Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Understanding terrestrial water storage variations in northern latitudes across scales

Tina Trautmann[1], Sujan Koirala[1], Nuno Carvalhais[1,2], Annette Eicker[3], Manfred Fink[4], Christoph Niemann[4], Martin Jung[1]

[1]Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, 07745, Germany
[2]CENSE, Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, 2829-516, Portugal
[3]HafenCity University, Hamburg, 20457, Germany
[4]Department of GIScience, Institute of Geography, Friedrich-Schiller University, Jena, 07745, Germany

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

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

To do so, we constrained a rather parsimonious hydrological model with multiple state-of-the-art Earth observation products including GRACE TWS anomalies, estimates of snow water equivalent, evapotranspiration fluxes, and gridded runoff estimates. The optimized model demonstrates good agreement with observed hydrological patterns, and was used to assess the relative contributions of solid (snow pack) versus liquid (soil moisture, retained water) storage components to total TWS variations. In particular, we analysed whether the same storage component dominates TWS variations at seasonal and inter-annual temporal scales, and whether the dominating component is consistent across small to large spatial scales.

Consistent with previous studies, we show that snow dynamics control seasonal TWS variations across spatial scales in the northern mid-to-high latitudes. In contrast, we find that inter-annual variations of TWS are dominated by liquid water storages, comprising mainly of soil moisture. However, as the spatial domain over which the storages are averaged becomes larger, the relative contribution of snow to inter-annual TWS variations increases. This is due to a stronger spatial coherence of snow anomalies as opposed to spatially more heterogeneous liquid water anomalies that cancel out over large spatial domains.

The findings first highlight the effectiveness of our model-data fusion approach that jointly interprets multiple Earth observation data streams with a simple model. Secondly, they reveal that the determinants of TWS variations in snow-affected northern latitudes are scale dependent. We conclude that inferred driving mechanisms of TWS cannot simply be transferred from one scale to another, which is of particular relevance for understanding the short and long-term variability of water resources.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences
Discussions

1 Introduction

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

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

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

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

On the one hand, this is due to the limitations and inherent uncertainties of each Earth Observation based product that need to be considered when comparing simulations and observations. For example, satellite-based soil moisture retrievals only capture the upper 5 cm of soil and therefore are difficult to compare to modelled soil water (Lettenmaier et al., 2015), while large scale observations of snow mass based on passive microwave sensors are known to suffer from uncertainties in deep

and wet snow conditions (Niu et al., 2007) and multispectral sensors solely provide estimates of snow cover in the absence of clouds (Lettenmaier et al., 2015).

Besides, the application of multi-criteria calibration approaches is limited by the increasing complexity of most macro-scale hydrological models over time (Döll et al., 2015). This high model complexity is not only associated with conceptual issues related to overparameterization (Jakeman and Hornberger, 1993) and large computational demand, but also has shown to not necessarily improve model performance (Orth et al., 2015). Therefore, it is desirable to implement rather parsimonious model structure (Sorooshian et al., 1993), especially in multi-criteria model-data fusion approaches.

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

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.

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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

2 Data and methods

2.1 Experiment design

To assess the composition of TWS variations in northern mid-to-high latitudes, we optimized a simple hydrological model on daily time steps at a 1° x 1° latitude/longitude resolution. We defined the area of interest as humid land surface north of 40° N, excluding Greenland as well as grids with > 90 % permanent snow cover, and > 50 % water fraction.

Forced with global observation-based climate data, the model parameters were constrained by multiple Earth observation data products using a multi-criteria calibration approach. These products include terrestrial water storage anomalies as seen by the GRACE satellites (Watkins et al., 2015;Wiese, 2015), measurements of snow water equivalent obtained in the GlobSnow project (Luojus et al., 2014), evapotranspiration fluxes based on FLUXCOM (Tramontana et al., 2016) and runoff estimates for Europe from EU-Grid (Gudmundsson and Seneviratne, 2016). Once the model parameters were calibrated, we evaluated the model against the same data, and finally applied it to quantify the contributions of snow and liquid water storages to the integrated TWS. Thereby we considered different spatial domains (local grid cell and spatially aggregated) and temporal scales (mean seasonal and inter-annual variations).

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

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

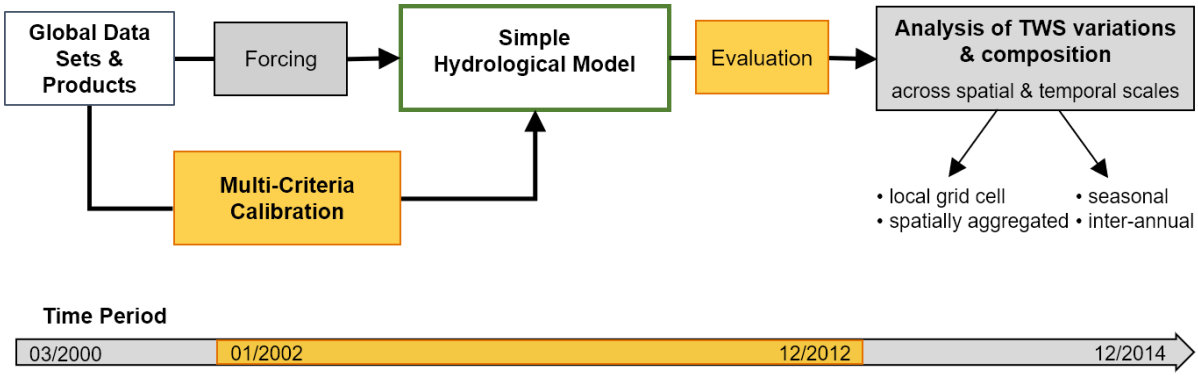

Figure 1. Experiment design and considered time periods for forcing/analysis (grey) and model calibration/evaluation (orange).

2.2 Model description

We designed a conceptual hydrological model with low complexity and a total number of 10 adjustable parameters. The model considers major hydrological fluxes as snow melt, sublimation, infiltration, evapotranspiration, and (delayed) runoff,

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences
Discussions

and includes water storages in the snow pack, the soil, and due to delay in runoff (Figure 2). It is forced by precipitation (P), air temperature (T) and net radiation (Rn), and calculates all hydrological processes on daily time steps for individual grid cells. A simple schematic diagram of the model is shown in Figure 2, while a detailed description of modelled processes is provided in S1.

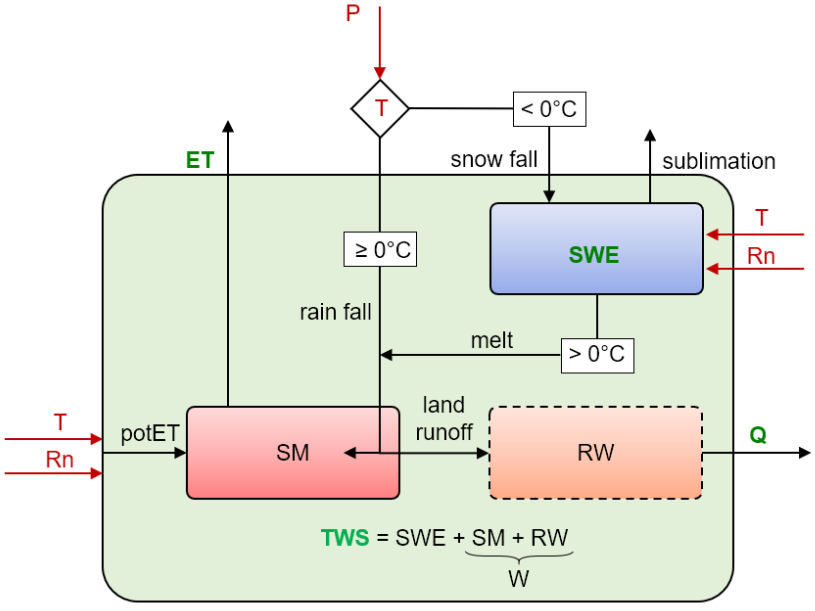

Figure 2. Schematic structure of the model with calculation of TWS. Boxes denote the water storages [mm]: snow water equivalent SWE, soil moisture SM, retained water RW, liquid water W and total terrestrial water storage TWS. Fluxes are represented by arrows. Red colour identifies forcing data: precipitation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

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

Similar to the WaterGAP model (Döll et al., 2003), incoming water from rain and snow melt is allocated to soil moisture (SM) and land runoff depending on soil moisture conditions (Bergström, 1991). SM is represented by a one-layer bucket storage that depletes by evapotranspiration (ET). We calculate ET as the minimum of demand-limited potential ET following the Priestley-Taylor formula (Priestley and Taylor, 1972) and supply-limited ET following Teuling et al. (2006).

To mimic runoff contributions from slowly varying reservoirs that are not explicitly represented in the model such as ground water and river storage, we use an exponential delay function (Orth et al., 2013). The amount of retained land runoff (RW) together with SM represents the total liquid water storage (W).

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

2.3 Input Data

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

Precipitation values originate from the 1° daily precipitation product version 1.2 of the Global Precipitation Climatology Project (GPCP-1DD) (Huffman et al., 2000;Huffman and Bolvin, 2013), that combines precipitation remote sensing and observations from gauges. Temperature was obtained from the CRUNCEP version 6.1 dataset (Viovy, 2015), which is a merged product of Climate Research Unit (CRU) TS.3.23 observation-based monthly climatology (1901-2013) (New et al., 2000) and the National Center for Environmental Prediction (NCEP) 6-hourly reanalysis data (1948-2014) (Kalnay et al., 1996). Net radiation is based on radiation fluxes of the SYN1deg Ed3A data product of the Clouds and the Earth's Radiant Energy Systems (CERES) program of the United States' National Aeronautics and Space Administration (NASA) (Wielicki et al., 1996).

Rather than using a single data stream, e.g. discharge measurements at the outlet of large continental catchments as used in traditional large-scale hydrological studies, we calibrated the model against multiple observation-based data streams on the grid scale. The integrated datasets include terrestrial water storage anomalies (TWSobs) [mm], snow water equivalent (SWEobs) [mm], evapotranspiration (ETobs) [mm d^{-1}], and gridded runoff estimates for Europe (Qobs) [mm d^{-1}].

TWSobs is derived from the GRACE Tellus Mascon product version 2 based on the GRACE gravity fields Release 05 processed at NASA's Jet Propulsion Laboratory (JPL) (Watkins et al., 2015;Wiese, 2015). The GRACE solutions were corrected for geocentric motion coefficients, according to Swenson et al. (2008) and for variations in Earth's oblateness (C20 coefficient) obtained from Satellite Laser Ranging (Cheng et al., 2013). The Glacial isostatic adjustment has been accounted for using the model by A et al. (2013). The dataset provides monthly anomalies of equivalent water thickness relative to the January 2004–December 2009 time-mean baseline for the period 2002–2016. Unlike previous GRACE products based on spherical harmonic coefficients, the JPL RL05M dataset uses equal area 3° x 3° spherical cap mass concentration blocks (mascons) to solve for monthly gravity field variation. To ensure a clean separation along coastlines within land/ocean mascons, a Coastline Resolution Improvement (CRI) filter has been applied (Watkins et al., 2015). For each mascon, uncertainties were estimated by scaling the formal covariance matrix. To enable hydrological analysis at sub-mascon resolution, we used the provided gain factors to scale the original TWSobs values.

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

The ET product is based on FLUXCOM (www.fluxcom.org), i.e. upscaled estimates of latent energy that were derived by integrating local eddy covariance measurements of FLUXNET sites, remote sensing, and meteorological data using the Random Forest (Breiman, 2001) machine learning algorithm (Tramontana et al., 2016). Although FLUXCOM ET performs relatively better than other gridded ET products, an underestimation in the order of 10–20 % can be expected owing to missing energy balance correction prior to upscaling. While the product captures seasonality and spatial patterns of mean annual fluxes well, predictions of inter-annual variations remain highly uncertain (Tramontana et al., 2016). To calculate ETobs [mm d^{-1}], we assume a constant latent heat of vaporization of 2.45 MJ m^{-2}.

Similar to TWS that represents the vertically integrated water storage, observations of river discharge spatially integrate hydrological processes within a basin. Thus, they provide an invaluable tool for model validation at large scales. However, it is desirable to apply gridded products to evaluate model performance at local (grid) scale. Therefore, we used the observation-based gridded runoff product E-RUN version 1.1 (Gudmundsson and Seneviratne, 2016) as constraint for runoff processes. This dataset is based on observed river flow from 2771 small European catchments that was spatially disaggregated to upstream grid cells using a machine learning approach. The data provides mean monthly runoff rates per unit area for each grid, so that river routing is not necessary to compare runoff estimates directly with modelled runoff. Similar to the ET data, gridded runoff estimates show high accuracy for the mean seasonal cycle across Europe, and poorer agreement regarding monthly time series and inter-annual variations (Gudmundsson and Seneviratne, 2016).

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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences Discussions

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

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

2.4 Multi-criteria calibration

In this study, calibration is intended to identify the set of 10 model parameters (Table 2) that achieves the best fit between simulations and observations for all grids cells and regarding all observational data simultaneously. Thereby, we aimed to exploit the strength of each data stream, while considering known uncertainties and biases. For this purpose, we defined a cost function that takes into account the weakness of each observed variable and evaluates the overall model fit with one value of total cost (see subsequent section). To minimize total costs and thus find the optimal parameter values, we applied the Covariance Matrix Evolution Strategy (CMAES) (Hansen and Kern, 2004) search algorithm. The CMAES, as an evolutionary algorithm, is a stochastic, derivative-free method for non-linear, non-convex optimization problems. Compared to gradient-based approaches, it performs superior on rough response surfaces with discontinuities, noise, local optima and/or outliers, and is a reliable tool even for global optimization (Hansen, 2014). Additionally, CMAES' guided search in the parameter space makes the algorithm less computationally demanding than other global optimization approaches which enumerate a large number of possible solutions (e.g. Monte Carlo Markov Chain methods) (Bayer and Finkel, 2007).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Within the iterative calibration process, the model simulations are carried out on daily time steps, while costs are calculated based on monthly values. Further, each model run includes an initialization based on 10 random years that were selected a priori.

Cost function

To objectively describe the goodness of fit, we defined a cost function based on model efficiency (Nash and Sutcliffe, 1970), but with explicit consideration of the uncertainty σ_i of the observed data stream as:

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

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

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

$$\mathbf{cost_{total}} = \mathbf{cost_{TWS}} + \mathbf{cost_{SWE}} + \mathbf{cost_{ET}} + \mathbf{cost_{Q}} , \quad (2)$$

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

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

For SWE, we applied an absolute uncertainty of 35 mm based on reported differences to ground-measurements (Liu et al., 2014;Luojus et al., 2014). Since GlobSnow SWE saturates above approx. 100 mm (Luojus et al., 2014), we do not penalize

model simulations when both, SWEobs and SWEmod, are larger than 100 mm in order to prevent the propagation of data biases to calibrated model parameters.

2.5 Evaluation of model performance

Once the parameters were optimized, we applied the model for the entire study domain, and evaluated its performance regarding all grids (6050) in terms of Pearson correlation coefficient r and root mean square error RMSE for each variable with observational data, respectively. On the one hand, the overall performance at local scale was assessed by calculating r and RMSE for the monthly time series of each grid individually. On the other hand, the model performance over the entire study domain was evaluated by comparing the seasonal and inter-annual dynamics of the regional average. Therefor we defined inter-annual variation (IAV) as the deviation of the monthly values from the mean seasonal cycle (MSC). As with the calibration, we focused on the common time period 2002–2012, and considered only the grid cells and time steps with available observations.

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

2.6 Analysis of TWS variations and composition

Finally, the contribution of snow and liquid water to seasonal and inter-annual TWS variability was quantified across spatial scales. For this, we ran the model with optimized parameters for the entire study domain from 2000 to 2014, and translated simulated storages as anomalies to the time-mean baseline. As in the model evaluation, the MSC and IAV of SWEmod, W and TWSmod anomalies were calculated at local scale for each grid individually and as spatial average over all grid cells. To assess storage variability, the variance in the MSC and the IAV of each storage component was computed. Assuming negligible covariance of snow and liquid water, their relative contribution to TWS variance was calculated as the contribution ratio CR:

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

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

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

3 Results and discussion

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

3.1 Model optimization

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

In detail, snow fall is reduced by p_{sf} to 67 % of precipitation occurring at T < 0 °C. This reduction agrees with Behrangi et al. (2016), who found GPCP to overestimate snowfall over Eurasian high latitudes. Similar, overestimation of precipitation undercatch correction in GPCP has also been reported by Swenson (2010). Therefore, p_{sf} allows to reduce inconsistencies between the precipitation forcing and the water storages as given by GlobSnow SWE and GRACE TWS.

Further, each grid is assumed to be completely covered by snow if SWE ≥ 80 mm. On the one hand, the snow pack can be reduced by sublimation, with sn_a = 0.44 indicating relatively high sublimation resistance, compared to a default of sn_a = 0.95 proposed by (Miralles et al., 2011a). The divergence probably results from interaction with snow melt, as net radiation also

contributes to melt with 0.9 mm MJ^{-1} (m_r) if T exceeds 0 °C. Nevertheless, melt is mainly induced by temperature, as the estimated day degree factor (m_t) is 2.63 mm K^{-1}, which is close to typical values of 3 mm K^{-1} (Müller-Schmied et al., 2014;Stacke, 2011). These parameter interactions underline that neither modelled snow melt nor sublimation are well constraint by data, resulting in uncertainty due to parameter equifinality between sn_a, m_r and m_t. However, for the objective of this study it's not primarily relevant whether sublimation or radiation induced melt decreases the snow pack, as the total melt amount remains relatively unchanged for different parameter combinations.

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

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

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

The uncertainty in the optimized parameter vector was estimated by quantifying each parameter's standard error as the square root of the product between the diagonal elements of the parameters' covariance matrix (calculated from the Jacobian matrix) and the sum of residual squares according to Omlin and Reichert (1999) and Draper and Smith (1981).

Most of the parameters have uncertainties smaller than 10 % (Table 2). This suggests that our model-data fusion method, fed by multiple observation streams, is able to reduce the initial theoretical parameter ranges (up to 500 %) to much narrower ranges. Nonetheless, some parameters have a larger uncertainty range than others (e.g. q_t, sn_c, m_t), which may highlight a limitation in suitable observations to constrain them, as well as a lower sensitivity of the model results and the cost function used. Further, given that the model only considers the spatial variability of climate, the uncertainty in global parameters obtained from inversion may reflect the natural variations in these parameters that arise from differences in local land surface characteristics such as topography or land cover.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences
Discussions

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

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

We adopted the calibrated parameter values as the global constants for model simulations over the entire study domain. Even though the globally uniform parameters may not provide perfect simulation for all grids over a large study domain, this approach represents a compromise between a priori parametrization of the model and its calibration at local or regional (e.g. basin) scale. While local and regional model calibration enables good adaption to geographic characteristics, it easily leads to overfitting of the model and thus propagates the constraints' inherent errors and uncertainties to the modelling result. As these uncertainties often vary in space, globally uniform parameter values diminish overfitting uncertainties. In addition, calibration for several independent grids is computationally demanding and subsequently requires a parameter regionalization approach (He et al., 2011). Since such approaches are not commonly accepted and not considered practical

(Sood and Smakhtin, 2015;Bierkens et al., 2015), macro-scale models mostly apply a priori parameters based on empirical values or on expert knowledge, that yet lead to suboptimal simulations (Beck et al., 2016).

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

3.2 Model performance

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

In general, all observed patterns are reproduced very well, taking into account the specific data weaknesses. We achieve a 'near perfect' correlation of 0.99 and 0.94 for seasonal variations of ET and Q, respectively. At the inter-annual scale, though, larger discrepancies exist, which at least partly arise from larger uncertainties in ETobs resp. Qobs (S2). Thus, we assume high confidence in modelled ET and Q fluxes and subsequently focus on evaluation of the water storages TWS and SWE.

3.2.1 Performance on local grid scale

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

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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences
Discussions

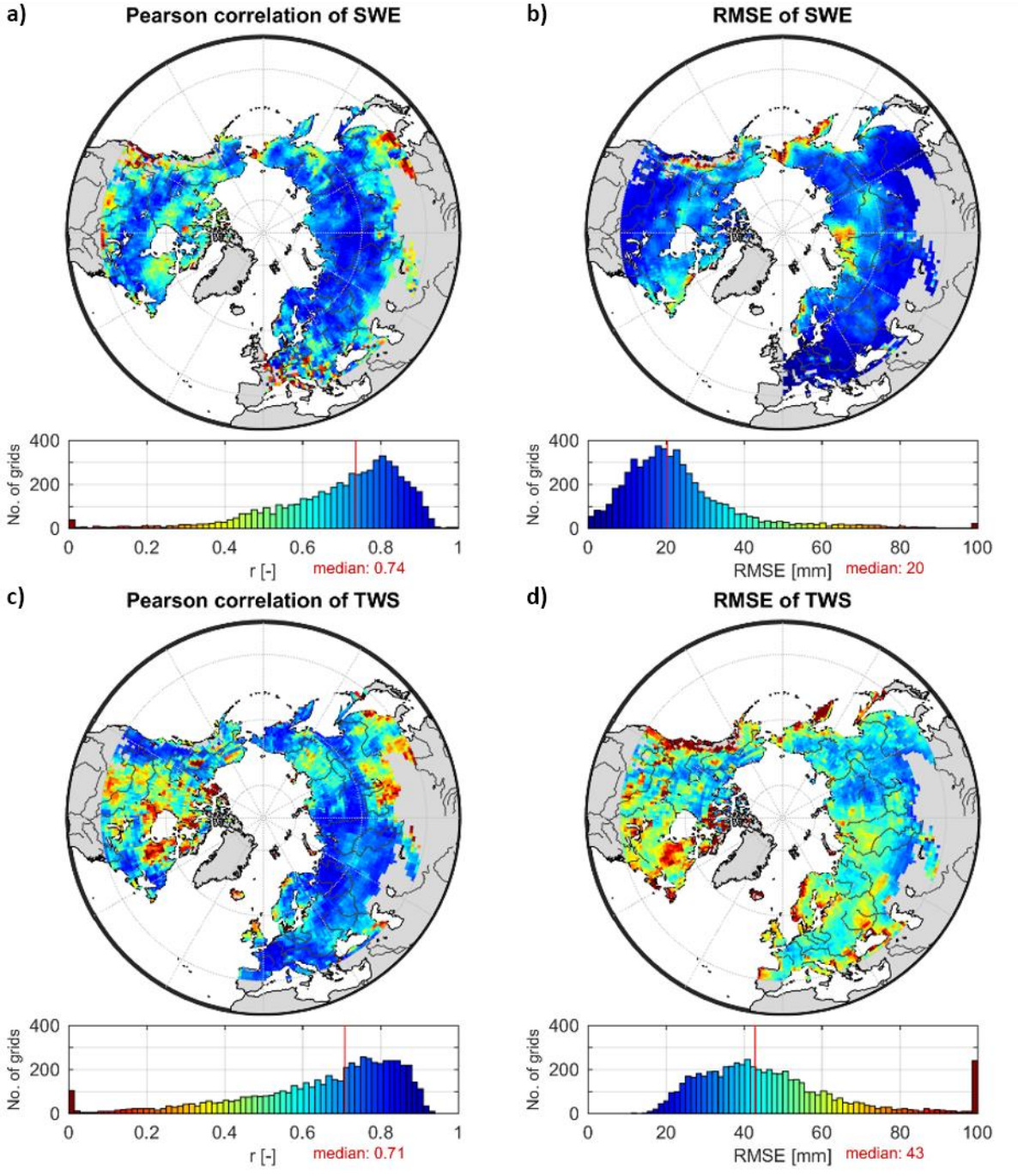

Figure 3. Pearson correlation coefficient r **(a,c)** and root mean square error RMSE **(b,d)** between monthly values of modelled SWE and GlobSnow SWE **(a,b)**, as well as modelled TWS and GRACE TWS **(c,d)** for the period 2002–2012 and for each 1° x 1° grid cell of the study domain.

3.1.2 Performance of the spatially integrated simulations

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

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

Similar to SWE, the spatial average TWS shows high correlation of 0.91 for seasonal variations, with positive anomalies from December to May/June and negative anomalies during summer and autumn months (Figure 4b). Even though the modelled amplitude is slightly larger than the observed one, it stays within the uncertainty range of TWSobs for most months, suggesting reliable simulations. However, TWSmod precedes TWSobs on average by one month, reaching the maximum in March instead of April, and the minimum in August instead of September. A similar phase shift of one month between GRACE TWS and modelled TWS has been reported by several state-of-the-art global models (Döll et al., 2014a;Schellekens et al., 2017). It should be noted that some areas such as East North America, Kamchatka, Scandinavia and Western Europe do not show phase differences, while the lag in South East Eurasia is even larger, as already suggested by lower overall correlation (Fig. S3). In general, the disagreement in timing is attributed to the lack of sufficient water storages and delay mechanism within the model, so that the modelled system reacts too fast (Schellekens et al., 2017;Döll et al., 2014a;Schmidt et al., 2008). Thus, we implemented model variants with an explicit groundwater storage to delay depletion of TWS, with spatially varying soil properties to better represent heterogeneous infiltration and runoff rates, as well as a variant that applied a more sophisticated approach to calculate snow dynamics based on energy balance. Despite the efforts, we achieved no improvement in terms of reducing the phase shift. Therefore, the question arose, whether it is not primarily the model formulation that prevents correction of the temporal delay, but rather the combination of forcing data respectively observational constraints. To further preclude possible errors due to such data inconsistencies e.g. between GRACE TWS and GlobSnow SWE, we excluded GlobSnow SWE data from calibration. Although this could slightly improve the agreement of TWS' MSC, it led to unrealistic behaviour of snow dynamics, and thus did not offer any advantages. Besides, we found no major differences in the magnitude or spatial distribution of the phase shift resulting from the precipitation forcing (GPCP vs. WFDEI), or compared to other GRACE solutions (S5). Since we obtained no general correlation to either elevation, land cover, soil properties, the presence of lakes and wetlands, or the distribution of

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences
Discussions

permafrost, we attribute the difference in TWS phase to a combination of missing, yet spatially varying processes and storage components. As suggested in previous studies, those processes likely relate to human impacts, glacier melt, groundwater reservoirs, wetlands, river storage and/or river routing. Additionally, high uncertainties of the precipitation forcing and GlobSnow SWE product in mountain (near) regions, as well as leakage errors in the GRACE signal influence the accuracy of both, TWSobs and TWSmod. Although these shortcomings should be kept in mind, we assumed that they do not affect our results regarding to the relative contributions of snow and liquid water to TWS significantly.

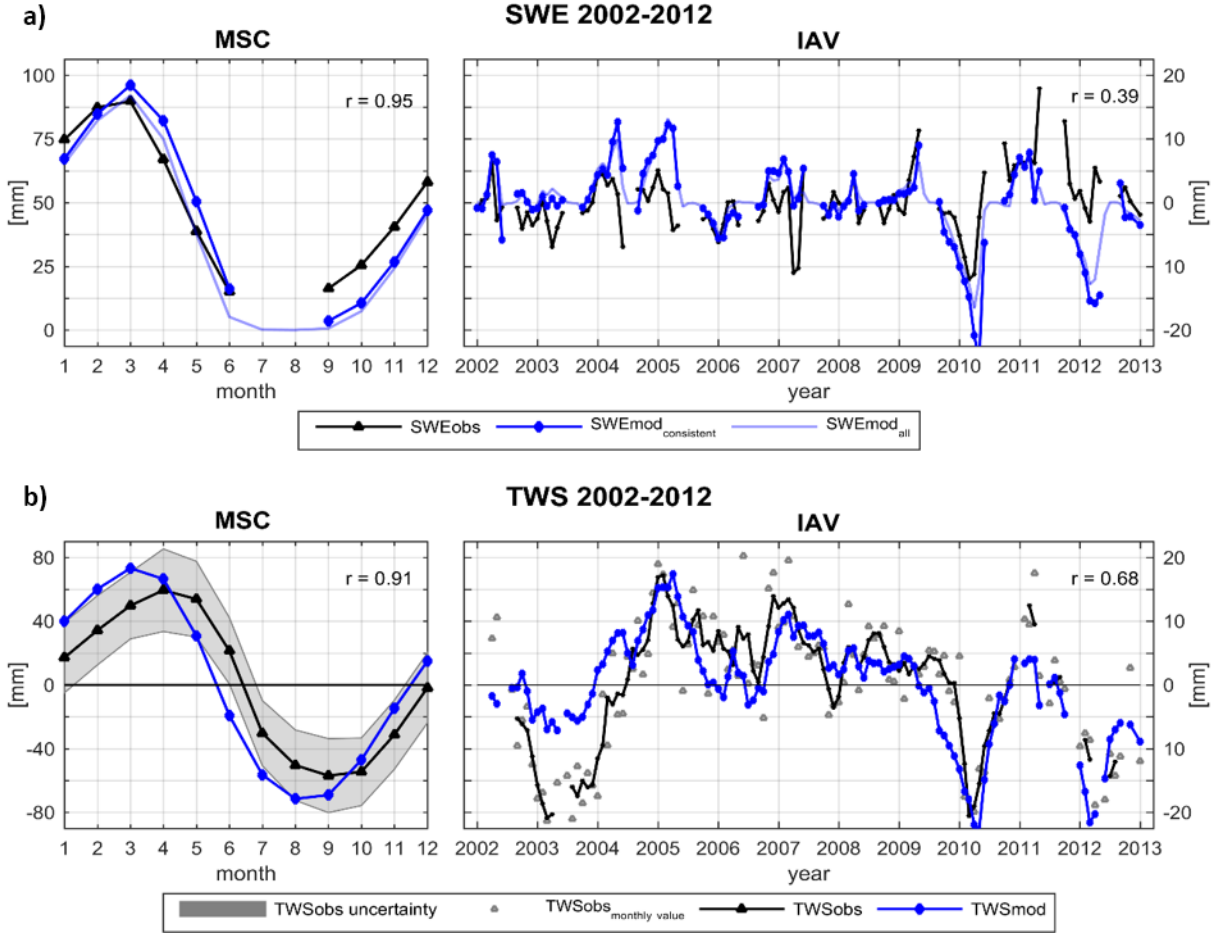

Figure 4. Spatially averaged mean seasonal cycle (MSC) of the period 2002–2012 as well as inter-annual variability (IAV, difference between monthly values and the MSC) for **a)** SWE and **b)** TWS. In a), $SWEmod_{consistent}$ refers to modelled SWE considering only data points with available SWEobs,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

In terms of inter-annual variations, the variance in monthly TWSobs values is highly underestimated, which on the one hand relates to noise within the GRACE signal, but on the other hand may again reflect missing process representation in the model. To reduce the noise, we applied a three-month-moving-average filter on the monthly time series. The smoothed time series then shows fairly good agreement of inter-annual dynamics, with correlation r = 0.68 (Figure 4b, solid lines). Solely the amplitude of the large negative anomaly in 2003 is not captured by the model.

3.1.3 Comparison with the eartH2Observe model ensemble

Compared to the model ensemble of the ertH2Observe dataset, we achieve equally good or better performance for the spatially integrated SWE and TWS on both, MSC and IAV, scales (Figure 5). Besides, the majority of the model ensemble obtains similar spatial patterns of performance criteria for SWE as well as for TWS (not shown).

The average observed MSC of SWE is captured with a correlation in the range of 0.79 (PCR-GLOBWB) to 0.99 (ORCHIDEE), whereby only ORCHIDEE outperforms our model (r = 0.95). However, modelled snow accumulation exceeds that of SWEobs for the majority of the models, which also reflects in higher RMSE (Fig. S5). On IAV scales, the correlation in general is lower, yet again we obtain a better fit (r = 0.39) than the model ensemble (r = 0.12 (ORCHIDEE) to 0.28 (LISFLOOD)). However, it remains uncertain, whether the discrepancies between SWEobs and SWEmod represent model deficiencies or evolve from issues related to the GlobSnow SWE retrieval (Schellekens et al., 2017).

Regarding average seasonal TWS variations, our model performs as well as the model ensemble (r = 0.91), with the range of the eartH2Observe ensemble spanning from r = 0.83 (ORCHIDEE) to r = 1.00 (PCR-GLOBWB). The amplitudes in the MSC of TWSmod (95 to 156 mm) are comparable to the observed amplitude of 118 mm, except for SWBM, whose amplitude is twice as large as that of TWSobs. This discrepancy is reflected in relatively high RMSE values for SWBM (Fig. S5). The model ensemble precedes observed seasonal TWS variations by 1 to 1.4 months, similar to our estimates of TWSmod (-1.1 month). Only PCR-GLOBWB, with a higher correlation than other models, shows a smaller average lag of less than 1 month (-0.3 months). This minor difference results from balancing out of preceding and succeeding in different regions over the study domain. Additionally, Schellekens et al. (2017) found that PCR-GLOBWB shows unrealistic snow accumulation over time in Europe and boreal North America. Except for PCR-GLOBWB, the majority of the models obtains comparable spatial pattern of preceding TWS, with small differences at regional scales. Even though the difference in the MSC is commonly attributed to the lack or inadequate size of water storages (Kim et al., 2009), a relationship between model performance and model complexity is not obvious. Relatively complex models, such as HTESSEL, SURFEX, and JULES, show similar phase differences as simpler models, such as SWBM and our model (-1.0 resp. -1.1 months). Since Schellekens et al. (2017) found the largest phase differences in cold regions, they postulate the implementation of processes associated with snow as important factor for this phase lag. In this context, constraining the model with snow observations as done in our study should increase confidence in the representation of snow processes. Nevertheless, we obtain a similar phase difference, which points to the importance of other hydrological processes and storages even in snow-affected regions.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

In summary, none of the models, regardless of its complexity, outperforms others in terms of all variables and at all spatial and temporal scales. Compared to the eartH2Observe model ensemble, our model performs equally well or better regarding all variables and performance criteria. However, we note that the comparison is not completely fair as the eartH2Observe models unlike our model were not informed by GRACE and/or GlobSnow data beforehand. Nevertheless, we only used a subset of 1000 random grids to calibrate the model parameters, which suggests that inclusion of multiple observations better constrains the model over a large domain (6050 grids).

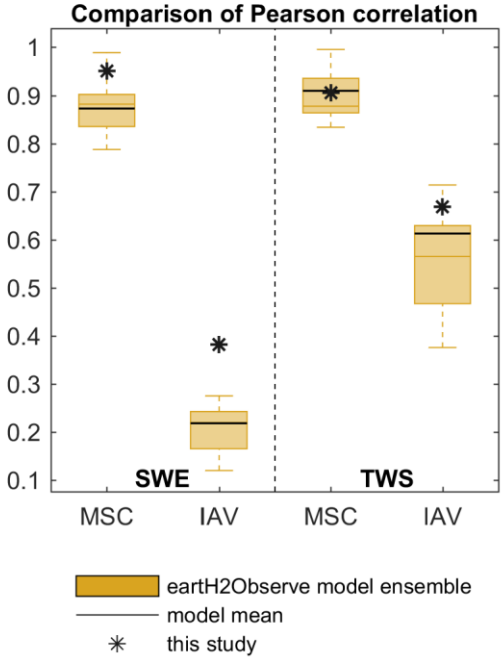

Figure 5. Pearson correlation for the spatially integrated SWE (left) and TWS (right) achieved by this study compared to the model ensemble of eartH2Observe dataset across temporal scales. In each box, the central orange line represents the median and the edges the 25 % and 75 % percentiles of the model ensemble, while the solid black line marks the performance of the ensemble mean.

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

3.3 TWS variation and composition

3.3.1 Spatially integrated

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

Regarding the MSC, all water storages show a clear seasonal pattern. The maximum TWSmod in March coincidences with the maximum in SWEmod. On contrary, W remains nearly constant throughout the winter, as related to the lack of evapotranspiration losses and missing infiltration due to prevailing solid precipitation. Starting from March, snow melt decreases SWEmod, and thus TWSmod, progressively. Thereby TWSmod declines with some delay, as positive W anomalies in April and May suggest buffering of melt water in the soil and on the surface before being transferred to runoff or evapotranspirated. During the summer months, snow is absent, while W decreases due to higher summertime evapotranspiration, and preceding runoff of temporarily stored water. With W and SWEmod being at their minimum in August/September, TWSmod reaches its minimum, too, before starting to increase again in September/October. This rise relates to dropping evapotranspiration rates in combination with more precipitation input (increasing W) and beginning snow accumulation (increasing SWEmod). Despite the interplay of SWEmod and W on seasonal variations of the integrated TWSmod, the amplitude of W (62 mm) is considerably lower than the one of SWEmod (92 mm) and TWSmod (144 mm). Thus, the seasonal accumulation of snow largely determines the magnitude of TWSmod. Additionally, W anomalies at least partly result from snow melt, whereas liquid water does not influence the snow pack. Thus, we conclude that average seasonal TWS variations in northern mid-to-high latitudes are mainly driven by annual snow accumulation and ablation processes. The Contribution Ratio CR (Eq.(3)) based on the spatially averaged MSC underlines this, as CR = -0.26 indicates that variations in SWEmod explain 63 % of seasonal TWSmod variability.

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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

anomalies in SWEmod implicate subsequent changes in W. As a result, W anomalies in any case affect TWSmod variability on IAV scales when analysing the spatial average composition.

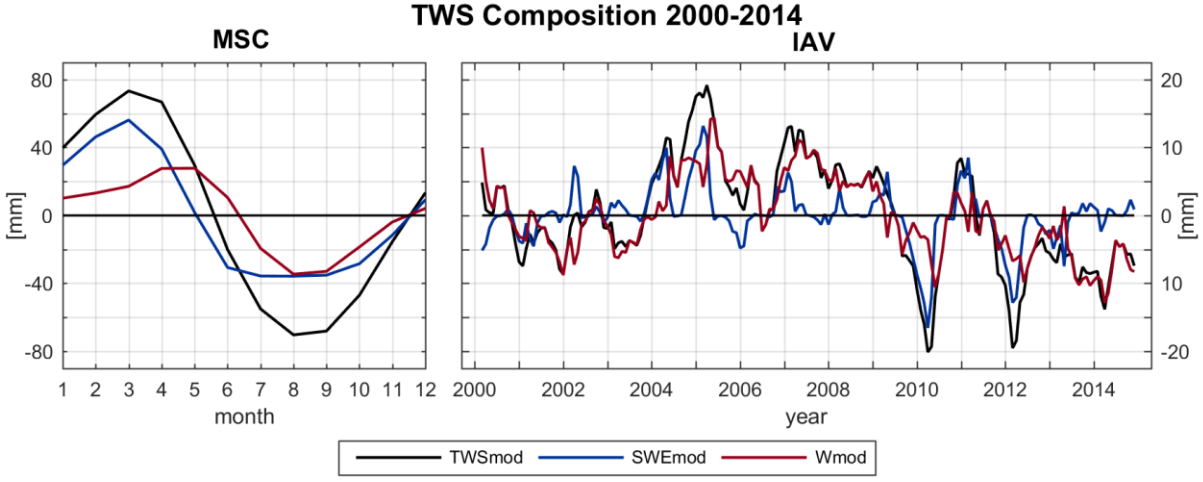

Figure 6. Spatially averaged mean seasonal cycle (MSC) of the period 2000–2014 as well as inter-annual variability (IAV, difference between monthly values and the MSC) for modelled TWS, SWE and W anomalies to the time-mean of 2000–2014.

3.3.2 Local grid scale

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

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

This obtained pattern confirms earlier studies, that received dominance of snow in northern latitudes in North America (Rangelova et al., 2007), and prevailing soil moisture dynamics further South e.g. in the Mississippi basin (Ngo-Duc et al., 2007). Besides, the north extent of dominating W reflects the temperature gradient in North America and Eurasia. Comparison with average annual temperature suggests, that for T > 10 °C variability of W dominates, while for T < 0°C SWEmod dynamics prevail. This is plausible, as temperature determines annual snow accumulation, and the relative

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

contribution of liquid water increases in the absence of snow. Yet, it further highlights the dependency on the used temperature data set, especially in a model that calculates snow fall and snow melt based on temperature thresholds as ours.

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

Since we already showed a link between average TWSmod IAV and previous precipitation anomalies, and as precipitation represents the main model forcing data, we investigated the relative contribution of rain and snow fall to inter-annual variability of total precipitation (Figure 8). Similar to the composition of TWSmod on local scale, rain anomalies prevail for most of the grid cells (mean CR = 0.68). This suggests that the greater contribution of W to inter-annual variations of TWSmod on local scale relate to highly variable (liquid) summertime precipitation events. On contrary, monthly snow fall seems less variable, resulting in less pronounced variations in SWEmod compared to W. Exceptions are regions of high maximum SWEmod, that accordingly show a considerable relative contribution of snow to the inter-annual TWSmod variability.

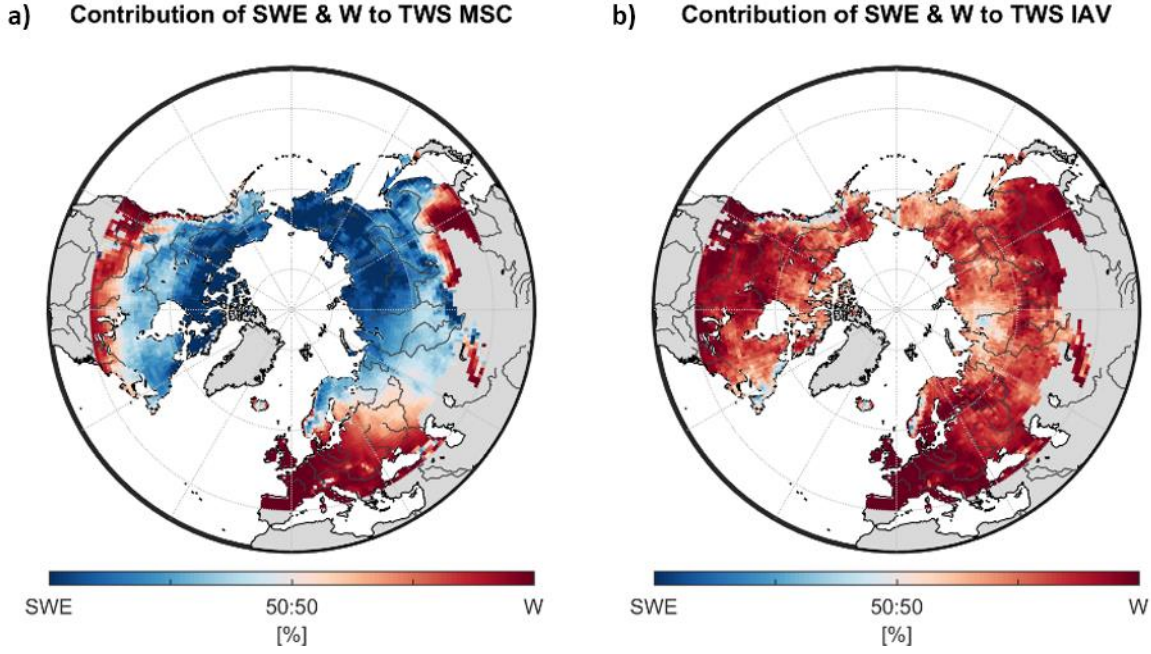

Figure 7. Relative contribution based on CR (Eq.(3)) of modelled snow (SWE) and liquid water (W) storage anomalies to **a)** mean seasonal variations from 2000–2014 of modelled TWS anomalies, and **b)** inter-annual variations of modelled TWS anomalies for each grid cell of the study domain, respectively.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

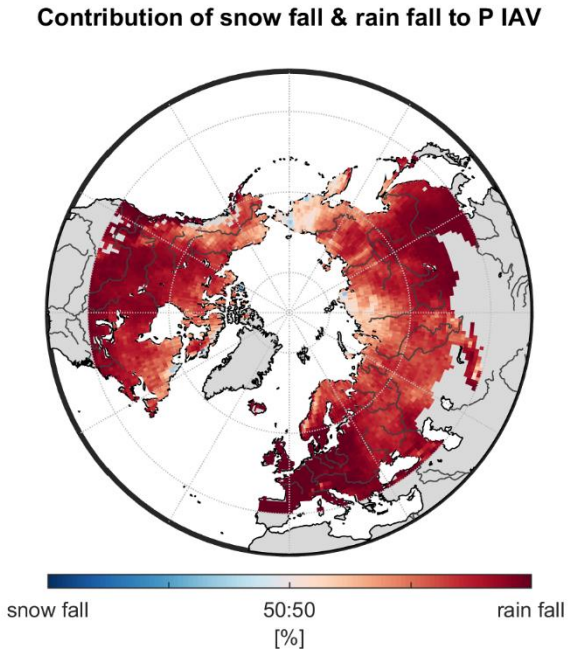

Figure 8. Relative contribution based on CR (Eq.(3)) of modelled snow fall and rain fall to total precipitation (P) anomalies on inter-annual (IAV) scales for each grid of the study domain.

3.3.3 Comparison of different scales

Figure 9 summarizes the above presented contributions to TWSmod variability across spatial and temporal scales. As explained in the previous sections, we obtained two scale dependent differences in the relative contribution to TWSmod variability: (1) in general between temporal scales, and (2) for inter-annual variability between spatial scales.

Regarding (1), Figure 9 emphasizes again that seasonal variations of TWSmod are mostly determined by seasonal snow dynamics, while on inter-annual scales TWSmod variability mainly originates from variations in liquid water. As previously stated, determination by SWEmod dynamics on MSC scales relates to the pronounced magnitude of seasonal snow variations in northern mid-to-high latitudes. In comparison, average monthly changes in W were found to be minor and additionally influenced by snow ablation. Thereby, the spatially integrated CR (black star) roughly agrees with the average of local contributions (dashed line).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-690
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 29 November 2017

Hydrology and Earth System Sciences
Discussions

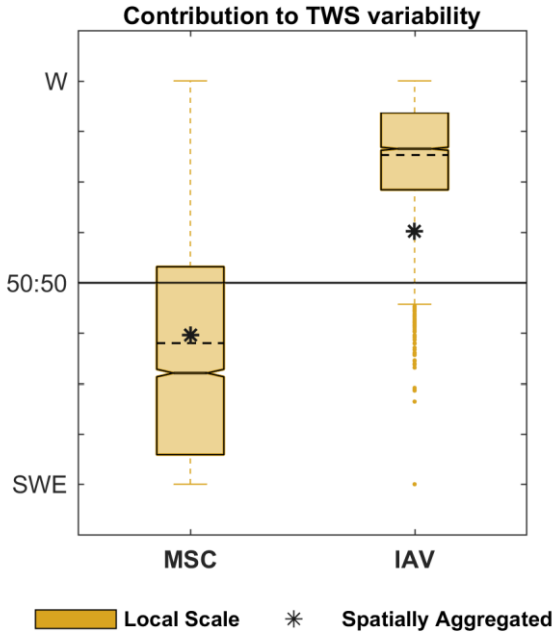

Figure 9. Relative contribution of snow (SWE) and liquid water (W) to TWS variability on different spatial (local grid scale, spatially integrated) and temporal (mean seasonal MSC, inter-annual IAV) scales based on CR (Eq.(3)). The boxplots represent the distribution of grid cell CR, with the dashed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

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

For inter-annual variations of SWEmod, the sum of positive covariances prevails (Figure 10a), whereas positive and negative values are more in balance for W (Figure 10b). This suggests SWEmod anomalies to be more spatially coherent than anomalies of W. Thus, when spatially averaging, the more homogeneous snow patterns maintain. On contrary, opposed anomalies of W compensate each other. This leads to a relatively larger influence of SWEmod to the spatially integrated inter-annual TWSmod variability compared to when analysing the local grid scale. Since positive covariation clearly

dominates for temperature anomalies, the spatial coherence of SWEmod relates to their homogeneous patterns (Figure 10c). Similar to W, positive covariances only slightly outweigh for year-to-year variations in rain fall (Figure 10d). The same is true for snow fall (not shown). Therefore, the spatial coherence of SWEmod anomalies is less pronounced than for temperature, as snow is additionally influenced by snow fall anomalies. Regarding W, this indicates that the spatial heterogeneity in our model, which misses explicit information on soils, topography, etc., mainly results from inhomogeneous patterns in rain fall anomalies. Thereby, the greater balance between positive and negative covariations for W compared to rain fall can be ascribed to the additional impact of evapotranspiration to SM. Since evapotranspiration anomalies, as influenced by temperature, do not necessarily show the same spatial pattern as rain fall anomalies, the combined effect introduces more spatial variability in W than rain fall or evapotranspiration obtain on their own.

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

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

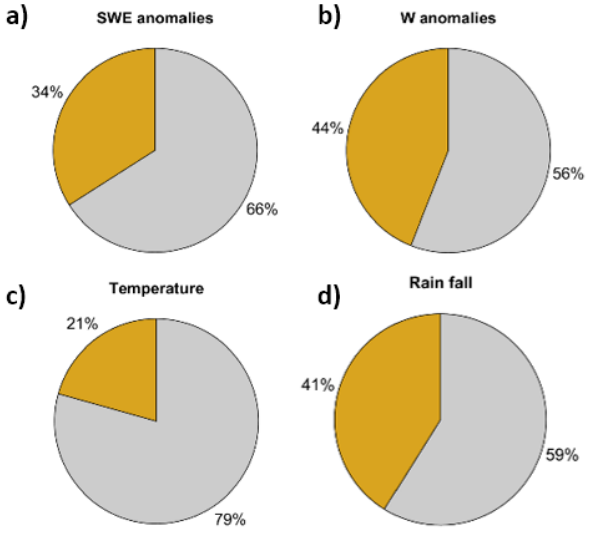

Figure 10. Proportion of total positive (grey) and negative (orange) covariances among grid cells for inter-annual variations of **a)** snow (SWE), **b)** liquid water storages (W), **c)** temperature, and **d)** rain fall.

3.4 Limitations of the approach

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

First of all, the simple structure only allows inferences on represented processes, that likely include effects of fluxes and storages not considered explicitly. For example, a more distinct partitioning of liquid water storages into its components such as soil moisture, deep groundwater and surface water, is not feasible with the current structure. As discussed previously, delayed land runoff comprises various (intermediate) storages and delay times, and thus cannot be associated with one distinct storage component. Even though soil moisture is distinguished from these slowly varying reservoirs, its quantity and pattern have not been directly validated. Future research is required to increase confidence by including remote sensing based data of soil moisture in calibration and/or validation. However, these satellite data still have limited value as the microwave signals can only capture moisture in the upper 5 cm of soil (Döll et al., 2015;Lettenmaier et al., 2015).

Further, the model does not include any human-induced changes of water storages, which yet contribute to observed TWS variability in many regions (Döll et al., 2015;Rodell et al., 2015). Other simplified or ignored hydrological processes include the coincident occurrence of rain and snow fall, liquid water capacity of snow, interception, capillary rise and other surface-groundwater interactions, as well as lateral flow from one grid cell to another. Besides, the model does not account for sub-grid heterogeneity of topography and land surface characteristics except for the fractional snow cover used to estimate snow melt and sublimation.

With regards to model parameter, we apply a global uniform parameter set and do not regionalize the parameters according to spatially distributed physio-geographical characteristics. In contrast, most macro-scale hydrological models include spatially distributed soil properties to define parameters related to infiltration, soil water holding capacity and percolation, as well as vegetation types to assess the effects of different plant functional types on evapotranspiration and canopy storage (Sood and Smakhtin, 2015). In contrast, our model only implicitly considers the effects of vegetation on ET, as the associated impacts are included in the observational constraint. However, we want to highlight that spatial distribution of model parameters depend on assumptions and some degree of simplification as well, and thus does not necessarily represent reality better than the global uniform parameter set obtained from multiple observational data.

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

Conclusion

In this study, we assessed the relative contributions of snow pack versus soil and retained water variations to the variability of total terrestrial water storage (TWS) for northern mid-to-high latitudes. To do so, we constrained a parsimonious hydrological model with multi-criteria calibration against multiple Earth observation data streams including TWS from

GRACE satellites and snow pack estimates from GlobSnow. The optimized model showed considerably good agreement with observed patterns of hydrological fluxes and states, and was found to perform comparable or better than simulations from state-of-the-art macro-scale hydrological models. This underlines the potential of simple hydrological models tied to observational data streams as powerful tools to diagnose and understand large scale water cycle patterns. Further, it highlights the benefits of considering multiple, complementary data constraints to overcome their individual shortcomings.

Consistent with previous studies, we found that seasonal TWS variations are dominated by the development of snow pack during winter months in most places of the mid-to-high northern latitudes. In contrast to this seasonal pattern, our study reveals that not snow but anomalies in liquid water storages, mainly comprising soil moisture, drive inter-annual TWS variations in almost the entire spatial domain. This counter-intuitive pattern was found to relate to larger rainfall anomalies as compared to snowfall anomalies.

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

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

Competing interests

The authors declare that they have no conflict of interest.

References

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

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

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

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system, Journal of Hydrometeorology, 10, 623-643, 10.1175/2008jhm1068.1, 2009.

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

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

Behrangi, A., Christensen, M., Richardson, M., 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.

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

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

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

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

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

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

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

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

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

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

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, Journal of Hydrology, 270, 105-134, 2003.

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

Döll, P., Fritsche, M., Eicker, A., and Müller-Schmied, H.: Seasonal water storage variations as impacted by water abstractions: Comparing the output of a global hydrological model with GRACE and GPS observations, Surveys in Geophysics, 35, 1311-1331, 10.1007/s10712-014-9282-2, 2014a.

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

Döll, P., Douville, H., Güntner, A., Müller-Schmied, H., and Wada, Y.: Modelling freshwater resources at the global scale: Challenges and prospects, Surveys in Geophysics, 37, 195-221, 10.1007/s10712-015-9343-1, 2015.

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

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

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

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

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

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

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

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

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

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

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

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

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

Huffman, G. J., Adler, R., Morrissey, M. M., Bolvin, D., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree resolution from multisatellite observations, Journal of Hydrometeorology, 2, 36-50, 2000.

Huffman, G. J., and Bolvin, D.: Version 1.2 GPCP One-Degree Daily precipitation data set: Documentation, NASA, 2013.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Orth, R., Koster, R. D., and Seneviratne, S. I.: Inferring soil moisture memory from streamflow observations using a simple water balance model, Journal of Hydrometeorology, 14, 1773-1790, 10.1175/jhm-d-12-099.1, 2013.

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

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

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

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

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

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F.: The observed state of the water cycle in the early twenty-first century, Journal of Climate, 28, 8289-8318, 10.1175/jcli-d-14-00555.1, 2015.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth System Science Data, 9, 389-413, 10.5194/essd-9-389-2017, 2017.

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

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

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

Sorooshian, S., Duan, Q., and Gupta, V. K.: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, Water Resources Research, 29, 1185-1194, 10.1029/92WR02617, 1993.

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

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

Swenson, S.: Assessing high-latitude winter precipitation from global precipitation analyses using GRACE, Journal of Hydrometeorology, 11, 405-420, 10.1175/2009jhm1194.1, 2010.

Syed, T. H., Famiglietti, J. S., and Chambers, D. P.: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales, Journal of Hydrometeorology, 10, 22-40, 10.1175/2008jhm993.1, 2009.

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

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

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

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

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

Trischenko, A. P.: Removing unwanted fluctuations in the AVHRR thermal calibration data using robust techniques, Journal of Atmospheric and Oceanic Technology, 19, 1939-1953, 2002.

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

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

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

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

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

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

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

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

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

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

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

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

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B. I., Smith, L. G., and Cooper, J. E.: Clouds and the Earths Radiant Energy System (CERES): An Earth observing system experiment, Bulletin of the Amercian Meteorological Society, 77, 853-868, 1996.

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

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

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

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