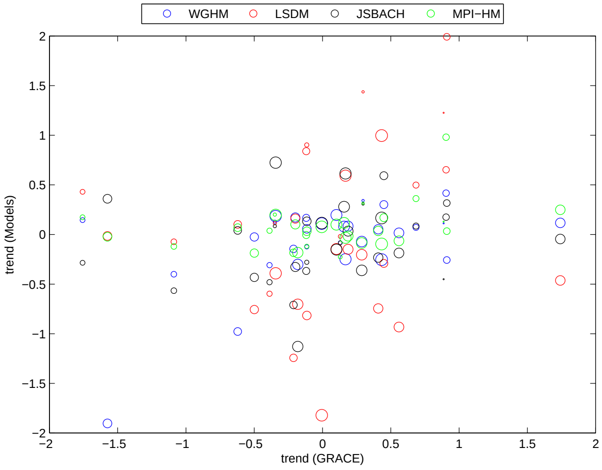

Figure 2. The scatter plot comparing trends from the models and those from GRACE in the 31 basins. Symbol size varies with rive basin area.

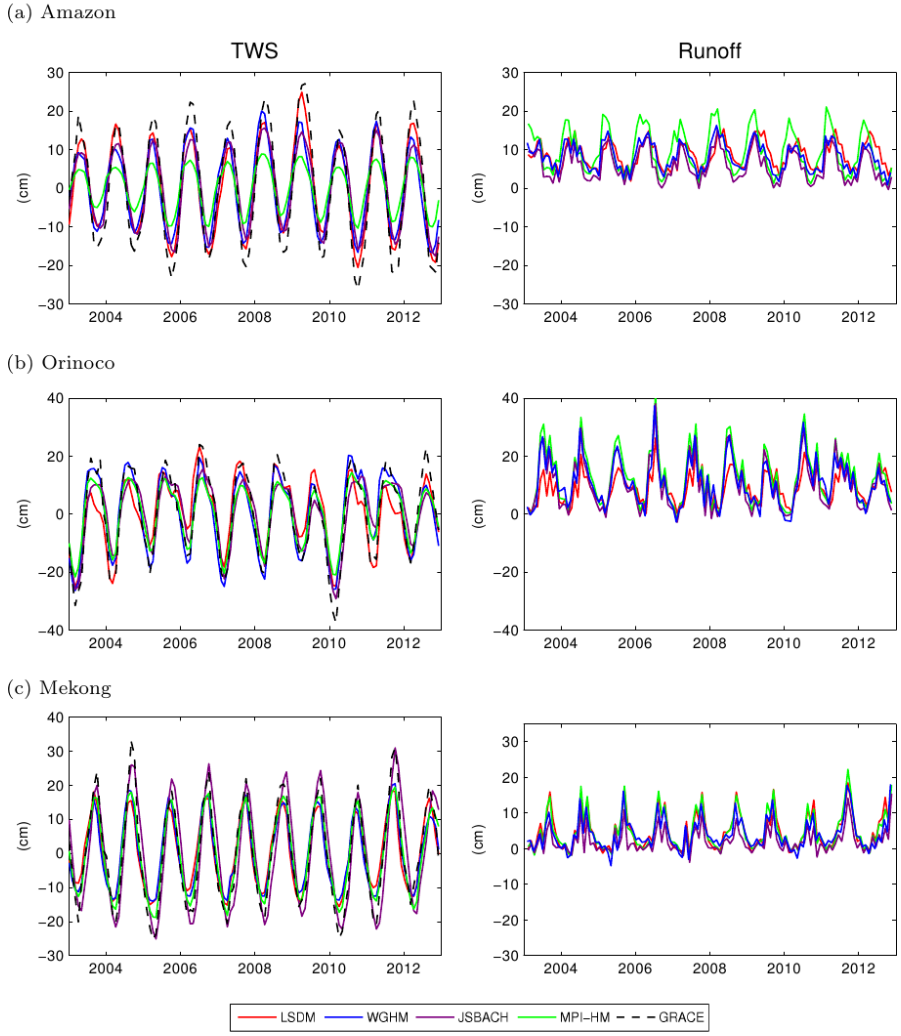

Figure 3. Time series of TWS (left) from GRACE and models and model simulated runoff time series (right); each for three different catchments in tropical zone: Amazon, Orinoco, and Mekong.

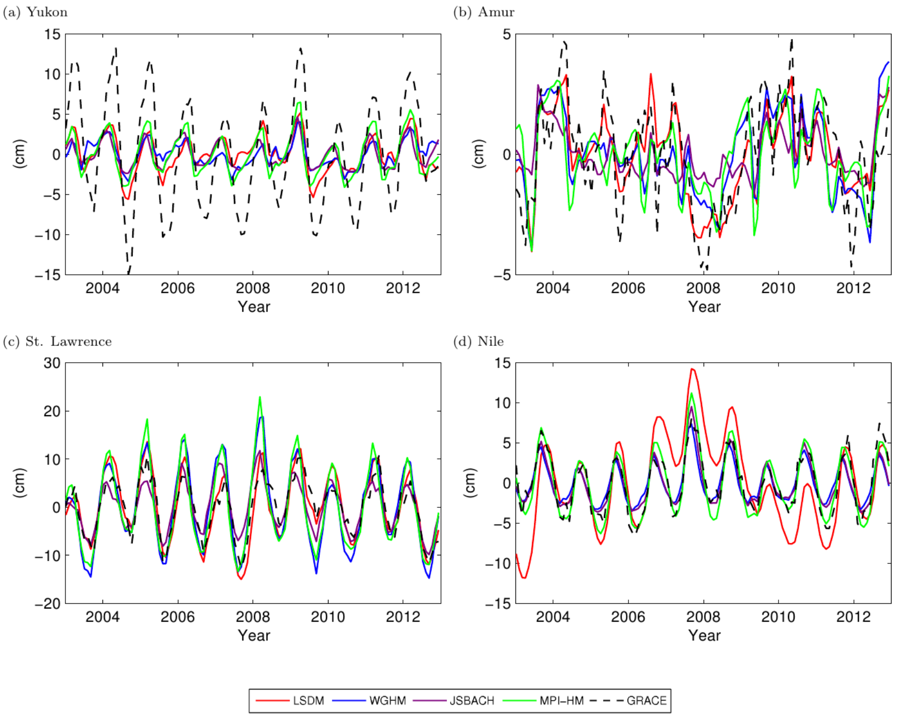

Figure 4. Examples of monthly TWS time series from GRACE and models for the basins with the largest deviation between model and GRACE in each of the four metrics: Relative amplitude differences (Yukon), phase differences (Amur), explained variance (St. Lawrence) and explained variance with annual signal removed (Nile).

Author response to Referee #2 comments from September 25th, 2016

We are thankful to Referee #2 for the constructive comments and suggestions, which will certainly improve our manuscript. In the following, the response to the individual comments is given. The original review is quoted in italics, whereas the author response is given in normal font.

The manuscript gives an excellent overview on the current performance of four global hydrological models validated by using the latest GRACE Release. 31 river basins worldwide covering different climate zones are assessed and deeper insights into a few basins in arid and snow dominated zones are given. The study addresses current scientific issues and goes further than previous works. It is well structures and the results are presented in a clear and comprehensible way. I think that the study will contribute to the improvement of hydrological models. In one point the paper could still be improved: especially in Section 3.2 and in Section 4 I miss some interpretation of the findings. Is it possible to discuss a few reasons for the different model behaviors? I have a few minor comments and questions which do not include the issues already discussed by the first reviewer.

-We would like to thank the referee for the positive comments on the manuscript. We add some discussions on the reasons for the different model behaviors in Line 9 of Page 10 (RM), which is further explained in the response to the specific comments in the following.

Page 1, Line 7: What is the meaning of 'different spatial characteristics' of the individual storage compartments? Do you mean that e.g. groundwater is simulated using a different number of layers?

-By 'different spatial characteristics', we mean that the spatial auto-correlation pattern of water storage is different for the different storage compartments. For instance, surface water exhibits a linear or point-like pattern as it is concentrated in areas such as rivers and lakes, whereas soil moisture and groundwater tend to have a smoother distribution in space with larger spatial correlation lengths. For hydrological models, a missing storage compartment, such as surface water in JSBACH, thus leads to a different spatial pattern of TWS variability compared to the other three models.

Page 2, Line 20: I think that an accuracy of 1 cm equivalent water height is a very optimistic estimate for areas as small as 100,000 km^2.

-Indeed, with this sentence we intend to indicate the limit of what could be achieved with GRACE, according to the cited reference.

Page 2, Line 24: You say, that there are more than 13 years of GRACE data available, but Figure 7 indicates that you use only 10 years of data instead of the full record. Why do you not use the whole time series of GRACE data? Is the data from the models missing?

-Yes, the 10 years chosen here is the common period of data from GRACE and the hydrological models available to us.

Page 3, Data sets: For WGHM you explicitly mention the water storage compartments. Please add information about the storage compartments for the other models (to some extend you already did this in your response to the first reviewer).

-The different water storage compartments simulated by the models are shown in Table 1 of the manuscript. Still, to be consistent and clear, some more sentences are added for each of the other models: The global water storage variations contain surface water in rivers, lakes and wetlands, groundwater and soil moisture, as well as water stored in snow and ice (LSDM). Snow is treated as external layers above the soil column, with maximum of five snow layers. Soil moisture in deep layers below the root zone is simulated and buffers extreme soil moisture conditions in the layers above (JSBACH). TWS from MPI-HM is simulated as the sum of soil moisture in the root zone, snow and surface water (MPI-HM).

Page 4, Line 27: Are the local re-scaling factors introduced for each grid cell?

-Yes, and 'for each grid cell' is added after the sentence.

Page 4, TWS Estimates from GRACE: Did you also remove the trend from the GRACE time series for the same period as for the models?

-Yes, and one sentence in Line 8 on Page 5 (RM) has been added to make it more clear: As for the model data, the linear trend is removed over the period Jan 2003 to Dec 2012.

Page5, Evaluation metrics: The relative annual amplitude differences and the phase differences are interesting metrics for river basins with a strong annual cycle that follow approximately a sine curve. Can you discuss the meaning of these metrics for river basins where interannual signals dominate (also with respect to Fig. 2 and Fig. 3)?

-The seasonal cycle and the inter-annual variations of the signal are investigated separately by different metrics in our study. The explained variance (eq. 4) is also calculated for de-seasonalized time series to evaluate model performance with respect to inter-annual variations. Fig. 2 and Fig. 3 (RM) just focused on the seasonal variations from the models where the inter-annual signals were not taken into account.

Page 5, Line 16: Why did you estimate the trend? You subtracted it in the preprocess- ing step.

-It is true that we removed the trend before the calculation of seasonal and inter-annual metrics to avoid any influence of the trend on them and it should be removed from the equation.

Page 5, Global evaluation: This paragraph gives an excellent overview on the current performance of the four models on a global scale. Is it possible to add some inter- pretation of the results, i.e. can you explain the results by the model structure or the parametrization? E.g. why do the models in general have an earlier seasonal storage maximum than GRACE, why does LSDM have smaller phase differences than the other models, what is the problem of the models in basins with small explained variance,. . .?

-As models perform differently in different areas and are affected by combined impacts from their parametrization, structure, and physical representation, it is hard to interpret the general global performance and to single out specific reasons for poor or bad model performance with respect to structure and parametrization. For instance, it is assumed that missing water storage compartments in a certain model is the main reason for the earlier seasonal storage maximum than GRACE in the previous work. Through our investigation, it is found that this is not totally true. Groundwater is missing in LSDM, but still it shows smaller phase differences than the other models. Besides, the negative phase difference in JSBACH is found to be more related to its snow water simulation. We discuss these and other related issues on model performance for specific regions in the manuscript, and also consider AET and runoff (besides TWS time series) for an interpretation.

Page 6, Line 24: Did you mean: 'most basins have low SNR values'?

-Yes, it has been changed accordingly.

Page 7, Line 8: Fig.7 shows Amazon, Zaire, Mekong, and Niger instead of Chari, Indus, Murray, and Niger. Is this intended?

-Fig. 7 is updated to Fig. 9 (RM) with focus on three basins in dry climate (Chari, Indus, and Mekong).

Page 7, Line 9: I do not think that the performances of the models at those four basins are quite consistent with each other. In fact, JSBACH performs quite differently.

-This sentence is indeed misleading and has been removed.

Page 9, Line 7-14: This is a good summary of the performance of the four models. Can you provide any reasons for the strengths and the deficiencies of the individual models?

-As noted in a comment above, sorting out the specific reasons for strengths and deficiencies of a specific model for a certain region is difficult due to model complexity. Nevertheless, we try to work out some reasons and extend the third paragraph in the Summary (RM) as follows: Model performance is also investigated in some snow dominated and dry catchments in more detail through time series comparison. The poor performance of JSBACH and MPI-HM in snow dominated regions is mainly related to negative phase shifts compared to GRACE. MPI-HM simulates identical snow variations as LSDM, however, the different simulations of subsurface water and especially surface water still lead to different TWS variations in snow dominated regions. Despite of the missing surface water compartment, the simulated snow variations in JSBACH already show smaller amplitude and negative phase differences compared with all the other models. This could be related to the fact that JSBACH simulates snow in a more physical way based on energy balance, which is totally different from the degree-day method applied by all the other models. The comparably better agreement of LSDM and WGHM with GRACE in terms of TWS in these snow dominated basins is partly caused by the realistic surface water compartment represented by these two models. In the dry catchments, the impact from AET on TWS is relatively strong. The smaller AET from MPI-HM also leads to better agreement with GRACE, whereas LSDM shows large differences with GRACE in terms of TWS especially at some dry basins in central Africa partly due to the too simple evaporation scheme. PET is simulated using a superior parametrization by MPI-HM, while LSDM applies still the traditional Thornthwaite method based solely on air temperature. The groundwater considered by WGHM also has some impact on the simulated TWS, especially at basins as Toscantins, Mekong, Niger and Mississippi. At Yukon basin, we found the bad performance of all models in terms of TWS when compared with GRACE, which could be due to the effects of atmospheric and oceanic de-aliasing errors not further discussed in our current study. In future, we would like to assess all possible errors of GRACE TWS through investigation of simulated GRACE-type gravity field time-series (Flechtner et al., 2016) based on realistic orbits and instrument error assumptions as well as background error assumptions out of the updated ESA Earth System Model (Dobslaw et al., 2015, 2016), which we believe will further help to explain the discrepancy between global models of the terrestrial water cycle and GRACE satellite observations.

Fig. 6: When the annual signal is removed the boxes become much smaller for all models except for LSDM. Do you have an explanation for this behavior?

-The comparably large spread of the variations for this metric (explained variance with the annual signal removed) from LSDM is related to the fact that the performance of the inter-annual signal simulation from LSDM varies among the different areas. When the annual signal is removed, the explained variance remains high at some basins in high latitudes of the Northern Hemisphere, whereas much larger inter-annual variations are simulated at central Africa which leads to quite low explained variance values. Furthermore, please note the different scale of the y-axis in both plots which may suggest a considerable reduction of the spread of the metric for the case with the annual signal removed, which actually is not that large in absolute values.

Technical comments Page 2, Line 12-16: Did you want to say that instead of using observations of precipi- tation (P), evapotranspiration (E), and runoff for closing the water budget equation, one can also use water vapor content and moisture flux convergence from atmospheric reanalysis data and river discharge? Is runoff and river discharge the same in both cases? Furthermore, P and E can also be taken from atmospheric reanalysis (Rodell et al. (2011) Estimating evapotranspiration using an observation based terrestrial water budget. Hydrological Processes, 25:4082–4092.) Maybe you would like to reformulate the sentence.

-Thanks for the suggestions. To be more clear, the sentence is changed to: The terrestrial water budget method estimates TWS by solving the terrestrial water balance equation through the data of precipitation, runoff and evapotranspiration from observations and atmospheric reanalysis (Zeng et al., 2008; Tang et al., 2010; Rodell et al. 2011), while TWS variations can also be derived from combined atmospheric and terrestrial water-balance computations, utilizing water vapor content and moisture flux convergence from atmospheric reanalysis data and river discharge measurements (Seneviratne et al., 2004; Hirschi et al., 2006).

Page 9, Line 24-27: The structure of the sentence is strange. Probably you should delete either 'In future' or 'in our next step'.

-It is true, and 'in our next step' is removed.

References

Rodell, M., McWilliams, E. B., Famiglietti, J. S., Beaudoing, H. K., and Nigro, J.: Estimating evapotranspiration using an observation based terrestrial water budget, Hydrological Processes, 25, 4082–4092, doi:10.1002/hyp.8369, http://dx.doi.org/10.1002/hyp.8369, 2011.

Validation of Terrestrial Water Storage Variations as Simulated by Different Global Numerical Models with GRACE Satellite Observations

Liangjing Zhang[1], Henryk Dobslaw[1], Tobias Stacke[2], Andreas Güntner[1], Robert Dill[1], and
Maik Thomas[1,3]

[1]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany
[2]Max Planck Institute for Meteorology, Hamburg, Germany
[3]Freie Universität Berlin, Institute of Meteorology, Germany

Correspondence to: Liangjing Zhang (liangjing.zhang@gfz-potsdam.de)

Abstract. Estimates of terrestrial water storage (TWS) variations from the satellite mission GRACE are used to assess the accuracy of four global numerical model realizations that simulate the continental branch of the global water cycle. Based on four different validation metrics, we demonstrate that for the 31 largest discharge basins worldwide all model runs agree with the observations to a very limited degree only, together with large spreads among the models themselves. Since we apply a common atmospheric forcing data-set to all hydrological models considered, we conclude that those discrepancies are not entirely related to uncertainties in meteorologic input, but instead to the model structure and parametrization, and in particular to the representation of individual storage compartments with different spatial characteristics in each of the models. TWS as monitored by the GRACE mission is therefore a valuable validation data-set for global numerical simulations of the terrestrial water storage since it is sensitive to very different model physics in individual basins, which offers helpful insight to modellers for the future improvement of large-scale numerical models of the global terrestrial water cycle.

1 Introduction

Growing observational evidence underlines the important role of the terrestrial water cycle in shaping the Earth's climate. For instance, soil moisture variability alters the atmospheric circulation through its impact on evaporation, that affects regional and global climate (Koster et al., 2004; Meehl et al., 2009; Seneviratne and Stöckli, 2007). Snow cover raises surface albedo and isolates the land surface from the atmosphere. Groundwater also shows a significant low-frequency variability that could have regional impacts on inter-annual climate variability (Bierkens and van den Hurk, 2007). Monitoring data on water availability from both in situ and remote sensing instruments is also essential for economic and societal development. It can be used to characterize extreme hydro-meteorological conditions as flood (Chen et al., 2010) and drought (Leblanc et al., 2009). Hydrological models are important tools to forecast water resources at both short and long-term perspectives. There is now an increasing number of models that simulate the terrestrial water cycle at large spatial scales, which generally fall into two categories: Land Surface Models (LSMs) and Global Hydrology Models (GHMs). LSMs focus on solving the surface-energy balance and can be coupled to atmospheric models, while GHMs rather focus on lateral water transfer and solving the water balance equation. Due to the different physical representation of land-surface processes, uncertainties in model structure, parameter values, and atmospheric forcing data, the performance of these models varies. There have been several model intercomparison projects, such as the Global Soil Wetness Project (GSWP; Dirmeyer et al., 2006; Dirmeyer, 2011), the Water Model Intercomparison Project (WaterMIP; Haddeland et al., 2011), and the Inter-Sectoral Impact Model Intercomparison Project (ISI-MPI; Schewe et al., 2014) which compare the results from a multitude of models to highlight shortcomings and inconsistencies. These projects have primarily focused on evapotranspiration or soil moisture content. Gudmundsson et al. (2012) has also evaluated nine large-scale hydrological models based on runoff observations.

The terrestrial water storage (TWS), which is understood here to contain all water compartments stored above and underneath the land surface including soil moisture, the water content of snow-pack, land ice, surface water, and groundwater in shallow and deep aquifers, forms an important compartment of the terrestrial water cycle. It is difficult to directly measure TWS on the ground due to insufficient in-situ observations of the very diverse hydrological stores and fluxes. The terrestrial water budget method estimates TWS by solving the terrestrial water balance equation through the data of precipitation, runoff and evapotranspiration from observations and atmospheric reanalysis (Zeng et al., 2008; Tang et al., 2010; Rodell et al., 2011), while TWS variations can also be derived from combined atmospheric and terrestrial water-balance computations, utilizing water vapor content and moisture flux convergence from atmospheric reanalysis data and river discharge measurements (Seneviratne et al., 2004; Hirschi et al., 2006). However, these methods are highly dependent on the accuracy of the reanalysis data which often contain systematic errors in particular at inter-annual time scales and longer. The Gravity Recovery and Climate Experiment (GRACE) launched in 2002 provides a unique data source to estimate spatio-temporal variations of the Earth's water storage at regional up to global scales (Tapley et al., 2004; Wahr et al., 2004). Averaged over an arbitrary area with a spatial extend of 100,000 km^2 and greater, TWS derived from GRACE is believed to reach an accuracy of better than 1 cm equivalent water thickness (Dahle et al., 2014). Although there is a mismatch between the spatial resolution of GRACE data and that of hydrological models, the effective spatial resolution can be extrapolated to finer spatial scales through proper post-processing (Landerer and Swenson, 2012). There are now more than 13 years of GRACE data available and this length of the time-series together with a recently completed reprocessing of the whole GRACE record (Dahle et al., 2012) motivates us to revisit the question of what can be learned from GRACE on the performance of global hydrological models in representing continental water storage variations.

Through the comparison of basin-averaged TWS from models with GRACE-based estimates, we intend to identify the advantages and deficiencies of a certain model and analyze the reasons for different model behaviors. Globally gridded TWS variations and uncertainties from GRACE estimated by the same post-processing procedure as described by Zhang et al. (2016) are applied. We ~~then~~ quantitatively analyze the correspondence between TWS estimates from 4 available hydrological models and GRACE in 31 of the world's largest river basins. To separate the effects of atmospheric input data, all the models apply the same meteorological forcing data-set. ~~Potential and actual evapotranspiration~~ Actual evapotranspiration and runoff rates calculated with the different models are also analyzed. Considering the diversity of the performance of the models in these 31 basins, we focus on time series of TWS variations in two regions which are characterized by different climate regimes, i.e., the snow-dominated catchments and the dry catchments by looking into the TWS variation time series from models and GRACE. Besides, snow, surface water and subsurface water including root zone or/and deep layer storage from the models are also compared in order to analyze the contribution of different storage compartments to the total water storage. By investigating the relative performance of these different models, we intend to contribute to the future model development of both LSMs and GHMs.

2 Data set

2.1 Hydrological model simulations

For this study, we selected four different models to represent a broad range from conceptual hydrological to complex land surface models (Table 1). In order to ensure that this spread between the simulations is indeed related to the different representation of physics in the model, all the models are forced with the WFDEI data-set based on ERA-Interim re-analysis data (Dee et al., 2011) that has been developed during the WATCH project (Weedon et al., 2011). This WFDEI meteorological forcing dataset is a quasi-observation which combines the daily variability of the ERA-Interim re-analysis with monthly in-situ observations such as temperature and precipitation (Weedon et al., 2014). There are two precipitation products available from WFDEI: (1) corrected by using the Climate Research Unit at the University of East Anglia (CRU) observations; and (2) corrected with the Global Precipitation Climatology Centre (GPCC) data-set. Since the WFDEI data sets incorporating the CRU-based precipitation products cover a longer time span, they are used in our study and referred to subsequently as WFDEI-CRU.

The WaterGAP Global Hydrological Model (WGHM) is part of the Water-Global Assessment and Prognosis model (WaterGAP; Döll et al., 2003). WGHM is a conceptual water balance model with grossly simplified process representations. It is calibrated by tuning a runoff generation parameter against observed river discharge in a station-based calibration approach (Hunger and Döll, 2008). The model simulates the continental water cycle including the water storage compartments soil moisture within the effective root zone of vegetated areas, groundwater, canopy water, snow and surface water in rivers, lakes, reservoirs and wetlands. The latest version of WGHM as calibrated for WFDEI-GPCC forcing (version 2.2 STANDARD; Müller Schmied et al., 2014) is used in this study. However, we run the model with WFDEI-CRU forcing without re-calibration.

The Land Surface Discharge Model (LSDM; Dill, 2008) is based on the Simplified Land Surface Scheme (SL-Scheme) and the Hydrological Discharge Model (HD-Model; Hagemann and Gates, 2003, 2001) from the Max-Planck-Institute for Meteorology. The global water storage variations contain surface water in rivers, lakes and wetlands, groundwater and soil moisture, as well as water stored in snow and ice. The code has been tailored to enable the simulation of continental water mass redistribution for geodetic applications, that include the derivation of Effective Angular Momentum Functions of the continental hydrosphere to interpret and predict changes in the Earth rotation (Dobslaw et al., 2010; Dill and Dobslaw, 2010); and of vertical crustal deformations as observed from GPS permanent stations (Dill and Dobslaw, 2013).

JSBACH (Raddatz et al., 2007; Brovkin et al., 2009) is a land surface model and forms together with ECHAM6 (Stevens et al., 2013) and MPIOM (Jungclaus et al., 2013) the current Max-Planck-Institute for Meteorology's Earth System Model (MPI-ESM). As part of the MPI-ESM, JSBACH includes interactive vegetation and a 5-layer soil hydrology scheme to provide the lower atmospheric boundary conditions over land, particularly the fluxes of energy, water and momentum. For this study, however, JSBACH was used in an offline mode without interactive coupling to the other MPI-ESM compartments, but driven by prescribed WFDEI-CRU atmospheric forcing. Snow in JSBACH is treated as external layers above the soil column, with maximum of five snow layers. Soil moisture in deep layers below the root zone is simulated and buffers extreme soil moisture conditions in the layers above.

Finally, the Max Planck Institute of Meteorology's Hydrology Model (MPI-HM; Stacke and Hagemann, 2012) is a global hydrological model. Its water flux computations are of similar complexity to land surface models, but it does not account for any energy fluxes. Additionally to precipitation and temperature, it requires potential evapotranspiration as input which also was derived from the WFDEI using the Penman-Montheith equation similar to the Weedon et al. (2011) study. TWS from MPI-HM is simulated as the sum of soil moisture in the root zone, snow and surface water.

Some of the main characteristics of the four numerical models are presented in Table 1, which provide more information of how models are different with each other. For instance, although soil moisture and snow water are included in all the models, surface water and groundwater are simulated differently. JSBACH is the only model which does not include surface water. Groundwater is simulated by WGHM, where the anthropogenic impact such as groundwater abstraction is also considered. JSBACH does not simulate groundwater directly but as the subsurface water in the deep layer, whereas groundwater is not considered by the other two models. We use the term subsurface water for both soil moisture and groundwater. But the impact from consideration of groundwater to TWS variations will be investigated in the following discussion.

LSDM, WGHM and MPI-HM are provided on a 0.5° by 0.5° grid, while JSBACH has a coarse resolution, with 1.875° spacing in longitude and irregular spacing in latitude. The mean values and the linear trends estimated over the period Jan 2003 to Dec 2012 - i.e., the common period of GRACE observations and model ~~results~~ experiments - are first removed for each grid cell. Then the TWS variations are averaged over the selected basins to obtain the basin-scale TWS. Since ice dynamics and glacier mass balance are not included in the numerical models applied in this study, water mass variations in Antarctic and Greenland are not considered throughout the reminder of this paper.

2.2 TWS Estimates from GRACE

The U.S.-German twin satellite mission GRACE provides since April 2002 estimates of month-to-month changes in the gravitational field of the Earth mainly based on precise K-band microwave measurements of the distance between two low-flying satellites (Wahr, 2009). After correcting for short-term variability due to tides in atmosphere (Biancale and Bode, 2006), solid earth (Petit and Luzum, 2010) and oceans (Savcenko and Bosch, 2012), as well as due to non-tidal variability in atmosphere and oceans (Dobslaw et al., 2013) from the observations, the resulting gravity changes mainly represent mass transport phenomena in the Earth system, which are - apart from long-term trends - almost exclusively related to the global water cycle.

We use the monthly GRACE release 05a Level-2 products from GFZ Potsdam (Dahle et al., 2012), which can be downloaded from the website of the International Centre for Global Earth Models (icgem.gfz-potsdam.de/ICGEM). The GRACE products are expressed in terms of fully normalized spherical harmonic (SH) coefficients up to degree and order 90, approximately corresponding to a global resolution of 2° in latitude and longitude. We apply the same post-processing steps to the GRACE data as described by Zhang et al. (2016). The degree-1 coefficients are added following the method of Bergmann-Wolf et al. (2014). The non-isotropic filter DDK2 corresponding to an isotropic Gaussian filter with 680 km full width half maximum (Kusche, 2007; Kusche et al., 2009) is applied to remove correlated errors at particular higher degrees of the spherical harmonic expension. In order to account for signal attenuation and leakage caused by smoothing and filtering, local re-scaling factors are introduced for each grid cell. We use median re-scaling factors obtained from a small ensemble of global hydrological models. The gridded TWS anomalies are then estimated which can be averaged over arbitrary basins. As for the model data, the linear trend is removed over the period Jan 2003 to Dec 2012. Error estimates as a quadrature of measurement error, leakage error and re-scaling error are also provided to assess the signal-to-noise ratio (SNR) of GRACE for particular basins (full details are given in Zhang et al. (2016)). In case of a small signal-to-noise ratio, discrepancies between TWS from GRACE and models might also be attributed to comparatively large GRACE TWS errors.

3 Evaluation of TWS from model realizations with GRACE

We compare the basin-averaged TWS from GRACE with the results of four different numerical model realizations introduced above. In total 31 globally distributed basins, where the GRACE SNR is larger than 2 (see Fig.1 and Table 2) are selected for further study. We first focus on the global statistical performance of the models compared to GRACE. For these basins, evaluation metrics as suggested by Gudmundsson et al. (2012) that focus both on seasonal signals and year-to-year variability are applied.

3.1 Evaluation metrics

First, relative annual amplitude differences are calculated according to

$$\Delta\mu = (\mu_M - \mu_O)/\mu_O, \tag{1}$$

where μ_O is the annual amplitude of the time series of TWS variations from GRACE, μ_M the annual ~~GRACE~~ TWS amplitudes from the different model realizations (Fig. 2). Second, the timing of the annual cycle is assessed using phase differences of the annual harmonic for models and observations according to

$$\Delta\phi = (\phi_M - \phi_O). \tag{2}$$

If the value of $\Delta\phi$ is negative, it implies that the seasonal maximum is earlier in the year in the model than in GRACE (Fig. 3. Annual amplitude and phase are calculated by least square regression as follows:

$$MIN \overset{!}{=} (\Delta\mathrm{TWS(t)} - (a + vt + A\sin(2\pi t/T + \phi))^T(\Delta\mathrm{TWS(t)} - (a + vt + A\sin(2\pi t/T + \phi)) \tag{3}$$

where $\Delta\mathrm{TWS}$ is the TWS anomaly time series, a is the constant, v is the trend, and T is the period of one year. Third, the explained variances for all the model realizations are calculated:

$$R^2 = (var(\mathrm{TWS_O}) - \mathrm{var}(\mathrm{TWS_O} - \mathrm{TWS_M}))/\mathrm{var}(\mathrm{TWS_O}) \tag{4}$$

where var denotes the variance operator. Fourth, we repeat the calculation of the explained variances for TWS time series from GRACE and the models with the mean seasonal variability removed.

3.2 Global evaluation

As shown in Fig. 2, the values of $\Delta\mu$ for WGHM and JSBACH are mostly negative. For JSBACH, these negative values mainly occur at mid to high latitudes of the Northern Hemisphere. WGHM underestimates the annual amplitude especially at the low latitudes. Contrarily, MPI-HM has more basins with positive $\Delta\mu$. For LSDM, most $\Delta\mu$ values lie between -0.3 and 0.3, indicating on average better agreement of annual amplitude with GRACE. The phase difference varies more among the different models, but in most cases an earlier seasonal storage maximum is shown for the model runs relative to GRACE. There are more basins with phase difference values near zero for LSDM, while WGHM, JSBACH and MPI-HM show large differences with respect to the GRACE result, especially in high latitudes of the Northern Hemisphere. (Fig. 3). LSDM explains the GRACE TWS variations relatively better than the other models at most basins (Fig. 4). Only in the Yukon, Nile, Zaire, Yangtze, Indus and the two basins at Australia, explained variances are less than 50%. Low values of explained variance also occur at the mid-latitude of the Northern Hemisphere for WGHM. JSBACH and MPI-HM perform generally better at basins in Africa but have worse results in Siberia. When the annual signal is removed, the explained variances for TWS time series from GRACE and the models are generally less than 60% (Fig. 5), indicating the models's poor ability to capture the inter-annual variations. LSDM shows especially low explained variance values for many basins in Africa.

The impact from consideration of groundwater to TWS variations in WGHM is investigated by showing the differences of explained variances with and without groundwater (Fig. 6). The positive values indicate that WGHM with groundwater exhibits better agreement with GRACE than the one without. The large impact mainly locates at basins such as Toscantins, Niger, Huang He, Mekong and Mississippi. Only in three basins (Lena, Indus and Yukon), the effect of groundwater consideration in the model is negative.

As each metric usually focuses only on one specific property of statistical performance and has its own limitations, the time series of TWS are given for some basins with the largest deviation between GRACE and the model. We shown Yukon basin, where both WGHM and JSBACH exhibit the largest deviation of annual amplitudes from GRACE. Although the annual amplitude is simulated better by LSDM and MPI-HM, apparent negative phase differences are shown. Amur basin is also shown, as LSDM, WGHM and MPI-HM all have the largest negative phase differences with GRACE here. Models generally capture the inter-annual signals but perform quite differently among each other and with GRACE in terms of seasonality. Almost opposite phase differences are found for these models. The smallest explained variance for MPI-HM happens at St. Lawrence basin, where a much larger amplitude and a negative phase difference compared with GRACE are found. When the annual signal is removed, models perform differently in terms of the explained variance. In Nile basin, large inter-annual variations simulated by LSDM lead to even negative explained variance compared with the other models.

Fig. 8 summarizes the overall performance of each statistical metric for all the basins considered by means of box plots. The median $\Delta\mu$ for MPI-HM is almost zero where the other three values are all negative, indicating an underestimation of the annual amplitude of TWS from LSDM, WGHM and JSBACH. As shown in Fig. 2d, MPI-HM overestimates the TWS variations at many basins, which compensate with those underestimated values and ~~come to a almost zero median value~~ lead to a median value at almost zero. All the models have a median phase difference below zero, with LSDM having the smallest bias and range, and MPI-HM the largest bias. This ~~shows~~ means that the TWS peaks of the models tend to proceed GRACE peaks~~, where LSDM performs best compared to other models~~. For the explained variance, LSDM shows the best median value, followed by WGHM, JSBACH and MPI-HM. However, when the annual signal is removed, many outliers appear in LSDM for the explained variances, while WGHM and MPI-HM show slightly better performances.

We also present the basin-averaged TWS errors from GRACE and the RMS differences between TWS variations from GRACE and from the hydrological model runs (Table 2), where the largest and smallest differences are shown in bold and underlined separately. The basins are grouped according to the Köppen climate zones (Kottek et al., 2006), which include Tropical climates, Dry climates, Temperate climates and Cold climates (see Fig. 1). For most of the basins, the GRACE errors are much smaller than the RMS differences, which indicates that the main contributions to the differences arise from model uncertainties. Out of the five basins in the tropical zone, three basins have largest differences between TWS variations from GRACE and models in LSDM. On the contrary, WGHM ~~shows~~ has no largest differences in this climate zone. The smallest value, however, seems to occur randomly among the models. In the dry zone, most ~~models~~ basins have low SNR values and the smallest RMS of the TWS differences are sometimes quite close to the GRACE TWS errors. For instance, at basins like Nile, Indus, and two Australian basins, the GRACE SNR estimates are all below 3. Thus, it is likely that the large uncertainty in GRACE TWS estimates contribute largely to the bad agreement in these basins. Still, MPI-HM and LSDM perform comparably better, showing a smaller number of largest differences and comparably more smallest differences. In the temperate zone, WGHM has most largest differences while MPI-HM has least. There is, however, no regular pattern of where the smallest difference occurs. In the cold zone, all the smallest differences happen in LSDM, whereas the largest differences mainly occur at MPI-HM and JSBACH.

The performance of the models varies from basin to basin, even within the same climate zone, which could be due to the model structure, parametrization, and also the different water storage compartments included in TWS. In order to find reasons for the different model performance, we focus on two specific areas that are dominated by snow and arid climates in more detail. There, we ~~first~~ assess actual evapotranspiration (AET) ~~which is one of the main drivers for differences in~~ and runoff which are the main components of the terrestrial water budget and subsequently look into the mean monthly time series of TWS and its individual storage compartments.

3.3 Actual evapotranspiration and runoff

As ~~a part of~~ part of the terrestrial branch of the water cycle, actual evapotranspiration (AET) and runoff may explain part of the differences among the models in terms of storage variations. ~~We choose four particularly affected basins and show the AET time series from all models (Fig. 7).~~ Although some large differences of AET are present, the effects on subsequently simulated TWS are damped. Especially in humid areas, no direct impact can be found. For arid basins, however, the impact from AET is more dominant. ~~For basins as Chari, Indus, Murray, and Niger~~We choose three particularly affected basins (Niger, Chari and Indus) and show the AET time series from all models (Fig. 9). For these basins, the time series comparison shows that the smaller (or larger) AET in wet season lead to higher (or lower) seasonal amplitude of TWS. Besides, in these dry areas, LSDM generally exhibits enhanced AET due to high temperatures and extremely low humidity which then lead to smaller TWS variations. As exemplarily demonstrated for the Niger basin, the relatively larger AET from LSDM covering the time period 2007 to 2009 are just correspondent to the comparably smaller TWS variations.

AET is calculated from the potential evapotranspiration (PET) as a function of the available amount of water. While starting with the same meteorological forcing data, PET is calculated differently by the models using various approaches. PET in the LSDM is calculated by the Thornthwait method, using only the daily temperature and a seasonal heat index that is based on monthly mean temperatures. In WGHM, PET is based on the Priestley-Taylor approach using net radiation, which in turn is computed as a function of incoming short-wave radiation, temperature and surface albedo. For MPI-HM, PET is computed in a pre-processing step based on Penman-Montheith using radiation, temperature, wind and humidity. JSBACH computes evaporation based on the energy balance by internally computing atmospheric water demand.

~~Fig. 7 also~~

Fig. 10 displays time series comparison of ~~PET from WGHM and LSDM. Some differences in PET are seen from these two models because of the different methods applied. These differences, however, are substantially modified when entering into AET due to the limitation of available water .~~ runoff from the models for three basins in the tropical zone (Amazon, Orinoco and Mekong). The runoff is calculated from the models following the equation:

$$R(t) = P(t) - ET(t) - TWSC(t), \tag{5}$$

where t is the time, P, ET and R are the basin-averaged precipitation, evapotranspiration and runoff, and TWSC is the terrestrial water storage change (Ramillien et al., 2006). It is seen that the performance of a certain model is connected with its differently simulated runoff. At Amazon basin, the comparably large runoff simulated from MPI-HM also leads to smaller variability in TWS, which is also shown at Orinoco basin. At Mekong basin, the larger amplitude in TWS from JSBACH compared with GRACE is related to the apparent small amplitude in its runoff.

3.4 Snow-dominated catchments

As highlighted in section 3.2, models perform quite differently in high latitudes of the Northern Hemisphere (cold zone) which are generally dominated by snow. Especially JSBACH and MPI-HM show large differences of the TWS when compared with

GRACE. We focus here on four basins in this area: Lena, Yenisei, Ob and Yukon, and look into the mean monthly time series of the TWS and its different compartments (Fig. ~~8~~11). For LSDM and MPI-HM, subsurface water ~~here~~ only includes the water storage in the root zone, while for WGHM and JSBACH, both root zone and deep layer water storage are included. ~~The performances of the models at those four basins are quite consistent with each other.~~ LSDM and WGHM show the smallest phase differences with GRACE in terms of TWS while the other two exhibit negative phase shifts. The subsurface water variations from WGHM and LSDM have very similar pattern, with an apparent peak usually in May. The phases of the snow water time series from LSDM and WGHM are also quite close, but LSDM always has a slightly larger amplitude. Since the two use the same snow scheme (degree-day method), this is certainly related to the different model parameters or sub-grid representation schemes. The surface water storage from these two models are sometimes different. For the Ob river, for instance, the different surface water storage also leads to the poor performance of WGHM in terms of TWS when compared with GRACE. The snow variations from LSDM and MPI-HM are almost identical with each other. However, the different subsurface and surface water simulated by MPI-HM causes a bad timing of the TWS peaks. For the Lena basin, although the snow variations from LSDM, WGHM and MPI-HM are quite close, MPI-HM simulates almost no surface water variations which leads to a poor agreement of TWS with GRACE estimates. For JSBACH, there is already a large phase difference in the snow storage, which is mainly due to the poor capture of the phase of the snow accumulation and onset of melting. This could be caused by the ~~different~~ specific snow scheme applied by JSBACH. Yukon, however, is quite different from the other snow-dominated basins. Here, all the models underestimate the annual amplitude of TWS when compared with GRACE. Since the basin-average TWS error from GRACE at Yukon is 1.19 cm and much smaller than the discrepancies between GRACE and the models (Table 2), it could be the case that all models fail to represent certain hydrological processes, or that our GRACE TWS errors are too optimistic here since the re-scaling errors are also estimated from a hydrological model ensemble. Besides, Seo et al. (2006) found also large TWS errors at Yukon basin and suggested that the atmosphere and ocean tidal and non-tidal de-aliasing errors might be a problem in this area. Investigating those discrepancies in full detail, however, is beyond the scope of our present paper and ~~we would like to assess the de-aliasing errors in a~~ will be left open for future study.

3.5 Dry catchments

We also focus on four catchments in the dry zone, which are characterized by annual precipitation smaller than annual potential evapotranspiration (McKnight and Hess, 2000). For the Nile and Niger basins, the subsurface water is the main contributor to the TWS changes (Fig. ~~9~~12). The TWS variations from JSBACH and MPI-HM show a quite similar annual cycle when compared to GRACE. MPI-HM generally exhibits a larger amplitude in simulated subsurface water and TWS. WGHM deviates considerably with a much smaller amplitude and a large phase shift in the subsurface water. The simulated surface water from WGHM brings TWS slightly closer to that from GRACE. LSDM, however, performs differently in these two basins. In Nile basin, although the subsurface water from LSDM is consistent with JSBACH and MPI-HM, the simulated surface water variations lead to a higher amplitude of TWS variations when compared with GRACE. In Niger, LSDM performs quite close to WGHM but with a slightly larger amplitude. All models tend to perform poorly in terms of TWS when compared with GRACE in Indus basin. We note a comparably low SNR (2.2 cm) for the GRACE estimated TWS here, which is mainly contributed by the large leakage error at this basin (Zhang et al., 2016). Besides, Indus basin is not only subject to large-scale groundwater depletion from intensive irrigation, but also affected by snow melting and glaciers melting from Himalaya. Here, the subsurface water simulated by the models show already large discrepancies. As in other basins affected by snow dynamics, JSBACH also fails to capture the snow variations properly. MPI-HM performs poorly in simulating the surface water with a delayed dynamics which leads to a preceded annual cycle. At Huang He basin, ~~as the main contributor to the TWS, the~~ the subsurface water from LSDM, WGHM and JSBACH as the main contributor to the TWS show similar annual variations as GRACE, while MPI-HM has a much larger amplitude. The surface water simulated differently by LSDM and WGHM then lead to different TWS variations.

4 Summary

We validate TWS variations simulated by four different global hydrological models with monthly GRACE gravity data. All the models are forced with the same WFDEI meteorological data-set to exclude the effect of meteorological forcing on the models. Four statistical metrics focusing on different aspects of model performance compared with GRACE have been applied. In addition, time series of TWS variations from GRACE and models are investigated, where different water storage compartments from models are shown as well.

At certain basins like Danube, Tocantins, Columbia, Ganges, Mekong, and Amazon, all numerical models show good agreement with GRACE. However, models still perform quite differently at many other basins, even though forced with the same meteorological data set. At Nile, Indus, Murray and Great Artesian Basin, large TWS errors and low SNR are found which suggests a major contribution from GRACE errors to the differences. A good capture of annual amplitude and phase at most basins leads to high values of explained variance in many basins for LSDM. However, serious problems are also found in the same model run in some central Africa basins, like Nile and Zaire, where TWS simulated by LSDM exhibits unusual large inter-annual variations. WGHM performs generally good at tropical and cold regions, but ~~poorly at~~ rather poorly in the temperate zone. JSBACH and MPI-HM show large discrepancies with GRACE at the basins in high latitudes of the Northern Hemisphere.

Model performance is also investigated in some snow dominated and dry catchments in more detail through time series comparison. The poor performance of JSBACH and MPI-HM in snow dominated regions is mainly related ~~with the~~ to negative phase shifts compared to GRACE. ~~Although~~ MPI-HM simulates identical snow variations as LSDM, however, the different simulations of subsurface water and especially surface water still lead to different TWS variations in snow dominated regions. ~~For JSBACH~~Despite of the missing surface water compartment, the simulated snow variations in JSBACH already show smaller amplitude and negative phase differences compared with all the other models~~, which also lead to the different performance of TWS. The~~ . This could be related to the fact that JSBACH simulates snow in a more physical way based on energy balance, which is totally different from the degree-day method applied by all the other models. The comparably better agreement of LSDM and WGHM with GRACE in terms of TWS in these snow dominated basins is partly caused by the realistic surface water compartment representing by these two models. In the dry catchments, the impact from AET on TWS is relatively strong~~in arid areas. For instance,~~ . The smaller AET from MPI-HM also leads to better agreement with GRACE, whereas LSDM shows large differences with GRACE in terms of TWS especially at some dry basins in ~~Africa, the smaller AET from JSBACH and~~ central Africa partly due to the too simple evaporation scheme. PET is simulated using a superior parametrization by MPI-HM~~compared with LSDM and WGHM also lead to better agreement with GRACE than LSDM~~, while LSDM applies still the traditional Thornthwaite method based solely on air temperature. The groundwater considered by WGHM also has some impact on the simulated TWS, especially at basins as Toscantins, Mekong, Niger and Mississippi. At Yukon basin, we found the bad performance of all models in terms of TWS when compared with GRACE, which could be due to the effects of atmospheric and oceanic de-aliasing errors not further discussed in our current study. In future, we would like to assess all possible errors of GRACE TWS through investigation of simulated GRACE-type gravity field time-series (Flechtner et al., 2016) based on realistic orbits and instrument error assumptions as well as background error assumptions out of the updated ESA Earth System Model (Dobslaw et al., 2015, 2016)~~in our next step~~, which we believe will further help to explain the discrepancy between ~~models and GRACE~~global models of the terrestrial water cycle and GRACE satellite observations.

Acknowledgements. This study has been supported by the German Federal Ministry of Education and Research within the FONA research program under grants 03F0654A and 01LP1151A.

References

Bergmann-Wolf, I., Zhang, L., and Dobslaw, H.: Global eustatic sea-level variations for the approximation of geocenter motion from GRACE, J. Geod. Sci., 4, 37–48, doi:10.2478/jogs-2014-0006, 2014.

Biancale, R. and Bode, A.: Mean annual and seasonal atmospheric tide models based on 3-hourly and 6-hourly ECMWF surface pressure data , Scientific Technical Report STR06/01, GFZ, Helmholtz-Zentrum, Potsdam, doi:10.2312/GFZ.b103-06011, 2006.

Bierkens, M. F. P. and van den Hurk, B. J. J. M.: Groundwater convergence as a possible mechanism for multi-year persistence in rainfall, Geophys. Res. Lett., 34, doi:10.1029/2006GL028396, http://dx.doi.org/10.1029/2006GL028396, 2007.

Brovkin, V., Raddatz, T., Reick, C. H., Claussen, M., and Gayler, V.: Global biogeophysical interactions between forest and climate, Geophys. Res. Lett., 36, doi:10.1029/2009GL037543, 2009.

Chen, J. L., Wilson, C. R., and Tapley, B. D.: The 2009 exceptional Amazon flood and interannual terrestrial water storage change observed by GRACE, Water Resour. Res., 46, doi:10.1029/2010WR009383, http://dx.doi.org/10.1029/2010WR009383, w12526, 2010.

Dahle, C., Flechtner, F., Gruber, C., König, D., König, R., Michalak, G., and Neumayer, K.: GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005 , Scientific technical report-data, GFZ, Helmholtz-Zentrum, Potsdam, Potsdam, doi:10.2312/GFZ.b103-1202-25, 2012.

Dahle, C., Flechtner, F., Gruber, C., König, D., König, R., Michalak, G., and Neumayer, K.-H.: GFZ RL05: An Improved Time-Series of Monthly GRACE Gravity Field Solutions, in: Observation of the System Earth from Space - CHAMP, GRACE, GOCE and future missions, edited by Flechtner, F., Sneeuw, N., and Schuh, W.-D., Advanced Technologies in Earth Sciences, pp. 29–39, Springer Berlin Heidelberg, doi:10.1007/978-3-642-32135-1_4, http://dx.doi.org/10.1007/978-3-642-32135-1_4, 2014.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteor., 137, 553–597, doi:10.1002/qj.828, http://dx.doi.org/10.1002/qj.828, 2011.

Dill, R.: Hydrological model LSDM for operational Earth rotation and gravity field variations , GFZ Scientific Technical Report-STR08/09, GFZ, Helmholtz-Zentrum, Potsdam, Potsdam, 2008.

Dill, R. and Dobslaw, H.: Short-term polar motion forecasts from earth system modeling data, J. Geodesy, 84, 529–536, doi:10.1007/s00190-010-0391-5, 2010.

Dill, R. and Dobslaw, H.: Numerical simulations of global-scale high-resolution hydrological crustal deformations, J. Geophys. Res.: B, 118, 5008–5017, doi:10.1002/jgrb.50353, 2013.

Dirmeyer, P. A.: A history and review of the global soil wetness project (GSWP), J. Hydrometeor, 12, 729–749, doi:10.1175/JHM-D-10-05010.1, 2011.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., , and Hanasaki, N.: GSWP-2: Multimodel analysis and implications for our perception of the land surface, Bull. Amer. Meteor. Soc., 87, 1381–1397, doi:110.1175/BAMS-87-10-1381, 2006.

Dobslaw, H., Dill, R., Grötzsch, A., Brzeziński, A., and Thomas, M.: Seasonal polar motion excitation from numerical models of atmosphere, ocean, and continental hydrosphere, J. Geophys. Res.: B, 115, doi:10.1029/2009JB007127, 2010.

Dobslaw, H., Flechtner, F., Bergmann-Wolf, I., Dahle, C., Dill, R., Esselborn, S., Sasgen, I., and Thomas, M.: Simulating high-frequency atmosphere-ocean mass variability for dealiasing of satellite gravity observations: AOD1B RL05, J. Geophys. Res.: Oceans, doi:10.1002/jgrc.20271, 2013.

Dobslaw, H., Bergmann-Wolf, I., Dill, R., Forootan, E., Klemann, V., Kusche, J., and Sasgen, I.: The updated ESA Earth System Model for future gravity mission simulation studies, Journal of Geodesy, 89, 505–513, doi:10.1007/s00190-014-0787-8, http://dx.doi.org/10.1007/s00190-014-0787-8, 2015.

Dobslaw, H., Bergmann-Wolf, I., Forootan, E., Dahle, C., Mayer-Gürr, T., Kusche, J., and Flechtner, F.: Modeling of present-day atmosphere and ocean non-tidal de-aliasing errors for future gravity mission simulations, Journal of Geodesy, 90, 423–436, doi:10.1007/s00190-015-0884-3, http://dx.doi.org/10.1007/s00190-015-0884-3, 2016.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, J. Hydrol, 270, 105–134, doi:10.1016/S0022-1694(02)00283-4, 2003.

Flechtner, F., Neumayer, K.-H., Dahle, C., Dobslaw, H., Fagiolini, E., Raimondo, J.-C., and Güntner, A.: What Can be Expected from the GRACE-FO Laser Ranging Interferometer for Earth Science Applications?, Surveys in Geophysics, 37, 453–470, doi:10.1007/s10712-015-9338-y, http://dx.doi.org/10.1007/s10712-015-9338-y, 2016.

Gudmundsson, L., Wagener, T., Tallaksen, L. M., and Engeland, K.: Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe, Water Resour. Res., 48, doi:10.1029/2011WR010911, W11504, 2012.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., Yehm, P., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., and Heinke, J.: Multimodel estimate of the global terrestrial water balance: Setup and first results, J. Hydrometeor, 12, 869–884, doi:10.1175/2011JHM1324.1, 2011.

Hagemann, S. and Gates, L.: Validation of the hydrological cycle of ECMWF and NCEP reanalyses using the MPI hydrological discharge model, J. Geophys. Res., 106, 1503–1510, doi:10.1029/2000JD900568, 2001.

Hagemann, S. and Gates, L.: Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations, Clim. Dyn., 21, 349–359, doi:10.1007/s00382-003-0349-x, 2003.

Hirschi, M., Seneviratne, S., and Schär, C.: Seasonal variations in terrestrial water storage for major midlatitude river basins, J. Hydrometeor, 7, 39–60, doi:10.1175/JHM480.1, 2006.

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydro. Earth Syst. Sc., 12, 841–861, 2008.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J. S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, JAMES, 5, 422–446, doi:10.1002/jame.20023, 2013.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y., and Yamada, T.: Regions of strong coupling between soil moisture and precipitation, Science, 305, 1138–1140, doi:10.1126/science.1100217, http://www.sciencemag.org/content/305/5687/1138.abstract, 2004.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15, 259–263, doi:10.1127/0941-2948/2006/0130, 2006.

Kusche, J.: Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, J. Geodesy, 81, 733–749, doi:10.1007/s00190-007-0143-3, 2007.

Kusche, J., Schmidt, R., Petrovic, S., and Rietbroek, R.: Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model, J. Geodesy, 83, 903–913, doi:10.1007/s00190-009-0308-3, 2009.

Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resour. Res., 48, doi:10.1029/2011WR011453, W04531, 2012.

Leblanc, M. J., Tregoning, P., Ramillien, G., Tweed, S. O., and Fakes, A.: Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia, Water Resources Research, 45, n/a–n/a, doi:10.1029/2008WR007333, http://dx.doi.org/10.1029/2008WR007333, w04408, 2009.

McKnight, T. L. and Hess, D.: Climate zones and types: Climate Zones and Types, Physical geography: A landscape appreciation. Upper Saddle River, NJ: Prentice Hall, pp. 223–6, 2000.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., and Stockdale, T.: Decadal prediction. Can it be skillful?, Bull. Amer. Meteor. Soc., 90, 1467–1485, 2009.

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, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, http://www.hydrol-earth-syst-sci.net/18/3511/2014/, 2014.

Petit, G. and Luzum, B.: IERS Conventions (2010), IERS Technical Note ; 36, Bundesamt für Kartographie und Geodäsie, Frankfurt am Main, 2010.

Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K. G., Wetzel, P., and Jungclaus, J.: Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, Clim. Dyn., 29, 565–574, doi:10.1007/s00382-007-0247-8, 2007.

Ramillien, G., Frappart, F., Güntner, A., Ngo-Duc, T., Cazenave, A., and Laval, K.: Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry, Water Resources Research, 42, n/a–n/a, doi:10.1029/2005WR004331, http://dx.doi.org/10.1029/2005WR004331, w10403, 2006.

Rodell, M., McWilliams, E. B., Famiglietti, J. S., Beaudoing, H. K., and Nigro, J.: Estimating evapotranspiration using an observation based terrestrial water budget, Hydrological Processes, 25, 4082–4092, doi:10.1002/hyp.8369, http://dx.doi.org/10.1002/hyp.8369, 2011.

Savcenko, R. and Bosch, W.: EOT11a - Empirical ocean tide model from multi-mission satellite altimetry, DGFI Report No. 89, Deutsches Geodätisches Forschungsinstitut (DGFI), München, 2012.

Schewe, J., Heinke, J., Gerten, D., Hadeland, I., Arnell, N. w., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X. C., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q. H., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Pointek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, PNAS, 111, 3245–3250., doi:10.1073/pnas.1222460110, 2014.

Seneviratne, S., Viterbo, P., and abd C. Schär, D. L.: Inferring changes in terrestrial water storage using ERA-40 reanalysis data : the Mississippi River basin, J. Climate, 17, 2039–2057, 2004.

Seneviratne, S. I. and Stöckli, R.: The role of land-atmosphere interactions for climate variability in Europe, vol. 33 of *Advances in Global Change Research*, Springer Netherlands, doi:10.1007/978-1-4020-6766-2_12, http://dx.doi.org/10.1007/978-1-4020-6766-2_12, 2007.

Seo, K.-W., Wilson, C. R., Famiglietti, J. S., Chen, J. L., and Rodell, M.: Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE), Water Resources Research, 42, n/a–n/a, doi:10.1029/2005WR004255, http://dx.doi.org/10.1029/2005WR004255, w05417, 2006.

Stacke, T. and Hagemann, S.: Development and evaluation of a global dynamical wetlands extent scheme, Hydrol. Earth Syst. Sci., 16, 2915–2933, doi:10.5194/hess-16-2915-2012, 2012.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, JAMES, 5, 146–172, doi:10.1002/jame.20015, 2013.

Tang, Q., Gao, H., Yeh, P., Oki, T., Su, F., and Lettenmaier, D.: Dynamics of terrestrial water storage change from satellite and surface observations and modeling, J. Hydrometeor, 13, 156–170, 2010.

Tapley, B. D., Bettadpur, S., Watkins, M., and Reigber, C.: The gravity recovery and climate experiment: Mission overview and early results, Geophys. Res. Lett., 31, doi:10.1029/2004GL019920, L09607, 2004.

Wahr, J.: Time-variable gravity from satellites, vol. 3, Elsevier, 2009.

Wahr, J., Swenson, S., Zlotnicki, V., and Velicogna, I.: Time-variable gravity from GRACE: First results, Geophys. Res. Lett., 31, doi:10.1029/2004GL019779, 2004.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., , and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, J. Hydrometeor, 12, 823–848, doi:0.1175/2011JHM1369.1, 2011.

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 Resour. Res., 50, 7505–7514, doi:10.1002/2014WR015638, http://dx.doi.org/10.1002/2014WR015638, 2014.

Zeng, N., Yoon, J.-H., Mariotti, A., and Swenson, S.: Variability of basin-scale terrestrial water storage from a PER water budget method: The Amazon and the Mississippi, J. Climate, 21, 248–265, doi:10.1175/2007JCLI1639.1, 2008.

Zhang, L., Dobslaw, H., and Thomas, M.: Globally gridded terrestrial water storage variations from GRACE satellite gravimetry for hydrometeorological applications, Geophys J Int., 206, 368–378, doi:10.1093/gji/ggw153, 2016.

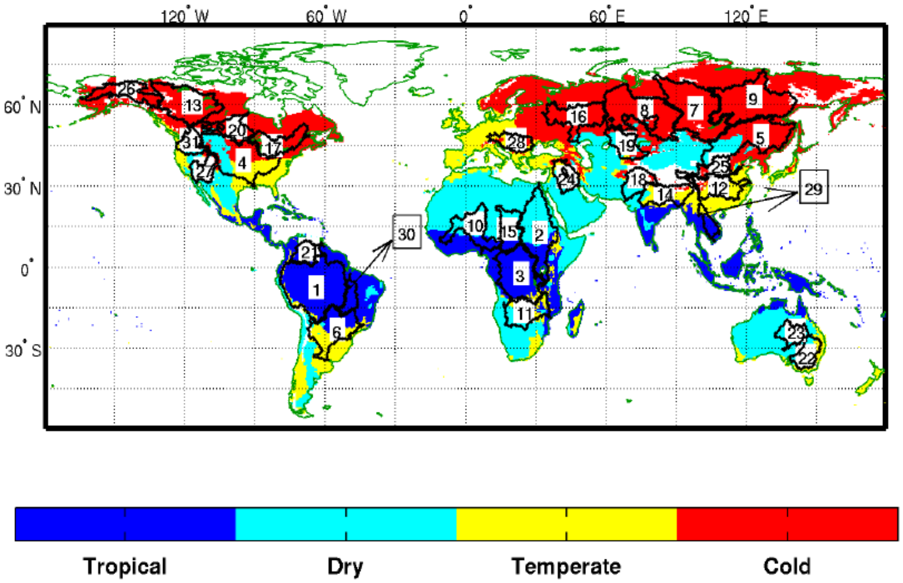

Figure 1. Locations of 31 globally distributed basins from the Simulated Topological Networks (STN-30p) with underlying Köppen-Geiger climate zones. Basins ID and names are indicated in Table 2.

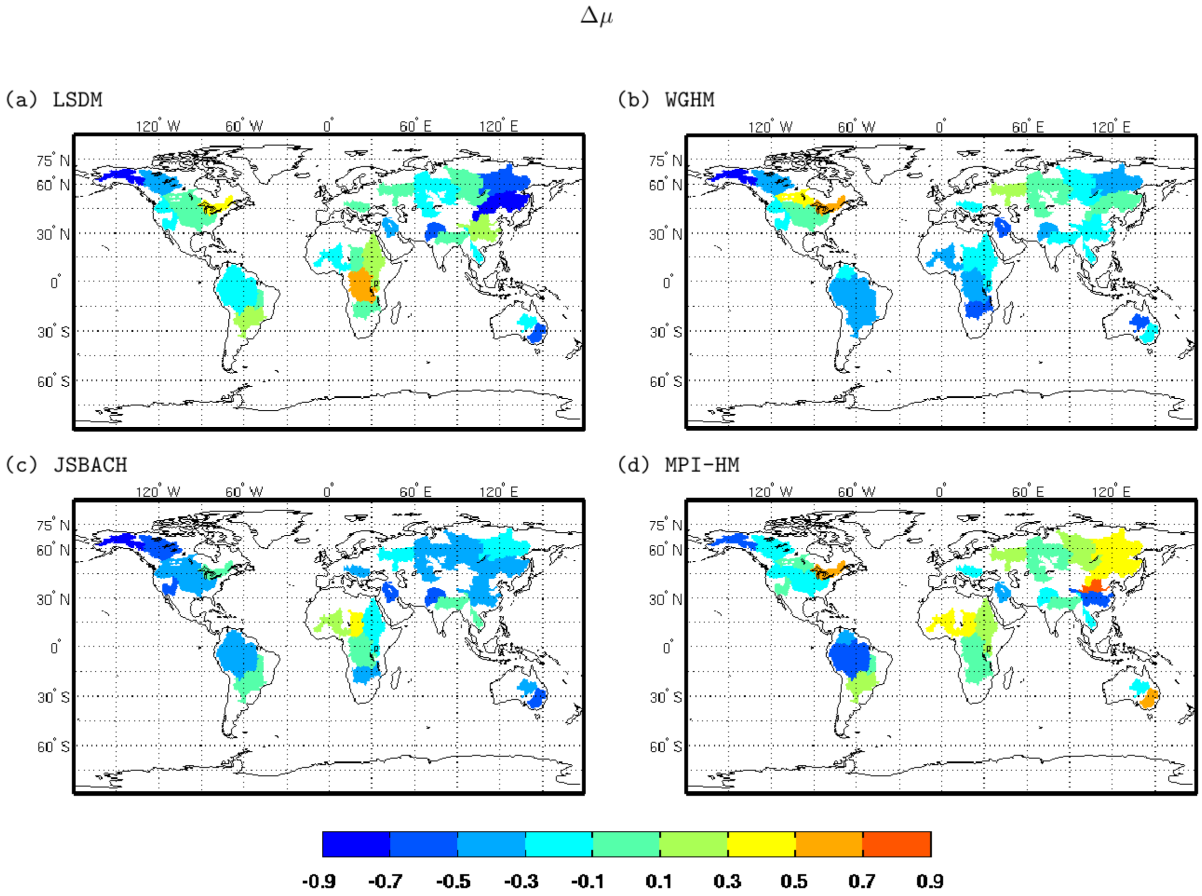

Figure 2. Relative amplitude differences of four hydrological model realizations with GRACE-based TWS observations.

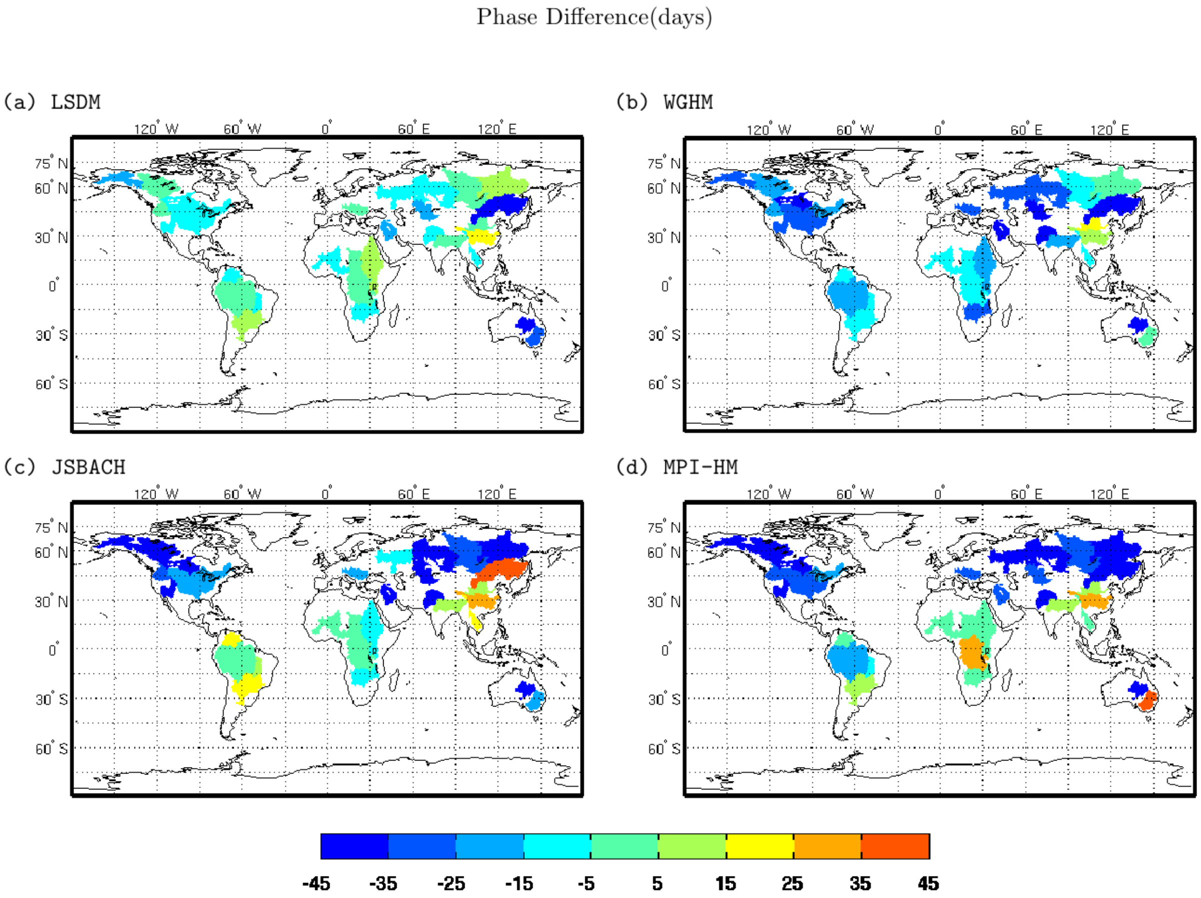

Figure 3. Phase differences for the annual signal of four hydrological model realizations with GRACE-based TWS observations.

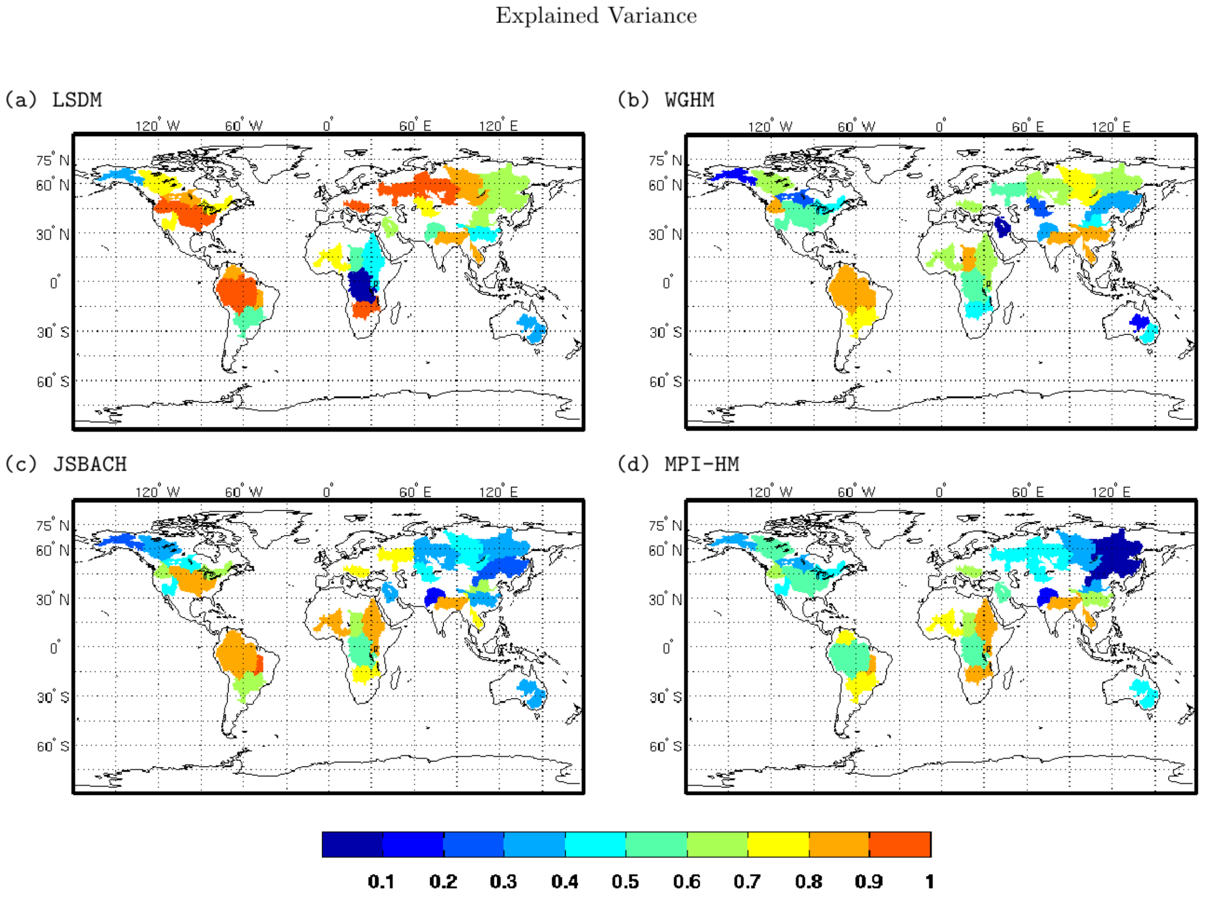

Figure 4. Variance of GRACE-based TWS observations that is explained by TWS as simulated in four hydrological model realizations.

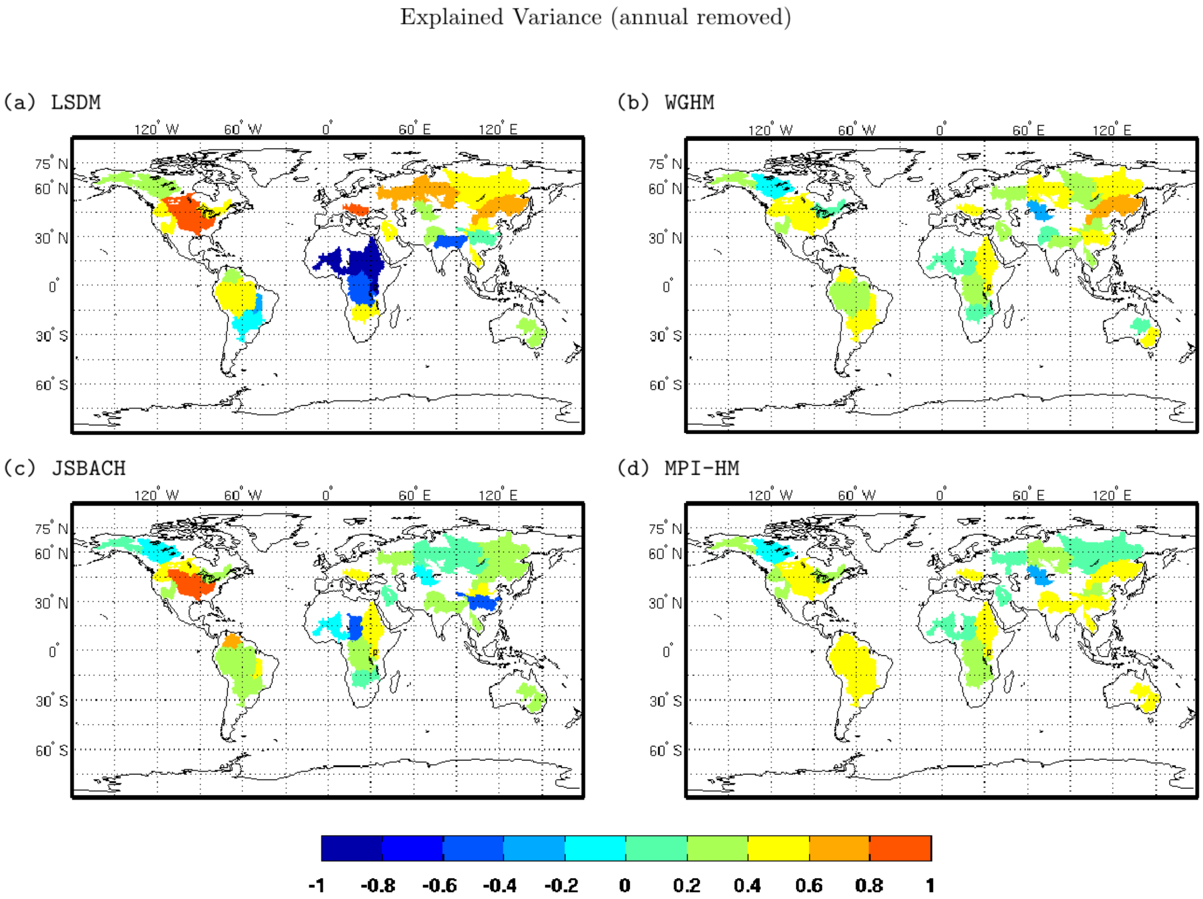

Figure 5. Variance of GRACE-based TWS observations that is explained by TWS as simulated in four hydrological model realizations. For both observations and model results, the annual signal has been removed.

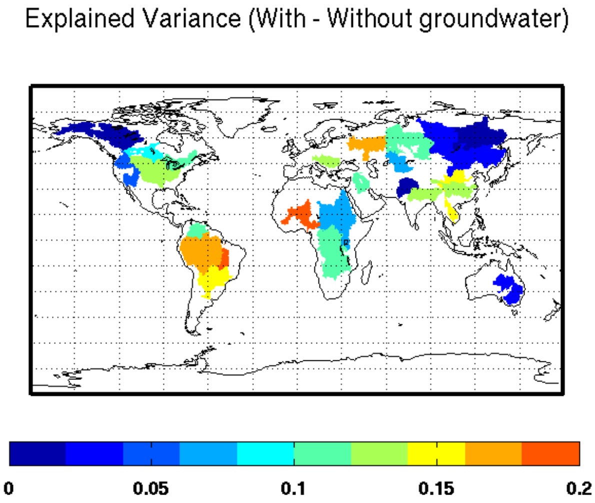

Figure 6. The differences between the explained variance values from WGHM with and without groundwater.

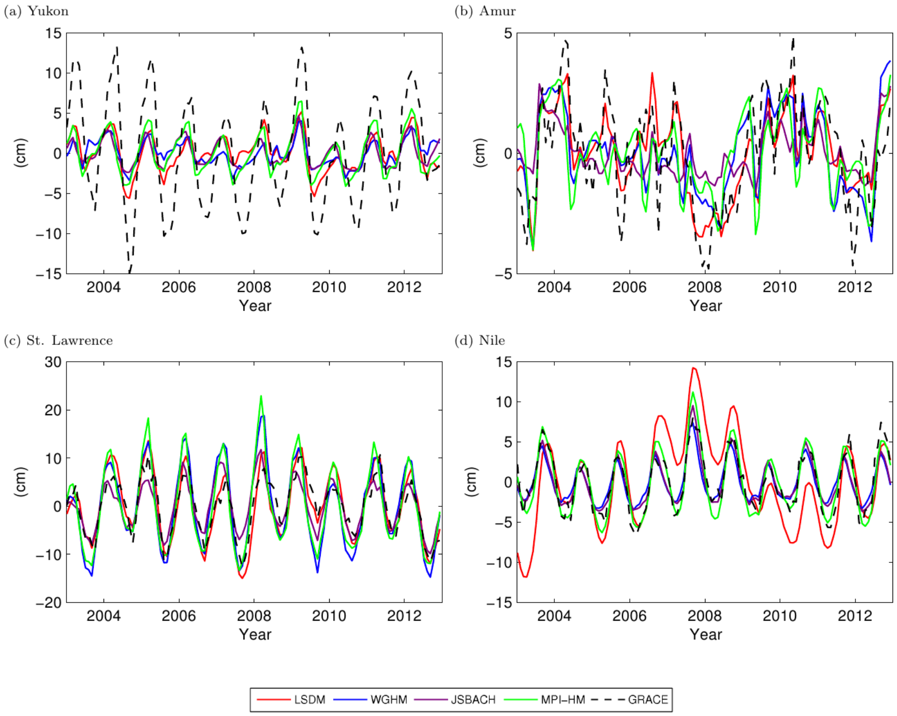

Figure 7. Examples of monthly TWS time series from GRACE and models for the basins with the largest deviation between model and GRACE in each of the four metrics: Relative amplitude differences (Yukon), phase differences (Amur), explained variance (St. Lawrence) and explained variance with annual signal removed (Nile).

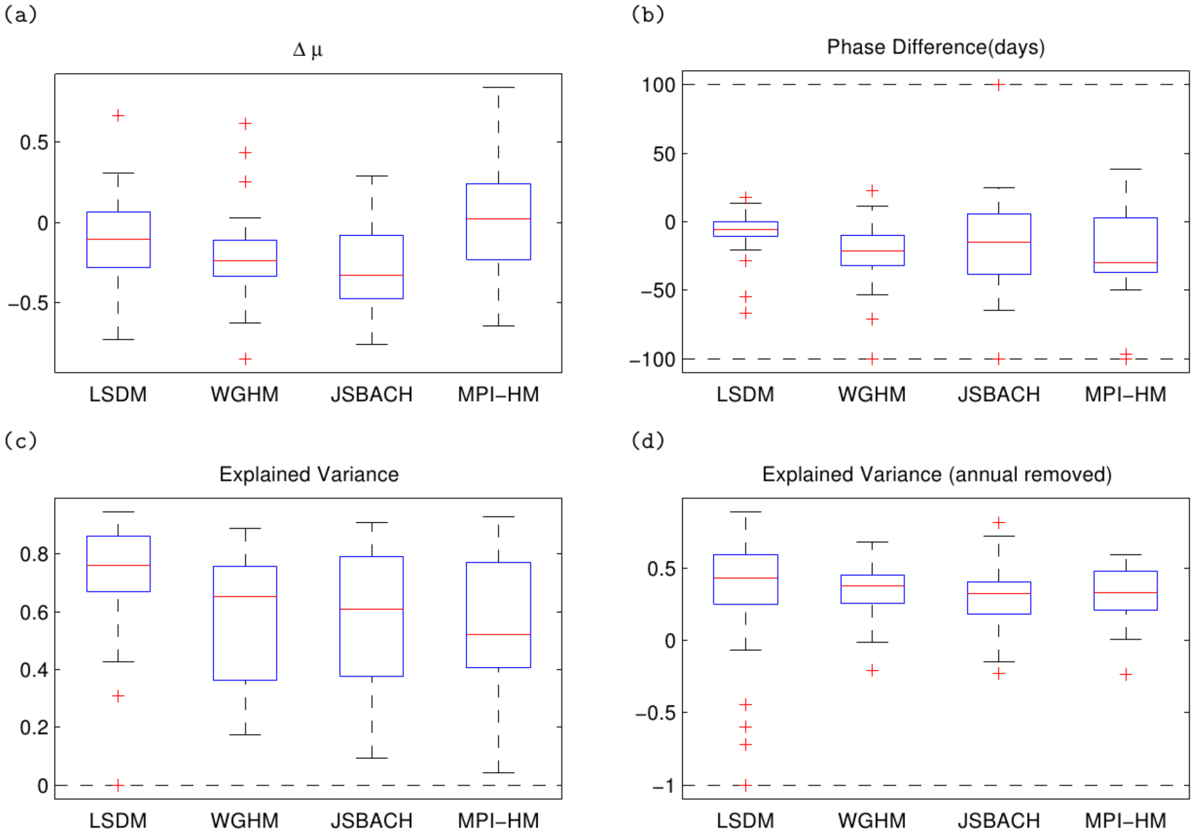

Figure 8. Box plots illustrating the $\Delta\mu$ (a), phase differences (b), Explained Variance (c) and Explained Variance with the annual signals removed (d) for the TWS from GRACE and models. The red horizontal line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually and set within the extreme data limits as indicated by the dashed line.

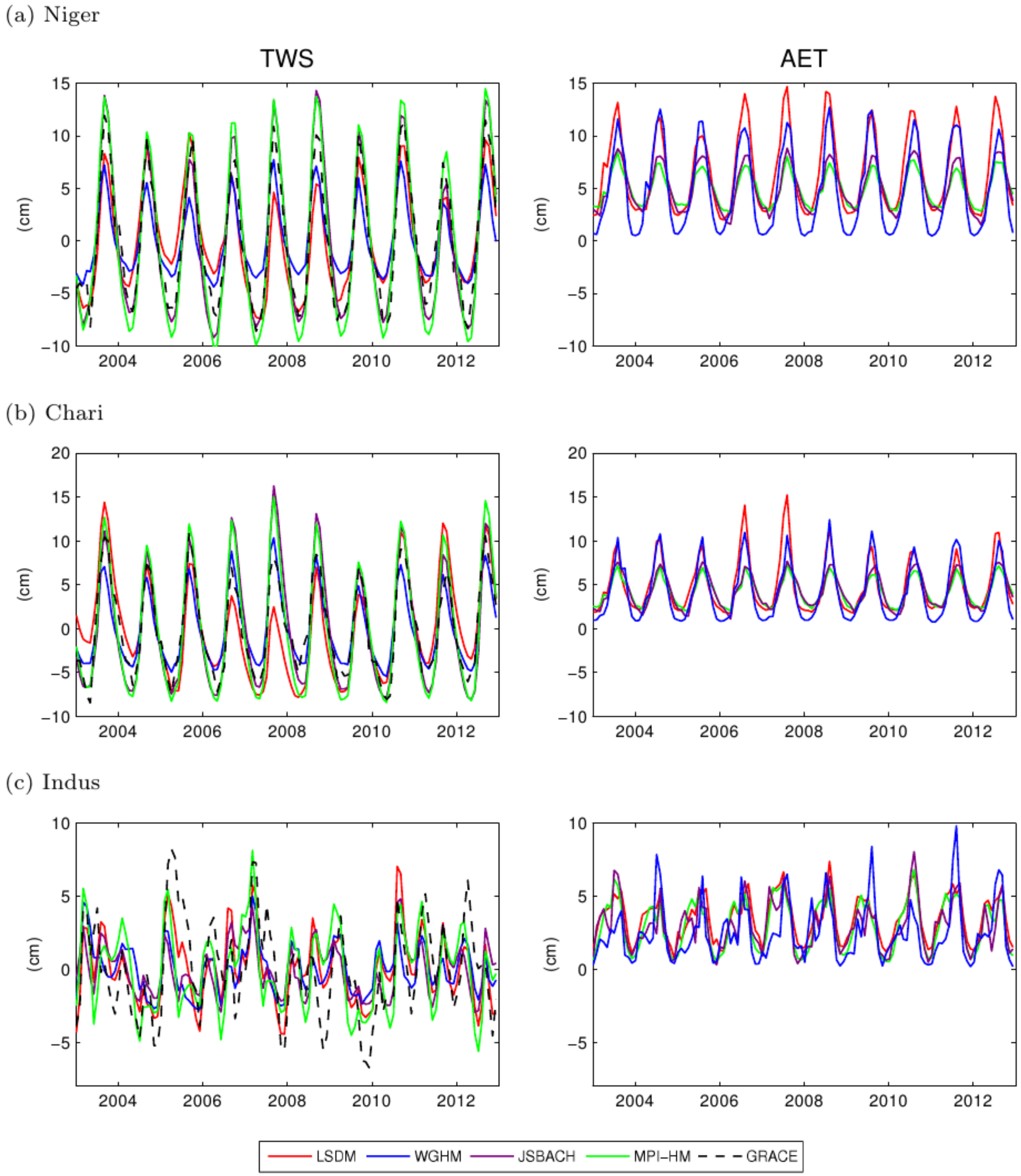

Figure 9. ~~Monthly time~~ Time series of TWS (~~first column~~left) from GRACE and models ~~,~~ and model simulated AET time series (~~second column~~right)~~and PET time series from LSDM and WGHM (third column)~~; each for ~~four~~ three different catchments in dry zone: ~~Amazon, Zaire, Mekong,~~ Niger, Chari and Indus.

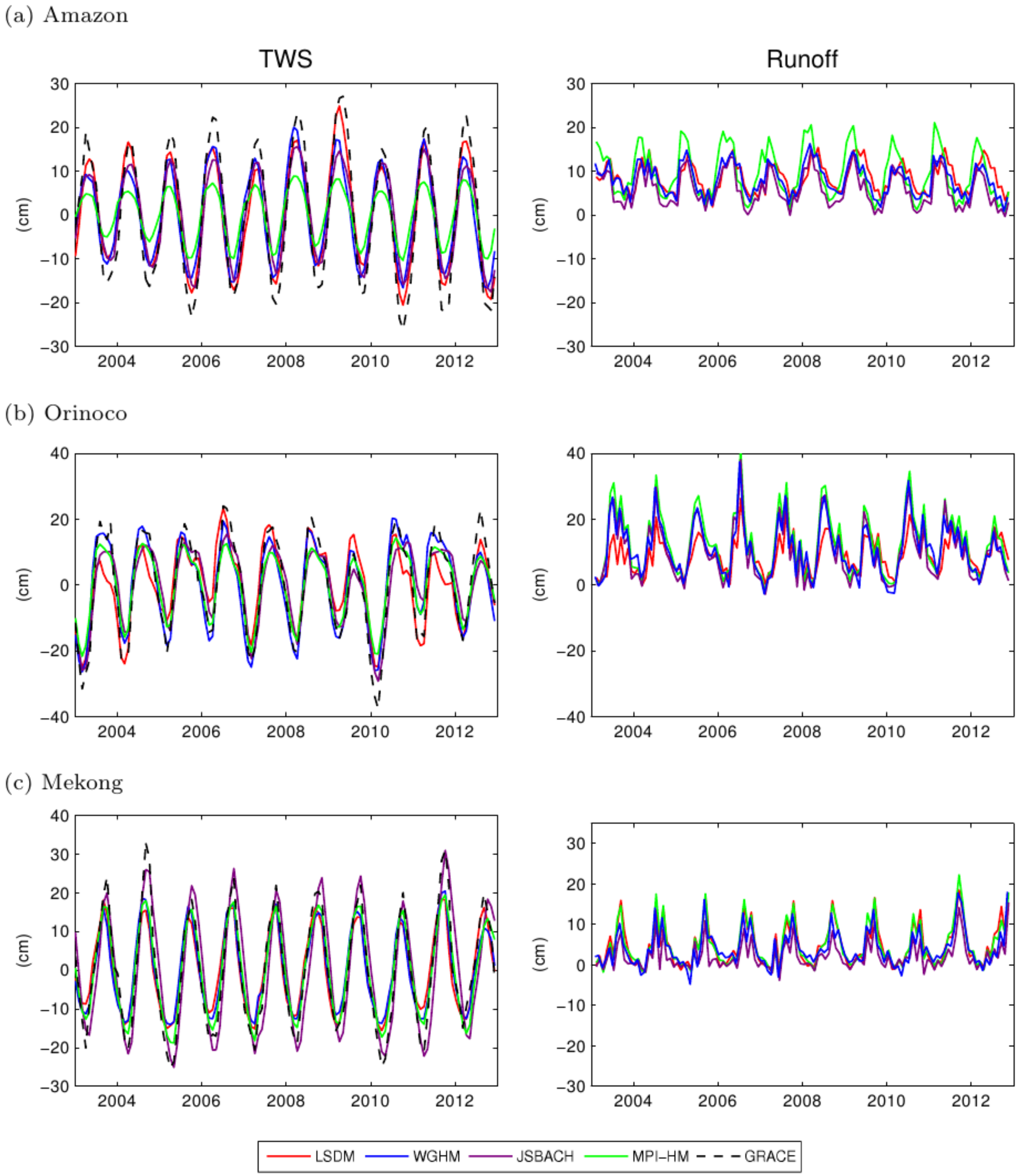

Figure 10. ~~Mean monthly time~~ Time series of TWS (~~first column~~left) ~~and the individual storage contributions~~ from ~~soil moisture (second column), snow water content (third column)~~ GRACE and ~~surface water~~ models and model simulated runoff time series (~~fourth column~~right); each for ~~five snowy~~ three different catchments in tropical zone: ~~Ob~~Amazon, ~~Lena~~Orinoco, ~~Yenizei~~ and ~~Yukon~~Mekong. ~~TWS from GRACE (dashed line) has been included into every sub-figure for reference.~~

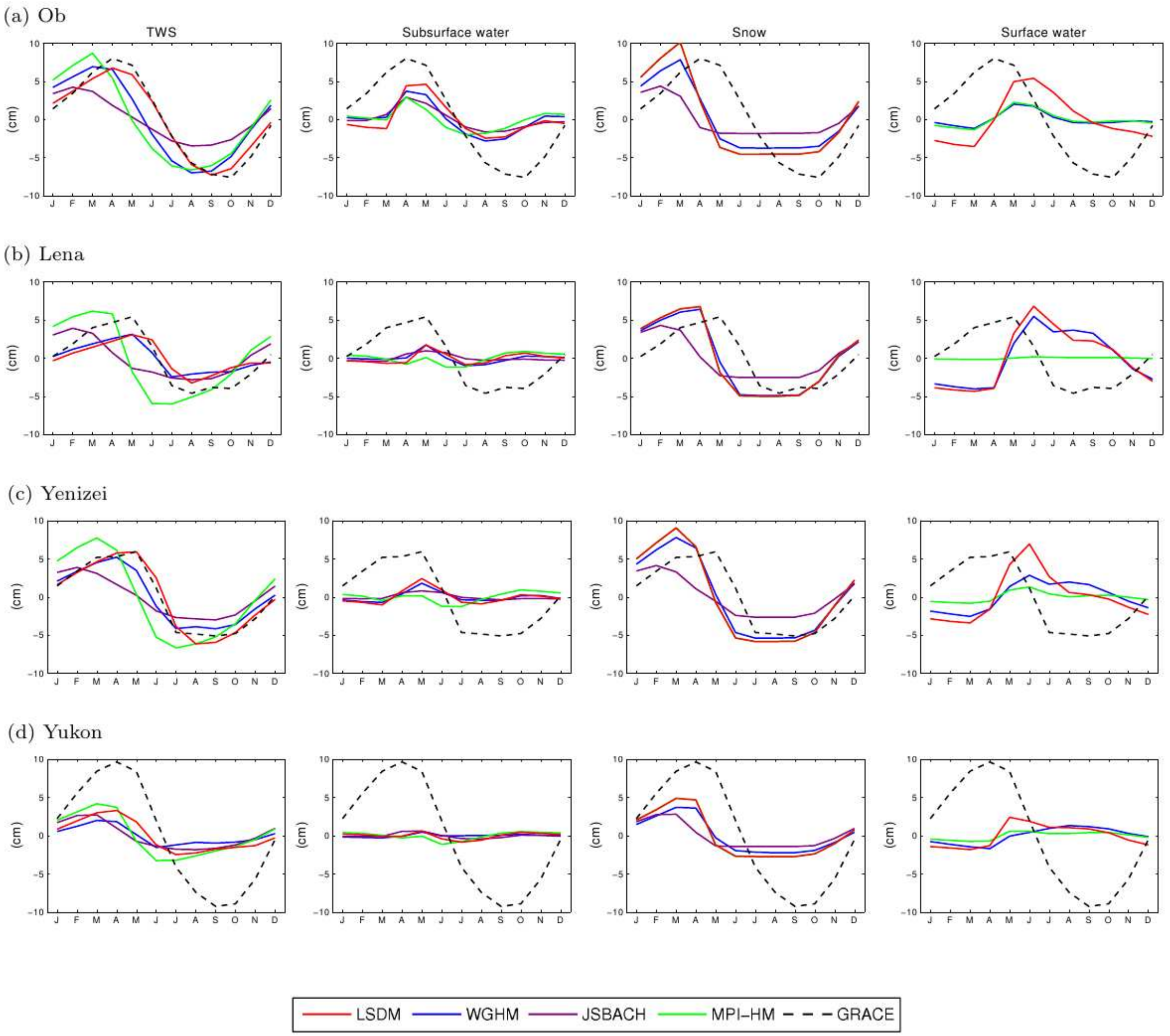

Figure 11. Mean monthly time series of TWS (first column) and the individual storage contributions from ~~soil moisture~~ subsurface water (second column), snow water ~~content~~ equivalent (third column) and surface water (fourth column); each for ~~five dry~~ four snowy catchments: ~~Nile~~Ob, ~~Niger~~Lena, ~~Indus~~ Yenizei and ~~Huang He~~Yukon. TWS from GRACE (~~black~~ dashed line) has been included into every sub-figure for reference.

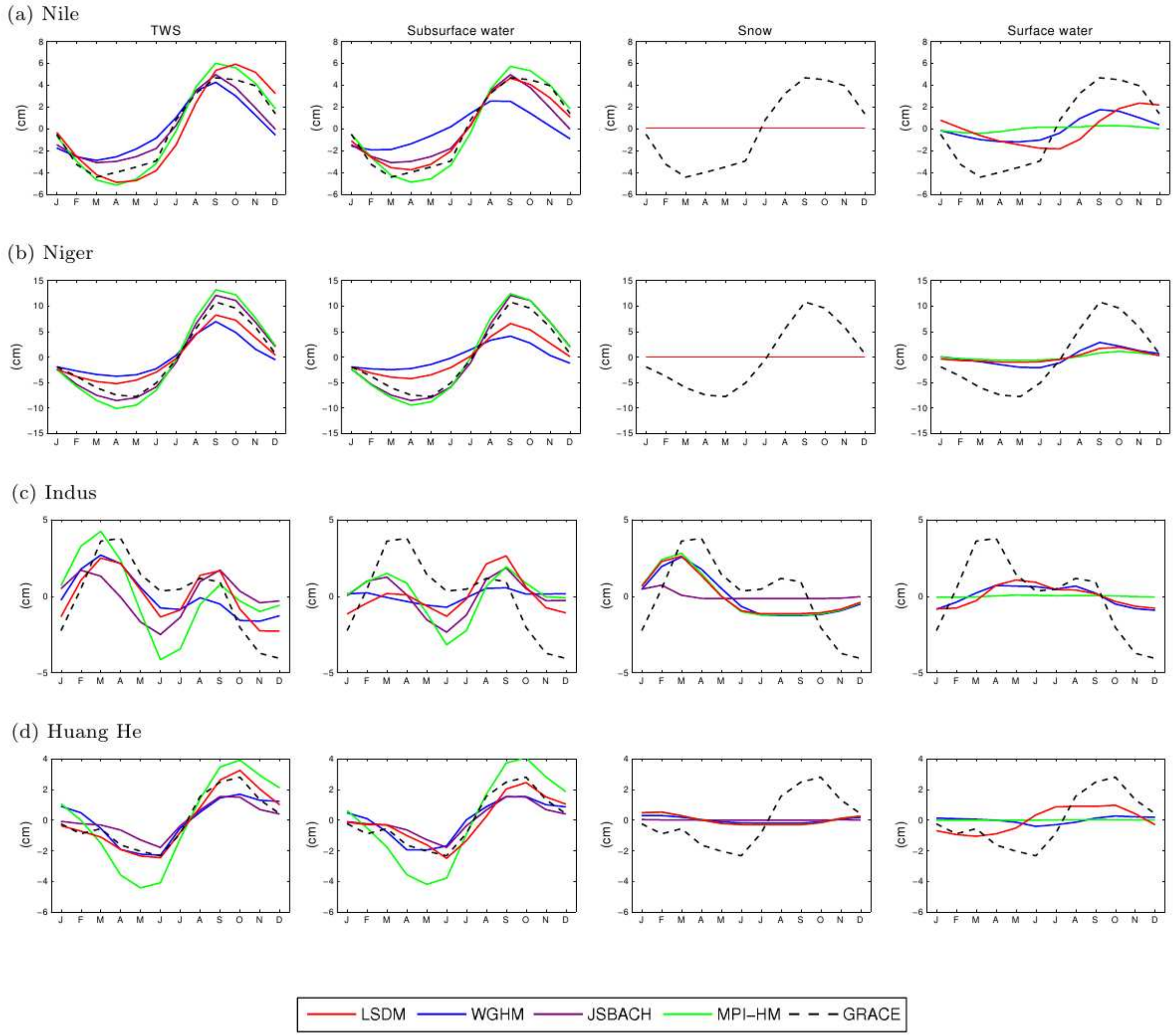

Figure 12. Mean monthly time series of TWS (first column) and the individual storage contributions from subsurface water (second column), snow water equivalent (third column) and surface water (fourth column); each for four dry catchments: Nile, Niger, Indus and Huang He. TWS from GRACE (black line) has been included into every sub-figure for reference.

Table 1. Overview of the main characteristics of the four numerical models particularly considered in this study.

Model name	Model type	Meteorological forcing variables	Storage compartments included	Soil moisture depth	Snow	Potential Evapotranspiration
LSDM	LSM	Precipitation, temperature	subsurface water (root zone), snow, surface water	bucket scheme without a depth	degree day	Thornthwaite
WGHM	GHM	Precipitation, temperature, shortwave incoming radiation	subsurface water (root zone+groundwater), snow, surface water	varies with rooting depth of land cover	degree day	Priestley-Taylor
JSBACH	LSM	Precipitation, temperature, wind, shortwave and longwave radiation, surface qair	subsurface water (root zone+deep layer), snow	down bedrock but at most 10 m	energy balance	physical parametrization
MPI-HM	GHM	Precipitation, temperature, wind, radiation, humility	subsurface water (root zone), snow, surface water	bucket scheme without a depth	degree day	Penman-Montheith

Table 2. Characteristics of the basins shown in Fig 1. Bold and underlined numbers are the largest and smallest RMS differences between GRACE and models separately.

Climate	Basin		Area	RMSE(cm) between TWS from GRACE and				GRACE TWS	
Zones	ID	Name	$(1000km^2)$	LSDM	WGHM	JSBACH	MPI-HM	error(cm)	SNR
Tropical	1	Amazon	5853	4.39	6.08	5.60	**9.53**	1.46	9.76
	3	Zaire	3699	**5.26**	3.36	3.08	3.49	1.32	3.82
	21	Orinoco	1039	**6.37**	4.96	6.21	5.79	3.14	4.74
	29	Mekong	774	5.87	5.60	**6.28**	4.51	3.86	3.73
	30	Tocantins	769	**7.69**	7.49	4.99	5.45	2.81	5.95
Dry	2	Nile	3826	**4.02**	1.85	1.61	1.39	1.06	3.26
	10	Niger	2240	2.53	**2.97**	1.87	2.23	1.29	4.93
	15	Chari	1571	**2.94**	1.96	2.40	2.50	1.50	3.42
	18	Indus	1143	2.17	2.61	**3.08**	3.04	1.54	2.42
	19	Syr-Darya	1070	2.00	**3.30**	3.07	2.89	1.12	3.65
	22	Murray	1031	3.45	3.61	**3.68**	3.38	1.88	2.73
	23	Great Artesian	977	2.44	**2.67**	2.36	2.22	1.33	2.67
	24	Shatt el Arab	967	2.28	3.64	**3.67**	2.85	1.49	3.81
	25	Huang He	894	1.52	2.09	1.74	**2.38**	1.28	2.35
	27	Colorado(Ari)	807	1.90	2.59	**2.98**	2.91	1.41	2.78
Temperate	4	Mississippi	3203	1.68	**3.54**	2.36	3.45	0.86	6.60
	6	Parana	2661	**4.17**	3.03	3.59	2.81	1.32	4.50
	11	Zambezi	1989	2.89	**7.05**	4.83	3.30	1.57	6.80
	12	Chang Jiang	1794	2.58	2.05	**3.24**	3.12	1.49	3.09
	14	Ganges	1628	4.04	**4.43**	3.73	2.90	1.94	5.99
Cold	5	Amur	2903	1.20	1.73	1.88	**2.05**	0.68	3.18
	7	Yenisei	2582	1.89	2.34	3.44	**3.54**	0.68	6.67
	8	Ob	2570	1.50	3.20	**4.35**	4.14	0.68	8.31
	9	Lena	2418	2.33	2.40	3.40	**3.99**	0.68	6.01
	13	Mackenzie	1713	2.67	2.83	**3.95**	3.39	0.83	6.20
	16	Volga	1463	2.11	4.55	3.28	**5.22**	0.84	8.43
	17	St.Lawrence	1267	2.59	4.74	3.42	**4.88**	1.14	4.94
	20	Nelson	1047	1.67	**3.87**	3.19	3.31	1.12	3.82
	26	Yukon	852	5.06	5.72	**5.74**	5.29	1.19	7.68
	28	Danube	788	1.72	4.18	4.03	**4.27**	1.50	4.96
	31	Columbia	724	2.69	4.75	**6.09**	5.71	1.85	5.32