Tina Trautmann

Understanding global Water Storage Variations using Model-Data Integration

Cumulative dissertation
for the degree ‘doctor rerum naturalium’ (Dr. rer. nat.)
in Hydrology

submitted to the
Faculty of Mathematics and Natural Sciences
Institute of Environmental Sciences and Geography
at the University of Potsdam, Germany

prepared at the
Max-Planck-Institute for Biogeochemistry, Jena
Department of Biogeochemical Integration

Date of Disputation: 15th September 2022

AUTHOR
Tina Trautmann
Max-Planck-Institute for Biogeochemistry, Jena, Department of Biogeochemical Integration
tina.trautmann0691@gmail.com

REVIEWERS
Prof. Dr. Andreas Güntner (1st Supervisor)
Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany and Institute of Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany
Dr. Martin Jung (2nd Supervisor)
Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany
Prof. Dr. Marc F. P. Bierkens
Department of Physical Geography, Utrecht University, Utrecht, the Netherlands

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-56595
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-565954

Acknowledgements

The last 5 years have been an exciting, challenging, fun and sometimes rough journey that has literally taken me around the globe. There are many people who have accompanied me on my path, be it just for a while or for the entire time, and for whom I am more than grateful.

First and foremost, there are my parents, grandparents and family, who have supported me constantly, even though they didn't quite understand what exactly I was doing.

Martin, who sometimes understood even better than I did what I was doing, and who taught me how to ‘condense and concise’ and be pro-active.

Andreas, who provided the often-needed objective view from a distance.

Sujan, who helped me navigate the rough seas with SINDBAD.

Hyungjun, who welcomed me kindly to experience the other side of the world.

BGI and the people at the institute, who contributed to a great working atmosphere, scientific discussions, lively meetings, relaxing coffee breaks and memorable summer parties.

All the colleagues, who have become amazing friends along the way, and who would always be up for awkward lunch conversations, ice cream dates and hikes, BBQs and brunches, morning swims and coffees, Ingwer-shots and dancing through nights.

Especially the book club, for thinking of creative ideas to escape pandemic boredom and for making me feel at home. S, for the artistic contribution to this thesis. JT, for always opening their home and garden, spontaneous Fenster-Schnapps, game nights and dinners, and making these years a real fun experience. And last but not least, the best office mate, for the constant supply of coffee, sliced fruit, getting me addicted to Sudokus and standing this end-of-thesis madness with me.

Danke euch.

Summary

Climate change is one of the greatest challenges to humanity in this century, and most noticeable consequences are expected to be impacts on the water cycle – in particular the distribution and availability of water, which is fundamental for all life on Earth. In this context, it is essential to better understand where and when water is available and what processes influence variations in water storages. While estimates of the overall terrestrial water storage (TWS) variations are available from the GRACE satellites, these represent the vertically integrated signal over all water stored in ice, snow, soil moisture, groundwater and surface water bodies. Therefore, complementary observational data and hydrological models are still required to determine the partitioning of the measured signal among different water storages and to understand the underlying processes. However, the application of large-scale observational data is limited by their specific uncertainties and the incapacity to measure certain water fluxes and storages. Hydrological models, on the other hand, vary widely in their structure and process-representation, and rarely incorporate additional observational data to minimize uncertainties that arise from their simplified representation of the complex hydrologic cycle.

In this context, this thesis contributes to improving the understanding of global water storage variability by combining simple hydrological models with a variety of complementary Earth observation-based data. To this end, a model-data integration approach is developed, in which the parameters of a parsimonious hydrological model are calibrated against several observational constraints, inducing GRACE TWS, simultaneously, while taking into account each data's specific strengths and uncertainties. This approach is used to investigate 3 specific aspects that are relevant for modeling and understanding the composition of large-scale TWS variations.

In the first study, the focus is on Northern latitudes, where snow and cold-region processes define the hydrological cycle. While this study confirms previous findings that seasonal dynamics of TWS are dominated by the cyclic accumulation and melt of snow, it reveals that inter-annual TWS variations on the contrary, are determined by variations in liquid water storages. Additionally, it is found to be important to consider the impact of compensatory effects of spatially heterogeneous hydrological variables when aggregating the contribution of different storage components over large areas. Hence, the determinants of TWS variations are scale-dependent and underlying driving mechanism cannot be simply transferred between spatial and temporal scales. These findings are supported by the second study for the global land areas beyond the Northern latitudes as well.

This second study further identifies the considerable impact of how vegetation is represented in hydrological models on the partitioning of TWS variations. Using spatio-temporal varying fields of Earth observation-based data to parameterize vegetation activity not only significantly improves model performance, but also reduces parameter equifinality and process uncertainties. Moreover, the representation of vegetation drastically changes the contribution of different water storages to overall TWS variability, emphasizing the key role of vegetation for water allocation among different storages, especially between sub-surface and delayed water storages. However, the study also identifies parameter equifinality regarding the decay of sub-surface and delayed water storages by either evapotranspiration or runoff, and thus emphasizes the need for further constraints hereof.

The third study focuses on the role of river water storage, in particular whether it is necessary to include computationally expensive river routing for model calibration and validation against the integrated GRACE TWS. The results suggest that river routing is not required for model calibration in such a global model-data integration approach, due to the larger influence other observational constraints, and the determinability of certain model parameters and associated processes are identified as issues of greater

relevance. In contrast to model calibration, considering river water storage derived from a routing scheme can already significantly improve modelled TWS compared to GRACE observations, especially in tropical regions and Northern lowlands and wetlands, and thus should be considered for model evaluation against GRACE data.

Beyond these specific findings that contribute to improved understanding and modeling of large-scale TWS variations, this thesis demonstrates the potential of combining simple modeling approaches with diverse Earth observational data to improve model simulations, overcome inconsistencies of different observational data sets, and identify areas that require further research. These findings are of interest for other large-scale hydrological studies as well, and encourage future efforts to take advantage of the increasing number of diverse global observational data.

Zusammenfassung

Der Klimawandel stellt eine der größten Herausforderungen für die Menschheit in diesem Jahrhundert dar, und es wird erwartet, dass die am ersichtlichsten Auswirkungen Veränderungen des Wasserkreislaufs sein werden - insbesondere hinsichtlich der Verteilung und Verfügbarkeit von Wasser, was von grundlegender Bedeutung für alles Leben auf der Erde ist. In diesem Zusammenhang ist es von entscheidender Bedeutung, besser zu verstehen, wo und wann Wasser verfügbar ist und welche Prozesse die Schwankungen in den Wasserspeichern beeinflussen. Diesbezüglich liefern die GRACE-Satelliten zwar Schätzungen der Gesamtschwankungen der terrestrischen Wasserspeicher (TWS), doch stellen diese das vertikal integrierte Signal über alles Wasser in Eis, Schnee, Bodenfeuchte, Grundwasser und Oberflächengewässern dar. Daher werden zusätzliche Beobachtungsdaten und hydrologische Modelle benötigt, um die Aufteilung des gemessenen Signals auf die verschiedenen Wasserspeicher zu bestimmen und die zugrunde liegenden Prozesse, die die beobachteten Schwankungen verursachen, zu verstehen. Die Anwendung von großmaßstäblichen Beobachtungsdaten ist jedoch durch ihre spezifischen Unsicherheiten und die Unfähigkeit, bestimmte hydrologische Flüsse und Speicher zu messen, begrenzt. Hydrologische Modelle hingegen variieren stark in ihrer Struktur und Prozessdarstellung und beziehen nur selten zusätzliche Beobachtungsdaten ein, um die sich aus ihrer vereinfachten Darstellung des komplexen Wasserkreislaufs ergebenden Unsicherheiten zu minimieren.

In diesem Zusammenhang trägt diese Arbeit dazu bei, die Variabilität der globalen Wasserspeicher besser zu verstehen, indem einfache hydrologische Modelle mit einer Vielzahl von sich ergänzenden Erdbeobachtungsdaten kombiniert werden. Zu diesem Zweck wird ein Ansatz zur Integration von Modellen und Daten entwickelt, bei dem die Parameter einfacher hydrologischer Modelle gleichzeitig gegen verschiedenen Beobachtungsdaten, einschließlich GRACE TWS, kalibriert werden, wobei die spezifischen Stärken und Unsicherheiten der einzelnen Daten berücksichtigt werden. Dieser Ansatz wird verwendet, um drei spezifische Aspekte zu untersuchen, die für die Modellierung und das Verständnis der Zusammensetzung von großräumigen TWS-Variationen relevant sind.

In der ersten Studie liegt der Schwerpunkt auf den nördlichen Breiten, wo der hydrologische Kreislauf durch Schnee und andere Prozesse, die spezifisch für kalte Regionen sind, bestimmt wird. Während diese Studie frühere Erkenntnisse darin bestätigt, dass die jahreszeitliche Dynamik des TWS von der zyklischen Akkumulation und Schmelze von Schnee dominiert wird, zeigt sie hingegen, dass die zwischenjährlichen TWS-Schwankungen durch Variationen der flüssiger Wasserspeicher bestimmt werden. Darüber hinaus wird festgestellt, dass es wichtig ist, die Auswirkungen kompensatorischer Effekte räumlich heterogener hydrologischer Variablen zu berücksichtigen, wenn der Anteil verschiedener Speicherkomponenten über große Gebiete aggregiert wird. Die Einflussfaktoren der TWS-Schwankungen sind daher skalenabhängig, und zugrunde liegenden Antriebsmechanismen lassen sich nicht einfach zwischen räumlichen und zeitlichen Skalen übertragen. Diese Ergebnisse werden durch die zweite Studie über nördlichen Breiten hinaus auch für die andere globalen Landflächen bestätigt.

Diese zweite Studie zeigt außerdem, dass die Art und Weise, wie Vegetation in hydrologischen Modellen dargestellt wird, einen erheblichen Einfluss auf die Aufteilung der TWS-Variationen hat. Die Verwendung von räumlich-zeitlich variierenden Erdbeobachtungsdaten zur Parametrisierung der Vegetationsaktivität verbessert nicht nur erheblich die Modellgüte, sondern verringert auch die Parameteräquifinalität und Prozessunsicherheiten. Darüber hinaus verändert die Berücksichtigung der Vegetation den Beitrag der verschiedenen Wasserspeicherkomponenten zum TWS drastisch und unterstreicht damit die Schlüsselrolle der Vegetation für die Wasserverteilung, insbesondere zwischen unterirdischen und verzögerten Wasserspeichern. Die Studie zeigt jedoch auch, dass die Parameter für die Verringerung von unterirdischen und verzögerten Wasserspeichern entweder durch Evapotranspiration oder Abfluss

schwer zu unterscheiden sind, und unterstreicht damit die Notwendigkeit einer weiteren und verbesserten Eingrenzung dieser Parameter.

Die dritte Studie befasst sich mit der Rolle des in Flüssen gespeicherten Wassers, insbesondere mit der Frage, ob es notwendig ist, rechenintensives Fluss-Routing für die Kalibrierung und Validierung des Modells gegen GRACE TWS zu berücksichtigen. Die Ergebnisse deuten darauf hin, dass das Flussrouting für die Modellkalibrierung in einem solchen globalen Modell-Daten-Integrationsansatz nicht erforderlich ist, da Einschränkungen durch andere Beobachtungsdaten einen größeren Einfluss haben und die Definierbarkeit bestimmter Modellparameter und damit verbundener Prozesse als von relevanterer Bedeutung identifiziert werden. Im Gegensatz zur Modellkalibrierung kann die Berücksichtigung der Wasserspeicherung in Flüssen den modellierten TWS im Vergleich zu den GRACE-Beobachtungen bereits erheblich verbessern, insbesondere in tropischen Regionen und nördlichen Tiefland- und Feuchtgebieten, und sollte daher bei der Modellevaluierung gegen GRACE Daten berücksichtigt werden.

Über diese spezifischen Ergebnisse, die zum besseren Verständnis und Modellierung großräumiger TWS-Variationen beitragen, hinaus, zeigt diese Arbeit das Potenzial der Kombination einfacher Modellierungsansätze mit verschiedenen Erdbeobachtungsdaten zur Verbesserung von Modellsimulationen, zur Überwindung von Unstimmigkeiten zwischen verschiedenen Beobachtungs-datensätzen und zur Identifizierung von Bereichen, die weitere Forschung erfordern. Damit sind diese Ergebnisse auch für andere großmaßstäbliche hydrologische Studien von Interesse und ermutigen künftige Bemühungen, die zunehmende Anzahl unterschiedlicher globaler Beobachtungsdaten zu nutzen.

Content

Chapter 1

Introduction

1.1 Background and Motivation

Water is the essence of all life and its availability shapes not only human existence, but also the functioning of ecosystems and the Earth's climate. Especially in the context of climate changes it is crucial to understand when and where water is available, and – equally importantly – what are the underlying processes and mechanism of these variations, in order to be able to estimate future changes in water availability and take actions accordingly.

To address this challenge, hydrological research has two tools at its disposal: (I) measurements and observations of water fluxes and storages, and (II) hydrological models that depict our process-understanding and combine the components of the water balance.

While scientist for years relied on point-measurements to derive, validate and extrapolate their hydrologic process-understanding in time and space, the increasing quantity, quality and temporal coverage of spatially continuous data from remote sensing-based observations along with increasing computational resources enabled a new era of possibilities in hydrological research.

Among all Earth-observation based data, the Gravity Recovery And Climate Experiment (GRACE) and its follow-on (GRACE-FO) take a unique role: for the first time they provide observations of one component of the water balance, that has been inaccessible before – the overall terrestrial water storage change. Although this enables various opportunities, GRACE observations are not a panacea, as they only provide overall water storage variability and, therefore, it's still necessary to understand the underlying processes, and which storages are the main contributors to the observed changes.

This thesis aims to improve our understanding of global water storage changes by combining GRACE observations with Earth-observation based data from multiple sources, and simple conceptual hydrological modeling approaches.

The following provides the background for this thesis. First, a brief overview on the global water cycle and its components is given. The subsequent sections provide a review on available large-scale Earth-observation based data and data products, evaluate state-of-the-art global hydrological modeling concepts and their challenges, and finally present common methods to combine both, data and models. Eventually, recent attempts to assess global water resources using the before introduced methods are summarized. Based on this, research gaps will be identified and the objectives of this thesis defined in detail.

1.1.1 The global Water Cycle

The global water cycle describes the occurrence and the flow of water within the Earth system, between atmosphere, lithosphere, hydrosphere, and back, driven by solar energy and gravitation (Dorigo et al. 2021) (Fig. 1.1). Following the principle of mass conservation, the terrestrial water cycle is characterized by the balance of incoming and outgoing fluxes, and its components can be estimated for a distinct area by the water balance equation (Sposób 2011):

$$\Delta TWS = Precip - ET - Q \tag{1.1}$$

where the change in water storages (ΔTWS) during a given time interval is determined by the inflow given by precipitation (Precip) minus the outflow by the fluxes of evapotranspiration (ET) and runoff (Q) (Schmidt et al. 2008a).

These water fluxes and storages are strongly interconnected and depend on various climatic and physio-geographic factors. For instance, precipitation, depending on temperature either in the form of snow or rain, falls on the land surface, where it is (temporarily) retained as snow or ice, on vegetation or on the land surface itself. Depending on various factors, such as characteristics of the surface (topography, land cover, vegetation) and subsurface (soil properties, soil saturation, geology), liquid water infiltrates into the soil and potentially percolates further into the groundwater, or accumulates in surface water bodies such as lakes. Eventually, water either returns to the atmosphere by the process of evapotranspiration or, following the gravitational gradient, leaves the terrestrial system as runoff in streams and rivers that finally discharge into the ocean (Trenberth & Asrar 2014).

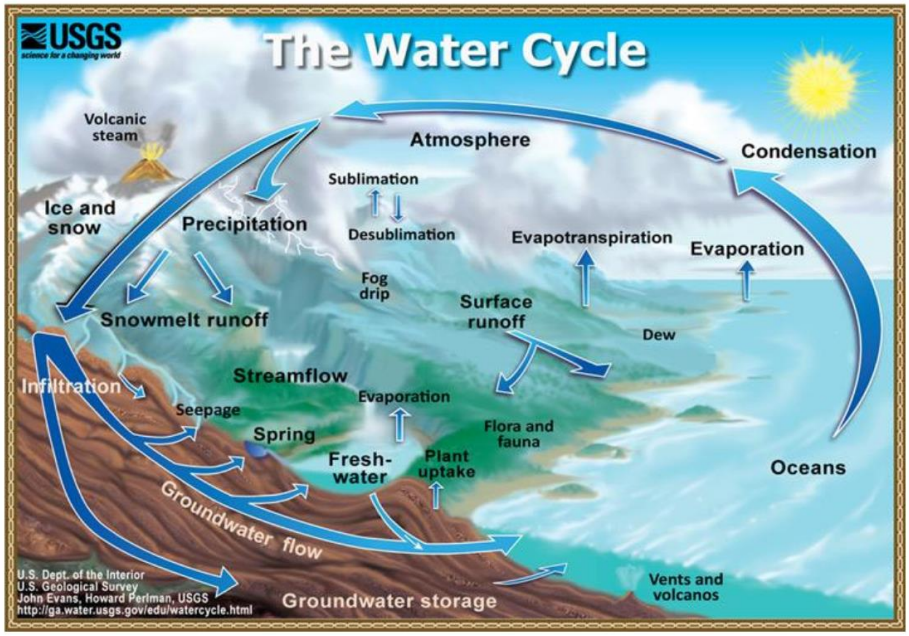

Figure 1.1. Schematic overview on the natural water cycle (Perlman & Evans 2000).

Evapotranspiration is the phase change of water into vapor and its transfer to the atmosphere (Novák 2011). As it is driven by energy, i.e., net radiation, it represents a key process within the Earth system that connects the global water and energy cycle. Evapotranspiration entails the solely physically driven evaporation from surfaces (such as soil and water bodies), and sublimation from snow surfaces, as well as transpiration, which is actively regulated through vegetation by opening and closing of their stomata for photosynthesis.

Runoff describes the lateral flow of water following the gravitational gradient out of the area of consideration. It encompasses any lateral movement of water on the land surface (overland flow), in the soil (inter-flow), and in the groundwater (groundwater flow). Commonly it's referred to as the lateral flow within river channels, that is composed of a fast, direct (surface) runoff component and a slow, delayed component (baseflow). While irrelevant for the global water balance, runoff represents a considerable water input at local and regional scale as it transports water from upstream to downstream areas.

The ***terrestrial water storage*** term in Eq. 1.1 includes all water stored in ice, snow, soil, groundwater, vegetation and surface water bodies such as lakes, rivers and reservoirs. Although its long-term net change can be assumed to be negligible in theory (Rodell et al. 2015), its daily, seasonal and inter-annual fluctuations as well as its varying distribution across space have a considerable impact on agriculture, vegetation, humanity and life itself. But not only the presence of water, also its actual form (ice, snow,

liquid) and where it is retained (soil, groundwater, surface) is relevant, as this defines its accessibility by vegetation and human kind. For instance, surface and river water are the main source of freshwater in many regions, and vegetation relies on the soil and groundwater their roots can access. Next to its accessibility, the kind of water storage further affects the hydrological and the ecosystem response to long-term climate change and extreme events, such as heat waves and heavy rainfalls. For instance, in cold regions, a large portion of annual rainfall is immobilized and plant-inaccessible in the snowpack, which is only released with increasing temperatures in spring, enabling vegetation growth but also (potentially) causing spring floods. Depending on the vegetation type specific rooting depth, plants access water from soil and (partly) groundwater, possibly maintaining photosynthesis during periods with no or little precipitation and even during heat waves.

Altogether, the global terrestrial water cycle is a complex and dynamic system with many interactions, feedbacks and dependencies (Hagemann 2011). For example, ***soil properties***, that define the infiltration capacity and maximum soil water storage, change over time due to the presence or absence of water. On much shorter time-scales, and arguably more dynamically, ***vegetation*** takes a central role at the interface between atmosphere and land surface, as its presence is not only defined by the availability of water, but it moreover itself shapes hydrological processes. Interception on the vegetation surface, for instance, reduces the quantity of water that reaches the land surface, while at the same time the interception storage is strongly exposed for evaporation, which compared to evaporation from soil is not limited by soil resistance (Fisher et al. 2014). Due to the canopy barrier, vegetation reduces the intensity of through fall and stemflow, what potentially increases infiltration and reduces surface runoff. Besides, vegetation cover changes the surface albedo and thus the radiation budget, which in turn affects evapotranspiration. Over time, root growth, nutrient uptake and decomposition of organic plant material also alters soil properties and thus e.g., soil water holding capacity and infiltration capacity. As mentioned above, plants take up water from soil (and groundwater) through their roots, and thus alter those storages. Part of this water is directly used for photosynthesis, while the remaining portion transpires through the plant's open stomata – a process plants can actively regulate. By the process of photosynthesis, vegetation additionally connects the terrestrial water and carbon cycles, introducing further feedbacks in the complex Earth's system.

1.1.2 Measuring the World – Large-scale observational Data

Characterizing the stocks and fluxes of the Earth's water budget at large scales poses considerable challenges to hydrological research (Rodell et al. 2015). While precipitation (e.g., with rain gauges and ground-based radar), evapotranspiration (e.g., with eddy co-variance measurements), runoff (in the form of river discharge) and the state of some water storages (e.g., soil moisture with lysimeters) can be - with more or less effort and accuracy - measured locally, these in-situ measurements only provide a local snapshot of the complex hydrological system, where non-linear processes, memory effects and processes happening far away from the point of measurement are influential. Over the last decades, spatially and temporal continuous information became available from satellite measurements and observation-integrating products that incorporate them (Dorigo et al. 2021). These data enable an observational-based assessment of hydrological processes and water storages not only in regions with no or sparse ground-based observations, but also at large regional and global scales. While being of undeniable value, such data products come with specific strengths and uncertainties that evolve from the characteristics of the underlying measurements, methods and assumptions, and that need to be considered when applying them for hydrological research (Dorigo et al. 2021). By that, for example, the possibilities to observe major fluxes of runoff and evapotranspiration are still limited, and also several key hydrologic states, such as groundwater, remain poorly measured in many regions (Famiglietti & Rodell 2013). Besides, the spatial resolution and temporal coverage of some satellite measurements might be insufficient. Additionally, the accuracy of large-scale data products is limited to the availability of in-situ observations for calibration and validation, which in most cases are not representative for the satellite's footprint, and suffer from a sparse observational network that is biased to temperate humid regions (Rodell et al. 2015). In particular the land surface hydrology in cold regions poses distinct challenges, due to the lack and low quality of observational data on the one hand, and deficiencies in the retrieval of remote-sensing based estimates

(e.g., limited applicability of optical sensors due to long polar nights, effects of permafrost and freeze/thaw dynamics that cannot be measured) on the other hand (Rodell et al. 2015).

The following provides an overview on available large-scale to global observational-based data of water balance components.

Precipitation is one of the most important hydrological variables, that is challenging to estimate due to its high spatio-temporal heterogeneity (Herold et al. 2016). Various efforts to estimate precipitation at global scale exist, ranging from upscaling of gauge-measurements (e.g., NOAA's Climate Prediction Center Unified precipitation product CPCU (Chen et al. 2008)), to radar remote sensing (e.g., Tropical Rainfall Measuring Mission TRMM (Huffman et al. 2010), Global Precipitation Measurement GPM (Huffman et al. 2012), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks PERSIANN (Nguyen et al. 2018)), products that combine both (e.g., Global Precipitation Climatology Project GPCP (Huffman et al. 2000)), and such that merge in-situ and satellite observations with re-analysis data (e.g., Multi-Source Weighted-Ensemble Precipitation MSWEP (Beck et al. 2017)). Gauge-based estimates suffer from the irregular density of the gauge network, as well as systematic biases in mountainous (due to elevation bias in gauge placement) and snow-dominated (due to wind-induced undercatch) regions (Rasmussen et al. 2012). Satellite estimates on the contrary observe large areas instantaneously, but are prone to systematic biases resulting from satellite operation, are insensitive to light rainfall, and have limited accuracy over snow and ice cover (Beck et al. 2017).

Estimating ***terrestrial evapotranspiration*** at large scales is particularly demanding, due to its high variability in space and time, and since it cannot be measured directly by remote sensing (Rodell et al. 2015, Dorigo et al. 2021). Existing estimates based on satellite retrievals, e.g., from the Global Land Evaporation Amsterdam Model (GLEAM, Miralles et al. 2011, Martens et al. 2017), and the evapotranspiration data sets from MODIS (Mu et al. 2011) and NOAA (Zhang et al. 2010), apply evapotranspiration models such as the Penman-Monteith or Priestley-Taylor formula that derive estimates of evapotranspiration based on vegetation indices, like Leaf Area Index (LAI), Normalized Vegetation Index (NDVI), and Vegetation Optical Depth (VOD), along with further satellite measurements, including land surface temperature, albedo, soil moisture and land cover (Cui et al. 2018). Other global estimates of evapotranspiration are based on re-analysis, that incorporate in-situ and satellite measurements with land surface models (e.g., MERRA (Rienecker et al. 2011), GLDAS (Rodell et al. 2004)), and thus rely largely on the underlying model. On the contrary, a purely data-driven estimate of evapotranspiration that is derived by merging remote sensing and meteorological data with eddy co-variance measurements from FLUXNET sites using machine learning algorithms is provided by the FLUXCOM initiative (Jung et al. 2019). However, these indirect estimates of evapotranspiration include systematic errors in semi-arid regimes and tropical rain forests and are known to imperfectly represent water stress and canopy interception (Dorigo et al. 2021).

Runoff is usually assessed as discharge that passes the cross-sectional area of a river per time unit. By that, discharge gives an integrated hydrological signal of the catchment area upstream of the gauge. A global database with discharge time series from thousands of gauging stations is provided by the Global Runoff Data Centre (GRDC). While this database is limited to station records of varying lengths and quality, a recent global database of river width from Landsat data (Allen & Pavelsky 2018), in combination with measurements of surface water elevation from the forthcoming Surface Water Ocean Topography (SWOT) mission, will allow to estimate river discharge independent from gauging stations in future (Tuozzolo et al. 2019). However, such estimates that rely on the surface water area resp. river width include uncertainties from riparian forest cover as well as ice cover and ice jams in winter (Hicks and Beltaos 2008). Accordingly, so far, spatially distributed information of discharge at global scale is lacking, whereas monthly gridded reconstructions of near-natural streamflow based on machine learning are available for Europe (E-RUN, Gudmundsson & Seneviratne 2016), and the global land area (GRUN, Ghiggi et al. 2019).

Since 2002, ***terrestrial water storage changes*** at global scale are available from the Gravity Recovery And Climate Experiment (GRACE) (Tapley et al. 2004, Wahr et al. 2004). The GRACE satellites measure the Earth's time-varying gravitational field, and because such variations are mainly caused by changes in the

hydrosphere, allow to infer overall monthly water storage variations. However, GRACE data cannot be feasibly replicated by ground-based observations (Rodell et al. 2015), as it comprises all water in and on land, including snow and ice, surface water, soil moisture, groundwater and biological water. Additionally, the native spatial resolution is rather broad (~250-300 km), and because only anomalies relative to a time-mean baseline are measured, GRACE data does not yield absolute water volumes. Therefore, further observations of its water storage components are required.

Among them, the presence of ***surface water*** can be determined relatively accurately from remote sensing due to the strong absorption of water in the visible, infrared and microwave wavelengths. However, it's difficult to distinguish between different types of surface water bodies, such as rivers, man-made reservoirs and paddy fields, which are affected by different hydrological processes. While global datasets of the location, seasonality and long-term trends of surface water bodies exist from statistical extrapolation of regional data (Downing et al. 2006), and from satellite imagery (Verpoorter et al. 2014, Yamazaki et al. 2015, Pekel et al. 2016), it's challenging to monitor the evolution of their depth and volume, and by that to give estimates of absolute water changes (Lu et al. 2013). In this context, the NASA's Surface Water Ocean Topography (SWOT) mission plans to provide an inventory and changes in freshwater bodies by using radar interferometry to measure the elevation of surface water.

Similar to surface water, ***snow*** cover can be effectively detected by optical remote sensing due to the high reflectance of snow in visible wavelengths, yet only under cloud free conditions. Besides, microwave remote sensors are able to monitor snow cover and snow water equivalent also in the presence of clouds, but the penetration depth of the radar signal is limited so that the signal saturates under deep snow conditions (Takala et al. 2011). Additionally, the spatial resolution of satellite imagery prohibits accurate estimates in mountainous regions, and alpine and high latitude regions usually have insufficient coverage by observational networks (Dorigo et al. 2021).

The same issue signal saturation affects the determination of ***soil moisture*** by remote sensing. Although various estimates of soil moisture from radar remote sensing, e.g., from ESA's Soil Moisture and Ocean Salinity (SMOS) mission (Kerr et al. 2012) and NASA's Soil Moisture Active Passive (SMAP) mission (Entekhabi et al. 2010), as well as from the combination of multiple sensors (e.g., ESA CCI Soil Moisture, Dorigo et al. 2017) exist, monitoring soil moisture from space is limited by the penetration depths of the radar signal to the top few centimeters, and becomes more difficult under dense vegetation cover (Dorigo et al. 2010). Next to satellite observations, globally continuous estimates of soil moisture up to the root zone are available based on machine learning (e.g., SoMo, O & Orth 2021), and modeling approaches (e.g., GLEAM root zone soil moisture, Martens et al. 2017). However, no purely observation-based estimate of root zone and deep soil moisture, that are relevant for many hydrological and biological processes, exist.

Likewise, ***groundwater*** storages cannot directly be monitored at large spatial scale (Dorigo et al. 2021), and thus total water storage variations from GRACE remain the only observational information on water storages below the top soil layer, while they do not allow to differentiate between the kind and potential accessibility of sub-surface storages by vegetation and human kind (Dorigo et al. 2021).

Similar to groundwater, the distribution, ice content and volume of ***permafrost*** cannot be directly mapped (Dorigo et al. 2021).

Next to components of the hydrological cycle per se, remote sensing first and foremost observes the ***Earth surface***, and based on the characteristic reflectance of different surfaces allows to map land cover classes (e.g., ESA CCI Land Cover time-series (ESA 2017)), and infer vegetation activity and health from different indices (e.g., Leaf Area Index LAI, Normalized Difference Vegetation Index NDVI, Enhanced Vegetation Index EVI).

Although the recent variety and quality of large-scale observational data is unprecedented, Rodell et al. 2015 showed that the global water balance cannot be closed by observations alone. Therefore, models are still needed to bridge the gap within and between different data streams, to reconstruct fluxes and storages that cannot be observed directly, and to extrapolate hydrological processes into the future.

1.1.3 Simulating the World – Global Hydrological Models

A myriad of models, that simulate the flow of water and estimate water resources, were designed for multiple purposes and application on various spatial scales. While they each are a simplified representation of reality, they are valuable tools to estimate future changes in water availability and to assess hydrological processes and storages that cannot directly be observed. At global scale, one can distinguish between 3 different kinds of models that simulate the global water cycle, that were developed for different purposes by different communities and thus focus on different processes and use different modeling approaches: (I) ***Macro-Scale Hydrological Models*** (MHMs) and ***Global Hydrology and Water Resources Models*** (GH-WRMs) that represent the vertical and lateral movement of water while focusing on streamflow, the water availability in surface and groundwater storages, and in the case of GH-WRMs, the human alteration of the water cycle; (II) ***Land Surface Models*** (LSMs) that evolved as components of Earth System Models and Global Circulation Models, and focus on the vertical flow of water by simulating the water and energy exchange between the land surface and atmosphere in a more physical manner; and (III) ***Dynamic Global Vegetation Models*** (DGVMs) that have their main focus on the global carbon cycle while modeling vegetation growth and distribution including the active response of vegetation to environmental conditions (Bierkens 2015, Telteu et al. 2021).

However, the distinction between these types is fluid, and this thesis, therefore, will refer to all kinds of models that simulate the large-scale and global water cycle as ***Global Hydrological Models*** (GHMs). Besides, differences not only exist between different types of global models, but also within each type, as different modeling groups consider different underlying hydrological processes that lead to varying model structures, apply different equations for the same hydrological process, and, even for the same equation, use different parametrizations. These differences lead to diverging simulations, that on the one hand, allow to assess the uncertainties that evolve from model structure and parametrization, by analyzing simulations of an ensemble of different models in Model Intercomparison Projects (MIPs). On the other hand, the variety of model structures and parametrizations hinders in-depth understanding of inter-model differences and identification of areas for future model development (Bierkens 2015, Telteu et al. 2015).

Initially, GHMs adopted rather conceptual process-representation from catchment-scale models, that became more process-based over time due to continuous adding of functionality (e.g., in terms of process representation) and the availability of global data (Döll et al. 2015). However, the grown-complexity does not necessarily improve model performance (Orth et al. 2015), but often impedes understanding of model behavior, and the majority of added and refined hydrological processes can rarely be validated by observational data. In general, existing GHMs differentiate between 3 to 11 water storages, usually including canopy, snow, soil water (with varying layers and depths), and groundwater storages. Some additionally simulate lakes, wetlands, river water and man-made reservoirs (Schellekens et al. 2017, Telteu et al. 2021). While all GHMs to some degree include vegetation and land cover to define processes such as transpiration and infiltration, they vary considerably in the detail and the way vegetation is represented. Hydrologic models usually include a rather simplified representation of vegetation (Quevedo et al. 2008, Weiss et al. 2012, Telteu et al. 2021), whereas other communities represent vegetation and its interaction with processes of the water and carbon cycle more detailed. Thus, the number of vegetation classes in GHMs can vary between 3 to 24, some prescribing vegetation characteristics with static parameter values, while others apply climatologies of, for example, remote sensing-based LAI (Telteu et al. 2021). Although simulating the lateral flow of water in streams using river routing schemes is an inherited feature of hydrological models, not all GHMs consider this process (Telteu et al. 2021). Besides, river routing is rather computational expensive, as it adds an iterative process in space, in which runoff produced in one modeling unit needs to be transferred to the adjacent downstream unit. Therefore, routing schemes are often an additional module independent from the model that simulates vertical hydrological fluxes (Sood and Smakhtin 2015).

Uncertainty of model simulations result from the simplified representation of hydrological processes but also from model parameterization, initial conditions and the used hydro-meteorologic forcing (Sood and

Smakhtin 2015). Therefore, it is essential, to validate model simulations against observational data and adjust them if necessary.

1.1.4 Combining the best of 2 Worlds – Model-Data Integration

The previous sections gave an overview on available large-scale observational data sets of hydrological variables, and models that simulate the global water cycle. While large-scale observational data possess a rather broad spatial and temporal resolution due to features of the underlying satellite measurements and do not provide (direct) estimates of all hydrological fluxes and storages, GHMs only present a simplified representation of reality and their accuracy depends on various factors, such as the process-representation, their parametrization and the quality of input data. Therefore, it is essential to combine both, models and data to: (I) bridge the spatio-temporal gaps in observations, overcome their inaccuracies, and estimate processes and storages that cannot be observed directly, and (II) to reduce model uncertainty and validate the simplified representation of reality in models (Lahoz & Schneider 2014). To do so, different methods, such as model calibration, data integration, data assimilation (Rodell et al. 2015), and recent efforts of hybrid-modeling (Reichstein et al. 2019) exist.

Especially in conceptual hydrological models, parameters often cannot be determined directly, as they not necessarily correspond to physically measurable properties. Therefore, they need to be estimated indirectly, by ***model calibration***. This procedure aims to derive the best possible match between simulated and observed variables by iteratively tuning model parameters and evaluating the resulting simulations against observations in terms of some efficiency metric, in order to infer the optimal parameter-set. The iterative permutation of parameter values can be either done manually by 'trial and error' and taking into account expert knowledge on plausible parameter values, or by applying automatic optimization techniques (Fischer 2013). The latter include a mathematical search algorithm, that searches the parameter space for the optimal solution (i.e., global minimum) regarding a cost function, that is usually based on one or more the efficiency metrics (Moriasi et al. 2007). The choice of efficiency metric depends on which aspect of the observed variable should be represented, as such the correlation coefficient emphasizes observed variations, whereas Nash-Sutcliff Efficiency (Nash & Sutcliffe 1970), or Kling-Gupta Efficiency (Gupta et al. 2009) consider correlation, bias and variance in the data. While model calibration holds the advantage to improve hydrologic process representation and increases confidence in model simulation, two potential issues exist: (I) overfitting of the model to the calibration data, and (II) parameter equifinality. To overcome (I), hydrological models are usually calibrated against a spatial or temporal sub-sample of the available data, so that the remaining data can serve as independent evaluation of the calibrated model (Xu 2002). (II) is caused by parameter interactions, that lead to similar good model performance regarding the used efficiency metric by different parameter sets. Parameter equifinality becomes especially relevant with increasing number of modelled processes that are insufficiently constrained by data (Beven 2006). Traditionally, hydrological models are calibrated against time-series of discharge measurements, that for the majority of time represented the only available and suitable observational constraint (Döll et al. 2015). Discharge measurements integrate over processes in the whole upstream catchment, and are a suitable constraint if, for example, the focus is on flood forecasting. However, discharge alone does not allow to infer to the processes that cause and contribute to the observed streamflow variations, and conformity with observed discharge does not guarantee proper representation of hydrological processes within the catchment (Güntner 2008). Therefore, and to mitigate issues of parameter equifinality, it's advantageous to not only calibrate model parameters against one single observed variable, but against multiple observations of hydrological variables simultaneously in a multi-criteria calibration approach (Syed et al. 2009, Sood & Smakthin 2015). Indeed, the benefits of using multiple constraints in model calibration has been shown by several large-scale hydrological studies (e.g., Livneh & Lettenmaier 2012, Lo et al. 2010, Rakovec et al. 2016, Bai et al. 2018, Mostafaie et al. 2018). Nevertheless, only few GHMs tune or calibrate model parameters, and even fewer studies consider multiple observational data sets (Long et al. 2015, Döll et al. 2015).

Next to the calibration of model parameters, observational data can directly be ***integrated*** into hydrologic models, as input to force, initialize and/or parameterize them. This includes for instance the use of static

or seasonally varying land surface characteristics, such as LAI to define certain vegetation parameters (e.g., Weiss et al. 2012). However, direct integration of observational data either limits the model run to the time period in which observations are available, or, if using static values or seasonal climatology, does not support delineation of (future) trends and feedbacks due to changing environmental conditions. Besides, this rather static approach does not necessarily consider model and data uncertainties.

Therefore, more sophisticated ***data assimilation*** techniques are used to combine observational data and models to an optimal estimate of hydrologic variables, while considering their respective uncertainties (Reichle 2008). Data assimilation methods, mostly applied with LSMs and forecast-models, include a process that sequential steps recursively through time, by continually comparing previous model forecasts with newly received data to update the model states for a new forecast (Kumar et al. 2008). Data assimilation techniques such as Kalman filters (e.g., Shamir et al. 2010, Eicker et al. 2014), particle filters (e.g., DeChant & Moradkhani, 2012) and 1-4 dimensional variational algorithms (e.g., Seo et al. 2009, Kumar et al. 2016) are useful tools to interpolate and extrapolate observations to the required scale, yet difficulties potentially arise from managing the complexities of necessary data processing, computational expenses and the trade-off between different assimilation algorithms (Kumar et al. 2008).

Recently, efforts in ***hybrid-modeling***, that combine the theoretical foundations and interpretability of process-based modeling approaches with deep-learning methods that represent poorly understood processes by data-driven machine learning (Kraft et al. 2020).

1.1.5 Assessing the World's Water Variations – Application of GRACE TWS

Over the last 2 decades, terrestrial water storage (TWS) variability observed by the GRACE satellites have become probably the most valuable and extensively used data set to assess the world's water variations and to challenge the assumption of negligible long-term net changes in TWS. GRACE data has been widely used to diagnose trends in water resources (Reager et al. 2015, Rodell et al. 2018, Scanlon et al. 2018, Tapley et al. 2019), to validate and improve GHMs (Güntner 2008, Werth and Güntner 2010, Döll et al. 2014, Eicker et al. 2014, Girotto et al. 2016, Kumar et al. 2016, Chen et al. 2017, Schellekens et al. 2017, Soltani et al. 2021), to enhance the understanding of the water cycle on regional to global scales and diagnose pattern of hydrologic variability (Rodell et al. 2009, Syed et al. 2009, Feng et al. 2013, Felfelani et al. 2017). While TWS variations can be analyzed using statistical methods or complementary observation-based data sets, in the end GHMs are needed to partition the TWS signal into different water storage components (Felfelani et al. 2017).

In this context, several studies suggest difficulties of GHMs to reproduce key pattern of TWS variations on the one hand, and identify partly diverging TWS partitioning between different models on the other hand (Schellekens et al. 2017, Zhang et al. 2017, Scanlon et al. 2018, Kraft et al. 2021). This uncertainty is a major obstacle for diagnosing and understanding global changes of the water cycle thoroughly. While most studies agree that snow dynamics are the primary component of TWS variations in Northern regions (Niu et al. 2007, Rangelova et al. 2007), results differ widely regarding the relevance of other storage components. For example, some models attribute seasonal TWS variations in the tropics to groundwater, while other models suggest they are mainly caused by soil moisture (Schellekens et al. 2017), whereas other studies highlight the importance of surface water to TWS variations in tropical regions (Güntner et al. 2007, Getirana et al. 2017). Since these results largely depend on model structure and parametrization, it is challenging to use models to decompose the integrated GRACE TWS signal and to draw implications of different processes and interactions (Schellekens et al. 2017). Therefore, the inclusion of additional information is required to reduce the uncertainty in partitioning of TWS (Güntner 2008). While observational data of surface water storages exists, especially the differentiation between soil moisture and groundwater is challenging due to the lack of appropriate observational data to constrain models. Hence, GRACE TWS remains the only large-scale observational estimate for sub-surface water storages below the top centimeters that are accessible with radar remote sensing. This imposes main challenges, as the differentiation between plant-accessible soil moisture and deep soil water resp. groundwater quantifies the available water for plant uptake and photosynthesis, and by that links the global water-carbon-and energy cycles. For example, Humphrey et al. (2018) showed co-variations between inter-

annual variations in GRACE TWS and CO_2 growth rate, based on the assumption that GRACE TWS represents fluctuations in plant-accessible water that influences the carbon uptake of land ecosystems. The validity of this assumption and its transferability between spatial scales however is relevant in the ongoing debates of whether temperature or water availability have a larger effect on global atmospheric CO_2 growth rates, and thus amplifying or decreasing of the observed greenhouse effect. Therefore, proper partitioning of TWS is not only relevant for hydrological purposes only, but also for improved understanding of the coupled carbon-water-energy cycle.

1.2 Research Gaps and Objectives

So far, the previous sections highlighted, that

I) it is important to understand TWS variability and its partitioning among different water storages, and how this varies in space and time, not only for hydrological research alone, but also for improved understanding of interactions and feedbacks of the coupled water-carbon cycles.
II) GRACE TWS and other Earth-observation based data provide information on various components of the water balance at global scale, yet their accuracy is limited and key hydrological processes and storages remain inaccessible.
III) existing GHMs are rather complex and differ significantly in their representation of hydrological processes, what hampers the interpretation of their behavior and increases the number of processes that are not supported by observational data.
IV) model-data integration methods are valuable tools to combine observations and process-understanding.

Within this context, this thesis focuses on three aspects that are relevant for understanding and modeling the composition of global TWS variability, while contributing to an overarching objective, as presented in the following:

(I) the potential to assess large-scale TWS variations by combining multiple data streams and simple hydrological modeling approaches

Large-scale observational data have a substantial potential to improve model simulations by using them as constraints in model calibration, yet so far only few modeling studies have used multiple data streams simultaneously to calibrate several parameters of GHMs (Long et al. 2015, Döll et al. 2015). This is surprisingly, because the accuracy and informative value of one observational data set alone is limited, but the combination of several data sets improves the information content and confidence in findings considerably. Especially when assessing large-scale water storage changes it is essential to incorporate several data sets and models, because GRACE data of the integrated TWS remain the only available data on major water storages so far, and rely on further information to distinguish between different water storage components. Therefore, the overarching objective of this thesis is to evaluate the potential of combining multiple Earth-observation based data sets with simple conceptual hydrological modeling approaches to improve our understanding of TWS variability and its composition among different water storages.

(II) the partitioning of TWS variations in cold regions on different spatio-temporal scales

Northern latitudes are among the areas most prone to climate change (Tallaksen et al. 2015), yet modeling the land surface hydrology in cold regions poses distinct challenges. While previous studies identified seasonal snow accumulation and melt as primary determinant for TWS variability and their high influence on hydrological dynamics in such regions (Niu et al. 2007, Rangelova et al. 2007), Jung et al. 2017 showed that the transferability of processes to other scales might be hampered by compensating and increasing effects due to the heterogeneity or coherence of spatial pattern in climatic signals. Hence, it is unclear whether snow is only relevant at local scale, or if the domination of snow holds across spatial scales and for inter-annual TWS variations as well. Therefore, the first objective is to investigate the partitioning of TWS variability among snow and other water storages across spatio-temporal scales by applying the developed model-data integration approach in the cold regions of the Northern Hemisphere.

(III) the impact of how vegetation is represented in GHMs on spatio-temporal TWS composition

While vegetation takes a central role in the hydrological and carbon cycle, its representation in GHMs differs significantly. Many GHMs prescribe vegetation with (static) parameters for distinct classes of land cover or plant functional type, and rarely take advantage of new large-scale remote sensing-based observations on vegetation activity and characteristics. Likewise, the dependency of vegetation presence and activity on water availability is subject of many studies (Wang et al. 2001, Porporato et al. 2004, Reyer et al. 2013, Yang et al. 2014), yet few hydrological studies investigated the opposing impact of vegetation on global TWS variability and composition. Thus, the second objective of this thesis is on the one hand to investigate the potential improvement in model performance by including spatially distributed and time varying vegetation parameters in a large-scale hydrological model, and, on the other hand, to assess the impact of how vegetation is represented on the composition of TWS variability. To do so, different spatially continuous data on vegetation characteristics are considered within the model-data integration approach.

(IV) the effect of river water on model calibration and validation against GRACE TWS

Since GRACE TWS represents the vertically integrated signal of all water storages, it potentially includes river water storage. Previous studies highlighted the relevance of river water for accurately representing observed TWS variability regionally (Güntner et al. 2007, Kim et al. 2009, Getirana et al. 2017), but simulating river routing is a computational demanding process, and it's unclear whether it is required for model calibration and validation against the integrated GRACE TWS at global scale. The third objective tackles this question by investigating the effect of river water storage on model calibration and validation against GRACE TWS.

According to these objectives, this thesis aims to answer the 4 main research questions shown in Fig. 1.2. To address these research questions, a model-data integration approach is developed, in which the parameters of a hydrological model are calibrated in a multi-criteria technique against several observational estimates of different water balance components simultaneously. In order to improve interpretability, the implementation of the hydrological model(s) follows the premise 'the-simpler-the-better', by applying rather simple modeling concepts and only add complexity when needed and supported by data. In this context, and since a profound understanding of the natural system is essential before adding an extra layer of complexity, this thesis doesn't consider human alterations of the water cycle, but concentrates on natural processes and is limited to regions under near-natural conditions.

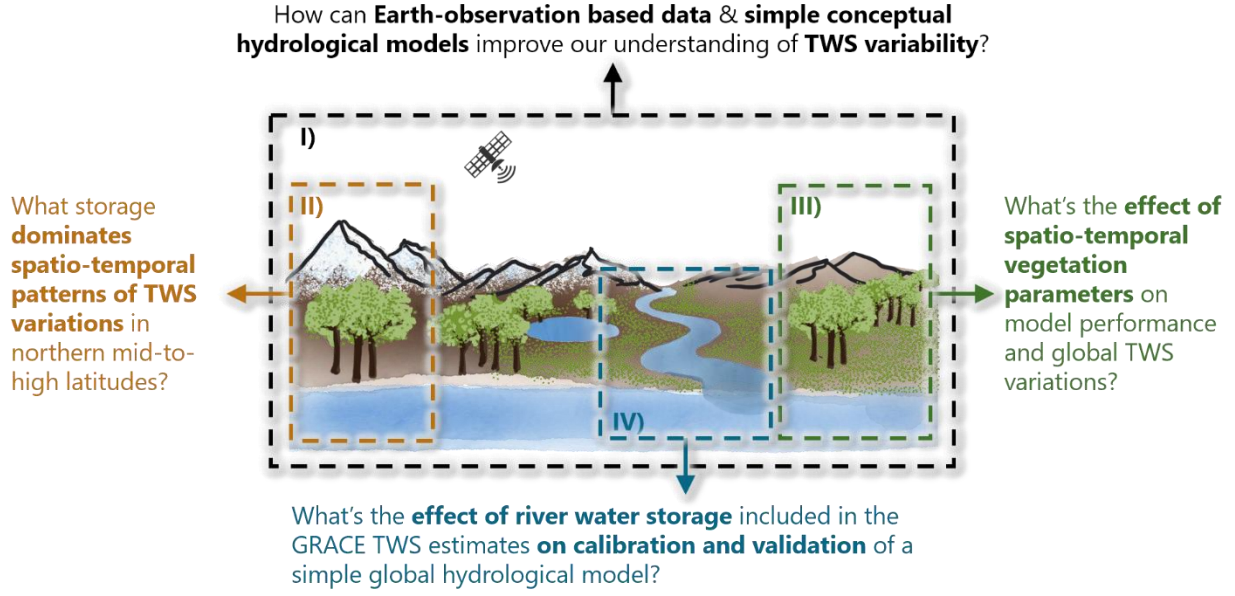

Figure 1.2. Overview on the main research questions addressed in this thesis, highlighting the aspect of the hydrological cycle that they focus on.

1.3 Thesis Outline and Author's Contributions

This cumulative thesis consists of three publications, which are published (**Chapters 2**, **3**) or submitted (**Chapter 4**) in peer-reviewed journals. Each of the publications addresses one of the aspects that have been identified as relevant for understanding and modeling the composition and the variability of TWS. While focusing on specific research questions, each study contributes to the overarching objective to assess the advantage of using multiple Earth-observation based data simultaneously to improve large-scale model simulations.

A conceptual overview on the chapters of this thesis, their main focus and addressed research questions, as well as the underlying methodical concept is shown in Fig. 1.3, while the following gives a more detailed outline on the subsequent chapters, including the author's own contributions.

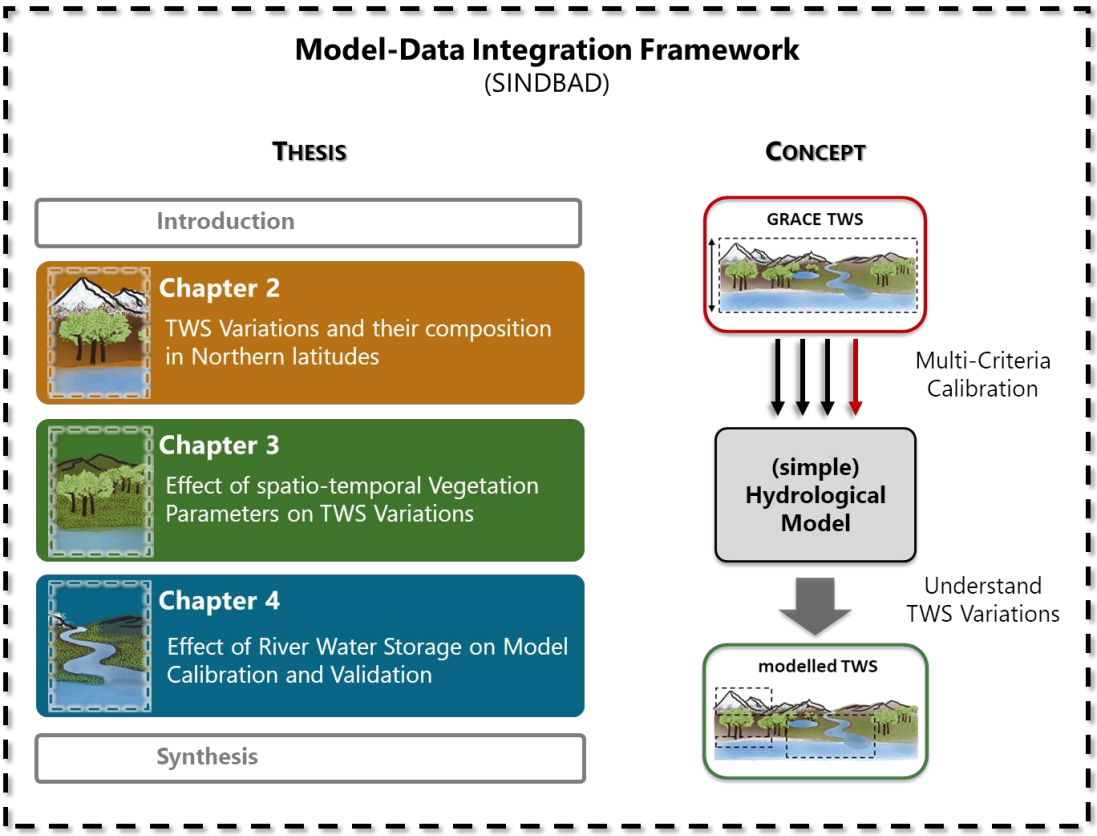

Figure 1.3. Conceptual outline of this thesis, showing the publications, their main focus and the concept of the underlying model-data integration framework*. Colors indicate the research questions that are addressed by the respective publication.

Chapter 2 introduces the model-data integration approach, that calibrates the parameters of a parsimonious conceptual hydrological model against multiple Earth-observation data, including GRACE TWS, simultaneously. While serving as a proof of concept for this approach, the focus is on simulating and evaluating hydrological dynamics in the humid Northern latitudes. Particular interest lies on the contribution of snow versus liquid water storages to overall TWS variability on different spatial (local grid-cell vs. spatially aggregated over the entire region) and temporal (mean seasonal vs. inter-annual) scales.

***Author's contribution**: design of the research in collaboration with the other co-authors; programming of the model and model calibration in the SINDBAD framework*, performing of the data analysis and preparation of the first draft of the manuscript, supported by recommendations and joint discussions with the other co-authors; corresponding author for manuscript submission and revision.*

In **Chapter 3**, the above introduced approach is extended from the Northern latitudes to the global land area. While still considering different spatio-temporal scales, the aim is to investigate the impact of how vegetation is represented on model performance and on simulated TWS composition. To this end, 2 model experiments are compared: a baseline experiment with globally uniform parameter values (as in **Chapter 2**), and an experiment in which vegetation-related parameters that define infiltration, root water uptake and transpiration processes are based on observational data (vegetation indices and estimates of rooting depth) and thus vary continuously in space and (partly) in time.

Author's contribution*: design of the research in collaboration with the other co-authors; programming and calibration of the model experiments in the SINDBAD framework*, performing of the data analysis and preparation of the first draft of the manuscript, supported by recommendations and joint discussions about the results by all co-authors; corresponding author for manuscript submission and revision.*

Chapter 4 builds on the previous findings, and investigates whether a computationally expensive river routing scheme is required when the integrated GRACE TWS is used for model calibration and validation in such a global model-data integration approach. For this purpose, we use either GRACE TWS or TWS estimates from which river storage is removed for model calibration and investigate differences in resulting simulations. Next to that, river routing is applied after model calibration, and estimated river storage added to simulated TWS. Doing so, the impact of explicitly accounting for river storage when validating modelled TWS against GRACE TWS is assessed on different spatial scales (local grid-scale, regionally and globally aggregated).

Author's contribution*: design of the research in collaboration with co-authors; programming of model experiments and transferring one of the routing schemes from Python to the SINDBAD framework*; performing the data analysis and preparation of the first draft of the manuscript, supported by recommendations and joint discussions about the results by co-authors; corresponding author for manuscript submission and revision.*

Chapter 5 provides the synthesis of the previous chapters, in which the main findings are summarized and discussed with respect to the research questions. Finally, it suggests future research possibilities and draws overall conclusions.

**The multi-criteria calibration approach, as well as the conceptual hydrological modeling concepts that are used in this thesis, have all been implemented in the SINDBAD model-data integration framework for the coupled water and carbon cycle, that was co-developed by the author as part of her work at the Max-Planck Institute for Biogeochemistry.*

Chapter 2

Understanding Terrestrial Water Storage Variations in Northern latitudes across Scales

Abstract

The GRACE satellites provide signals of total terrestrial water storage (TWS) variations over large spatial domains at seasonal to inter-annual timescales. 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 snowpacks, 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 spatiotemporal patterns and was used to assess the relative contributions of solid (snowpack) versus liquid (soil moisture, retained water) storage components to total TWS variations. In particular, we analyzed whether the same storage component dominates TWS variations at seasonal and inter-annual temporal scales, and whether the dominating component is consistent across small to large spatial scales.

Consistent with previous studies, we show that snow dynamics control seasonal TWS variations across all spatial scales in the Northern mid- to high latitudes. In contrast, we find that inter-annual variations of TWS are dominated by liquid water storages at all spatial scales. The relative contribution of snow to inter-annual TWS variations, though, increases when the spatial domain over which the storages are averaged becomes larger. This is due to a stronger spatial coherence of snow dynamics that are mainly driven by temperature, as opposed to spatially more heterogeneous liquid water anomalies, that cancel out when averaged over a larger spatial domain. The findings first highlight the effectiveness of our model–data fusion approach that jointly interprets multiple Earth observation data streams with a simple model. Secondly, they reveal that the determinants of TWS variations in snow affected Northern latitudes are scale-dependent. In particular, they seem to be not merely driven by snow variability, but rather are determined by liquid water storages on inter-annual timescales. 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.

This chapter is based on:

Trautmann, T., Koirala, S., Carvalhais, N., Eicker, A., Fink, M., Niemann, C., Jung, M. (2018): Understanding terrestrial water storage variations in northern latitudes across scales, Hydrology and Earth System Sciences, 22(7): 4061-4082, doi: 10.5194/hess-22-4061-2018

2.1 Introduction

Since the start of the mission in 2002, measurements from the Gravity Recovery and Climate Experiment (GRACE) provide unprecedented estimates of changes in the terrestrial water storage (TWS) across large spatial domains (Tapley et al. 2004; Wahr et al. 2004). Due to their global coverage and independence from surface conditions, the data represent 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 have been widely 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. 2014; 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 snowpacks (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 macroscale models rely on a priori parametrization. So far, only a few models used to assess hydrological processes on continental to global scales are constrained by observations, and if so, they are mainly calibrated against the 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 the 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 the constraint of model simulations with multiple data streams simultaneously. While this can provide a more robust understanding of how variations in water storages translate into the observed TWS (Werth and Güntner, 2010), it is very challenging in practice and has rarely been implemented.

On the one hand, this is due to the limitations and inherent uncertainties of each Earth-observation-based product that need to be considered when comparing simulations and observations. For example, satellite-based soil moisture retrievals only capture the upper 5 cm of soil under snow-free conditions and therefore are difficult to compare to modelled soil water (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).

On the other hand, 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 over-parametrization (Jakeman and Hornberger, 1993) and large computational demand, but has also been shown to not necessarily improve model performance (Orth et al. 2015). Therefore, it is desirable to implement a rather parsimonious model structure (Sorooshian et al. 1993), especially in multi-criteria model-data fusion approaches.

Applying multiple observational constraints is particularly 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 be exacerbated 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 timescales in these regions. Moreover, most analysis has so far focused 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 (SWE), evapotranspiration, and runoff into a rather simple hydrological model. This model is designed as a combination of standard model formulations yet aims to maintain low complexity in order to facilitate multi-criteria calibration and to focus on variables that can be constrained by observations.

First, we explain the applied methods, including the implemented model, the data used, 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.

2.2 Data and Methods

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

2.2.1 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°x1° latitude–longitude resolution. We defined the area of interest as humid land surface north of 40°N, excluding Greenland as well as grids with > 90 % permanent snow cover and > 50 % water fraction. Humid areas are derived based on an aridity index AI = 0.65, which was calculated as the ratio of precipitation and potential evapotranspiration (United Nations Environment, 1992). Therefore, we used the same precipitation and potential evapotranspiration data as for model forcing (see Sect. 2.3). To mask out grids with > 90 % permanent snow cover and > 50 % water fraction, we applied the SYNMAP land cover classification (Jung et al. 2006). This dataset has an original resolution of 1 km and was used to determine the fraction of land cover classes within each 1°x1° grid cell.

Forced with global observation-based climate data, the model parameters were constrained for a subset of the study domain 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 E-RUN based on E-OBS (Gudmundsson and Seneviratne, 2016). Once the model parameters were calibrated, we evaluated the model against the same data, taking into account the entire study domain. Finally, we applied the calibrated model to quantify the contributions

of snow and liquid water storages to the integrated TWS. Thereby we considered different spatial domains (local grid cell and spatially aggregated) and temporal scales (mean seasonal and inter-annual variations).

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

An overview on the experiment design and the selected time periods is provided by Fig. 2.1, while the following sections give a detailed description of the individual steps.

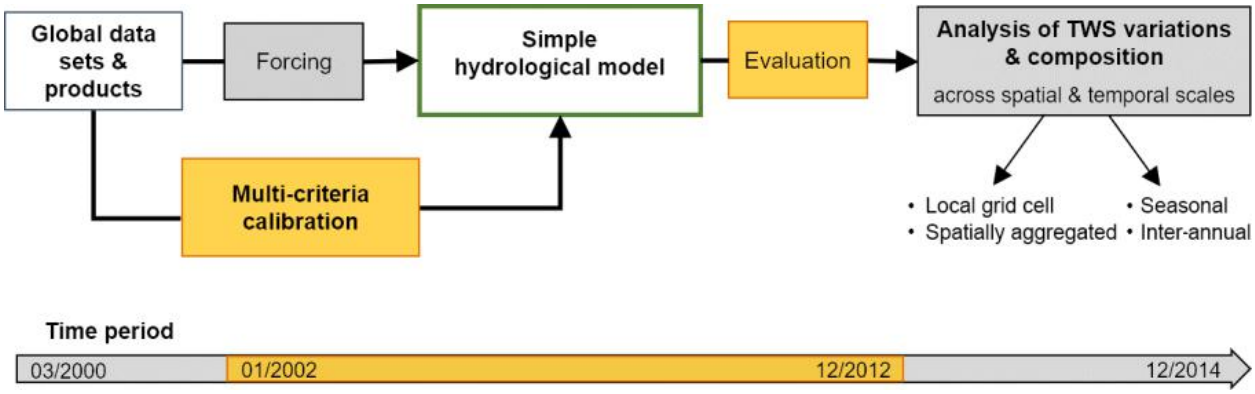

Figure 2.1. Experiment design and considered time periods for forcing and analysis (grey) as well as model calibration and evaluation (orange).

2.2.2 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 such as snowmelt, sublimation, infiltration, evapotranspiration, and (delayed) runoff and includes water storages in the snowpack, in the soil, and due to delay in runoff (Fig. 2.2). It is forced by precipitation (P), air temperature (T), and net radiation (R_n) and calculates all hydrological processes on daily time steps for individual grid cells. A simple schematic diagram of the model is shown in Fig. 2.2, while a detailed description of modelled processes is provided in Sect. A.1 in the Supplement.

In the first step, precipitation P is partitioned into liquid precipitation (rainfall) and snowfall based on a temperature threshold of 0 C. Accumulating snowfall increases the snowpack represented by the snow water equivalent (mm), which depletes by sublimation and melt if T exceeds 0 C. We calculate sublimation based on the GLEAM model (Miralles et al. 2011) and apply an extended degree-day approach to estimate snowmelt (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 snowmelt and sublimation.

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

As land runoff results from an effective soil water recharge formulation, the calculated runoff is essentially all the water that cannot be stored in the soil. Thus, it implicitly contains both surface and subsurface runoff as well as the percolation to deeper water storages such as groundwater, as well as contributions from surface water bodies. To account for runoff contributions from slow-varying storages, we calculate runoff from each grid cell (Q) by applying an exponential delay function on Q_S (Orth et al. 2013). Based on mass balance, we derive the amount of retained land runoff (RW), which implicitly accounts for the effects

of several water pools that are not explicitly represented in the model (groundwater, lakes, wetlands, and the river storage). The sum of RW and SM is then taken as the total liquid water storage (W). Frozen soil water is not explicitly included in the model. Further, the model does not account for lateral flow of water among grid cells and does not consider river routing explicitly. While the effect of the routing can be significant in large river basins of humid regions (Kim et al. 2009), it is negligible on the spatial scale of a grid cell (as also shown by small influence of the delayed storage component), and at the temporal scale of monthly aggregated values. To ensure that the model calibration is not affected by river routing, we do not compare simulated runoff to measured river discharge of large basins in our model–data fusion approach.

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

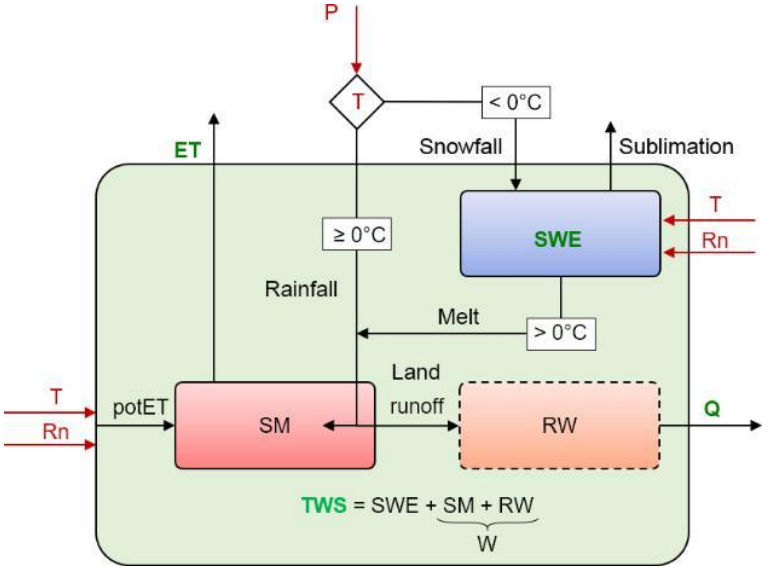

Figure 2.2. Schematic structure of the model with calculation of TWS. Boxes denote the water storages (mm): snow water equivalent SWE, soil moisture SM, retained water RW, liquid water W and total terrestrial water storage TWS. Fluxes are represented by arrows. Red color identifies forcing data: precipitation P (mm day^{-1}), air temperature T (°C), and net radiation R_n (MJ m^{-2} day^{-1}); green color indicates variables constrained by observations: evapotranspiration ET (mm day^{-1}), runoff Q (mm day^{-1}), SWE (mm), and TWS (mm).

2.2.3 Input Data

As meteorological forcing we used globally available, daily cumulated gridded precipitation sums (mm day^{-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), which combines remotely sensed precipitation 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 US 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 (TWS_{obs}) (mm), snow water equivalent (SWE_{obs}) (mm), evapotranspiration (ET_{obs}) (mm day^{-1}), and gridded runoff estimates for Europe (Q_{obs}) (mm day^{-1}).

Table 2.1. Overview on data applied for meteorological forcing and multi-criteria calibration and model evaluation (NH: Northern Hemisphere).

	Variable	Dataset	Coverage and resolution		Reference
			Spatial	Temporal	
Meteorological forcing					
P	precipitation	GPCP 1dd v1.2	1°x1° global	daily 1996–present	Huffman et al. (2000), Huffman et al. (2016)
T	air temperature	CRUNCEP v6.1	0.5°x0.5° global	daily 1901–2014	Viovy (2015)
R_n	net radiation	CERES SYN1deg Ed3A	1°x1° global	3-hourly Mar 2000–May 2015	Wielicki et al. (1996)
Calibration and evaluation					
TWS	terrestrial water storage anomalies	GRACE Tellus JPL-RL05M v2	0.5°x0.5° global	monthly 2002–2016	Watkins et al. (2015), Wiese et al. (2016b)
SWE	snow water equivalent	GlobSnow v2.0	0.25°x0.25° non-alpine NH	daily 1979–2012	Luojus et al. (2014)
ET	evapotranspiration	FLUXCOM	0.5°x0.5° global	daily 1982–2013	Tramontana et al. (2016)
Q	runoff	EU-RUN v1.1	0.5°x0.5° Europe	monthly 1950–2015	Gudmundsson and Seneviratne (2016)

TWS_{obs} 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°x3° 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 TWS_{obs} values.

To gain confidence in the 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) for the non-alpine Northern Hemisphere based on assimilating passive microwave satellite data and observed snow depth from weather stations by applying a semi-empirical snow emission model. 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 behavior of the snowpack (Derksen et al. 2014, Takala et al. 2011, Luojus et al. 2014).

The ET product is based on FLUXCOM (http://www. fluxcom.org, last access: 8 April 2016), i.e., upscaled estimates of latent energy that were derived by integrating local eddy covariance measurements of FLUXNET sites, remote sensing, and meteorological data using the Random Forest (Breiman, 2001) machine learning algorithm (Tramontana et al. 2016). In this study, we apply the Random Forest (Breiman, 2001) realization of FLUXCOM-RS + METEO (see Tramontana et al. 2016 for details). While the product captures seasonality and spatial patterns of mean annual fluxes well, predictions of inter-annual variations

remain highly uncertain (Tramontana et al. 2016). In addition, the performance of FLUXCOM ET was found to be lower in extreme environments that are not well represented by FLUXNET sites such as those in the Arctic. An underestimation of the order of 10 %–20 % of ET can be expected owing to missing energy balance correction prior to upscaling for this respective FLUXCOM ET realization. To calculate ET_{obs} (mm day^{1}), we assume a constant latent heat of vaporization of 2.45 MJ kg^{1}.

Similar to TWS that represents the vertically integrated water storage, observations of river discharge spatially integrate hydrological processes within a basin. Thus, they provide an invaluable tool for model validation at large scales. However, it 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 a 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 provide mean monthly runoff rates per unit area for each grid, so that river routing is not necessary to directly compare runoff estimates 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°x1° latitude-longitude grid and common daily (meteorological forcing) and monthly (calibration data) time steps. Data preparation further included extraction of the relevant, overlapping time period and area under consideration.

Table 2.2. 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			Eq.
				value	± uncertainty (%)		
Snow							
p_{sf}	scaling factor for snowfall	–	0–3 (∞)	0.67	±1 x 10^{-3}	(<1%)	(S2)
sn_C	minimum SWE that ensures complete snow cover of the grid	mm	0–500 (∞)	80	±19	(24 %)	(S3)
m_t	snowmelt factor for T	mm K^{-1} day^{-1}	0–10	2.63	±0.26	(10 %)	(S4)
m_r	snowmelt factor for R_n	mm MJ^{-1} day^{-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)
eta	alpha coefficient in Priestley–Taylor formula	–	0–3	1.20	±0.01	(1 %)	(S14)
et_{sup}	ET sensitivity and/or SM fraction available for ET	day^{-1}	0–1	0.02	±6 x 10^{-5}	(<1%)	(S18)
Runoff							
q_t	recession timescale for land runoff	d	0.5 (0)–100	13	±4	(31 %)	(S20)

2.2.4 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 grid cells while 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 better on rough response surfaces with discontinuities, noise, local optima, and/or outliers and is a reliable tool even for global optimization (Hansen and Kern, 2004). Additionally, the 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).

In order to keep computational demands low and to avoid overfitting by a very small sample size, we perform calibration for a subset of 1000 randomly chosen grid cells. Within this iterative 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 follows:

$$cost = \frac{\sum_{i=1}^{n} \frac{\left(x_{obs,i} - x_{mod,i}\right)^2}{\sigma_i}}{\sum_{i=1}^{n} \frac{\left(x_{obs,i} - \bar{x}_{obs,i}\right)^2}{\sigma_i}} \quad (2.1)$$

where $x_{obs;i}$ is the observed data, $\bar{x}_{obs}$ is the average of x_{obs}, and $x_{mod;i}$ is the modelled data of each space-time point i. 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 to ∞. 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 5th-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. 2.2). Since a perfect simulation would yield a total cost of 0, calibration aims to find the global minimum of $cost_{total}$.

$$cost_{total} = cost_{TWS} + cost_{SWE} + cost_{ET} + cost_{Q} \quad (2.2)$$

As the uncertainty σ of observational data in Eq. (2.1) 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 ET_{obs} and Q_{obs} on inter-annual scales, we only employed the grid's mean seasonal cycles, while the full monthly time series of gridded TWS_{obs} and SWE_{obs} were taken into account.

As ET_{obs} and Q_{obs} do not explicitly provide uncertainty estimates, we assume an uncertainty of 10 % and minimum of 0.1 mm, respectively. In order to define of TWS_{obs} 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 1 January 2002–31 December 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 SWE_{obs} and SWE_{mod} are larger than 100 mm in order to prevent the propagation of data biases to calibrated model parameters.

For maps of the temporal average uncertainties see Sect. S2.

2.2.5 Evaluation of Model Performance

Once the parameters were optimized, we applied the model for the entire study domain and evaluated its performance regarding all grid cells (6050) in terms of Pearson correlation coefficient r and root mean square error RMSE for each variable with observational data. 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. Therefore, we defined inter-annual variation (IAV) as the deviation of the monthly values from the mean seasonal cycle (MSC). As with the calibration, we focused on the common time period 2002–2012 and considered only the grid cells and time steps with available observations.

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. 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.

2.2.6 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 SWE_{mod}, W, and TWS_{mod} anomalies were calculated at local scale for each grid individually and as spatial average over all grid cells. To assess storage variability, the variance in the MSC and the IAV of each storage component was computed. Assuming negligible covariance of snow and liquid water (see Sect. S8), their relative contribution to TWS variance was calculated as the contribution ratio CR:

$$CR = \frac{var(W)}{var(TWS_{mod})} - \frac{var(SWE_{mod})}{var(TWS_{mod})} \tag{2.3}$$

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 (snowpack), with CR = +1 meaning that all variation is explained by liquid water and CR = -1 suggests determination solely by snow. From var(SWE) = var(TWS), the percentage contribution of liquid water storages to the variability of TWS can be inferred as CW:

$$CW = \frac{var(W)}{var(TWS_{mod})} = \frac{CR+1}{2} \tag{2.4}$$

As this study intends to analyze the effects of storage components on TWS at different spatial scales (local grid scale and large (regional) spatial averages), the difference in spatial heterogeneities of these components is considered. Some storage components, e.g., soil moisture anomalies, have much larger spatial variability than others. Due to this high small-scale heterogeneity, the effect on larger regional scale might be smaller than expected, as different local scale heterogeneities compensate for each other when the regional averages are calculated (Jung et al. 2017). Thus, we assessed the spatial coherence of

simulated patterns of SWE and W by calculating the proportion of total positive and total negative covariances among grid cells (Eqs. 4 and 5 in Jung et al. 2017). If the sum of positive covariances outweighs the sum of negative covariances, it implies some degree of spatial coherence of the anomalies. Spatial coherence of anomalies then causes a larger variance in the averaged anomalies compared to the sum of the variances of individual grid cells. This assessment of spatial coherence of SWE and W anomalies allows for understanding different contributions of SWE and W to TWS variability at local scale compared to the regional scale.

2.3 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.

2.3.1 Model Optimization

Optimization of the model identifies the parameter values listed in Table 2 as being most suitable regarding all data constraints simultaneously. The CMAES search algorithm converged after 3272 function evaluations as no further improvement of $costs_{total}$ could be achieved, which suggests a reliable estimate of the global optimal parameter set. The individual cost terms obtained with default and optimized parameter values are contrasted in Table A.1 in the Supplement.

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, snowfall is reduced by p_{sf} to 67 % of precipitation occurring at T < 0°C. This reduction agrees with Behrangi et al. (2016), who found that GPCP overestimated snowfall over Eurasian high latitudes by about 20 % compared to other precipitation products. Similar, overestimation of precipitation undercatch correction in GPCP has been reported by Swenson (2010). Taking into account the mismatch in temporal and spatial domains, as well as the experimental definitions, reducing GPCP snowfall in our study by 33 % is roughly consistent with both studies. Therefore, p_{sf} allows the reduction of 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 (sn_c). On the one hand, the snowpack 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. (2011). The divergence probably results from interaction with snowmelt, as net radiation also contributes to melt with 0.9 mm MJ^{-1} (m_r) if T exceeds 0°C. On the other hand, melt is mainly induced by temperature, as the estimated degree-day 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 an equifinality issue between modelled snowmelt and sublimation due to missing data constraints, resulting in larger parameter uncertainties for sn_a, m_r, and m_t. However, for the objective of this study it is not primarily relevant whether sublimation- or radiation-induced melt decreases the snowpack, as the total snow loss amount remains relatively unchanged for different parameter combinations. The maximum soil water holding capacity is set to 515 mm after calibration, a comparatively high value that is likely to include storages in surface water bodies such as lakes and wetlands within our study domain. The optimized value of s_{exp} is 1.46, which suggests a non-linear relationship between soil moisture storage and runoff generation. For the same amount of incoming water (rainfall and snowmelt), the non-linear relationship produces a smaller runoff and larger infiltration than a linear relationship (s_{exp} = 1).

Regarding evapotranspiration, the alpha coefficient (et_a) in the Priestley–Taylor formula is generally taken as 1.26 for well-watered crops based on experimental observations (Priestley and Taylor, 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 to 0.5 day^{-1} and is close to the median value (5 %) (Teuling et al. 2006).

Finally, the calibrated recession timescale that delays land runoff is 13 days (q_t). Compared to much smaller alpine catchments for which Orth et al. (2013) reported q_t of 2 days, the longer delay coefficients are reasonable at a spatial resolution of 1°x1° grids, because the elevation gradients are much smaller within a large spatial area. 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, which is recharged with some proportion of land runoff and linearly depletes over time, led to similar retardation times.

The uncertainty in the optimized parameter vector was estimated by quantifying each parameter's standard error as the square root of the product between the diagonal elements of the parameters' covariance matrix (calculated from the Jacobian matrix) and the sum of residual squares according to Omlin and Reichert (1999) and Draper and Smith (1981). The resulting relative parameter uncertainty is particularly instructive for comparing how well individual parameters could be constrained.

Most parameters were well constrained (Table 2), suggesting that our model–data fusion method, fed by multiple observation streams, succeeded in reducing 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.

We adopted the calibrated parameter values as 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 in the modeling 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 (Sood and Smakhtin, 2015, Bierkens et al. 2015), macro-scale models mostly apply a priori parameters based on empirical values or on expert knowledge, which may yet lead to suboptimal simulations (Beck et al. 2016, Sood and Smakhtin, 2015).

2.3.2 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 mean seasonal variations of ET and Q, respectively. The median RMSE of mean seasonal ET is 11 and 9.5 mm $month^{-1}$ for Q, which represent 15 % resp. 17 % of the average observed annual amplitude. At the inter-annual scale, though, larger discrepancies exist, which at least partly arise from larger uncertainties in ET_{obs} and Q_{obs} (Sect. S4). Thus, we assume high confidence in modelled ET and Q fluxes and subsequently focus on evaluation of the water storages TWS and SWE.

2.3.2.1 Performance on local Grid-Scale

Overall, the model performs well compared to the observations of monthly time series of SWE and TWS (Fig. 2.3). More than half of the grid cells obtain correlation values higher than 0.74 between SWE_{obs} and SWE_{mod}. In general, the median RMSE is 20 mm, which is smaller than the average uncertainty of 35 mm in SWE_{obs}. 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 or 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, which may be significant in near-mountain or coastal regions. Additionally, the accuracy of observed large snow accumulation is limited as the radar-retrieval methods tend to saturate at large SWE_{obs} 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 TWS_{obs} and TWS_{mod}, 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 TWS_{obs} of the study domain (Wiese, 2015). However, when comparing GRACE TWS with model simulations, several aspects have to be considered. First, TWS_{obs} 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 preprocessing steps, the models that account for postglacial rebound or leakage between neighboring grid cells, for example, 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. 2016a). Altogether this 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 are both relatively dry regions that are rather dominated by inter-annual TWS variations (Humphrey et al. 2016). Discrepancies between TWS_{obs} and TWS_{mod} thus relate to a low signal-to-noise ratio in TWS_{obs} due to small seasonal TWS variations. However, 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.

2.3.2.2 Performance of the spatially integrated Simulations

Since the aim of this study is to analyze the composition of TWS across temporal scales, we additionally evaluated average (spatially integrated) MSC and IAV of SWE and TWS (Fig. 2.4). 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. 2014, Schellekens et al. 2017, Zhang et al. 2017), their inter-annual variations are more uncertain and contain considerable noise. This clearly reduces the information content in the observational data, so that we evaluate the IAV in more qualitative terms.

As with the comparison at grid scale, the spatially averaged SWE_{mod} compares well to SWE_{obs}, with a correlation of 0.95 suggesting a good reproduction of seasonal snow accumulation and ablation processes (Fig. 2.4a). Owing to the high uncertainty of SWE_{obs} peaks due to signal saturation, the higher amplitude of SWE_{mod} seems reasonable. Although inter-annual variations are not as well represented as the MSC, general tendencies, e.g., increasing/decreasing positive/negative anomalies, coincide.

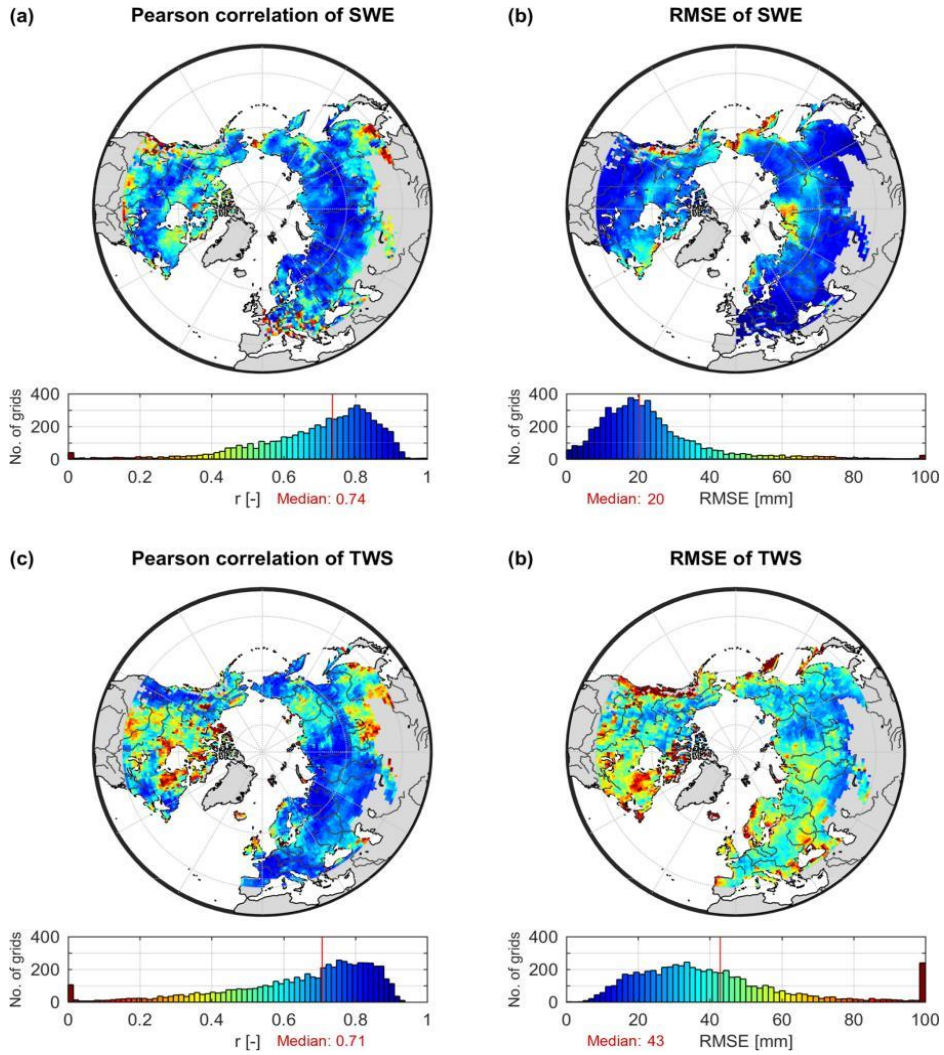

Figure 2.3. Pearson correlation coefficient r (a, c) and root mean square error RMSE (b, d) between monthly values of modelled SWE and GlobSnow SWE (a, b), as well as modelled TWS and GRACE TWS (c, d) for the period 2002–2012 and for each 1 1 grid cell of the study domain. Values of r are truncated to the range 0–1 (a, c), and values of RMSE to the range 0–100 mm (b, d).

Similar to SWE, the spatial average of TWS shows high correlation of 0.91 for seasonal variations, with positive anomalies from December to May–June and negative anomalies during summer and autumn months (Fig. 2.4b). Even though the modelled amplitude is slightly larger than the observed one, it stays within the uncertainty range of TWS_{obs} for most months, suggesting reliable simulations. However, TWS_{mod} precedes TWS_{obs} on average by 1 month, reaching the maximum in March instead of April, and the minimum in August instead of September. A similar phase shift of 1 month between GRACE TWS and modelled TWS has been reported by several state-of-the-art global models (Döll et al. 2014, Schellekens et al. 2017). It should be noted that some areas such as eastern North America, Kamchatka, Scandinavia, and western Europe do not show phase differences, while the lag in south-eastern Eurasia is even larger, as already suggested by lower overall correlation (Fig B.5). In general, the disagreement in timing is attributed to the lack of sufficient water storages and the delay mechanism within the model, so that the modelled system reacts too fast (Schellekens et al. 2017, Döll et al. 2014, Schmidt et al. 2008b). 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 as to whether it is not primarily the model formulation that prevents correction of the temporal delay, but rather the combination of forcing data and 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 behavior 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 (Sect. S6). Further, the lag in TWS simulation can occur due to several mechanisms and processes that are not yet considered in the current model structure, such as lateral flow and surface storages (wetland and lakes), vegetation processes, glacier melt, and human influence with dams and reservoirs. However, we do not observe a general or systematic relationship with either elevation, land cover type, soil properties, or the occurrence of lakes and wetlands. There is a tendency that larger negative lags occur more frequently in regions with sporadic permafrost, but the ranges of permafrost fractions are large for both short and long lags in TWS, suggesting a complex interaction between permafrost extent and its effect on lag in seasonal TWS dynamics. Finally, potential biases in the timing of ET due to snow cover and/or vegetation processes may also affect the timing of the depletion of SM and TWS. Additionally, high uncertainties of the precipitation forcing and GlobSnow SWE product in (near-)mountain regions, as well as leakage errors in the GRACE signal influence the accuracy of both TWS_{obs} and TWS_{mod}. Although these shortcomings should be kept in mind, we assumed that they do not significantly affect our results regarding to the relative contributions of snow and liquid water to TWS.

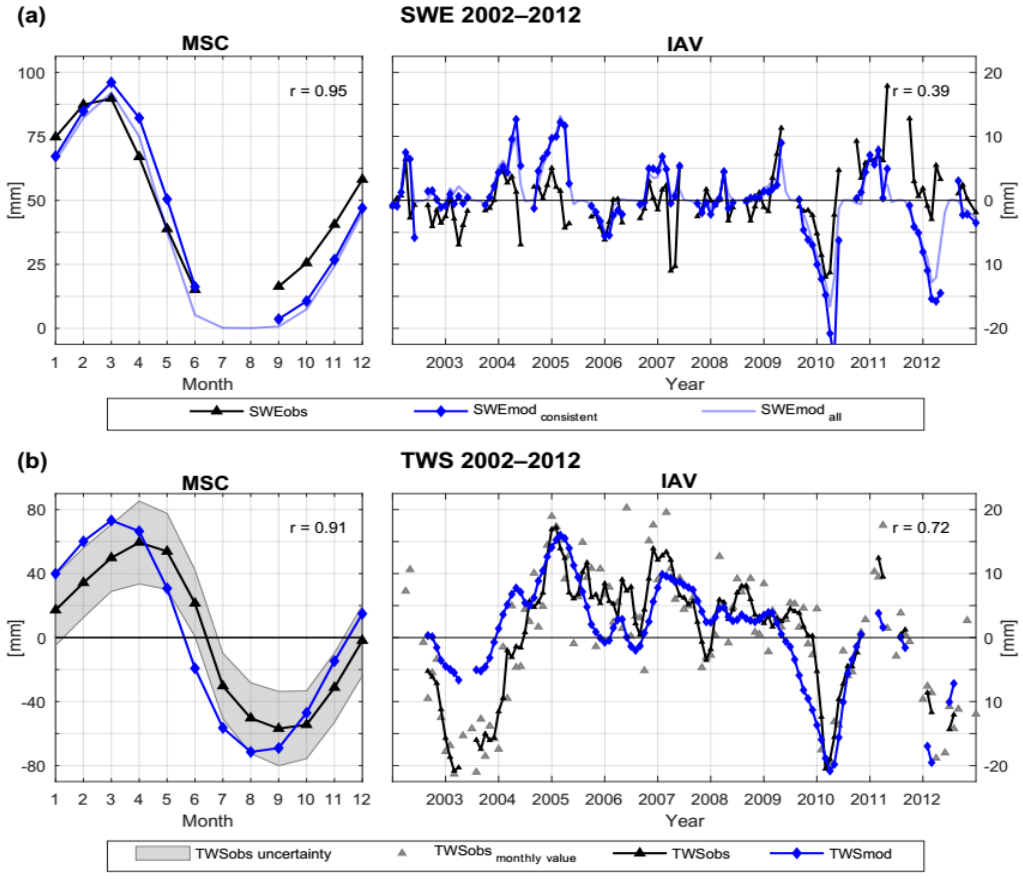

Figure 2.4. Spatially averaged mean seasonal cycle (MSC) of the period 2002–2012 as well as inter-annual variability (IAV, difference between monthly values and the MSC) for (a) SWE and (b) TWS. In (a), SWE_{mod} consistent refers to modelled SWE considering only data points with available SWE_{obs}, while SWE $_{mod}$ all incorporates all time steps for all grids of the study domain. Correlation r is calculated only for consistent data points. In (b) IAV, TWS_{obs} monthly value shows the original IAV of individual TWS_{obs} months, while TWSobs and TWS_{mod} 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 TWS_{obs} values is highly underestimated by modelled TWS, which on the one hand relates to noise within the GRACE signal, but on the other hand may again reflect missing process representation in the model. To reduce the noise, we applied a 3-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 (Fig. 2.4b, solid lines). Only the amplitude of the large negative anomaly in 2003 is not captured by the model. While the spatial pattern of this negative TWS anomaly can be simulated, the forcing data do not show large anomalies in 2003, so that the model fails to reproduce the magnitude of observed TWS, especially in North America. Issues with the precipitation forcing are further suggested by a negative SWE anomaly of on average 5 mm (see Fig. 2.4a), indicated in the GlobSnow data, that is not captured by the model, either. The reason why this snow anomaly is not captured by the forcing remains unclear at this point – it persists when using the WFDEI forcing dataset. Besides, the agreement between GRACE and modelled TWS IAV gets substantially better when isolating inter-annual variations by removing the trends in both TWS time series (increase in correlation r from to 0.77). This suggests that the trend in GRACE TWS is to some extent either subject to observational issues or represents a process that is not captured by the model.

2.3.2.3 Comparison with the eartH2Observe Model Ensemble

Compared to the model ensemble of the eartH2Observe dataset, we achieve an equally good or better performance for the spatially integrated SWE and TWS on both MSC and IAV scales (Figs. 2.5 and S6). 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 shows a better correlation than our model (r = 0:95). However, modelled snow accumulation exceeds that of SWE_{obs} for the majority of the models, which is also reflected in higher RMSE (Figs. S6-S8). On IAV scales, the correlation is lower in general, 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 SWE_{obs} and SWE_{mod} represent model deficiencies or evolve from issues related to the GlobSnow SWE retrieval (Schellekens et al. 2017).

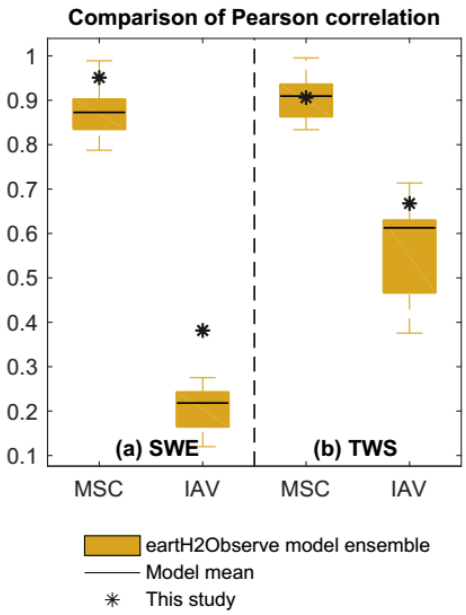

Figure 2.5. Pearson correlation for the spatially integrated SWE (a) and TWS (b) achieved by this study compared to the model ensemble of the eartH2Observe dataset across temporal scales. In each box, the edges represent the 25 % and 75 % percentiles of the model ensemble, while the solid black line marks the performance of the ensemble mean.

Regarding average seasonal TWS variations, our model performs as well as the model ensemble (r D 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 TWS_{mod} (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 TWS_{obs}. This discrepancy is reflected in relatively high RMSE values for SWBM (Fig B.8). The model ensemble precedes observed seasonal TWS variations by 1 to 1.4 months, similar to our estimates of TWS_{mod} (-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 the preceding and succeeding of TWS_{mod} 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 obtain comparable spatial pattern of preceding TWS, with small differences at regional scales. Even though the difference in the MSC is commonly attributed to the lack or inadequate size and number of water storages (Kim et al. 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 to simpler models, such as SWBM and our model (-1.0 and -1.1 months, respectively). Since Schellekens et al. (2017) found the largest phase differences in cold regions, they postulate the implementation of processes associated with snow as an 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.

Although our modeling framework assimilates information from more data streams compared to the model simulations in the eartH2Observe ensemble, e.g., GRACE and GlobSnow data, we only used a subset of 1000 random grid cells to constrain the model parameters. Despite this, our model performs better than the eartH2Observe ensemble over the whole domain (6050 grids). This improvement in model performance is also consistent among several modelled variables and not limited to storage components only. This suggests that remote sensing data, with larger spatial coverage than site measurements, have large potential to improve hydrological simulations over a large domain. In addition, remote sensing data also hold potential beyond their use as an observational constraint and can provide information on identifying and formulating functional relationships across several spatial and temporal scales, which should be addressed in future efforts.

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 analyze the composition of TWS across spatial and temporal scales.

2.3.3 TWS Variation and Composition

2.3.3.1 Spatially integrated

To assess the average composition of seasonal and inter-annual TWS variations, we spatially integrated modelled TWS anomalies as well as modelled anomalies of snow (SWE) and liquid water storages (W) across all grids of the study domain (Fig. 2.6).

Regarding the MSC, all water storages show a clear seasonal pattern. The maximum TWS_{mod} in March coincides with the maximum in SWE_{mod}. On the contrary, W remains nearly constant throughout the winter, related to the lack of evapotranspiration losses and missing infiltration due to prevailing solid precipitation. Starting from March, snowmelt decreases SWE_{mod}, and thus TWS_{mod}, progressively. Thereby TWS_{mod} 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 summer-time evapotranspiration, and preceding runoff of temporarily stored water. With W and SWE_{mod} being at their minimum in August–

September, TWS_{mod} 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 SWE_{mod}). Despite the interplay of SWE_{mod} and W on seasonal variations of the integrated TWS_{mod}, the amplitude of W (62 mm) is considerably lower than the one of SWE_{mod} (92 mm) and TWS_{mod} (144 mm). Thus, the seasonal accumulation of snow largely determines the magnitude of TWS_{mod}. Additionally, anomalies at least partly result from snowmelt, whereas liquid water does not influence the snowpack. 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 CR (Eq. 2.3) based on the spatially averaged MSC underlines this, as CR D 0:26 indicates that variations in SWE_{mod} explain 63 % of seasonal TWS_{mod} variability (CW D 37 %). This agrees with previous findings of Güntner et al. (2007), who found that SWE contributes to 68 % of seasonal TWS variations in cold regions using the WaterGAP model.

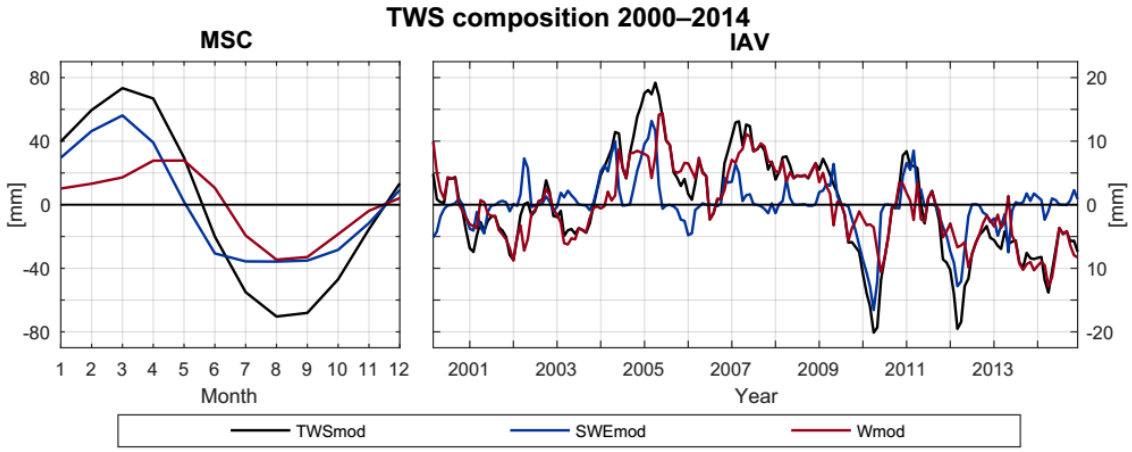

Figure 2.6. Spatially averaged mean seasonal cycle (MSC) of the period 2000–2014 as well as inter-annual variability (IAV, difference between monthly values and the MSC) for modelled TWS, SWE, and W anomalies to the time-mean of 2000–2014.

On IAV scales, the pattern seems less homogeneous (Fig. 2.6). In contrast to the MSC, CR D 0:25 suggests a larger influence from liquid water anomalies (CW = 62.5 %) than snow anomalies on inter-annual TWS variations. Thereby, we found the main contributor to TWS_{mod} anomalies being dependent on the phase of previous precipitation anomalies, in that they define whether snowfall anomalies lead to anomalies in the SWE_{mod}, 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 TWS_{mod} anomalies, such as in 2005, 2010, and 2012, follow anomalies in wintertime precipitation and go along with anomalies in SWE_{mod} (Fig. 2.6). On the contrary, summertime anomalies related to W are usually less pronounced in their magnitude (e.g., 2003, 2006). We attribute this to the 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, the largest TWS_{mod} anomalies are obtained in early spring before snowmelt starts and when snow accumulation is highest. Nevertheless, since W is influenced by the quantity of snowmelt, anomalies in SWE_{mod} implicate subsequent changes in W. Additionally, snowpack anomalies are eliminated each summer, while W anomalies dominate the summer. As a result, W anomalies in any case affect TWS_{mod} variability on IAV scales when analyzing the spatial average composition.

Besides, Güntner et al. (2007) demonstrated a shift from short-term storages with high seasonality such as SWE on MSC scales towards storages with longer delay times on IAV scales. Although modelled W mainly represents soil moisture, it implicitly includes surface water and groundwater storages. Thus, its contribution of CW = 62.5 % to inter-annual TWS variations is roughly comparable to 55 % contribution

from soil moisture (27 %) and surface water (28 %) in cold regions, found by Güntner et al. (2007). Despite this, the relatively large influence of surface water bodies shown by Güntner et al. (2007) suggests that the lack of explicit surface water storages in this study may contribute to the remaining discrepancies with GRACE and the lower magnitude of modelled inter-annual TWS variability compared to GRACE estimates.

2.3.3.2 Local Grid-Scale

Based on CR (Eq. 2.3), Fig. 2.7 shows the relative contribution of SWE_{mod} and W variances to total TWS_{mod} variability on MSC and IAV timescales for each grid. Thereby, blue colors represent prevailing SWE_{mod} variations as indicated by CR < 0, while red colors show the dominance of variations in W (CR > 0).

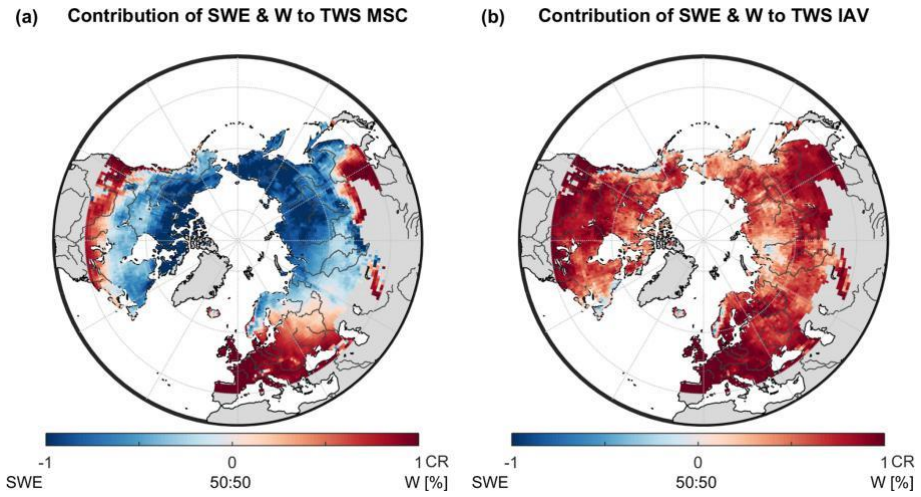

Figure 2.7. Relative contribution based on CR (Eq. 2.3) of modelled snow (SWE) and liquid water (W) storage anomalies to (a) mean seasonal variations from 2000 to 2014 of modelled TWS anomalies, and (b) inter-annual variations of modelled TWS anomalies for each grid cell of the study domain.

Accordingly, variations in the MSC of TWS_{mod} are mainly influenced by snow in northern regions, with the mean CR = -0.30 indicating that on average 65 % of seasonal TWS_{mod} variability can be explained by SWE_{mod} (CW = 35 %) (Fig. 2.7a). The contribution of the variation in liquid water in general increases southwards and prevails seasonal TWS_{mod} variability south of approximately 50 latitudes. An exception is Europe, where the influence of W reaches up to 60 latitudes, 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 showed the dominance of snow in northern latitudes in North America (Rangelova et al. 2007), and prevailing soil moisture dynamics further south, e.g., in the Mississippi basin (Ngo-Duc et al. 2007, Güntner et al. 2007). Besides, the north extent of dominating W reflects the temperature gradient in North America and Eurasia. Comparison with average annual temperature suggests that for T > 10°C variability of W dominates, while for T < 0°C SWE_{mod} dynamics prevail. This is plausible, as temperature determines annual snow accumulation, and the relative contribution of liquid water increases in the absence of snow. Yet, it further highlights the dependency on the temperature dataset used, especially in a model that calculates snowfall and snowmelt based on temperature thresholds as ours does.

Opposed to the MSC, the variability of W dominates TWS_{mod} variations on IAV scales in the entire study domain, as is clearly indicated by average CR D 0:63 (Fig. 2.7b). Inter-annual variations of SWE_{mod} seem to be relevant only in regions that receive the highest annual snow amounts, such as the Canadian Arctic Archipelago, the northern west coast of North America, north-eastern Siberia, and the northern West Siberian Plain. Due to a prolonged cold period there, the time span with rainfall and evapotranspiration is short, decreasing the occurrence of potential variability in W. However, even in these regions the

influence of SWE_{mod} is low compared to the MSC. This reduced importance of snow on inter-annual scales again agrees with the findings of Güntner et al. (2007).

Apart from that, and since we already showed a link between average TWS_{mod} IAV and previous precipitation anomalies, and as precipitation represents the main model forcing data, we investigated the relative contribution of rainfall and snowfall to inter-annual variability of total precipitation (Fig. 2.8). Similar to the composition of TWS_{mod} on local scale, rain anomalies prevail for most of the grid cells (mean CR D 0:68). This is consistent when snowfall is not scaled by p_{sf} and thus suggests that the greater contribution of W to inter-annual variations of TWS_{mod} on local scale relates to highly variable (liquid) summertime precipitation events. On the contrary, monthly snowfall seems less variable between years, resulting in less pronounced variations in SWE_{mod} compared to W. Exceptions are regions of high maximum SWE_{mod}, which accordingly show a considerable relative contribution of snow to the inter-annual TWS_{mod} variability.

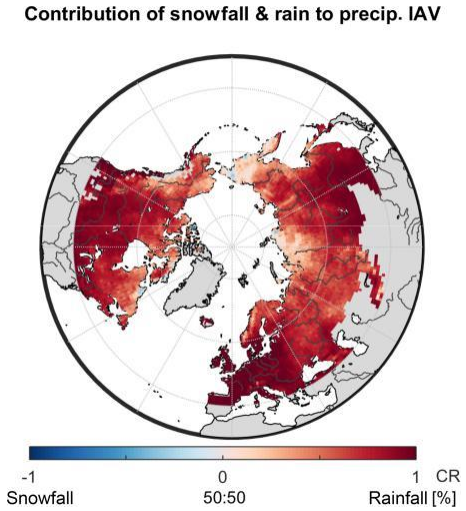

Figure 2.8. Relative contribution based on CR (Eq. 2.3) of modelled snowfall and rainfall to total precipitation (P) anomalies on inter-annual (IAV) scales for each grid cell of the study domain.

2.3.3.3 Comparison of different Scales

Figure 2.9 summarizes the above-presented contributions to TWS_{mod} variability across spatial and temporal scales. As explained in the previous sections, we obtained two scale-dependent differences in the relative contribution to TWS_{mod} variability: (1) in general between temporal scales, and for inter-annual variability between spatial scales. Regarding (1), Fig. 2.9 emphasizes again that seasonal variations of TWS_{mod} are mostly determined by seasonal snow dynamics, while on inter-annual scales TWS_{mod} variability mainly originates from variations in liquid water. As previously stated, determination by SWE_{mod} 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).

Concerning IAV scales, we found that the determination of TWS_{mod} variability by W relates to larger inter-annual variations in liquid precipitation compared to snowfall. However, considerable differences between spatial scales exist (Fig. 2.9). Opposed to the MSC, the spatially integrated CR (black star) for IAV is not within the interquartile range of local CR. This indicates a relatively larger effect of SWE_{mod} variations when looking on the spatially integrated values compared to local values. Liquid water storages are highly heterogeneous in space, mainly due to heterogeneous rainfall anomalies. On the contrary, snow variability is affected by fewer factors, and mainly regulated by a range of temperature values that control freezing and melting. Temperature anomalies in turn show sizeable spatial coherence across large

areas. To assess the spatial coherence of W compared to SWE_{mod}, we calculated the proportion of total positive and total negative covariances among grid cells (Fig. 2.10).

For inter-annual variations of SWE_{mod}, the sum of positive covariances prevails (Fig. 2.10a), whereas positive and negative values are more in balance for W (Fig. 2.10b). This suggests that SWE_{mod} anomalies are more spatially coherent than anomalies of W. Thus, when spatially averaging, the more homogeneous snow anomaly patterns gain importance. On the contrary, opposed anomalies of W compensate for each other more strongly. This leads to a relatively larger influence of SWE_{mod} to the spatially integrated inter-annual TWS_{mod} variability compared to when analyzing the local grid scale. Since positive covariation clearly dominates for temperature anomalies, the spatial coherence of SWE_{mod} relates to their homogeneous patterns (Fig. 2.10c). Similar to W, positive covariances only slightly outweigh year-to-year variations in rainfall (Fig. 2.10d). The same is true for snowfall (not shown). Therefore, the spatial coherence of SWE_{mod} anomalies is less pronounced than for temperature, as snow is additionally influenced by snowfall anomalies. Regarding anomalies, this indicates that the spatial heterogeneity in our model, which misses explicit information on soils, topography, etc., mainly results from inhomogeneous patterns in rainfall anomalies. Thereby, the slightly more balanced positive and negative covariations for W compared to rainfall can be ascribed to the additional impact of primarily radiation-driven evapotranspiration to SM. 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 snowfall estimates of the WFDEI product (Weedon et al. 2014). Since we observed the same patterns, we assume our findings to be robust (Sect. S7.1).

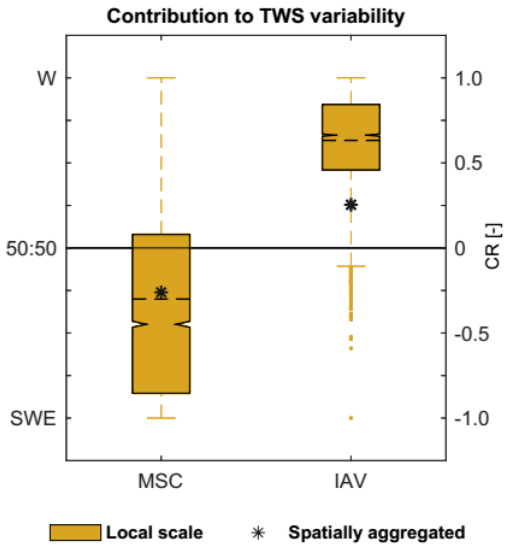

Figure 2.9. Relative contribution of snow (SWE) and liquid water (W) to TWS variability on different spatial (local grid scale, spatially integrated) and temporal (mean seasonal MSC, inter-annual IAV) scales based on CR (Eq. 2.3). The box plots represent the distribution of grid cell CR, with the dashed line marking the corresponding average. The star represents the CR calculated for the spatially integrated values.

2.3.4 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 are associated with limitations in terms of process understanding and predictive power.

First of all, the simple structure only allows inferences on represented processes, which likely include effects of fluxes and storages not considered explicitly. For example, the model does not resolve individual liquid water storages such as deep groundwater and surface water explicitly. As discussed previously, our delayed land runoff comprises various (intermediate) storages and delay times, and thus cannot be associated with one distinct storage component. Even though soil moisture is distinguished from these

slowly varying reservoirs, its quantity and pattern have not been directly validated. Future research is required to increase confidence by including remote-sensing-based data of soil moisture in calibration and/or validation. However, these satellite data still have limited value as the microwave signals can only capture moisture in the upper 5 cm of soil and do not provide estimates under snow cover and dense vegetation (Döll et al. 2015, Lettenmaier et al. 2015). Therefore, a multi-layer soil scheme is needed to compare model outputs to satellite-derived soil moisture estimates, as was successfully demonstrated by Albergel et al. (2017) for example.

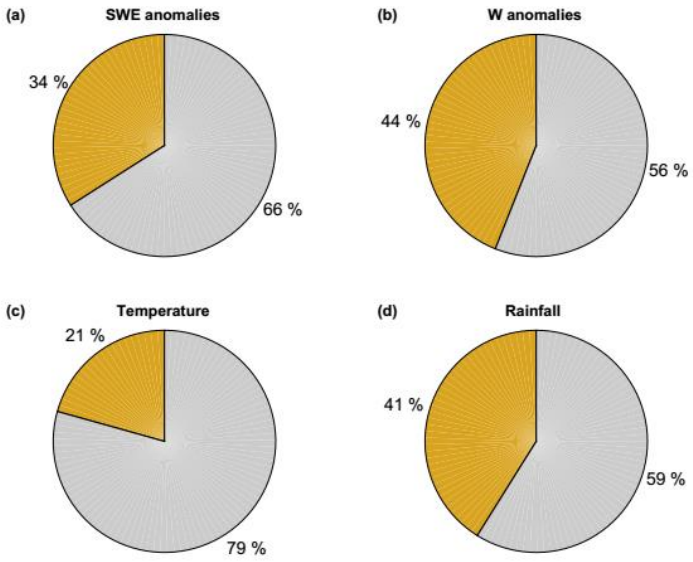

Figure 2.10. Proportion of total positive (grey) and negative (orange) covariances among grid cells for inter-annual variations of (a) snow (SWE), (b) liquid water storages (W), (c) temperature, and (d) rainfall.

Further, the model does not include any human-induced changes in 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 snowfall, liquid water capacity of snow, interception, freeze–thaw dynamics within the soil, capillary rise, and other surface-groundwater interactions, the effect of vegetation or other surface properties, and lateral flow from one grid cell to another. In the downstream areas of large basins especially, the latter represents a potential input that may significantly affect total TWS (Kim et al. 2009) and thus may contribute to the discrepancy between TWS_{obs} and TWS_{mod} in some regions. Besides, the model does not account for spatial variability of topography and land surface characteristics.

With regards to model parameters, 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). Our model only implicitly considers the effects of vegetation, for example on ET, but not its spatial variability, as the associated impacts are included in the observational constraint. Spatial variability of model parameters might affect the relative contributions of different storage components to TWS variability at different spatial scales. However, the comparison with eartH2Observe models, which generally involve spatial heterogeneity in model parameters, suggests that the main conclusions remain unchanged. Additionally, we want to highlight that the spatial distribution of model parameters depends on assumptions and some degree of simplification as well and thus does not necessarily improve model performance compared to a global uniform parameter set obtained from multiple observational data.

Further, as we encountered issues with parameter equifinality, especially between modelled snowmelt and sublimation, future efforts should include a stronger utilization of runoff data in the calibration and validation process. This would help to better constrain water fluxes to the atmosphere and liquid water fluxes, which can contribute to the runoff.

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

2.4 Conclusion

In this study, we assessed the relative contributions of snowpack 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 snowpack estimates from GlobSnow. The optimized model showed considerably good agreement with observed patterns of hydrological fluxes and states, and was found to perform comparably to 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 snowpacks 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 counterintuitive pattern was found to relate to larger rainfall anomalies compared to snowfall anomalies.

Apart from the timescale-dependent dominant controls on TWS variations, we additionally observed different behavior 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 snowpack 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 snowpack 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 snowpack 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.

Chapter 3

The Importance of Vegetation in Understanding Terrestrial Water Storage Variations

Abstract

So far, various studies have aimed at decomposing the integrated terrestrial water storage variations observed by satellite gravimetry (GRACE, GRACE-FO) with the help of large-scale hydrological models. While the results of the storage decomposition depend on model structure, little attention has been given to the impact of the way that vegetation is represented in these models. Although vegetation structure and activity represent the crucial link between water, carbon, and energy cycles, their representation in large-scale hydro-logical models remains a major source of uncertainty. At the same time, the increasing availability and quality of Earth-observation-based vegetation data provide valuable information with good prospects for improving model simulations and gaining better insights into the role of vegetation within the global water cycle.

In this study, we use observation-based vegetation information such as vegetation indices and rooting depths for spatializing the parameters of a simple global hydrological model to define infiltration, root water uptake, and transpiration processes. The parameters are further constrained by considering observations of terrestrial water storage anomalies (TWS), soil moisture, evapotranspiration (ET) and gridded runoff (Q) estimates in a multi-criteria calibration approach. We assess the implications of including varying vegetation characteristics on the simulation results, with a particular focus on the partitioning between water storage components. To isolate the effect of vegetation, we compare a model experiment in which vegetation parameters vary in space and time to a baseline experiment in which all parameters are calibrated as static, globally uniform values.

Both experiments show good overall performance, but explicitly including varying vegetation data leads to even better performance and more physically plausible parameter values. The largest improvements regarding TWS and ET are seen in supply-limited (semi-arid) regions and in the tropics, whereas Q simulations improve mainly in northern latitudes. While the total fluxes and storages are similar, ac-counting for vegetation substantially changes the contributions of different soil water storage components to the TWS variations. This suggests an important role of the representation of vegetation in hydrological models for interpreting TWS variations. Our simulations further indicate a major effect of deeper moisture storages and groundwater–soil moisture-vegetation interactions as a key to understanding TWS variations. We highlight the need for further observations to identify the adequate model structure rather than only model parameters for a reasonable representation and interpretation of vegetation-water interactions.

This chapter is based on:

Trautmann, T., Koirala, S., Carvalhais, N., Güntner, A., Jung, M. (2022): The importance of vegetation in understanding terrestrial water storage variations, Hydrology and Earth System Sciences, 26 (4): 1089-1109, doi:10.5194/hess-26-1089-2022

3.1 Introduction

Since 2002, the Gravity Recovery and Climate Experiment (GRACE) mission has facilitated global monitoring of terrestrial water storage (TWS) variations from space – a milestone of global hydrology (Rodell, 2004, Famiglietti and Rodell, 2013). Observed TWS variations from GRACE have since become a cornerstone for diagnosing trends in water resources due to climate change or anthropogenic activities (Rodell et al. 2018, Reager et al. 2015, Scanlon et al. 2018, Syed et al. 2009, Tapley et al. 2019), as well as for benchmarking and improving global hydrological models (GHMs) (Scanlon et al. 2016, Döll et al. 2014, Werth et al. 2009, Zhang et al. 2017, Kumar et al. 2016, Eicker et al. 2014). Significant covariations between GRACE TWS and the global land carbon sink (Humphrey et al. 2018) and surface temperatures (Humphrey et al. 2021) highlight the importance of the water cycle as the nexus in the Earth system.

However, GRACE TWS estimates represent a vertically integrated signal of all water stored in snow, ice, soil, surface water, and groundwater. Thus, understanding processes and mechanisms of TWS variations requires attribution of TWS variations to individual storage components. Despite advancements in remote sensing, largescale quantification of these components based on observations remains challenging. For example, remote sensing-based estimates of soil moisture only capture depths up to 5 cm and do not necessarily reflect the moisture availability in the deeper soil column (Dorigo et al. 2015). While these observations can be extrapolated to derive estimates of root zone moisture, either by using statistical relationships (Zhuang et al. 2020) or by data assimilation into land surface models (Reichle et al. 2017, Martens et al. 2017), such products rely on many assumptions. Therefore, GHMs have been necessary to interpret TWS variations in terms of contributions by snow, soil moisture, groundwater, or surface water. However, several studies suggested that current state-of-the-art GHMs cannot reproduce key patterns of observed TWS variations and show partly diverging TWS partitioning (Scanlon et al. 2018, Schellekens et al. 2017, Zhang et al. 2017, Kraft et al. 2021). This uncertainty of the available tools to interpret TWS variations is clearly a major obstacle for diagnosing and understanding global changes of the water cycle.

To improve model performance and reliability, GHMs are traditionally calibrated against measured discharge time series at the outlet of catchments (Müller-Schmied et al. 2021, Telteu et al. 2021). However, discharge provides an integrated response of an entire catchment with very limited information on the interplay of different processes and spatial heterogeneities. In fact, the use of spatio-temporal data, for example, from remote sensing, has been suggested for model calibration (Su et al. 2020). While using spatio-temporal vegetation data, for example, the normalized difference vegetation index (NDVI) seemed promising for this at the catchment scale (Ruiz-Perez et al. 2017), many GHMs still have a limited usage of such data in their modeling approach. Some large-scale studies have shown clear improvements in model performance when a larger number of observational constraints are used to constrain the model parameters, especially when using TWS variations from GRACE (e.g., Lo et al. 2010, Rakovec et al. 2016, Bai et al. 2018, Mostafaie et al. 2018, Trautmann, 2018). Among them, Trautmann et al. (2018) contributed insights into the drivers of TWS variations across spatial and temporal scales in northern high latitudes, in particular with respect to contributions by snow vs. liquid water storages. In this study, we follow a similar framework of using multiple observational data streams to constrain a simple hydrological model to understand the role of varying vegetation characteristics for the partitioning of TWS components at the global scale.

Among liquid water storages, especially the differentiation between soil moisture and groundwater poses a challenge. Reflecting on the determinants of rather shallow soil moisture vs. deeper groundwater storage variations, it is apparent that under most conditions, the soil moisture state itself is the first order control valve. In particular, it determines the amount of water that is available for soil water uptake for evapotranspiration but also for percolation into deeper soil layers and consequently recharge into the groundwater storage. The two key processes that shape soil moisture dynamics, infiltration and evapotranspiration (ET), are strongly mediated by the presence and properties of vegetation (Wang et al. 2018). For example, vegetation promotes infiltration over surface runoff due to larger surface roughness, dampened precipitation intensities, and more soil macropores due to rooting and biological activity. In fact, such roles of vegetation in a global climate model were already envisioned and evaluated almost 4 decades ago (Rind, 1984). In addition, vegetation alters soil properties like soil texture and organic matter

content. Such soil properties together with rooting depth control the size of the soil moisture reservoir that is available for ET and how plants respond to drought stress conditions (Baldocchi et al. 2021, Yang et al. 2020). Furthermore, deep roots may connect to the groundwater and provide access to the deeper moisture storages and thus have wider implications on the hydrological cycle. Rooting depth is species-specific and, in addition, determined by the infiltration depth and groundwater table depth and thus has a high spatial heterogeneity both across the globe and at the local scale (Fan et al. 2017). The significance of interactions between vegetation and soil moisture is at the heart of ecohydrology (Rodriguez-Iturbe et al. 2001) and has become evident in many theoretical and experimental studies. Many studies analysed the effects of water availability on vegetation functioning (Porporato et al. 2004, Reyer et al. 2013, Wang et al. 2001, Yang et al. 2014) and the effect of changing vegetation cover on ecosystem water consumption (Du et al. 2021). While large-scale hydrologic models usually apply simplified and static vegetation characteristics (Quevedo et al. 2008, Weiss et al. 2012, Telteu et al. 2021), spatio-temporal variations of vegetation pattern are vital for good predictions of available water resources (Andersen et al. 2010). On the ecosystem scale, Xu et al. (2016) showed the advantage of accounting for different plant hydraulic traits in an ecosystem model. And on a global scale, Weiss et al. (2012), for instance, showed the positive influence on modelled evaporation when static vegetation characteristics are replaced by monthly LAI (leaf area index) estimates in a climate model. However, how the representation of vegetation affects global water storages and in particular the partitioning of TWS in large-scale hydrological models has received surprisingly little attention so far.

Therefore, the objective of this study is to investigate the effect of vegetation-dependent parameterizations of key hydrological processes on TWS partitioning at the global scale using a multi-criteria model data fusion approach. The model, an expanded version of Trautmann et al. (2018), is a simple conceptual four-pool water balance model. Model parameters are calibrated against TWS variations from GRACE (Wiese et al. 2018), ET from FLUXCOM (Jung et al. 2019), runoff from GRUN (Ghiggi et al. 2019), and ESA CCI soil moisture (Dorigo et al. 2017).

We contrast two experiments which differ only with respect to how vegetation-related parameters are defined: (1) a baseline experiment with global uniform parameters and (2) a vegetation experiment where vegetation parameters vary in space and partly in time. In contrast to the traditional approach of spatializing vegetation parameters by plant functional type or land cover class, and keeping this a priori parameterization fixed during model application, we take advantage of continuous information on few key properties that link vegetation and hydrological processes: (1) spatially distributed and time-varying active vegetation cover that influences transpiration demand and interception storage, (2) the spatial pattern of soil water supply for transpiration via roots, and (3) the spatially distributed and time-varying influence of vegetation cover on infiltration and runoff generation. Specifically, we are addressing the following questions:

1) Where, when, and by how much are global hydrological simulations improved by spatially distributed and time-varying vegetation parameters?
2) To what extent do the attribution and interpretation of TWS variations for individual storage components change when introducing spatial and temporal variation of vegetation parameters?

3.2 Methods

In the first section, we give a general overview on the design of this study. Subsequently, the model and data streams used as well as the calibration and evaluation approach are explained in more detail.

3.2.1 Overview

To assess the potential effect of including continuous information on vegetation, we compare two model variants that are based on the same conceptual structure: (1) a base model with static, globally uniform parameter values (B) and (2) a model variant that includes spatially (and temporally) varying vegetation characteristics by defining vegetation parameters as a function of global data products (VEG). We additionally performed an experiment that discretizes vegetation parameters for distinct classes of plant

functional types, similarly to some other GHMs. This PFT experiment is explained and shown in Sect. B.9 in the Supplement.

Forced with global climate data, the parameters of each variant are calibrated for a spatial subset against multiple Earth observation-based data. In the B experiment, the parameters themselves are calibrated, and globally constant parameter values are obtained. While the optimized parameters implicitly account for the effect of the nearly ubiquitous presence of vegetation, they cannot represent effects of spatially and/or temporally varying vegetation properties. In the VEG experiment, we describe vegetation-related parameters as a linear function of spatio-temporal varying vegetation variables; i.e., we calibrate scalars representing the slopes of these functions. By calibrating the slope, we include the continuous pattern from the data but scale it to best fit the observational constraints. Hence, vegetation-related parameters vary explicitly spatially and partly temporally.

Once the parameters are calibrated, the simulations for the whole domain (global) are used to evaluate the model performance at different spatial and temporal scales. To finally delineate the effect of including varying vegetation characteristics on the composition of simulated TWS across temporal (mean seasonal, inter-annual) and spatial (local grid scale, spatially aggregated) scales, we use the impact index as defined by Getirana et al. (2017).

The model is run on daily time steps at a 1°x1° latitude/longitude resolution, focusing on vegetated regions under primarily natural conditions. To avoid biases of the calibrated model parameters due to processes that are not represented in the model structure, we exclude grid cells with10 % permanent snow and ice cover, > 50 % water fraction, > 20 % bare land surface, and > 10 % artificial land cover fraction. These grid cells are masked out using the GlobeLand30 fractional land-cover map v2 (Chen et al. 2014). Additionally, we exclude regions with a large human influence, mainly related to groundwater extraction, on the trend in GRACE TWS variations (Rodell et al. 2018) (see Fig. 3.2). The final study area comprises 74 % of the global land area. All other datasets used in this study were resampled to the 1°x1° grid and subset to the same grid cells.

Due to the temporal coverage of forcing data and observational constraints, we calibrate the model for the period January 2002–December 2014, while the global-scale model runs and analyses are performed for the period March 2000–December 2014. Prior to each model run, all states are initialized by a 8-year spin-up period. The forcing for the spin-up period is assembled by randomly rearranging complete years of the forcing data.

3.2.2 Model Description

The conceptual hydrological model is forced by daily precipitation, air temperature, and net radiation (Table 3.1). It includes a snow component (see Trautmann et al. 2018), a two-layer soil water storage (wSoil), a deep soil water storage (wDeep), and a delayed, slow water storage (wSlow). The schematic structure of the model is shown in Fig. 3.1, and calibration parameters are explained in Table 3.2.

Depending on air temperature (T_{air}), precipitation (Precip) is partitioned into snowfall (Snow), that accumulates in the snow storage (wSnow), and rainfall (Rain), that is partly retained in an interception storage. Interception throughfall together with snowmelt is partitioned into infiltration and infiltration excess depending on the ratio of actual soil moisture and maximum soil water capacity following Bergström (1991):

$$I_{exc} = I_{in} \cdot \left[\frac{\sum_{l=1}^{2} wSoil(l)}{\sum_{l=1}^{2} wSoil_{\max(l)}} \right]^{p_{berg}} \quad (3.1)$$

where I_{exc} is the infiltration excess, I_{In} is the incoming water from throughfall and snowmelt, wSoil(l) is the soil moisture and $wSoil_{max(l)}$ the maximum soil water capacity of each soil layer l, and p_{berg} is a calibration parameter. While $p_{berg} < 1$ allocates a small fraction of the incoming water to the soil water pool even if it is nearly empty, $p_{berg} > 1$ allows a large fraction of incoming water to infiltrate into the soil

when soil saturation is already high, and p_{berg} = 1 describes a linear relationship between soil water saturation and the amount of incoming water that infiltrates.

A fraction of the infiltration excess (defined by the global calibration parameter rf_{Slow}) then replenishes a delayed water storage (wSlow) that acts as a linear reservoir and generates slow runoff (Q_{slow}). The remaining infiltration excess represents fast direct runoff (Q_{fast}). Q_{fast} and Q_{slow} together represent total runoff Q that flows out of the system, i.e., grid cell.

Infiltrated water is distributed among two soil layers following a top-to-bottom approach, where the maximum capacity of the first soil layer is prescribed as 4 mm, in order to match the tentative depth of satellite soil moisture observations, while the storage capacity of the second soil layer is a calibration parameter ($wSoil_{max(2)}$). The second soil layer is connected with a deeper water storage (wDeep). The size of wDeep is defined as a multiple of $wSoil_{max(2)}$ by the calibrated scaling parameter s_{deep}. Depending on the moisture gradient between the two storages, water either percolates from the second soil layer to the deeper soil, or it rises from the deeper storage into the second soil layer, by scaling to a maximum flux rate (defined by the global calibration parameter, f_{max}). The deeper storage therefore acts as a storage buffer that linearly discharges further to the delayed water storage (wSlow), which also receives part of the infiltration excess.

Evapotranspiration (ET) is represented by a demand– supply approach that is driven by a potential ET demand following Priestley–Taylor and is limited by the available soil moisture supply. ET is partitioned into interception evaporation (E_{Int}), bare soil evaporation from the first soil layer (E_{Soil}), and plant transpiration from the two soil layers (E_{Transp}). Interception and plant transpiration are only calculated for the vegetated fraction of each grid cell, while bare soil evaporation is limited to the non-vegetated fraction of each grid cell.

While water in wSoil is directly available for ET, wDeep is only indirectly accessible by capillary rise, and the water stored in wSlow is not plant-accessible. Total water storage is the sum of all water storages, including wSnow, wSoil, wDeep, and wSlow. Although groundwater and surface water storages are not implemented explicitly, they are effectively included in wDeep and wSlow, especially after calibration of associated storage parameters against GRACE TWS.

3.2.3 Vegetation Characteristics

We include three aspects of vegetation influence on hydrological processes: (1) the specific transpiration demand by vegetation, (2) the soil water supply for transpiration via roots, and (3) the influence of vegetation on infiltration and runoff generation. These three aspects are controlled by three corresponding model parameters, namely the grid cell's vegetation fraction (p_{veg}), the maximum plant-available soil water ($wSoil_{max(2)}$), and the runoff generation-infiltration coefficient (p_{berg}). In the VEG experiment, scalar parameters are used as linear multipliers of observation-based spatio-temporal patterns to harvest the information of spatial and temporal patterns from the continuous data products.

3.2.3.1 Vegetation Fraction

The parameter p_{veg} reflects the vegetation cover of each grid cell that influences the grid's interception storage, transpiration demand, and partitioning of evapotranspiration components. To describe its spatial and seasonal variations, we include the mean seasonal cycle (MSC) of the Enhanced Vegetation Index (EVI). Therefore, p_{veg} at each time step is defined as a linear function of EVI, where s_{EVI} is the calibrated scaling parameter:

$$p_{veg} = s_{EVI} \cdot EVI \tag{3.2}$$

with $0 \leq p_{veg} \leq 1$.

EVI data are calculated via the MODIS standard formula (Didan and Barreto-Munoz, 2019) using the daily BRDF, nadir BRDF-adjusted reflectance values MCD43C1 v6 (Schaaf and Wang, 2015) for the period January 2001–December 2014:

$$EVI = 2.5\frac{NIR-red}{NIR+6\cdot red-7.5\cdot blue+1} \quad (3.3)$$

Since the daily EVI time series are not continuous due to noise and missing values during cloudy conditions, snow, and darkness, the data were preprocessed to be used in the model. For each grid cell, we calculate the median seasonal cycle, fill long gaps during wintertime with a low value, interpolate missing values, and smooth the time series. Therefore, winter is defined as days with negative net radiation, and gaps are considered long when 10 consecutive days of EVI data are missing. The wintertime gaps are filled with the 5th percentile of available wintertime data. The remaining missing values are linearly interpolated, and finally the resulting seasonal cycle is smoothed by a local regression with weighted linear least squares and a first-order polynomial model.

Table 3.1. Data used for model forcing, for description of vegetation characteristics, and for model calibration.

	Product	Space	Time	Data uncertainty	Reference
Forcing					
Precip	GPCP 1dd v1.2	global	daily		Huffmann et al. (2000)
T_{air}	CRUNCEP v6	global	daily		Viovy et al. (2018)
Rn	CERES Ed4 A	global	daily		Loeb et al. (2018); NASA/LARC/SD/ASDC (2017)
Vegetation characteristics					
EVI	based on MCD43C1 v6 (MODIS daily BRDF), calculated via MODIS standard EVI formula		daily climatology		Schaaf and Wang (2015)
RD1	maximum rooting depth		static		Fan et al. (2017)
RD2	effective rooting depth		static		Yang et al. (2016)
RD3	maximum soil water storage capacity		static		Wang-Erlandsson et al. (2016)
RD4	maximum plant-available water capacity		static		Tian et al. (2019)
Calibration					
TWS	GRACE mascon RL06	global	monthly	with product	Wiese et al. (2018)
wSoil	ESA CCI SM v4.04 (combined product)	global	daily	with product	Dorigo et al. (2017)
ET	FLUXCOM RS ensemble	global	daily	with product	Jung et al. (2019)
Q	GRUN v1	global	monthly	50 %	Ghiggi et al. (2019)

3.2.3.2 Plant-available Soil Water

In order to determine the soil water supply for transpiration as a function of vegetation, we define the maximum soil water capacity of the second soil layer $wSoil_{max(2)}$ based on rooting depth and soil water storage capacity data. We include the maximum rooting depth by Fan et al. (2017) (RD1), effective rooting depth by Yang et al. (2016) (RD2), maximum soil water capacity by Wang-Erlandsson et al. (2016) (RD3), and maximum plant-accessible water capacity by Tian et al. (2019) (RD4). Due to our definition of $wSoil_{max(2)}$ as maximum plant-accessible water, all four data are, theoretically, suitable when focusing on spatial patterns. Practically, though, they vary in their definition, underlying approaches, spatial coverage and derived spatial pattern. The RD1 and RD2 are based on principles of vegetation optimality and plant adaptation, and RD3 and RD4 are based on a water balance perspective but using Earth observations and/or data assimilation techniques. Therefore, we employ an approach in which we obtain a linear combination of the four products where the weights of each product are calibrated during the multi-criteria parameter optimization:

$$wSoil_{\max(2)} = \sum_{d=1}^{4} s_{RD(d)} \cdot RD(d) \quad (3.4)$$

where RD(d) is the data from each data stream d, and $s_{RD(d)}$ denotes the corresponding scaling factors that are calibrated. As RD4 from Tian et al. (2019) is only available for arid to moderately humid vegetated land area and excludes tropical forests (Tian et al. 2019), resulting gaps in the study area are filled by the calibration parameter $wSoil_{max(RD4)}$ prior to scaling RD4.

3.2.3.3 Runoff/Infiltration Coefficient

Finally, vegetation structure also affects the infiltration and runoff generation process as it alters the surface and sub-surface characteristics. To reflect this influence, we describe the infiltration–runoff parameter p_{berg} (Eq. 3.1) as a linear function of the vegetation fraction p_{veg}:

$$p_{berg} = s_{berg} \cdot p_{veg} \tag{3.5}$$

where s_{berg} is the calibrated scaling parameter.

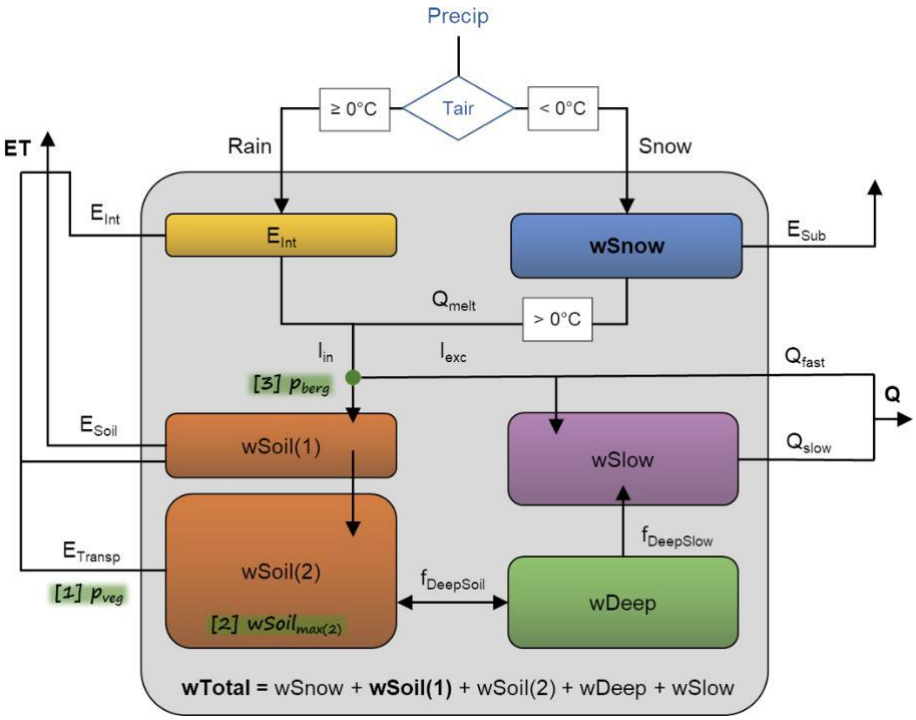

Figure 3.1. Schematic of the underlying model structure, with blue font denoting forcing data: Precip is precipitation, and T_{air} is air temperature. Boxes represent states as follows: E_{int} is interception storage, wSnow snow water storage, wSoil(1) upper soil layer, wSoil(2) second soil layer, wDeep deep water storage, and wSlow slowly varying water storage. Arrows denote fluxes as follows: Rain is rainfall, Snow snowfall, E_{Sub} sublimation, Q_{melt} snowmelt, I_{in} incoming water from throughfall and snowmelt, I_{exc} infiltration excess, Q_{fast} fast direct runoff, Q_{slow} slow runoff, Q total runoff, E_{Int} evaporation from interception storage, E_{Soil} soil evaporation, E_{Transp} plant transpiration, ET total evapotranspiration, $f_{DeepSoil}$ the flux between wSoil and wDeep (percolation resp. capillary rise), and $f_{DeepSlow}$ the flux from wDeep to wSlow. Bold print highlights model variables that are constrained in the calibration. Green highlights show where vegetation influence is included explicitly: [1] the parameter p_{veg} to define each grid cell's vegetation fraction, [2] the parameter $wSoil_{max(2)}$ that defines the maximum plant-available soil water, and [3] the parameter p_{berg} to define the infiltration and runoff generation partitioning.

Table 3.2. Calibrated model parameters and their description, range, and calibrated values for experiments B and VEG. Text in italics highlights calibrated values at the predefined parameter bounds.

Parameter	Description	Units	Default value	Range	Calibrated values ± uncertainty (%)			
					B		VEG	
Vegetation fraction								
p_{veg}	active vegetation fraction of the grid cell		0.5	0.3 - 1	0.37	± 0.05		
s_{EVI}	scaling parameter to derive active vegetation fraction from EVI data		1	0 - 5			3.89	± 0.05
Evapotranspiration								
p_{Int}	interception storage	mm	1	0 - 10	1.0	± 0.08	0.6	± 0.02
k_{Soil}	fraction of 1st soil layer available for evaporation		0.5	0.1 - 0.95	*0.1*	± 0.01	0.4	± 0.08
α_{veg}	α parameter of the Priestley-Taylor equation		1	0.2 - 3	2.25	± 0.15	0.92	± 0.00
k_{Transp}	fraction of soil water available for transpiration		0.02	0 - 1	0.12	± 0.32	0.48	± 1.76
Infiltration-runoff								
p_{berg}	runoff-infiltration coefficient		1.1	0.1 - 5	1.32	± 0.02		
s_{berg}	scaling parameter to derive the runoff-infiltration coefficient from p_{veg}		3	0.1 - 10			3.08	± 0.02
Soil moisture								
$wSoil_{max\,(2)}$	maximum (available) water capacity of the 2nd soil layer	mm	300	10 - 1000	752	± 0.02		
$s_{RD(1)}$	weight to include maximum rooting depth by Fan et al. (2017)		0.05	0 - 5			0.01	± 0.00
$s_{RD(2)}$	weight to include effective rooting depth by Yang et al. (2016)		0.05	0 - 5			*0.00*	± 0.00
$s_{RD(3)}$	weight to include maximum soil water storage capacity by Wang-Erlandson et al. (2016)		0.05	0 - 5			0.15	± 0.06
$s_{RD(4)}$	weight to include plant available water capacity by Tian et al. (2019)		0.05	0 - 5			0.15	± 0.07
$wSoil_{max\,(RD4)}$	maximum (available) water capacity of the 2nd soil layer for grids with missing estimates in Tian et al. (2019)	mm	50	0 - 1000			145	± 0.08
Deep soil								
s_{deep}	scaling parameter to derive the maximum deep soil storage from $wSoil_{max\,(2)}$		0.5	0 - 50	9.1	± 461317	5.6	± 0.21
f_{max}	maximum flux rate between deep soil and the 2nd soil layer	mm d^{-1}	10	0 - 20	1.5	± 0.00	5.1	± 0.01
d_{Deep}	depletion coefficient from deep soil to delayed water storage		0.5	0 - 1	*1.0*	± 5.61	0.01	± 0.00
Delayed water storage								
rf_{Slow}	recharge fraction of infiltration excess into delayed water storage		0.5	0 - 1	0.78	± 1.72	0.68	± 0.01
d_{Slow}	depletion coefficient from delayed water storage to slow runoff		0.01	0 - 1	*1.0*	± 2329	0.02	± 0.03

3.2.4 Model Calibration

In order to keep computational costs low and to avoid overfitting of model parameters, we perform model calibration for a subset of 904 (8 %) grid cells. Since model parameters are expected to vary much more in space than in time (between years), and due to the rather short time period of available constraints, we build two subsets of data for calibration and validation data in the spatial domain rather than in time (spatial split-sample approach). Calibration grid cells are chosen by a stratified random sampling method that maintains the overall proportion of different climate and hydrological regimes defined by Köppen-Geiger climate regions (Kottek et al. 2006).

Since this study focuses on the impact of vegetation, and in order to keep the number of calibration parameters low, we do not optimize snow-related parameters and use the optimized snow parameters from Trautmann et al. (2018). This results in a total of 11 calibration parameters for the B model and a total of 16 parameters for the VEG model (Table 3.2).

In order to constrain different aspects of the water cycle, we use a multi-criteria calibration approach similar to Trautmann et al. (2018). The parameters of each model variant are simultaneously optimized against multiple observational constraints, including monthly TWS anomalies from GRACE (Wiese et al. 2018, Watkins et al. 2015), ESA CCI soil moisture (Dorigo et al. 2017), evapotranspiration estimates from the FLUXCOM-RS ensemble (Jung et al. 2019), and gridded runoff from GRUN (Ghiggi et al. 2019) (Table 3.1).

When using observational datasets from several sources, it is essential to consider possible inconsistencies between them that arise from their respective characteristics and uncertainties (Zeng et al. 2015, 2019). Therefore, we derived the monthly water (im)balance of the observations following a similar approach to Rodell et al. (2015) (see Sect. B.10 in the Supplement). Although we did not find major systematic inconsistencies at the global scale, we take into account each dataset's characteristics and uncertainties in model calibration via the cost term at the grid cell level. To this end, we only use grid cells and time steps with available observations, which vary for the different data streams. To retrieve one cost term per observational constraint, we concatenate the time series of all grid cells into a single vector for which costs are calculated. The individual cost terms are considered to have the full weight of 1, resulting in a total cost value $cost_{total}$ as the sum of individual costs. The total cost is then minimized during the optimization process using a global search algorithm, the covariance matrix evolutionary strategy (CMAES) algorithm (Hansen and Kern, 2004).

$$cost_{\text{total}} = \sum_{ds=1}^{n} cost(ds) \tag{3.6}$$

where cost(ds) is the cost for each data stream ds. For TWS, ET, and Q, the cost terms are based on the weighted Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970), which explicitly considers the observational uncertainty σ:

$$cost = \frac{\sum_{i=1}^{n} \frac{\left(x_{obs,i} - x_{mod,i}\right)^2}{\sigma_i}}{\sum_{i=1}^{n} \frac{\left(x_{obs,i} - \bar{x}_{obs,i}\right)^2}{\sigma_i}} \tag{3.7}$$

where $x_{mod;i}$ is the modelled variable, $x_{obs;i}$ is the observed variable, $\bar{x}_{obs}$ is the average of x_{obs}, and σ_i is the uncertainty of x_{obs} of each data point i. The cost criterion reflects the overall fit in terms of variances and biases, with an optimal value of 0 and a range from 0–1.

Owing to the larger uncertainties of Q_{obs} on inter-annual scales (Ghiggi et al. 2019), we only use the monthly mean seasonal cycle, while for the other variables, full monthly time series were used.

To define of ET_{obs}, we utilize the median absolute deviation of the FLUXCOM-RS ensemble. For Q_{obs}, we assume an average uncertainty of 50 % based on values reported in Ghiggi et al. (2019). For TWS_{obs}, the spatially and temporally varying uncertainty information provided with the GRACE data is used. In addition, the largest monthly values of TWS_{obs} (< 500 and > 500 mm) were masked out to avoid the effect of outliers on optimization results. Note that these outliers represent less than 0.5 % of the data and are mainly located in coastal arctic regions and are, thus, potentially related to land and sea ice and/or leakage from neighboring grid cells over ocean. Before calculating $cost_{TWS}$, the monthly means of observed and modelled TWS are respectively removed to calculate anomalies over a common time period 1 January 2002–31 December 2012.

Since remote sensing-based soil moisture only captures the top few centimeters of soil depth, usually about 5 cm, $cost_{wSoil}$ is calculated based on the modelled soil moisture in the first soil layer. As the combined ESA CCI soil moisture imposes absolute values and ranges from GLDAS Noah (Dorigo et al.

2015), we use Pearson's correlation coefficient as $cost_{wSoil}$ and focus on soil moisture dynamics that is most reflective of the original remote sensing observation. Only estimates from 1 January 2007 onwards are considered, as data before that period are sparse. Further, $cost_{wSoil}$ is calculated from the monthly averaged values to circumvent the large noise in the daily data. Thereby, only months with observations available for at least 10 d are considered. Due to snow cover, the temporal coverage of the product decreases with increasing latitude. Therefore, to prevent a bias towards northern summer months, we also exclude grid cells that lack more than 40 % of monthly estimates. After filtering for missing data, monthly surface soil moisture time series for 56 % of the total study area and 51 % of the calibration grid cells are available.

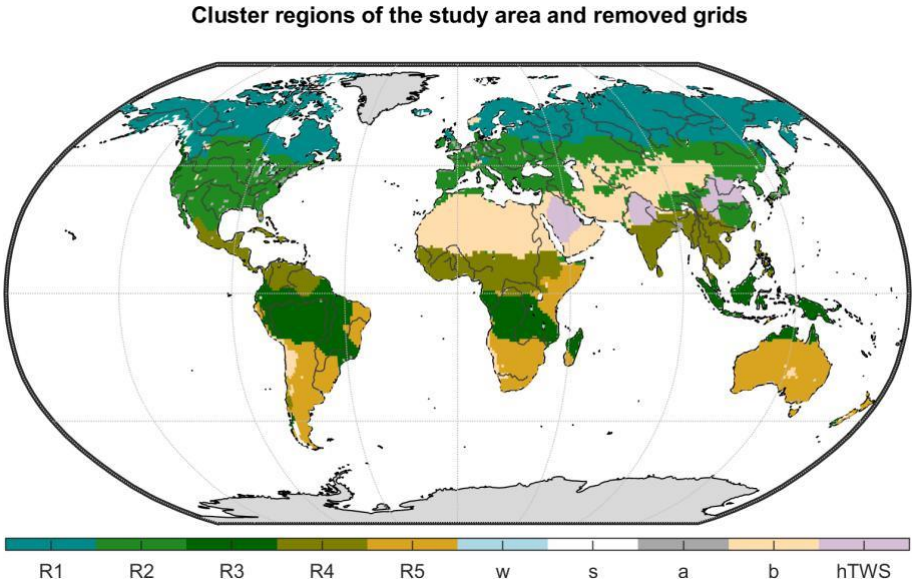

Figure 3.2. Hydroclimatic cluster regions of the study area (R1 – cold, R2 – temperate, R3 – humid, R4 – sub-humid, R5 – semi-arid) and grid cells that have been excluded from this study (*w* is the water fraction > 50 %, *s* the permanent snow and ice cover > 10 %, *a* the artificial land cover fraction > 10 %, *b* the bare land surface > 20 %, and *hTWS* the direct human impact on the trend in GRACE TWS).

3.2.5 Model Evaluation and Analysis

For model evaluation, we contrast the optimized parameter values and their uncertainties. The relative uncertainty in the optimized parameter vector is estimated by quantifying each parameter's standard error according to Omlin and Reichert (1999) and Draper and Smith (1981), similar to Trautmann et al. (2018).

For each experiment, the optimized parameter sets are used to produce model simulations for the global study area. Their performances are then evaluated using Pearson's correlation coefficient and the uncertainty-weighted Nash-Sutcliffe efficiency (wNSE) for TWS, ET, and Q observations (Eq. 3.8). The performances are evaluated on local (for each grid cell individually), regional, and global scales

$$wNSE = 1 - \frac{\sum_{i=1}^{n} \frac{\left(x_{obs,i} - x_{mod,i}\right)^2}{\sigma_i}}{\sum_{i=1}^{n} \frac{\left(x_{obs,i} - \bar{x}_{obs,i}\right)^2}{\sigma_i}} \tag{3.8}$$

For the regional analysis, we derive five hydroclimatic regions by performing a cluster analysis using the spatio-temporal characteristics of TWS, ET, and Q observations, as well as each grid cell's latitude. By that, each zone is characterized by similar seasonal dynamics and amplitudes of the water cycle variables, allowing for a better comparison of regional averages than, for example, the commonly used Köppen–

Geiger regions which lump regions with very different amplitudes and phasing of the water cycle variables. The resulting regions are shown in Fig. 3.2. Region 1 comprises the snow-dominated northern latitudes (cold), while region 2 includes the moderate mid-latitudes (temperate). Very humid and mostly tropical regions are combined in region 3 (humid). Region 4 is characterized by a distinct rain season (sub-humid), while region 5 includes semi-arid areas in low latitudes (semi-arid). Although we hereafter use these hydro-climatic cluster regions for model evaluation, the same analysis for Köppen-Geiger climate zones is presented in Fig B.11 in the Supplement to facilitate comparison with other studies.

Finally, we assess the contributions of the four water storage components, wSnow, wSoil, wDeep and wSlow, to seasonal and inter-annual variations of the total water storage across spatial scales, i.e., the local grid cell, the regional and the global average. To do so, we apply the impact index I following Getirana et al. (2017). The metric describes the contribution C of each water storage s as the sum of its absolute monthly anomaly:

$$C_s = \sum_{t=1}^{nt} |s_t - \bar{s}| \tag{3.9}$$

where $\bar{s}$ is the average storage of the time steps t–nt, with nt = 12 for mean seasonal and nt = 178 for inter-annual dynamics.

The impact index I_s is then defined as the ratio of each water storage component contribution C_s to the total contributions from all storage components:

$$I_s = \frac{C_s}{\sum_s^n C_s} \tag{3.10}$$

The value of I_s range from 0–1, with 0 indicating no impact and 1 indicating full control of all variations.

3.3 Results

In the following section we first evaluate both calibrated model variants by comparing their calibrated model parameters and by comparing modelled TWS, ET and Q against observations at global, regional and local scale. Subsequently, we show the contribution of individual storage components to TWS variability for B and VEG on different spatial and temporal scales.

3.3.1 Model Evaluation

3.3.1.1 Calibrated Parameters

Table 3.2 summarizes the calibrated parameters and their uncertainties for the B and VEG model experiments. Overall, including varying vegetation characteristics leads to more plausible parameter values after calibration, while in B several parameters hit their prescribed bounds. Furthermore, very high parameter uncertainties present in B, that indicate poorly constrained values, could be strongly reduced in VEG (Fig B.3 in the Supplement).

For B, p_{veg} suggests that on average only 37 % of each grid cell is covered with vegetation globally. This low vegetation fraction is counteracted by a high $_{veg}$ value (2.25), which is much higher than commonly used coefficients of the

Priestley–Taylor equation of around 1.2 (Lu et al. 2005), to yield good performance of modelled ET (Fig. 3.3). At the same time, a very low fraction of the first soil layer is available for soil evaporation, as k_{Soil} hits its lower bound of 10 %. In addition, the parameters controlling the drainage from deep and slow water storage (d_{Deep}, d_{Slow}) are high, resulting in a fast drainage and effectively discarding any influence of these water pools.

For VEG, the median vegetation fraction is 73 %, leading to a more realistic fraction of soil moisture being available for evaporation (k_{Soil} D 0:4), which is similar to the modal value of 0.33 reported by McColl et al. (2017), and a more realistic $_{veg}$ value of 0.92, that effectively leads to the median Priestley–Taylor coefficient of 0.81 (Fig B.2 in the Supplement). In comparison to B, the resulting $wSoil_{max(2)}$ of VEG with a median value of 52 mm is considerably lower. Its spatial pattern mainly originates from RD3 (Wang-Erlandsson et al. 2016) and RD4 (Tian et al. 2019) data, while RD1 (Fan et al. 2017) contributes only little, and RD2 (Yang et al. 2016) data are negligible. The resulting spatial patterns of the maximum soil water capacity from the combination of all datasets (Sect. B.2) are still consistent with those from other estimates and patterns of rooting depth (e.g., Schenk and Jackson, 2005). We note here that the soil water capacity data are favoured over the rooting depth data. This agrees with Küçük et al. (2022), who suggest that estimating plant storage capacity based on Earth observation data may be more suitable than those using optimality principles. Related to the limited size of wSoil, calibration enforces a deeper and a slow water storage with reasonable depletion parameters (d_{Deep}, d_{Slow}) in order to match observed TWS variations. By that, the considerable low $wSoil_{max(2)}$ parameter is counteracted by refilling wDeep, which indirectly provides plant-accessible water via capillary rise. Likewise, k_{Transp}, which describes the fraction of the second soil layer that is available for transpiration, is relatively high, as a larger fraction of the small soil water storage needs to transpire to match observed ET. Hence, calibrated k_{Transp} is higher than empirical values of ET decay between 0.02–0.08 that are based on assuming one soil water pool (Teuling et al. 2006).

3.3.1.2 Model Performance

Table 3.3 contrasts the overall model performance metrics for TWS, ET, and Q for the two experiments for the calibration subset of 8 % grid cells (opti) and the entire study domain (global). The metrics are calculated in the same way as during optimization, i.e., by concatenation of the time series of all grid cells into a single vector for which statistics are calculated. In general, the differences between opti and global, as well as between B and VEG, are marginal. For VEG, results mainly improve for TWS and slightly for ET. Although the models were only calibrated for the spatial subset in opti, equally good or even better performances are obtained when the calibrated parameters are applied over the entire study domain. This suggests that the calibration subset was representative of the entire study domain, and the calibration did not overfit model parameters.

Table 3.3. Overall model performance metrics in terms of weighted Nash–Sutcliffe efficiency (wNSE) and Pearson's correlation coefficient (corr) of total water storage (TWS), evapotranspiration (ET), and runoff (Q) in B and VEG experiments for the calibration subset (opti) and the entire study domain (global).

	TWS				**ET**				**Q**			
	wNSE		corr		wNSE		corr		wNSE (MSC)		corr (MSC)	
	opti	global	opti	global	opti	global	opti	global	opti	global	opti	global
B	0.33	0.33	0.69	0.69	0.97	0.97	0.90	0.90	0.63	0.63	0.86	0.86
VEG	0.38	0.41	0.71	0.72	0.98	0.98	0.90	0.91	0.60	0.57	0.85	0.85

Among the variables, the best model performance in terms of wNSE and corr is obtained for ET. While the correlation between observed and simulated TWS is high, the overall wNSE is relatively low, which mainly results from higher uncertainties in TWS_{obs} and a larger variance error, likely originating from grid cells with low observed TWS variance.

Similar to the global metrics, the average mean seasonal cycle of different regions shows an equally good or slightly better performance of VEG compared to B regarding all variables (Fig. 3.3). At regional scale (Fig. 3.4), the general pattern of grid-wise Pearson correlation is similar for both experiments. However,

the difference between the correlation coefficients highlights an improvement using VEG for a large proportion of grid cells and regarding all TWS, ET, and Q (indicated by brown color).

For TWS, the amplitude at the global scale is well captured yet with a phase difference of 1 month in both model variants, where both model variants show an earlier timing of peak storage (Fig. 3.3). The phase shift is also apparent in the temperate and cold regions, while the seasonal dynamics in sub-humid and humid regions is captured well, yet with an underestimation of the amplitude. Though differences are small, VEG obtains higher correlation, except for the semi-arid region. At local scale, correlation with GRACE TWS is lowest in rather semi-arid grid cells (Fig. 3.4), where TWS variation is low. However, including spatial patterns of vegetation improves TWS mainly in these (semi-)arid regions.

Regarding ET, both experiments reproduce seasonal dynamics in all regions quite well yet tend to underestimate ET in the semi-arid, sub-humid, and humid regions, especially in months with low ET (Fig. 3.3). At grid scale (Fig. 3.4), correlation of ET is very high, except for tropical regions due to low seasonality. Compared to B, VEG improves correlation here, as well as in some (semi-)arid regions such as the Sahel zone and the Western United States.

In contrast to ET, performance for Q is generally the best in regions with poorer model performance in terms of ET (semi-arid, sub-humid and humid regions) (Fig. 3.3), suggesting a trade-off between the two different observation data streams, i.e., the inability of matching both observed fluxes simultaneously. Nonetheless, including varying vegetation characteristics improves peak runoff in all regions and reduces the underestimation of Q especially in the cold region. While the improvement of Q simulations in northern latitudes gets even more obvious at grid scale, B shows higher correlation with observations in Africa and the Mediterranean (Fig. 3.4).

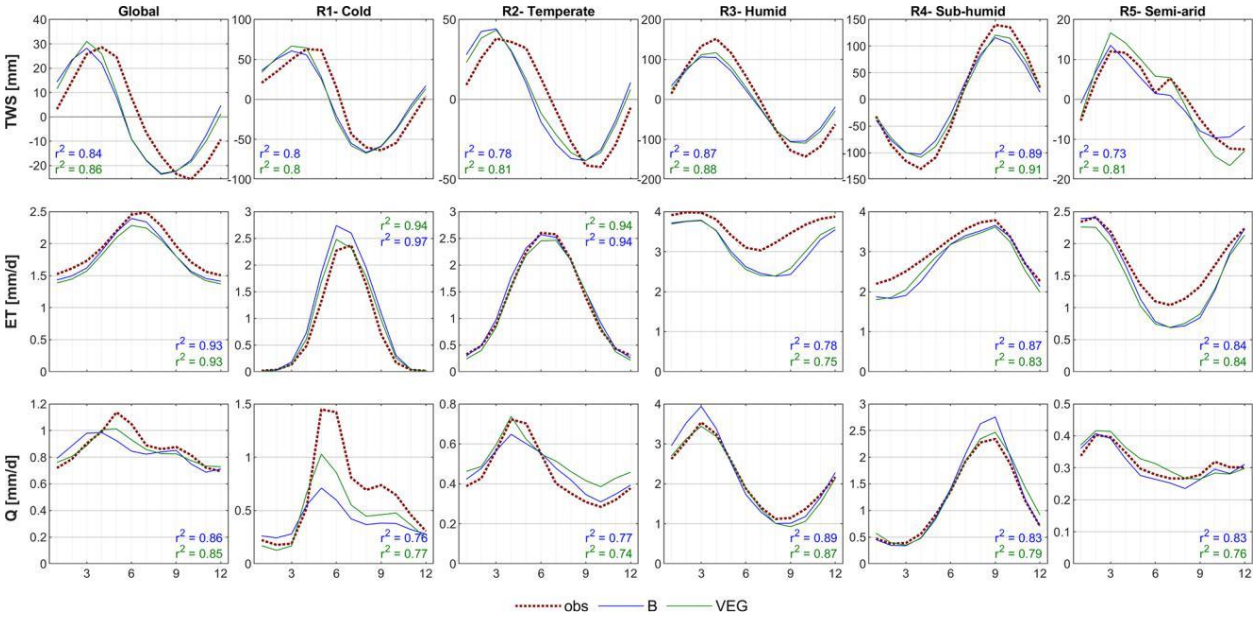

Figure 3.3. Global and regional mean seasonal cycles of total water storage (TWS), evapotranspiration (ET), and runoff (Q) for the B and VEG experiments compared to the observational constraints by GRACE (TWS), FLUXCOM (ET), and GRUN (Q).

3.3.2 Importance of varying Vegetation Properties to TWS Variability

In this section, we present the influences of vegetation on TWS partitioning into snow (wSnow), plant-accessible soil moisture (wSoil), not directly plant-accessible deep soil water (wDeep), and non-plant-accessible slow water storages (wSlow) at different spatial and temporal scales. We first focus on mean seasonal dynamics and continue with the contribution of each component to inter-annual TWS variability at local grid-cell and regional scales, respectively, before presenting the analysis at the global scale.

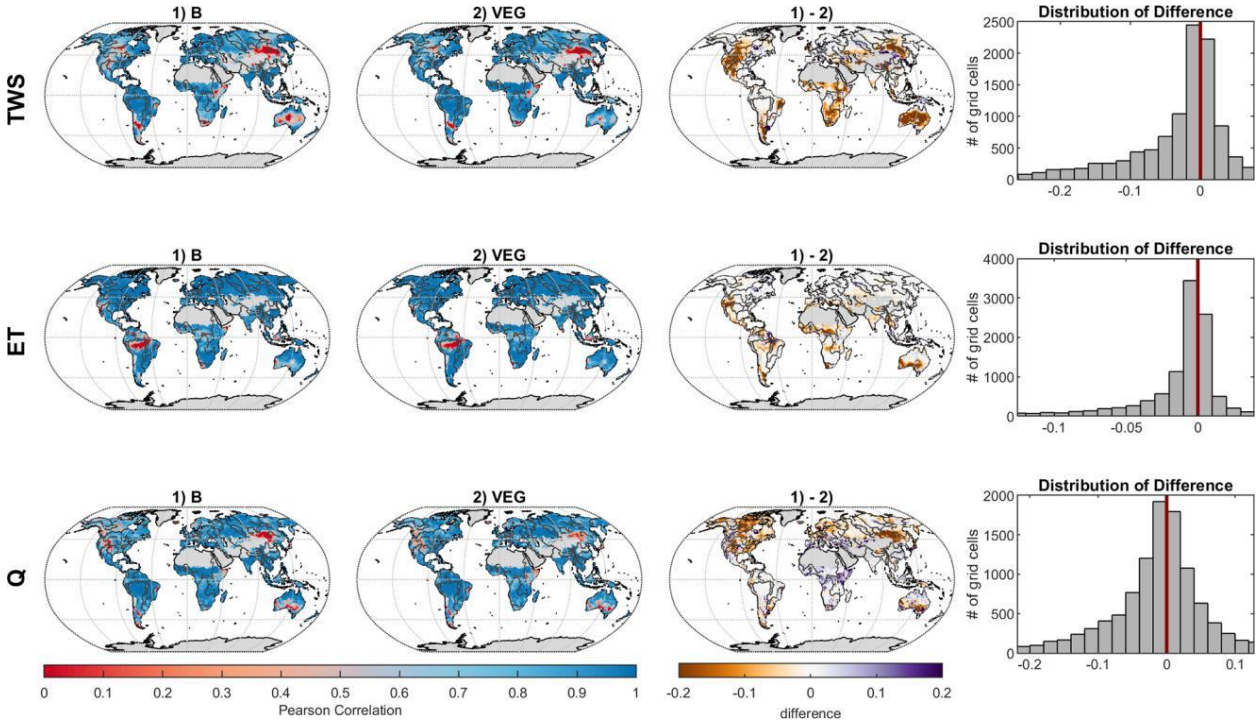

Figure 3.4. Grid-wise Pearson's correlation coefficient for total water storage (TWS), evapotranspiration (ET), and runoff (Q) between (1) observations and B and (2) observations and VEG, as well as differences between (1) and (2) (brown color, i.e., negative values, indicates higher correlations for VEG, while purple color, i.e., positive values, indicates better correlation values for B).

3.3.2.1 Local and regional Scale

Figure 3.5 shows the contribution of individual water storages to mean seasonal TWS variations at local grid scale. For both B and VEG, wSnow has the highest impact in northern latitudes and high altitudes where snowfall occurs regularly. Locally, the contribution of liquid water increases gradually with decreasing latitude and, finally, causes all TWS variations south of 45 N. Within the liquid water storages, B attributes nearly all variations to directly plant-accessible soil moisture wSoil, with an average of 76 % over all grid cells. While showing a similar pattern of increasing contribution towards lower latitudes, the VEG experiment only has an average of 17 % contribution from wSoil. Instead, most variations (40 %) are due to variability in the deeper soil storage, wDeep. In addition, the average impact of slow water storages wSlow (20 %) is comparable to that of wSnow (22 %) in VEG, though it is spatially much more limited to tropical regions, such as the Amazon basin.

Mean seasonal dynamics averaged globally and for different regions are shown in Fig. 3.6. As indicated by the grid-scale results, wSnow dominates TWS variations in the northern cold region (73 % in B, resp. 69 % in VEG) and plays a considerable role in the temperate region (28 % resp. 26 %). For the other regions, B attributes nearly all remaining variability to wSoil, while in VEG wDeep has the highest impact index (59 % in semi-arid, 50 % in sub-humid, and 43 % in humid).

At the inter-annual scales, the impact of wSnow decreases to 10 % (B) and 12 % (VEG) locally (Fig. 3.7). For most of the grid cells, all inter-annual TWS variations are caused by wSoil in B. In VEG, however, the deeper soil layer wDeep is again the most important storage, with an average impact index of 53 % for all grid cells. The contribution of wSoil and wSlow remains more or less the same as that for seasonal TWS variations.

Average contributions for different regions and globally (Sect. B.4 in the Supplement) show again that, in B, nearly all inter-annual TWS variability is caused by wSoil (87 %– 99 %). Only in the cold region does the impact of wSoil decrease to 69 % in favor of wSnow (31 %). Similar to the local scale, in VEG, wDeep explains > 50 % of TWS variability in most regions. Only in the cold region is the contribution of wDeep

similar to wSnow (39 % vs. 38 %). The contribution of wSoil ranges from 9 % (cold) to 19 % (semi-arid), while the impact of wSlow is between 16 %–18 % in most regions and increases in sub-humid (24 %) and humid (34 %) regions.

3.3.2.2 Global Scale

Finally, Fig. 3.8 contrasts the impact of water storage components to the total storage, in B and VEG, at the global scale. As with the local and regional scales, including varying vegetation characteristics differentiates the composition of global TWS variations drastically. In both experiments, wSnow clearly dominates the spatially aggregated mean seasonal cycle with an impact index of 71 % (B) and 61 % (VEG). These contributions are considerably higher than the average local impact index over all grid cells (B 24 %, VEG 22 %; Fig. 3.5). As already seen at local scale, liquid water storages dominate the inter-annual TWS variability, whereby B and VEG differ in the attribution to different components of the liquid water storage. In B, all variations other than wSnow originate from wSoil, but wDeep dominates in VEG. Especially at inter-annual scales, wDeep accounts for half of all TWS variations. In contrast to B, in VEG, wSoil only has a minor impact of 7 % at seasonal and 13 % at inter-annual scale. Instead, wSlow has a moderate contribution of 11 % (mean seasonal) and 17 % (inter-annual). In contrast to the mean seasonal dynamics in which the dominating storages are different at local and global scales, the inter-annual dynamics are consistent across scales, with the same storage component dominating at both local and global scale (Figs. 3.5, 3.7, and 3.8).

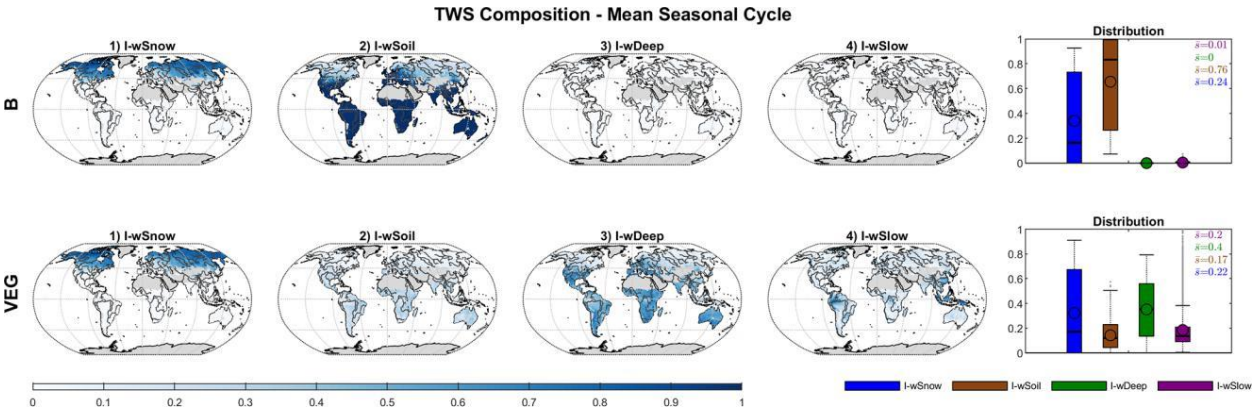

Figure 3.5. Global distribution of the impact index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep), and delayed water storage (wSlow) to the mean seasonal cycle of total water storage, for B and VEG.

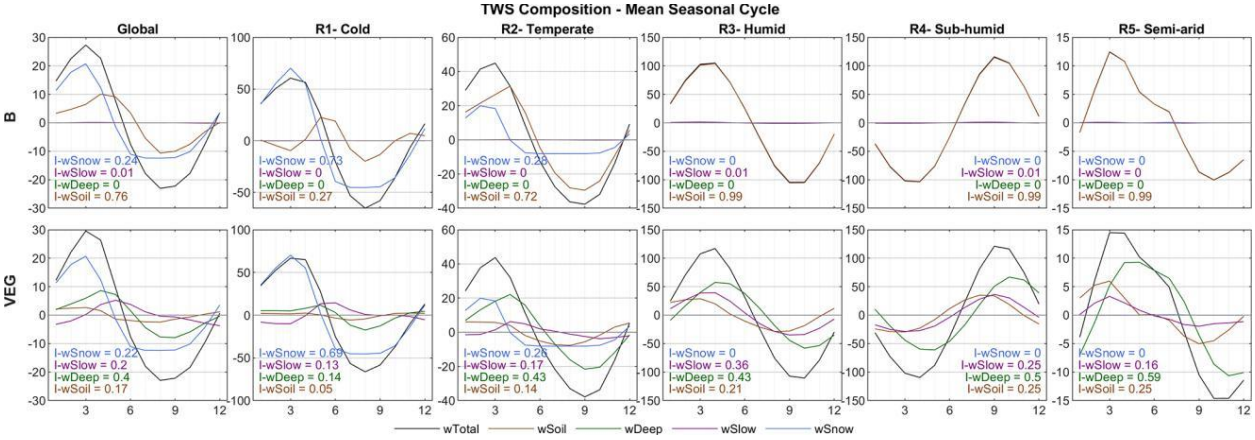

Figure 3.6. Global and regional average mean seasonal cycles of simulated total water storage and its components for B and VEG, including the regional impact index I for each storage.

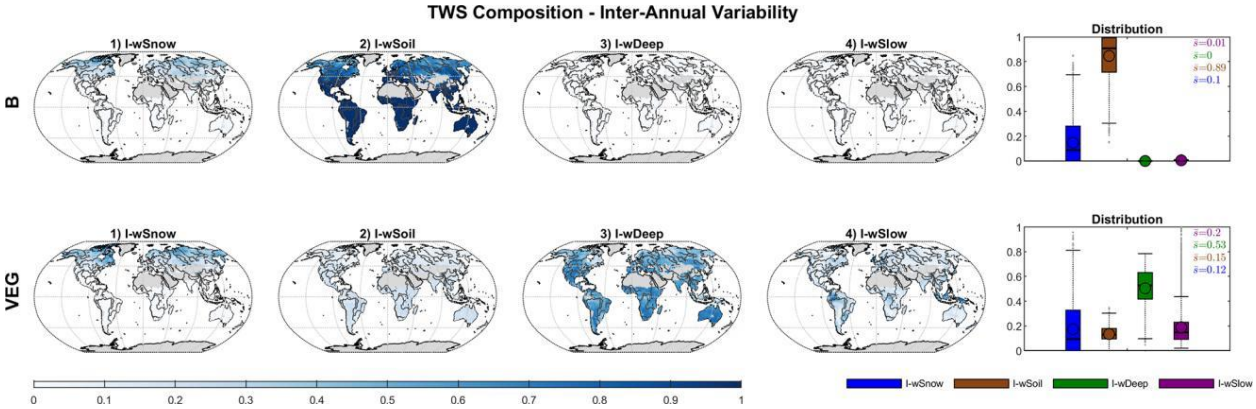

Figure 3.7. Global distribution of the impact index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep), and delayed water storage (wSlow) to the inter-annual variability of total water storage, for B and VEG.

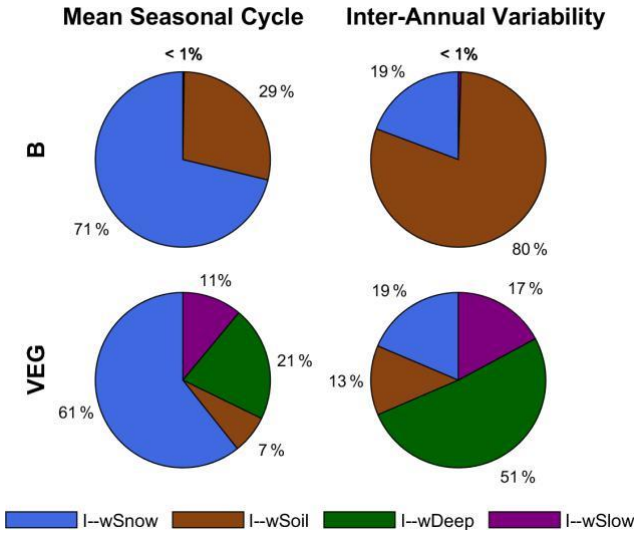

Figure 3.8. Impact index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep), and delayed water storage (wSlow) to the global average mean seasonal cycle and inter-annual variability of total water storage, for B and VEG.

3.4 Discussion

In order to address the two main research questions of this study, the following section discusses the above-shown differences between B and VEG, first regarding model performance and finally regarding the modelled partitioning of TWS.

3.4.1 Model Performance

Both experiments show good performance against the observational constraints, and the differences between B and VEG are relatively small at the global scale. However, there are systematic improvements for VEG at the regional and local scale, and calibrated parameter values for VEG are more realistic and better constrained. This suggests a more realistic representation of fluxes and states in VEG overall. Remaining discrepancies compared to observations can be associated with shortcomings and uncertainties in the observational data as well as to the processes that are not represented in the rather simple model structure.

The differences in the seasonal phase of global TWS in both model experiments mainly originate from the temperate and cold regions, and such model simulation differences have been reported previously (Döll

et al. 2014, Schellekens et al. 2017, Trautmann et al. 2018). One of the potential reasons is the temporary storage of meltwater during spring in rivers and other surface water bodies, which occurs contiguously over large areas in mid-latitudes to high latitudes (Döll et al. 2014, Schellekens et al. 2017, Schmidt et al. 2008, Kim et al. 2009) and which delays the storage decay. In this context, lateral water transport may also additionally affect the TWS variations in downstream grid cells. Yet, such processes and conditions are neither represented in B nor VEG.

Weaker performance of TWS in (semi-) arid regions is likely mainly due to low observed TWS variations and a low signal-to-noise ratio (Scanlon et al. 2016). Hence, less weight is also given to those grid cells in the cost component during calibration due to their small variations. In addition, alteration by human activities like groundwater withdrawal, dams, and irrigation to overcome the natural water shortage in such regions as Northeast China and the American (Mid)West can be regionally large in relative terms. While we aimed to exclude grid cells with large human impact a priori, we cannot completely exclude the influence of the aforementioned anthropogenic processes that are not explicitly represented in our model experiments. It should, however, be noted that the observational EVI data used in the VEG experiment do have an imprint of, for example, irrigated agriculture, as the measured surface reflectance includes the higher vegetation activity due to irrigation. The better representation of ET in semi-arid regions due to the EVI constraint contributes to the improved simulation of TWS variations in the VEG experiment.

While overall ET performance is good, tropical regions show low correlation. These areas are associated with higher uncertainties in the FLUXCOM ET estimates (Jung et al. 2019) due to underlying data uncertainties of the eddy covariance observations. Those uncertainties are related to poor station coverage and energy balance closure gap but also to issues of the satellite data inputs caused by cloud coverage. Nonetheless, including varying vegetation characteristics data improves simulated ET here, suggesting a better representation of the characteristic highly active vegetation compared to other regions and to global averages. In addition, VEG improves ET mainly in water supply-limited regions for the reasons already presented above for improved TWS performance in (semi-) arid regions.

The trade-off between the performances, in particular in terms of the bias of Q and ET, suggests larger uncertainties in one of the data streams for these regions, inconsistencies between the ET and Q constraints from independent sources, and/or model structure deficits. A small tendency to a negative water balance in the consistency checks of the observational data for these regions (Sect. B.10) implies either underestimation of the precipitation forcing or overestimation of FLUXCOM ET or GRUN Q. Global precipitation datasets tend to underestimate precipitation (Trenberth et al. 2007, Contractor et al. 2020) due to limitations of the satellite retrieval, gauge measurements, and, if combined, the combination method (Fekete et al. 2004). Validation of the GPCD 1DD data used in this study showed an underestimation of precipitation in complex terrain and regionally during spring and autumn, while precipitation in wintertime tends to be overestimated (Huffman et al. 2000). While we accounted for the latter by reducing snowfall (via a scaling parameter that was calibrated in Trautmann et al. 2018), we do not consider potential underestimation in the rainfall forcing. Therefore, precipitation forcing may not provide sufficient water input for ET and Q in the model to achieve the magnitudes given by the observation-based products. Lastly, some deterioration of performance of Q in VEG may originate from deficiencies in the GRUN product itself, which was generated with climatic drivers only, disregarding information on spatio-temporal variations in vegetation (Ghiggi et al. 2019).

The improvement of Q in northern latitudes is associated with the activation of the slow and delayed storage in the VEG experiment, with spatial varying parameterization of soil water storage capacity. The relatively low storage capacity in these regions facilitates faster saturation excess runoff. In addition, the slow storage better represents the runoff delay in surface water and rivers in these regions that results in improvements of low flow during winter as well as the increase of runoff during spring (Fig. 3.3). Such delayed runoff also improves the simulation of peak runoff in the sub-humid and humid regions.

The remaining deficiencies in model performance, especially in the cold region, indicate missing processes in the simple model structure. Such processes include freeze–thaw dynamics and permafrost (Yu et al. 2020) as well as ice jam in river channels that would increase surface water storage and allow for high spring flood (Kim et al. 2009). In addition, snow parameters have been calibrated against remote sensing-

based GlobSnow snow water equivalent, that is known to saturate for deep snow conditions (Takala et al. 2011) (see Trautmann et al. 2018). Although the calibration process considered this shortcoming, an underestimation of modelled snow accumulation is possible – leading to an underestimation of peak snowpack in winter that would result in an underestimation of runoff due to lower snowmelt in spring.

While the VEG experiment presented here considers all three aspects of vegetation influences on hydrological processes explicitly (see Sect. 2.2.1), we also run experiments that include these aspects separately in the model calibration (not shown). These analyses found that the largest improvement was obtained when including soil water storage capacity as a function of rooting depth and storage capacity data and a rather low impact when considering the runoff/infiltration partitioning as a function of the vegetation fraction. This highlights the central role of soil water storages and the importance of adequately describing soil moisture pattern and dynamics in hydrological models.

3.4.2 Contribution to TWS Variability

Albeit their global coverage, the above-presented results agree with the previous regional study that focused on northern mid-latitudes to high latitudes (Trautmann et al. 2018). Similarly, both model experiments show a dominating role of snow accumulation and depletion on global seasonal TWS variability, whereas liquid water storages determine inter-annual TWS variations. At the same time, the contribution of individual storages to TWS variations differs at the local grid scale compared to when they are averaged over a region or globally. The stronger contribution of snow on spatially aggregated signals can be explained by the spatial coherence of snow accumulation over larger areas. Liquid water storages, on the other hand, are more spatially heterogeneous, with increasing and decreasing dynamics across regions that cancel out and compensate for each other when spatially aggregated (Trautmann et al. 2018, Jung et al. 2017). In contrast to the mean seasonal dynamics, the inter-annual impact indices of the storage components at the global scale are similar to the average local impact indices (Figs. 3.7 and 3.8). This suggests that at inter-annual timescales, there is no spatially coherent pattern of one single storage component that leads to higher accumulated impact indices than the local averages. However, while both experiments agree in the general pattern of the impact of snow versus liquid water storages, they systematically differ in the allocation of water among liquid storage compartments. In B, all variations other than wSnow originate from directly plant-accessible soil moisture, whereas, in VEG, the deeper soil storage wDeep becomes the most important. Therefore, including observation-based information on vegetation changes the attribution of TWS variations drastically, while the variations of total TWS themselves do not change significantly.

Differences in the composition of TWS variability between B and VEG are effectively reflected in the differences of calibrated parameters. In B, the directly plant-accessible soil water storage is larger, due to a higher effective $wSoil_{max(2)}$, while delayed water storages are “turned off” because of increased drainage (d_{Deep}, d_{Slow}), reducing the variations in wDeep and wSlow. Although VEG has been calibrated in the same way as the same observational constraints, calibrated model parameters differ as the data on vegetation characteristics included provide complementary information on spatial and temporal patterns. Therefore, the resulting calibrated parameters can be assumed to be more realistic. For example, they enable (delayed) longer term water storage as well as capillary rise from the deeper soil water storage when the directly plant-accessible storage dries out. Due to this process, TWS variations are mainly controlled by wDeep in VEG.

In detail, the increased importance of the indirect plant-accessible storage wDeep in VEG can be related to the limited maximum soil water capacity $wSoil_{max(2)}$ that is constrained by rooting depth–soil water capacity data and to a higher k_{Transp} parameter. The smaller wSoil storage increases percolation to wDeep, but the water is still available when needed due to the capillary rise from wDeep to wSoil.

Removing capillary flux from wDeep to wSoil in fact increases the contribution of wSoil to seasonal variability, while the impact of wDeep remains high on inter-annual scales (Sect. B.7 in the Supplement). While the contribution of capillary rise to total ET is < 20 % for most grid cells, it becomes more important in arid-to-wet transition regions, for example, sub-Saharan Sahel, Savannas, northern Australia, and the Indian subcontinent (Fig. 3.9). These are regions with high precipitation seasonality, where vegetation

often grows deep roots to access deep unsaturated zone storage and groundwater during the dry season. The spatial patterns of ET supported by capillary rise agree with the findings of Koirala et al. (2014), who applied the physically based model MATSIRO to investigate the effect of capillary flux on hydrological variables. The spatial patterns are also in line with the predicted probability of deep rooting by Schenk and Jackson (2005) and are supported by Tian et al. (2019), who found that vegetation remains active long into the dry season in Africa, suggesting that soil–deep soil–groundwater interaction plays a considerable role. Therefore, the spatial pattern of the interactions of wDeep with wSoil in VEG seems reasonable, and our results indicate that capillary rise appears to be a process of large-scale relevance.

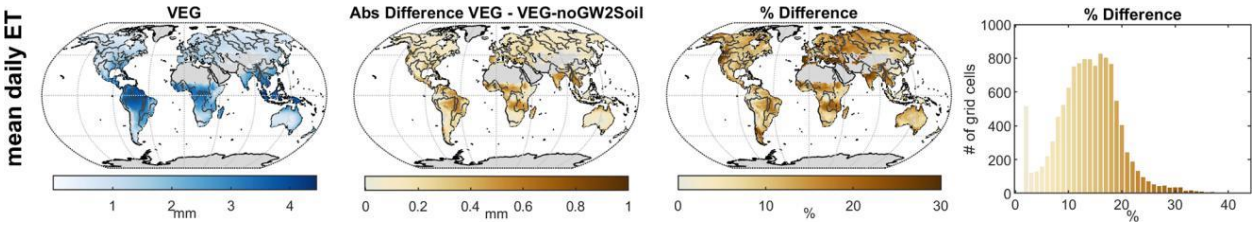

Figure 3.9. Total evapotranspiration (ET) of VEG with capillary flux from the deep soil water storage (left) and difference compared to a model version without capillary flux in millimetres (right map) and as a percentage difference (right).

While defined as the "fraction of soil water available for transpiration", k_{Transp} is an effective decay parameter for the depletion of wSoil via transpiration processes under water-limited conditions. Plausible values derived from eddy covariance observations of ET are in the order of 10^{-3}–10^{-1} (Teuling et al. 2006), similar in magnitude to delay coefficients for baseflow. By calibrating a model against GRACE TWS, it is difficult to decide whether water leaves the system slowly via ET or by Q, especially during dry-down periods. In B, k_{Transp} is much smaller than in VEG and more consistent with expected magnitudes, yet other slow depleting storages are effectively "turned off". In contrast, VEG with additional vegetation data simulates an important slow storage that contributes to Q and also to soil moisture via capillary rise and has a rather high calibrated k_{Transp}. To better understand the implications of parameterizing supply-limited ET decay in the model, we conducted another experiment where we fixed k_{Transp} in VEG to 0.05 (about the median value of empirically derived k_{Transp} from Teuling et al. 2006) and optimized all other parameters again. This caused most TWS variations to originate from wSoil but with less improvement in model performance compared to B (Sect. B.8 in the Supplement). Therefore, TWS decomposition is very sensitive to parameters controlling ET under water-limited conditions. However, VEG and VEG with fixed k_{Transp} qualitatively agree in the importance of the slow water storage in humid regions, which was also shown by Getirana et al. (2017). Overall, our results imply that the representation of ET under water-limited conditions in the models plays a decisive role in the simulated partitioning of TWS in soil moisture and slow water pools.

The large impact of the role of vegetation and of transpiration water supply within the model is also supported by a complementary experiment in which vegetation parameters were discretized for plant functional type classes and calibrated with the same multi-criteria approach (Sect. B.9).

As with the presented model variants, TWS composition simulated with existing large-scale hydrological models differs widely (Scanlon et al. 2018, Schellekens et al. 2017, Zhang et al. 2017). For example, PCR-GLOBWB and W3RA attribute seasonal TWS variations in the tropics to groundwater, while other models suggest they are mainly caused by soil moisture. These results are largely dependent on model structure and parametrization, which is potentially a challenge when models are used to decompose the integrated GRACE TWS signal and when implications of different processes and interactions are drawn. For example, Humphrey et al. (2018) analyzed how the CO_2 growth rate, a proxy for the land carbon balance fluctuations, is affected by inter-annual variations in GRACE TWS, assuming that these represent fluctuations in plant-accessible water that influence the carbon uptake of land ecosystems. In contrast, our study, along with previous reports, shows that a significant proportion of the GRACE TWS signal in the

tropics is not directly plant-accessible soil moisture but deeper soil water and slow storage components. The latter comprises surface water storage, whose importance for TWS variations in tropical regions has been shown by several studies (e.g., Güntner et al. 2007, Getirana et al. 2017).

Although VEG can be considered more reliable because of more realistic parameter values and better model performance, the current study still has some shortcomings. Despite using a multi-criteria calibration, individual component fluxes and states may not necessarily be well constrained. To further improve and solidify conclusions, especially on TWS partitioning, more constraints, such as deep soil moisture estimates or high-quality observations of surface water, are needed. Furthermore, spatial constraints for defining the depletion of water storages via ET and Q – either with spatial information on the delay parameters (k_{Transp} for ET, d_{Slow} for Q) or on their sub-fluxes (transpiration or evaporation, baseflow or direct runoff) – would be beneficial. In this context, runoff characteristics as the baseflow index or the baseflow recession coefficient provided by Beck et al. (2015) are potentially useful to define the spatial pattern of the slow runoff component. In addition, a GRACE product with daily resolution (Eicker et al. 2020) could enable better decomposition and differentiation of fast and slow storages whose short-term imprints are lumped in the monthly TWS signal.

3.5 Conclusions

In this study, we investigated the effect of varying vegetation characteristics on global hydrological simulations and in particular on the partitioning of TWS variations among snow, plant-accessible soil moisture, a deep soil water storage, and a slowly varying water pool that represents groundwater, surface and near-surface water storage. To do so, we included observation-based continuous vegetation information to parameterize the hydrological processes of evapotranspiration, soil water storage, and runoff generation in a large-scale hydrological model. With the parsimonious model that was constrained against multiple observations, we highlight the value of observation-based datasets in constraining model parameters of global hydrological models while maintaining simple model formulations to evaluate the influences of vegetation in the global hydrological cycle. First, we find that using a multi-criteria calibration approach allows for different model variants to perform relatively well despite major differences in model parameterization among them. In fact, even without accounting for dynamics and patterns of vegetation explicitly, the model performance can be interpreted as reasonable and more so at the global scale. However, including the spatial pattern of vegetation further improved the model performance. For example, large improvements were found in supply-limited regions, i.e. (semi-) arid regions (TWS and ET) and in tropical regions (ET), and Q simulations both globally and regionally in the Northern Hemisphere. Undoubtedly, spatio-temporal variations of vegetation characteristics are relevant for regional and global hydrological simulations. Interestingly, we find that the calibrated parameter values are also more reasonable when the model is fed with the vegetation information. In particular, parameter interactions and equifinality were reduced even though the same observational constraints were used for calibration. Lastly, we show how the representation of vegetation can modulate surface and subsurface hydrological process representation in the model, changing the spatial–temporal dynamics of individual storage components while maintaining the same overall response of total hydrological fluxes and storage variations. With or without accounting for varying vegetation characteristics explicitly, seasonal storage variations are dominated by snow at the global scale. However, including varying vegetation characteristics drastically changes the attribution of TWS variations among soil moisture, deep soil water, and slow water storages. Without varying vegetation parameters, the soil moisture effectively controls most of the TWS variation, but with varying vegetation characteristics, the role of deeper and delayed water storage becomes prominent. In particular, the representation of water-limited ET by the interplay of its sensitivity to soil moisture, maximum plant-accessible water storage capacity, and interactions with deep soil moisture or groundwater seem to play a decisive role in TWS partitioning in the simulations.In summary, this study highlights the value of including varying vegetation characteristics to further constrain model parameters with a parsimonious model structure. The findings further suggest an important role of groundwater–soil moisture–vegetation interactions in TWS variations. Since the representation of vegetation-related processes in global hydrological models seems to be a key factor for

controlling TWS partitioning, we emphasize the need for further studies and improvements of global water cycle models with respect to the role of vegetation by utilizing observational constraints on ecohydrological functioning in multi-criteria model calibration exercises.

Chapter 4:

Implications of River Storage for Integrating GRACE Terrestrial Water Storage observations into a Global Hydrological Model

---

Abstract

Although river water storage contributes to Total Terrestrial Water Storage (TWS) variations obtained from GRACE satellite gravimetry, it is unclear if computationally expensive river routing schemes are required when GRACE data is used to calibrate and validate global hydrological models (GHMs). Here, we investigate the role of river water storage on calibration and validation of a parsimonious GHM. In a multi-criteria calibration approach, the model is constrained against either GRACE TWS or TWS from which river water storage is removed. While we find that removing river water storage changes the TWS constraint regionally and globally, there are no significant implications for model calibration and the resulting simulations. However, adding modeled river water storage a-posteriori to calibrated TWS simulations improves model validation against GRACE TWS globally and regionally, especially in tropics and Northern low- and wetlands. While our findings justify the exclusion of explicit river routing for global model calibration, we find that the inclusion of river water storage is relevant for model evaluation.

This chapter is based on:

Trautmann, T., Koirala, S., Güntner, A., Kim, H., Jung, M. (2022): Implications of river storage for integrating GRACE TWS observations into a global hydrological model, *submitted to Environmental Research Communications*.

4.1 Introduction

Over the last decade, terrestrial water storage variations from GRACE and GRACE-FO satellite gravimetry provided valuable information for calibration and validation of global hydrological models (GHMs) (Werth et al., 2009, Döll et al., 2014, Scanlon et al., 2016, Kumar et al., 2016, Mostafaie et al. 2018, Trautmann, 2018). However, satellite gravimetry measures the vertically integrated total water storage (TWS), that includes water stored in ice, snow, canopy, soil moisture, groundwater, but also in wetlands, surface water bodies and river channels (Watkins et al. 2015). GHMs, on the contrary, do not necessarily simulate all these storages, and also vary significantly in their complexity and the represented hydrologic processes (Schellekens et al. 2017, Telteu et al. 2021). Among others, their inability to correctly simulate GRACE TWS variations is often attributed to neglected processes, such as river and floodplain storage dynamics (Kim et al. 2009). Although regionally, the relevance of river storage variability for GRACE TWS variations has been shown (Kim et al., 2009; Getirana et al. 2017), many GHMs don't include a river routing scheme to explicitly estimate river water storage (Telteu et al. 2021), even if they are calibrated against GRACE TWS. This is, on the one hand, due to historical reasons as GHMs rather focused on the vertical water balance (Shaad 2018), and, on the other hand, due to the high computational demand that river routing schemes require, especially if they need to be applied globally and in thousands of iterations during the model calibration process. While many efforts to reduce the computational costs of global routing schemes exist (Yamazaki et al. 2013, Mizukami et al. 2021), they remain a considerable factor for time and computational performance when compared to a model run without river routing. At the same time, the actual relevance of river routing for model calibration and validation against GRACE TWS at a global scale is rather unclear, also given the broad spatial and temporal resolution of GRACE TWS.

In the context of the development of new land surface models and data assimilation schemes, it's essential to know whether computational resources need to be invested in river routing during model calibration, and what are the consequences if routing is only applied as a post-processing, i.e., after defining model parameters, to validate model simulations.

Therefore, we specifically investigate the need for consideration of river storage and its potential effect on model calibration and validation in global hydrological studies that apply GRACE TWS.

To do so, we use a parsimonious hydrological model that does not explicitly account for river dynamics. We constrain the model in a multi-criteria calibration approach either against original GRACE TWS estimates, or against TWS estimates from which river storage was removed, and compare the resulting simulations. In the second step, we apply a routing scheme on the calibrated model and validate the performance with and without additional consideration of river storage compared to the original GRACE TWS. Specifically, we focus on:

I) the sensitivity of model calibration and resulting hydrological simulations to the removal of river storage from GRACE TWS.
II) the need of river routing for validation of hydrological simulations with GRACE TWS observations at regional and global scales.

In the following, we provide an overview on the methodology of this study. In section 4.3, we present and discuss the results regarding the effect of river storage on model calibration, followed by its influence on validation against GRACE TWS. Finally, section 4.4 summarizes the implications for future global hydrological studies.

4.2 Data and Methods

This study consists of 2 parts:

I) the effect of river storage on model calibration, and
II) the effect of river storage on model validation.

An overview on the methodologies and data for both parts is given in Fig.4.1, and described in detail in section 4.2.2 and 4.2.3. For all analysis, we apply the same hydrologic model, which is introduced in the following section.

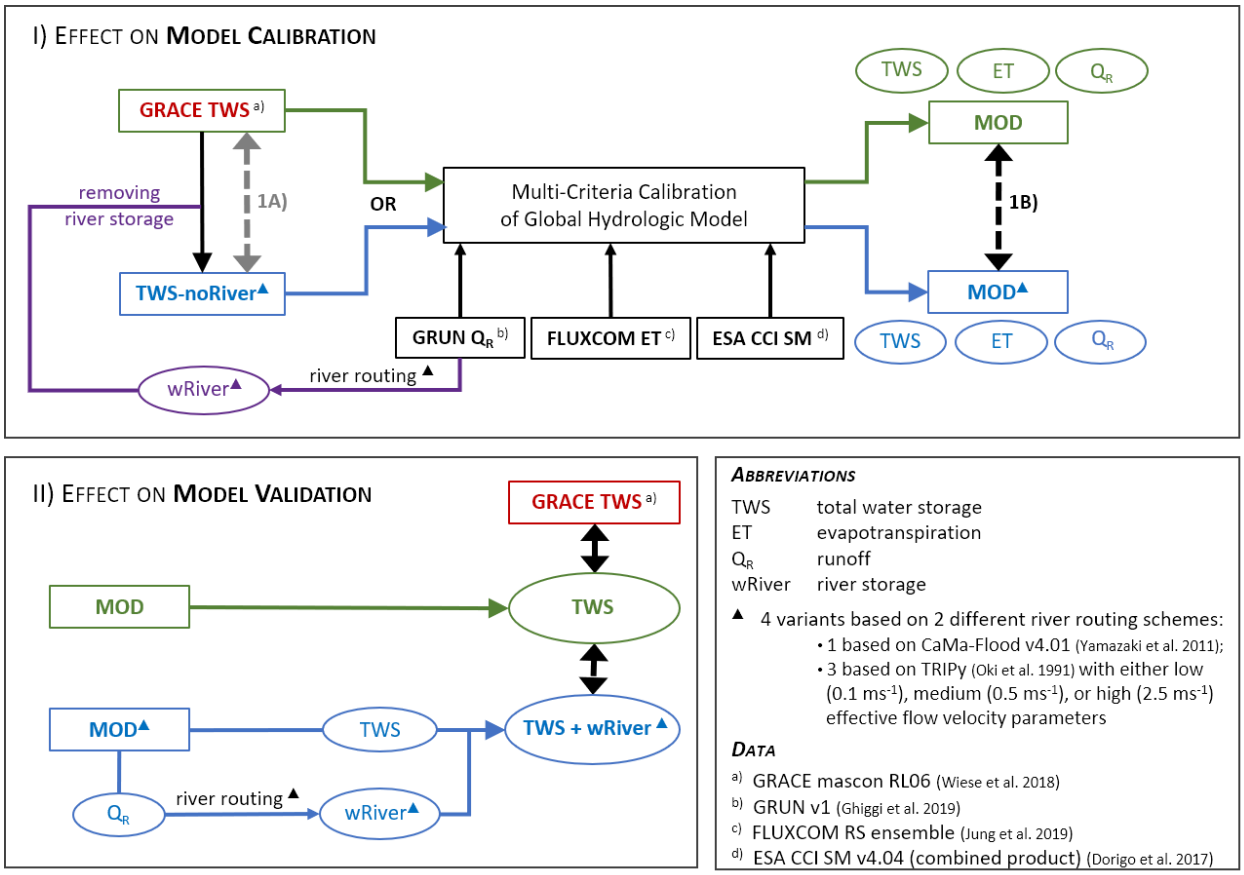

Figure 4.1. Schematic structure of the applied data and methodologies to assess the effect of river storage on model **I)** calibration, and **II)** validation. In **I)** the TRIPy and the CaMa-Flood routing schemes are forced by GRUN Q_R to estimate wRiver, which is then removed from GRACE TWS. Either GRACE TWS or TWS-noRiver data is then used along with other observational data to calibrate a global hydrologic model. In the first step 1A) we compare GRACE TWS against TWS-noRiver to identify the wRiver signal in GRACE TWS data (section 4.3.1.1), and in the second step 1B) simulations from different calibrations are compared against each other to assess the effect of wRiver on model calibration (section 4.3.1.2). In **II)** we force the TRIPy and the CaMa-Flood routing scheme with modeled Q_R from I) and add it to modeled TWS to assess the effect of accounting for wRiver on model validation against the original GRACE TWS wRiver.

4.2.1 Hydrological Model

Exemplary for the variety of global hydrological models, we apply the conceptual hydrologic model introduced in Trautmann et al. 2022. While being more parsimonious than its established counterparts, its structure reflects classical process representation of GHMs and the calibrated model achieves equally good and partially better performance as, e.g., models from the EartH2Observe model ensemble (Schellekens et al. 2017), regarding different observational data (Fig. C.5).

Forced by precipitation, air temperature and net radiation, the model simulates evapotranspiration (ET) and runoff (Q_R), and considers 4 water storages: a snow component, a 2-layer soil water storage, a delayed water storage component, and a deep soil water storage that interacts with the soil and delayed storage components. Simulated total water storage (TWS) is the sum of these 4 storages. While groundwater, surface water and river water are not implemented explicitly, they are assumed to be effectively included in the deep and slow storage components after calibration of associated model parameters against GRACE TWS.

We run the model on a 1°x1° latitude/longitude spatial resolution on daily time steps for the period 03/2000 to 12/2014, focusing on vegetated land area under near-natural conditions.

For regional analysis, we consider hydro-climatic regions obtained from cluster analysis of latitude, mean seasonal dynamics and amplitudes of TWS, ET and Q_R observational data (Fig. 4.2).

Further details are available in Trautmann et al. 2022.

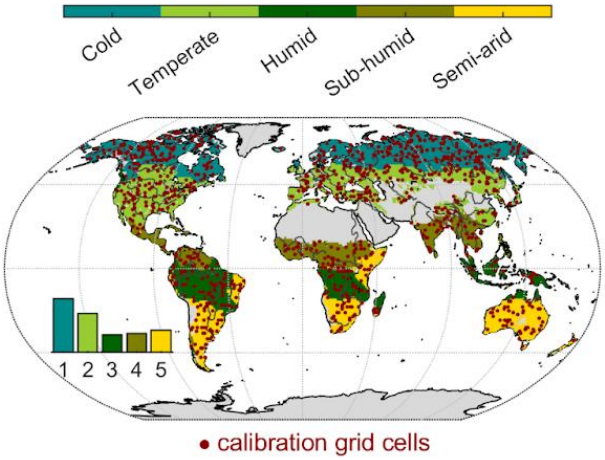

Figure 4.2. Hydro-climatic clusters of the study area with red dots indicating grid cells used in model calibration.

4.2.2 Effect of River Storage on Model Calibration

To assess the effect of river water storage (wRiver) included in GRACE TWS estimates on model calibration, we constrain the model against monthly GRACE TWS variations of the JPL mascon solution (RLM06v2; Wiese et al. 2018), from which estimates of wRiver were removed or not. To do so, we first estimate river storage variations, and then calibrate the hydrological model against GRACE data with or without river storage.

To estimate wRiver variations, we force river routing schemes with observation-based runoff Q_R reconstructions from GRUN v1 (Ghiggi et al. 2019). To do so, the monthly average gridded Q_R estimates are resampled to daily modeling time steps by replicating the monthly value.

Since the choice of the river routing scheme essentially affects simulated wRiver and discharge (Q_{Dis}) (Zhao et al., 2017) we consider 2 different river routing schemes: 1) the simple routing scheme TRIPy (Oki et al., 1999), and 2) the more sophisticated, widely used Catchment-based Macro-scale Floodplain (CaMa-Flood) river routing scheme (Yamazaki et al., 2011).

TRIPy calculates Q_{Dis} from each grid cell along the river network based on Q_R and maps of flow direction and river sequence, using a linear reservoir algorithm. Thereby, the parameter effective flow velocity (*eff_vel*) [ms^{-1}] defines how fast Q_R is discharged from one grid cell to the next, i.e., how long water is stored in the grid cell's wRiver. While in reality, flow velocity varies spatially as it depends on land surface characteristics such as slope, TRIPy uses a globally uniform value for simplicity. To yet assess the sensitivity to *eff_vel* in TRIPy, we consider a range of global *eff_vel* values in different experiments, from low (0.1 ms^{-1}) to medium (0.5 ms^{-1}) to high (2.5 ms^{-1}) values, to derive corresponding estimates of wRiver that produce a range of river storage dynamics with large and fast variability for high *eff_vel*, and small and slow variability for low *eff_vel*.

Similar to TRIPy, CaMa-Flood v4.01 derives the time evolution of water storage from the water balance equation, considering the inflow from upstream grid cells, the input from runoff forcing generated at the respective grid cell, and the outflow to downstream grid cells. However, in contrast to TRIPy, it, next to Q_{Dis}, explicitly calculates flow velocity along a prescribed river network that is automatically generated with the Flexible Location of Waterways (FLOW) (Yamazaki et al. 2009) method. Utilizing a parametrization based on the sub-grid topography obtained from HydroSHEDS, CaMa-Flood simulates

water storage within the river channel, but also water storage in flood plains. Thus, CaMa-Flood allows a more dynamic simulation of Q_{Dis} and wRiver while considering the spatial variability of discharge-generating characteristics.

In the following, the experiments with river routing from CaMa-Flood and the 3 experiments from TRIPy are summarized with ▲, or denoted with *CaMa, 01*, *05*, or *25* if referred to explicitly.

Estimates of wRiver▲ are removed from GRACE TWS to obtain new estimates of TWS-noRiver▲. Before using either GRACE TWS or TWS-noRiver▲ for model calibration, we compare them to assess the contribution of wRiver to the TWS constraint regionally and globally (Fig. 4.1 1A; section 4.3.1.1). For this purpose, we calculate the Nash-Sutcliffe Efficiency (MEF, Eq. 4.1) between GRACE TWS and each TWS-noRiver▲ to quantify their similarity for different spatial scales.

$$MEF = 1 - \frac{\sum_{i=1}^{n}(x_{obs,i} - x_{mod,i})^2}{\sum_{i=1}^{n}(x_{obs,i} - \bar{x}_{obs,i})^2} \quad (4.1)$$

where $x_{mod,i}$ corresponds to TWS-noRiver▲, $x_{obs,i}$ corresponds to GRACE TWS, and $\bar{x}_{obs}$ is the average of GRACE TWS at each data point i.

Either GRACE TWS or TWS-noRiver▲ estimates are then used along with other observational data including GRUN Q_R, FLUXCOM ET (Jung et al. 2019), and ESA CCI soil moisture (Dorigo et al. 2017) to constrain model parameters in a multi-criteria calibration approach. The approach, described in Trautmann et al. 2022, aims to derive the globally best performing parameter set regarding all constraints simultaneously, while considering each data stream's strengths and uncertainties. For each observational constraint we calculate a cost metric that is summed up to a total cost value which is optimized (minimized) using the CMAES algorithm (Hansen & Kern 2004) to derive the globally best performing parameter set. We perform calibration for a spatial subset of grid cells that is obtained by stratified random sampling among Koeppen-Geiger zones. The calibration subset mirrors the global and regional distribution of observed TWS, ET, Q_R, and wRiver, and therefore allows for efficient calibration of parameter values that are globally applicable. To appraise parameter equifinality and uncertainties of the optimization procedure, 10 calibration runs are performed for each experiment.

Finally, we analyze and compare the simulations that were calibrated against GRACE TWS (MOD) and those calibrated against the 4 different TWS-noRiver▲ estimates regarding TWS, ET and Q_R (Fig. 4.1 1B) for global and regional mean seasonal dynamics. While taking into account the spread between 10 different calibration runs of each experiment, we focus on the best performing calibration run when calculating MEF between MOD (x_{obs} in Eq.4.1) and each MOD-R▲ (x_{mod} in Eq.4.1).

4.2.3 Effect on Model Validation

To assess the relevance of wRiver for validation of TWS simulations against GRACE TWS, we add wRiver to model simulations after model calibration. For this purpose, we apply the CaMa-Flood and the TRIPy routing scheme for each calibrated MOD-R▲, i.e., using the calibrated Q_R as forcing for the routing schemes. Thereby, we use the same routing scheme and parametrization as for the TWS-noRiver▲ constraint that was used for the respective calibration (e.g., TRIPy with *eff_vel* of 0.1 ms^{-1} for MOD-R01 that was calibrated against TWS-noRiver-01; and CaMa-Flood for MOD-CaMa that was calibtated against TWS-noRiver-CaMa).

We then add simulated wRiver to TWS of each MOD-R▲ and compare it against the TWS simulations from MOD without additional wRiver, as well as against the original GRACE TWS. For model validation, we calculate MEF between GRACE TWS and MOD resp. MOD-R▲ at the local grid-cell scale, as well as for global and regionally aggregated mean seasonal dynamics.

4.3 Results and Discussion

4.3.1 Effect of River Storage on Model Calibration

4.3.1.1 Effect of River Storage on TWS Patterns

Figure 4.3 compares the decrease in similarity between the original GRACE TWS and estimates of TWS from which wRiver is removed.

As expected, low *eff_vel* increases wRiver, and therefore, the lowest correspondence with the original GRACE TWS can be seen for TWS-noRiver-01, while the similarity increases with increasing *eff_vel*, so that TWS-noRiver-25 is nearly identical with the original GRACE TWS. Spatially, differences between GRACE TWS and TWS-noRiver-CaMa are similar to those from TRIPy with a medium *eff_flow* (TWS_noRiver-05).

Overall, spatial correspondence with GRACE TWS mainly decreases along larger river networks with a large water accumulation (Fig. 4.3a). The largest grid-wise changes are obtained in the *Cold* and *Humid* regions, where river networks are dense and streamflow and, thus, wRiver is large in absolute terms. On the contrary, removing wRiver from GRACE TWS has little effect in the *Semi-arid* region, where water accumulation is smaller than in humid regions.

Regarding seasonal dynamics (Fig. 4.3b), removing wRiver mainly changes the amplitude of TWS variations, which increases in snow affected regions and tends to decrease otherwise. In the *Cold* region, gradual snow melt, retarded infiltration due to shallow and frozen soils, slow discharge to downstream areas, as well as additional water input from upstream areas in large river networks dampen TWS variations. Removing wRiver attenuates these delaying effects, causing increased TWS variations and shifts the TWS peak to one month earlier. While in the *Cold* region seasonal TWS variations are mainly affected by snow accumulation and melt, TWS in other regions is dominated by liquid water storages (Trautmann et al. 2018). In non-snow affected regions, Q_R increases with wetness, i.e., with TWS, which in turn reduces TWS. Due to this negative feedback, removing wRiver from TWS reduces the TWS amplitude. Due to the spatial variability of parametrization and flow velocity in CaMa-Flood, TWS-noRiver-CaMa agrees with different TRIPy TWS-noRiver estimates in different regions, e.g., rather with high *eff_vel* parametrization in *Temperate*, *Humid* and *Sub-humid* regions, but with low *eff_vel* parametrization in the *Semi-arid* region.

4.3.1.2 Effect of River Storage on Model Calibration

Figure 4.4 shows the mean seasonal dynamics of TWS, ET and Q_R simulated by MOD and MOD-R$^{\blacktriangle}$ after model calibration. Respective observations are plotted for better evaluation of the calibration results, yet the following focuses on differences between MOD and MOD-R$^{\blacktriangle}$. A detailed evaluation of performance of MOD against TWS, ET and Q_R observations can be found in Trautmann et al. 2022.

Despite being calibrated against either GRACE TWS or TWS-noRiver$^{\blacktriangle}$, overall little variance between calibrated MOD and MOD-R$^{\blacktriangle}$ is evident. While we expected the model parameters to adapt to differences in the TWS constraint, the mean seasonal TWS simulations are nearly identical among experiments, globally and regionally. The same holds for ET. Some differences between calibration runs are obtained regarding Q_R, especially for *Cold*, *Temperate* and *Semi-arid* regions. However, while the spread is larger than for the other simulated variables, the differences of the best performing calibration runs are still negligible.

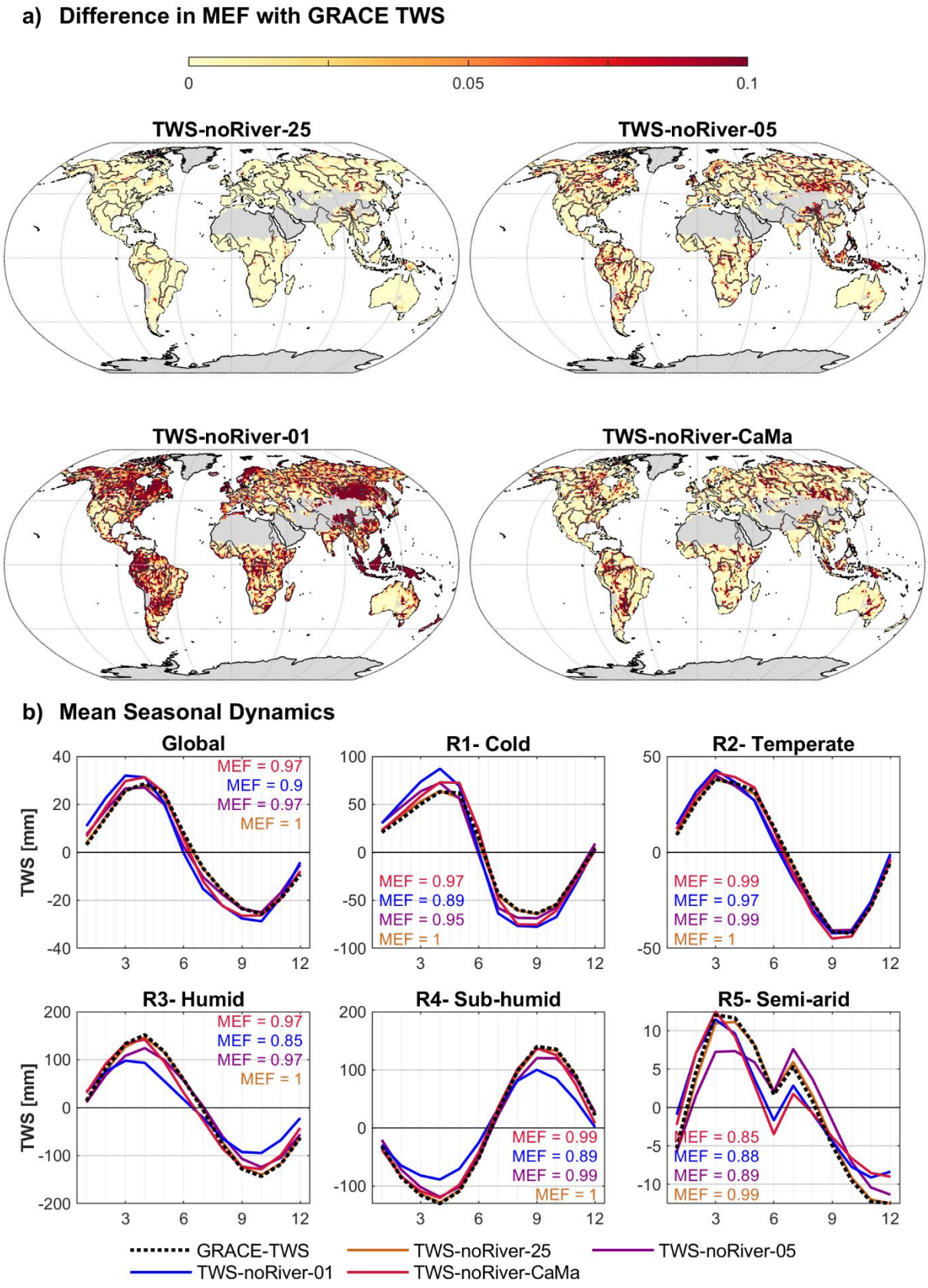

Figure 4.3. Comparison of GRACE TWS and TWS-noRiver▲. **a)** similarity between the original GRACE TWS and TWS-noRiver▲ in terms of gridwise Nash-Sutcliff efficiency (MEF); **b)** global and regional mean seasonal dynamics of GRACE TWS and TWS-noRiver▲. For **a)** and **b)** MEF of TWS-noRiver▲ with the original GRACE TWS is calculated and subtracted from the optimal MEF of 1. Note the different ranges on the y-axis for TWS variations in panel **b)**.

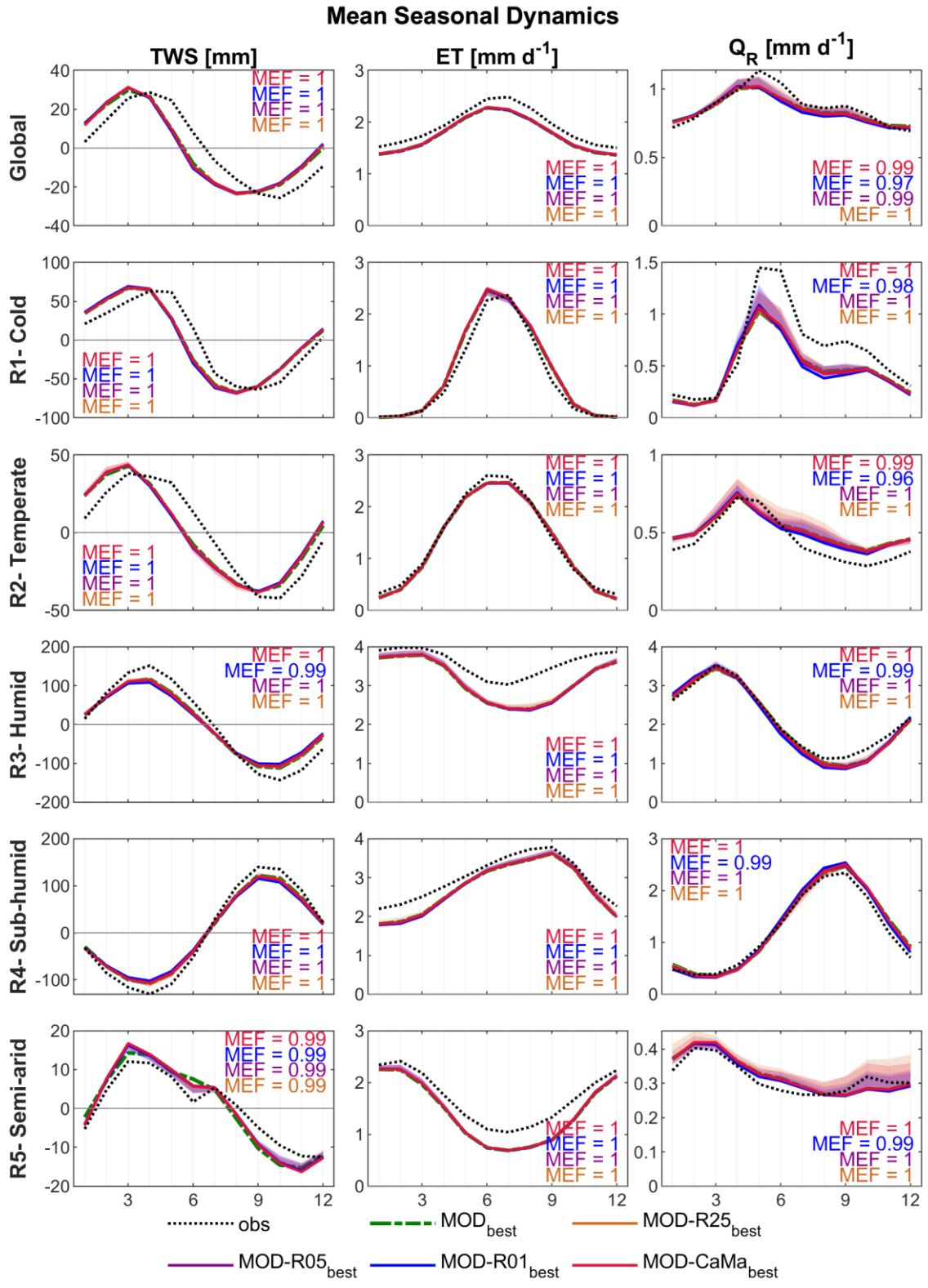

Figure 4.4. Mean seasonal dynamics of simulated and observational TWS, ET, and Q_R summarized globally and for different regions. Uncertainty bands denote the spread between 10 different calibration runs of each experiment, while dashed/solid lines indicate the best performing calibration run of each experiment. Black dotted lines denote observational GRACE TWS, FLUXCOM ET, and GRUN Q_R. Listed MEF compares the best performing calibration run MOD▲ against the best performing calibration run of MOD (dashed line).

High agreement in simulated fluxes and TWS variations goes along with no systematic differences between calibrated parameter values of MOD and different MOD-R▲. Hence, we do not find any evidence for biases between experiments that result from either considering wRiver or not in model calibration. While the calibrated parameter values don't vary significantly between different experiments, different calibration runs point to two parameter sets that achieve (nearly) equal good performance (Fig. C.1). Affected are parameters that regulate the size of soil water storage and it's depletion by ET. The interplay of these parameters with the other outgoing flux Q_R also explains the spread of simulated Q_R. This parameter equifinality is not related to wRiver, but to the equifinality of baseflow and ET decay, especially under water limitation as discussed in Trautmann et al. 2022.

The absence of a qualitative effect of using GRACE TWS or TWS-noRiver▲ for model calibration is also evident when inspecting the total costs and cost components among the experiments (Fig. C.1). They are fairly similar on average, while the spread of costs between different calibration runs of one experiment tends to be larger than the differences between experiments. This underlines that uncertainties arise mainly from other aspects of hydrological modeling than from the effect of wRiver included in the TWS constraint. However, we do see a tendency for elevated TWS costs for MOD-R01 and MOD-R05. Higher TWS costs and thus total costs suggest difficulties of the model to adapt to the TWS constraint, when the removed wRiver is large. This suggests that the comparatively large removal of wRiver is harder to reconcile with the other constraints of ET and Q_R, and thus may indicate that such large wRiver based on low *eff_vel* is not plausible. While it is notable that the spread in total, TWS, and Q_R costs of MOD-CaMa is comparatively smaller and it achieves the overall lowest total costs due to low soil moisture and ET costs, the general difference to the other calibrated models remains small, indicating no significant impact of the chosen routing scheme on the global multi-criteria calibration.

4.3.2 Effect of River Storage on global Model Validation

While we did not find systematic differences between TWS simulations of MOD and MOD-R▲ after calibration, explicitly accounting for wRiver and adding it to TWS of MOD-R▲ causes systematic differences of model performance against the original GRACE TWS, locally as well as for regional and global mean seasonal dynamics (Fig. 4.5). The largest differences are notable for considering wRiver based on low *eff_vel*, while the largest improvement of model performance relative to the original GRACE TWS is achieved when adding wRiver based on medium *eff_vel* - globally and in most regions (Fig. 5b). While the choice of the routing scheme does not significantly affect model performance regarding seasonal dynamics of TWS for large spatial regions (Fig. 4.5b), it is relevant for simulating hydrological dynamics at smaller, e.g., catchment, scale. As such, Q_{Dis} from MOD-CaMa provides consistently good estimates of Q_{Dis} at various GRDC stations, while different MOD-R▲ from TRIPy perform better for different stations, highlighting the benefit of spatially distributed river flow parameters as opposed to global average parameter values of *eff_vel* in TRIPy (Fig. C.6). However, locally, the differences in model performance with GRACE TWS are less pronounced when wRiver from CaMa-Flood is added (MOD-CaMa). Overall, including wRiver improves TWS simulations at local scale especially in the tropics and Northern low- and wetlands where rivers accumulate water over large catchments (Fig. 4.5a). The importance of wRiver in the tropics has already been shown by previous studies (Getirana et al. 2017). Similarly, the inability to reproduce observed TWS variations by models in the *Cold* region is among others attributed to missing representations of floodplain and river flow processes (Kim et al. 2009). Indeed, in the *Cold* region, accounting for wRiver improves MEF for the majority of grid cells (Fig. 4.5a), highlighting the importance of additional inflow from upstream grid cells and the delay of water outflow in these regions. While MOD-R01 matches the timing of TWS variations slightly better, it underestimates the seasonal TWS amplitude (Fig. 4.5b). On the contrary, MOD-R05 and MOD-CaMa reproduce TWS amplitude, yet still precede TWS variations, although not as much as MOD or MOD-R25. Hence, the phase-shift issue of simulated TWS in *Cold* regions, which is prevalent in many GHMs (Schellekens et al. 2017), is unlikely to arise from unaccounted river storage variations only. The underestimation of TWS amplitude (and peak spring discharge, Fig. C.6)) can also be affected by deficiencies in the precipitation forcing (Huffman et al. 2001, Contractor et al. 2020), but remaining difficulties in reproducing the timing of TWS dynamics indicate the

potential importance of other missing processes such as freeze/thaw dynamics and permafrost (Yu et al. 2020), and ice jam in river channels (Kim et al. 2009).

The preceding of simulated TWS compared to GRACE TWS can also slightly be reduced at global scale, when wRiver is considered (Fig. 4.5b). However, a preceding of simulated TWS variations is still apparent for large areas, especially the *Temperate* region (Fig. 4.5b), indicating again the relevance of other processes than water delay in wRiver, such as irrigation, land cover changes and interactions between groundwater and surface water dynamics.

While including wRiver improves agreement with GRACE TWS over large areas, MEF decreases notably in the *Semi-arid* region when the TRIPy routing scheme is used (Fig. 4.5). The slightly better performance of MOD-CaMa in semi-arid regions locally (Fig. 4.5a) as well as for the average seasonal dynamics in the *Semi-arid* region (Fig. 4.5b) indicates the benefits of the variable flow velocity and the accounting of flooding in CaMa-Flood, which is relevant in such regions that are characterized by rain- and dry-seasons. However, especially the regional average dynamics are not notably different from the other simulations by TRIPy and the local improvement remains low. Already MOD does not agree well with GRACE TWS in the *Semi-arid* regions (Fig. 4.5a), where TWS variability is sensitive to parameters controlling ET under water limited conditions, which are poorly constrained (Trautmann et al. 2022). In addition to the parameter equifinality issues, poor (initial) performance points to model structure uncertainties and deficiencies, such as missing processes of evaporation and percolation to groundwater from open water surfaces. Besides, the GRUN Q_R constraint is known for larger uncertainties in arid regions (Ghiggi et al. 2019) and tends to inconsistencies with the other observation-based data in the *Semi-arid* region (Trautmann et al. 2022). Therefore, comparatively good MEF of modeled and observed Q_R (Fig. 4.4) does not necessarily reflect good representation of Q_R, which is also underlined by poor representation of observed discharge at semi-arid GRDC stations (Fig. C.6). As modeled Q_R is used to derive modeled wRiver, this causes poorer model performance in semi-arid regions when wRiver is added to TWS.

Overall, a general improvement of TWS when wRiver is added to modeled TWS, except for in semi-arid regions, highlights potential room for improvements in our approach of modeling the hydrology in semi-arid regions. Among others, the improved model performance by using CaMa-Flood in such regions indicates the importance of seasonal flooding.

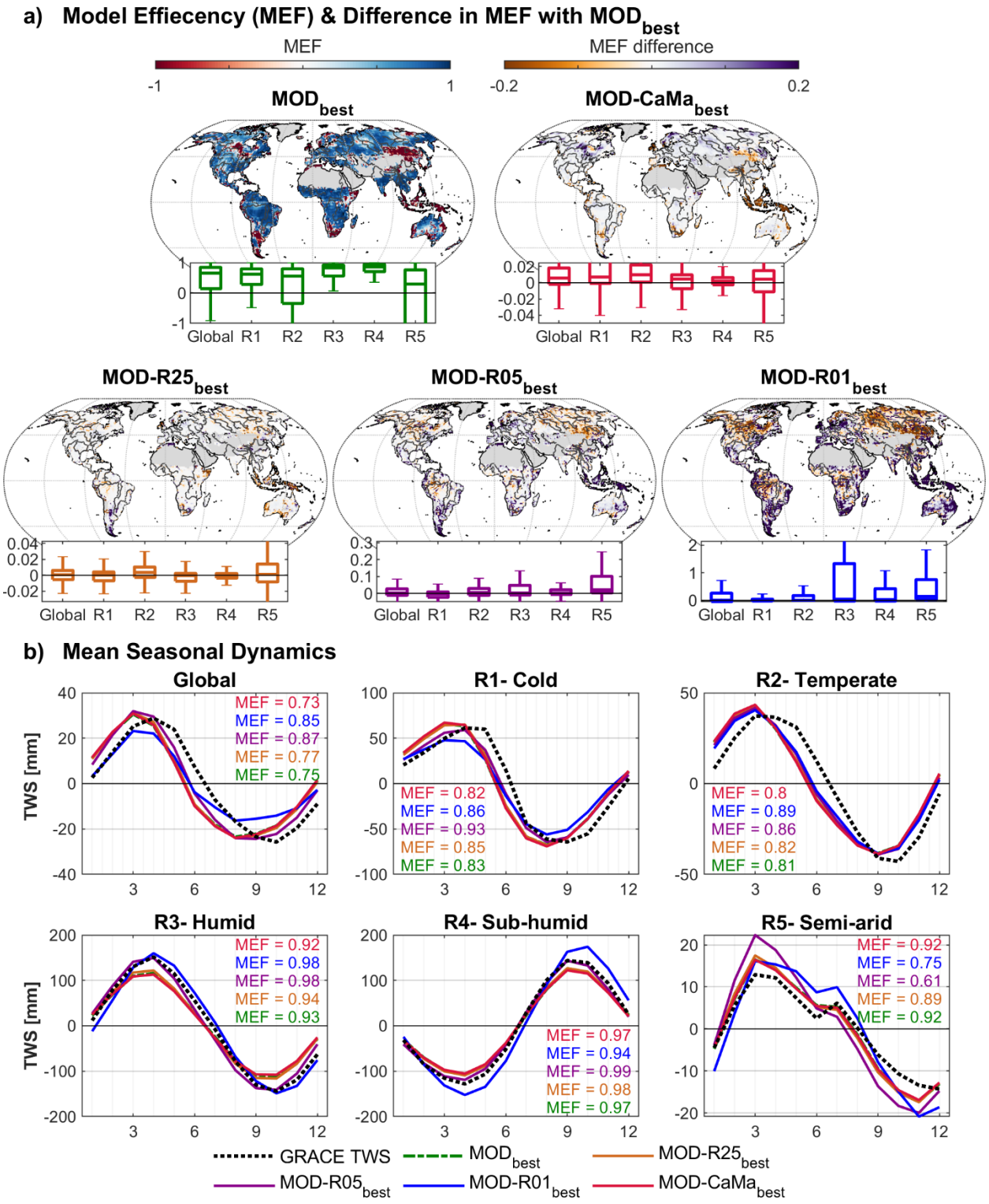

Figure 4.5. Comparison of the original GRACE TWS, TWS simulated by MOD and TWS+wRiver simulated by MOD▲. **a)** gridwise Nash Sutcliff efficiency (MEF) between GRACE TWS and MOD (upper left map), and differences of MEF obtained by MOD and MEF obtained by MOD-CaMa, MOD-R01, MOD-R05, MOD-R25, respectively; including their global and regional distributions (outlier are not plotted). **b)** mean seasonal dynamics averaged globally and for different regions. Listed MEF compares MOD resp. MOD▲ with the original GRACE TWS.

4.4 Implications

We showed that river storage has a relevant impact on seasonal TWS dynamics, in particular in cold and humid regions, and accounting for it when simulating TWS improves performance against GRACE TWS observations. However, we did not find a systematic impact of either including or excluding river storage in TWS for global-scale model calibration. Compared to river storage, restrictions from other observational constraints seem more relevant to define model parameters in our model-data integration approach. For example, the main discrepancies may be related to missing processes (especially in cold and semi-arid regions), and issues of parameter identifiability due to insufficient and partly conflicting data constraints. Our findings hold across sensitivity experiments with different routing schemes and effective flow velocity parameters that produce a wide range of river storage dynamics. While we do not argue that river routing is of relevance at local and smaller regional scales, the impact at larger regional to global scales vanishes. Especially when using GRACE TWS for model calibration, the effect of small rivers and less dense river networks is likely smoothed out by its 250 km native resolution (Wiese et al. 2018).

While we cannot exclude that these findings are conditional on the specific model structure, calibration approach and data used in this study, we argue that our findings are of general relevance for global hydrological modeling studies across a spectrum of GHMs that ignore or include river routing schemes.

The model used in this study is based on classic hydrologic process representation and despite its simple structure achieves good performance that is comparable to more complex state-of-the-art GHMs (Fig. C.5). The identified key issues of model-data mismatches in cold and semi-arid regions are also shared among GHMs in general (Schellekens et al. 2017) and appear unrelated to river storage variations. The identified problem of parameter and thus process identifiability is due to insufficient observational constraints, and this problem is expected to be even larger for more complex models with more parameters. Given limited computational resources, it seems advisable to save the comparatively large costs of river routing schemes when calibrating against GRACE TWS observations in order to address the bigger issues more efficiently. Computational costs can be saved significantly, not only by omitting the routing itself but also because we can subsample grid cells for calibration instead of demanding a full global simulation in each calibration iteration, which would be required when simulating river routing.

Our findings may be of particular relevance for future global hydrological modeling in the era of Earth observations, that support modeling approaches which complement calibration against a single river discharge constraint, that is often regulated by complex anthropogenic influences, by multiple large-scale constraints from GRACE TWS and other observational data streams on complementary water fluxes and storages.

Chapter 5

Synthesis

5.1 Summary of the Main Findings

This thesis aims to contribute to our understanding of global water storage variations and their composition by combining simple modeling approaches with observations of terrestrial water storage (TWS) variations by GRACE and complementary Earth-observation based data sets. To this end, a model-data integration approach, that calibrates simple hydrological models against multiple observational data constraints simultaneously, is developed. This approach is used to address three aspects that are relevant for understanding and modeling the composition of global TWS variations. In **Chapter 2**, the focus is on the contribution of snow versus liquid water storages to TWS variations in cold regions across spatio-temporal scales. **Chapter 3** analyses the impact of how vegetation is represented within a global hydrological model on model performance and simulated TWS composition. Finally, **Chapter 4** investigates the effect of river water storage that is included in GRACE TWS observations on model calibration and validation in such a global model-data integration approach.

In the following, the main findings of this thesis are summarized with respect to the research questions framed in the Introduction. First, the specific research questions are answered, and thereafter put in context regarding the overarching objective of this thesis. Subsequent sections discuss these findings, including the limitations and potentials of the underlying approach, give prospects for future research, and finally provide a general conclusion of this thesis.

5.1.1 What Storage dominates spatio-temporal Patterns of TWS Variations in Northern Latitudes?

Consistent with previous studies, **Chapter 2** and **Chapter 3** identify the seasonal accumulation and melt of snow during winter months as the dominant control on seasonal TWS variations in most places of the Northern mid-to-high latitudes. However, while snow dynamics control seasonal TWS variations from local (grid-cell) to regionally and globally aggregated scales, liquid water storage variations are found to determine the inter-annual TWS variability across all spatial scales. This is caused by larger inter-annual anomalies of rain fall than of snow fall. Despite these general findings, **Chapter 2** shows that the relative contribution of snow to inter-annual TWS variability increases as the spatial domain over which storages are aggregated increases. This can be attributed to the stronger spatial coherence of snow dynamics that are mainly driven by temperature, that itself features strong spatial coherent patterns. In comparison, highly heterogeneous rain fall patterns determine liquid water anomalies, so that positive and negative anomalies of these storages are cancelled out by aggregating over larger spatial domains.

Chapter 3 additionally allows to distinguish liquid water storages of plant-accessible soil moisture, deep soil water, and slowly varying, delayed water storages. These analyses reveal that, among liquid water storages, the deep soil water and the slowly varying water pools are most influential. Although the relative contribution of river water storage to TWS variations is quantified, **Chapter 4** further highlights the importance of water delay in river channels and the additional water input provided from upstream areas in the large catchments for reproducing (seasonal) TWS variations in cold Northern regions.

Overall, different chapters of this thesis highlight that the determinants of TWS variations are scale-dependent, in Northern latitudes as well as globally. Accordingly, the driving mechanisms of TWS variability cannot simply be transferred from one scale to another, and especially pattern from large-scale integrated signals should not be associated with locally operating processes due to the confounding

impact by spatial covariations of climate variables. This fact is of particular relevance for assessing the short- and long-term variability of water resources.

5.1.2 What's the Impact of spatio-temporal Vegetation Parameters on Model Performance and global TWS Variations?

To analyze the effect of varying vegetation characteristics on model performance and TWS partitioning, different methods to parametrize the maximum plant available soil water storage, and the processes of evapotranspiration and runoff generation are implemented within the developed model-data integration approach. Exemplary, **Chapter 3** compares two model variants with the same underlying structure: a model variant B, in which globally uniform parameter values are calibrated, and a model variant VEG, in which scalars that amplify or attenuate the pattern of continuous vegetation information are constrained in the model-data integration approach. Besides, an additional variant follows the 'traditional' approach of representing vegetation by distinguishing between different classes of plant functional types (PFT) and calibrating different parameter values for each class (**Appendix B**).

Despite their alternative ways to represent vegetation, all model variants perform reasonably well regarding major hydrological variables after model calibration. However, including varying vegetation data improves model performance at regional and local scale. Largest improvements were found in tropics (evapotranspiration), where dense vegetation cover and high activity throughout the year certainly require different parametrization than in other regions, and in supply-limited regions (TWS, evapotranspiration), that as well differ from humid regions due to the adaption of vegetation to regular dry conditions. Besides, spatial and temporal pattern of the Enhanced Vegetation Index, that are included in the VEG model variant, implicitly represent effects of varying vegetation activity due to irrigation, and thus allow to indirectly account for processes not directly incorporated in the simplified model structure, leading to improved simulation of evapotranspiration and water storage changes. Additionally, including spatio-temporal varying vegetation data leads to more reasonable parameter values, while simultaneously reducing parameter equifinality. This suggests a more realistic representation of water fluxes and storages when continuous fields of vegetation data are used for model parametrization. Therefore, defining spatio-temporal varying vegetation parameters based on continuous vegetation characteristics is preferable – not only compared to globally uniform parameter values, but moreover also compared to the traditional PFT approach, because the larger number of parameters that results from defining parameter values for each PFT classes is not in relation to the improvement in model performance, but rather increases parameter uncertainty due to over-parametrization.

While all model variants maintain the same overall response of total hydrological fluxes and storage variations, the way how vegetation is represented in the model changes the spatio-temporal dynamic of individual hydrological processes and storage components considerably, and substantial differences between PFT, B and VEG in terms of TWS composition underline that the representation of vegetation is critical for interpreting and partitioning TWS. Along with **Chapter 2**, all model variants identify snow as dominant factor for seasonal TWS variations – locally in Northern latitudes, as well as globally and in most regions. However, including spatio-temporal varying vegetation parameters changes the contribution of different liquid water storage components to TWS drastically: while with globally uniform parameter values one storage component appears to be sufficient and all TWS variations other than by snow are caused by soil moisture, the deep, not directly plant-accessible water storages, and delayed water storages become prominent, when including varying vegetation characteristics. Thus, in particular the representation of water-limited evapotranspiration by the interplay of its sensitivity to soil moisture, maximum plant-accessible water storage capacity and interactions with deep soil moisture and groundwater appears to play a critical role in TWS partitioning.

Beside the improved understanding of TWS composition, **Chapter 3** further highlights the relevance of slowly varying storages for reproducing runoff dynamics in cold regions, by including spatially varying soil water capacities as defined by observation-based data and replacing the runoff estimate for Europe with a global runoff estimate in model calibration.

By that means, spatio-temporal varying vegetation parameters are undoubtedly relevant for global and regional hydrological simulations. Moreover, the representation of vegetation-related processes is found to be a key factor that controls the partitioning of TWS among different water storage components. Hence, **Chapter 3** emphasizes the need for further studies and improvements of GHMs with respect to the role of vegetation.

5.1.3 What's the Effect of River Water Storage included in the GRACE TWS Estimates on Calibration and Validation of a simple Global Hydrological Model?

To assess the effect of river water storage on model calibration in **Chapter 4**, a widely used and a simple routing scheme are forced with observation-based runoff estimates and the derived river water storages then removed from GRACE TWS variations. Either GRACE TWS or TWS estimates without river storage are then used to constrain the VEG model (introduced in **Chapter 3**) in the developed multi-criteria calibration approach. While removing river water storage considerably changes the TWS constraint regionally and globally, there are no significant implications of either including or excluding river storage on calibrated model parameters, and thus neither on resulting model simulations. Moreover, the constraints on other water balance components by complementary observational data (i.e., evapotranspiration and runoff) are found to be more influential on defining model parameters. Therefore, the effect of river water storage included in GRACE TWS is found to be negligible when using multiple observational data sets for global model calibration, and it seems not necessary to invest in computational expensive river routing during a model calibration that requires thousands of model runs. This additionally holds the benefit that it's valid to use a subset of spatially independent grid cells (i.e., not necessarily connected by a river network) for model calibration. Instead of investing computational resources in river routing, **Chapter 4** identifies the determinability of certain model parameters and associated processes of greater relevance for similar hydrological studies.

However, in contrast to model calibration, **Chapter 4** shows a clear effect of including river storage when hydrological simulations are evaluated against GRACE TWS. Adding river storage to simulated TWS after model calibration improves model performance regarding GRACE TWS regionally and globally, especially when river storage is derived by either a more sophisticated routing scheme or a simple routing scheme with a medium effective flow velocity of rivers. Main regions of improvement are the tropics and Northern low- and wetlands, where rivers accumulate much water over large catchments. These findings highlight once more the role of river and surface water storages in tropical regions (see **Chapter 3**, Getirana et al. 2017), as well as the importance of water delay in river channels and of the additional water input from upstream areas in the large catchments in cold regions (see **Chapter 2**). Accordingly, **Chapter 4** stresses the relevance of explicitly simulating river water storage when evaluating the performance of large-scale hydrological models against GRACE TWS.

5.1.4 How can Earth-observation based Data and simple conceptual hydrological Modeling Approaches improve our Understanding of TWS Variability?

As underlying framework for all chapters of this thesis I developed a model-data integration procedure, which includes low-complexity models that are based on a combination of standard hydrologic model formulations, and a sophisticated multi-criteria calibration approach that constrains model parameters against several observed variables simultaneously, while considering each data streams strengths and uncertainties. The effectiveness of this model-data integration approach is first proven in **Chapter 2**. Despite the model's low complexity, and the calibration of a uniform parameter set for a random spatial subset of grid-cells, this approach improved model performance over the entire mid-to-high Northern latitudes regarding several observed variables. Likewise, it achieved equally good or better performance and similar spatial pattern regarding TWS and SWE compared with the eartH2Observe ensemble of more complex state-of-the-art hydrological models (Schellekens et al. 2017).

In **Chapter 3**, the approach is extended from the Northern latitudes to the global land area under near-natural conditions, and the additional value of using Earth-observation based data to represent vegetation characteristics is shown. On the one hand, the multi-criteria calibration approach enables different model

variants to reach relatively good performance despite major differences in model parametrization. At the same time, including continuous fields of observation-based data to define spatio-temporal pattern of model parameters reduces parameter equifinality and uncertainty while keeping the number of calibration parameters low, and making a categorical classification of the land surface redundant. Hence, it is found to be preferable to the parametrization with globally uniform static values or with values for distinct landcover classifications. On the other hand, additionally including the spatially continuous data on vegetation characteristics allows to depict the interaction of vegetation with deep soil and groundwater during dry down periods – a process that wasn't accessible without applying the observational data and which changes the contribution of different water storages to overall TWS variability. By that means, integrating this observation-based data revealed that the representation of vegetation-related processes is a key factor for controlling simulated TWS partitioning, and allows to differentiate between (sub-surface) water storages that are unidentifiable by observational data alone.

Already by using a simple approach for river routing in **Chapter 4** TWS simulations could further be improved regionally, and thus suggests a considerable contribution of river water storage to TWS variability at local and regional scale. However, at the same time, **Chapter 4** shows that restrictions from other constraints make the implementation of a computational expensive routing scheme in – at least – model calibration redundant, as the relative effect of river water storage is found to be weaker than the confinement by the other observations.

Overall, throughout this thesis, the value of large-scale observation-based data to improve simulations of TWS variations and their partitioning in hydrological models using a multi-criteria calibration approach has been proven. Considering multiple complementary data sets on the one hand allows to infer to processes that are not explicitly represented in the model and reduce parameter uncertainty and equifinality. On the other hand, the multi-criteria calibration approach helps to overcome the data's individual shortcomings, and prevent overfitting of the model to a single, uncertain data set. By that, this concept helps to overcome data inconsistencies, which are a considerable issue when combining multiple data sets from various sources. All this underlines the potential of simple hydrological models that are tied to observational data streams as powerful tools to diagnose and understand large-scale water cycle patterns. In particular, the used model-data integration approach improves confidence in simulated TWS variability and partitioning, emphasizes the scale-dependencies of TWS determinants, highlights the representation of vegetation as a key factor that controls TWS partitioning, and shows the important role of the interaction of sub-surfaces water storages, as well as river and surface water storages for TWS dynamics.

5.2 Discussion and Prospects

As any research that applies models and data, the presented results and drawn implications are subject to certain limitations regarding their validity and transferability to other research frameworks. The following discusses such limitations that arise from the used model-data integration approach. The potential and restrictions of the presented work are elaborated in more detail by means of two challenges that emerged in several chapters of this thesis. Finally, the last section summarizes potential solutions to counteract the discussed limitations and shows opportunities for future research.

5.2.1 Limitations and Potentials of the Model-Data-Integration Approach

Limitations of the model-data integration approach emerge from several factors, such as the used data sets, the underlying model structures and their parametrization, as well as the implemented calibration approach. However, some of these limitations offer specific opportunities of such a model-data integration concept as opposed to traditional global hydrologic modeling.

Data

Since the implemented approach is by definition tied to observational data, its merits rise and fall with the used data sets. While the strengths and uncertainties of forcing and calibration data depend on the specific variable and product used, the common limitation is the limited data availability. Large-scale

observational data are only available for recent decades, so that the applicability of such a model-data integration approach is restricted to this period. Additionally, the short temporal coverage restrains the delineation of trends, and observations cannot monitor processes that occur over long time periods, such as the long-term changes in surface and groundwater storages and the melting of glaciers and permafrost. Therefore, the approach is unsuitable to resolve such long-term processes, and its advantages lie in diagnosing and understanding processes that are relevant on shorter time scales for which observational data are available. When combining data sets from various sources that are based on different methods and assumptions, impacts of possible data inconsistencies need to be considered. An in-depth analysis of data inconsistencies, as provided in **Appendix B**, allows to emphasize certain data sets and re-evaluate desirable and less desirable model behavior. In this context, considering multiple data sets on complementary variables simultaneously helps to identify inconsistent behavior of one and to mitigate its effect on model calibration and resulting simulations, and thus represents one major advantage of the used model-data integration approach.

Model Structure

Since one of the main objectives of this thesis is to investigate the potential of combining Earth-observation based data with simple conceptual modeling approaches, the implemented models have a low complexity and parsimonious process representation. The focus is on major hydrological processes and storages that can (in-) directly be constrained by data, in order to increase interpretability and confidence in model simulations. Hence, the used models cannot directly be compared to state-of-the-art large-scale hydrological models, that include a profound process-representation, in terms of their applicability and level of detail. Among potentially relevant, yet not represented processes, are such related to permafrost, permanent ice and snow cover, and human alterations of the water cycle, such as irrigation, groundwater withdraw and the creation of water reservoirs. To minimize effects of such processes on the results, certain grid-cells are excluded from all analyses. Nevertheless, despite these precautions, not all effects resulting from associated processes can be excluded completely, and, thus, potentially influence model simulations, either in terms of decreased performance or by being implicitly included in simulated fluxes and storages after model calibration.

Model Parametrization

Next to deficiencies in model structure, certain limitations are imposed by model parametrization. Parameters in conceptual models, such as the models used in this thesis, are usually not physically measurable (and such measurements in most cases would neither be available at global scale), and thus require adjustment by model calibration. Since in the introduced multi-criteria calibration approach only globally uniform parameter sets are calibrated, these parameters have to balance the spatial and temporal heterogeneity across multiple physio-geographic and climatic conditions on the one hand, and compensate for missing processes on the other hand in order to achieve the best possible fit with observational data. In **Chapter 2** no spatial or temporal variation of calibrated parameter values is included, yet the focus is on Northern mid-to-high latitudes, i.e., regions with similar main driving hydrological mechanisms. When extending to the global scale in **Chapter 3** and **Chapter 4**, spatial and temporal patterns from data are used for model parametrization, and a global constant scalar for this pattern is calibrated, what keeps the computational demand of the calibration process relatively low. By that, the parametrization is still tied to the pattern of the integrated data, but a clear improvement in model performance and decrease in parameter uncertainty is achieved. While this suggests that including continuous observational pattern improves model parametrization, remaining deficits due to parameter equifinality are identified in **Chapter 3** and **Chapter 4**. Such effects of deficient observational constraints on parameter and process identifiability are presumably even larger for more complex models, that include more processes, wherein this very complexity impedes the attribution of such deficiencies.

Multi-criteria Calibration Approach

Compensatory effects are not only introduced in model parametrization by applying and calibrating globally uniform parameter sets, but also emerge from optimizing one total cost value in the multi-criteria calibration approach. Doing so, for instance, less weight is put on fluxes and grid-cells with low variability,

and relatively high costs regarding one constrained can be balanced out by improving the costs regarding another constraint. This is partly counteracted by the choice of efficiency metric that defines the cost component, but a more sophisticated analysis of the Pareto front to evaluate the trade-offs between costs for different constraints, and, based on such analyses, a more sophisticated weighting of individual cost components has the potential to decrease this issue. Next to that, the presented approach calibrates parameter values for a spatial subset of grid-cells, that by their hydro-climatic and physio-geographic characteristics, substantially influences the calibrated parameter values and their applicability for other regions (**Chapter 4**, **Appendix C**). In order to ensure that the subsets represent the global distribution of major hydrological fluxes and storage variations, the calibration grid-cells are selected based on a stratified random sampling among Koeppen-Geiger climate zones. While this is an appropriate method for a global analysis such as in this thesis, a different sampling of calibration grid-cells might be appropriate if the focus is on other purposes.

Despite these points that must be taken into account, a general advantage of such a multi-criteria calibration approach is, that it allows to mitigate the impact of uncertainty and inconsistency of and between data sets, to overcome limited model structures, and solidify the confidence (at least) in the calibrated variables. In this regard, the calibrated models presented in this thesis achieve equally good or better performance compared to more complex hydrological models, despite their simple process representation and the calibration of a globally uniform parameter set.

5.2.2 Challenge 1: Preceding of observed seasonal TWS Variations in cold Regions

In all chapters of this thesis, model simulations precede the peak in seasonal TWS variations as compared to GRACE by appr. 1 month, especially in cold and temperate regions. A comparable (spatial) pattern exists in many other models and modeling studies (Döll et al. 2014, Schellekens et al. 2017). In general, this mismatch is attributed to an inadequate size and number of water storages (Kim et al. 2009), and as largest differences exist in cold regions, Schellekens et al. 2017 proposed the implementation of snow associated processes. In this context, also the temporary storage of meltwater during spring in rivers and surface water bodies, which occurs contiguously over large areas in mid-to-high latitudes and delays the overall storage decay, is mentioned (Döll et al. 2014, Schellekens et al. 2017, Kim et al. 2009, Schmidt et al. 2008b). To increases the confidence in modelled snow processes, the model is constrained against observations of snow water equivalent in **Chapter 2**. However, in the resulting simulations the TWS phase lag is still apparent, and, similarly, no relationship between model performance and model complexity is found. Based on this, other dynamics appear to be more important, even in snow-affected regions. As such, **Chapter 4** applies river routing and considers water delay in river storages. In fact, this reduces the discrepancy between observed and simulated TWS in cold regions, highlighting the importance of the delay of snow melt in rivers, and of additional meltwater input from upstream grid-cells for TWS dynamics. However, the mismatch is not completely resolved, and still evident in temperate regions. Therefore, even other processes and, in particular, their interaction are relevant. In cold regions, such processes may include the effects of permafrost and freeze/thaw dynamics (Yu et al. 2020), as well as ice jam in river channels that increases surface water storages temporarily (Kim et al. 2009). Besides, land cover changes, interactions between groundwater and surface water storages, as well as human alterations, such as irrigation, potentially effect seasonal TWS variations. While a first comparison with irrigation area did not find any relationship, further data is needed to attribute the preceding of seasonal TWS. As such, a representation of the effects from permafrost would be desirable. Besides, data of surface water storages, such as from the forthcoming SWOT mission, or by daily GRACE TWS estimates (Eicker et al. 2020), might help to detangle the short-term imprints of slow and fast water storages that are lumped in the monthly GRACE resolution, and by that identify possible lacks of process representation.

5.2.3 Challenge 2: TWS Variability in semi-arid Regions

Semi-arid regions are especially prone to droughts and are among the regions that are expected to experience large changes due to climate change. Therefore, it is essential to properly assess water resources and their variations in these regions. However, several chapters of this thesis show difficulties

to reproduce hydrological processes especially in semi-arid regions. While including river routing in **Chapter 4** improved or didn't affect model performance in most regions, the representation of TWS in semi-arid regions decreased when the simple routing scheme was used, and a trade-off between evapotranspiration and runoff in model performance is evident at several occasions throughout this thesis. These difficulties to model hydrology in semi-arid regions can partly be related to the aforementioned limitations of the here presented model-data integration approach. However, at the same time, the used approach helps to identify the origins of such deficits. On the one hand, the overall TWS variability and the signal-to-noise ratio is low in semi-arid regions (Scanlon et al. 2016), so that such grid-cells receive little weight in the calibration approach, not only regarding the TWS cost, but also regarding the total cost value, for which the TWS cost is the largest contributor. In this context, the compensatory effects that result from optimizing one global cost value, come into play and prevent appropriate adaption to the unique characteristics of semi-arid regions. Improved parametrization and a more sophisticated calibration approach would be beneficial at this point. Issues of parameter equifinality are identified regarding the definition of soil water, root zone and slowly varying water storages (**Chapter 3**) as well as regarding parameters that define their depletion by either evapotranspiration or runoff, especially under water limited conditions (**Chapter 3**, **Chapter 4**). The need for better parametrization of the deeper sub-surface on large to continental scales has been postulated before (Bierkens 2015, Clark et al. 2017), yet so far observational data is missing. To better constrain the outflow by either evapotranspiration or baseflow, and by that the partitioning among sub-surface and slowly varying water storages, it is essential to increase parameter identifiability by adding constraints. In this context, spatial pattern of decay parameter to better distinguish between baseflow and transpiration are potentially valuable, such as a baseflow index to define slow runoff (e.g., Beck et al. 2015) or metrics to define the decay by evapotranspiration (e.g., Küçük et al. 2022). Next to parameter equifinality per se, **Chapter 3** resp. **Appendix B** identifies data inconsistencies in semi-arid regions, that reveal themselves in form of tentatively negative water balance, indicating either an underestimation by the precipitation data set, or an overestimation of either evapotranspiration and runoff. While precipitation underestimation is a common issue of remote sensing-based data products (Huffman et al. 2000), the used runoff product has larger uncertainties in semi-arid regions and is generated with climatic drivers only while disregarding spatio-temporal variations in vegetation (Ghiggi et al. 2019). Since this very variability is shown to be essential for reproducing hydrological dynamics in semi-arid regions, less weight should be given to the model performance regarding the runoff data, and further complementary data that refines evapotranspiration and runoff processes need to be included. In this context, already using EVI data for defining vegetation activity in **Chapter 3**, has a positive impact on model performance regarding TWS and evapotranspiration in semi-arid regions, because it reflects the access of vegetation to water from sub-surface storages and implicitly includes irrigation by (remaining) high vegetation activity also during dry down events. By that, the used model-data integration approach not only helps to identify potential issues of data and model structure, but moreover allows to implicitly include processes not covered by the model structure, and to mitigate the effect of inconsistent behavior of one data stream.

5.2.4 Prospects for Future Research

Based on the presented findings and their limitations, several ideas and suggestions for future research emerge – some concerning specific questions and methods, others of a more general nature:

- ***Assessing the contribution of river storage to TWS variability across scales***

Chapter 4 investigates the effect of river water storage on model calibration and validation, and shows its importance for model validation against GRACE TWS in several regions. However, only the relevance for improving model performance regarding mean seasonal dynamics is analyzed. Therefore, the logical implication is to directly assess the relative contribution of river water storage to TWS variability compared to other storage components, also considering different time-scales. In this context, the application of more sophisticated routing schemes, that take into account spatial (and temporal) variations of effective river flow velocity, are preferable. Such estimates can either be calculated as part of the routing scheme itself, e.g., following the Manning-Strickler equation or based on a slope-dependent

approach (e.g., Sausen et al. 1994), or, more preferably in order to keep computational costs low, by including (pre-calculated) spatio-(temporal) variable flow velocity estimates as input data.

- ***Assessing trends and future changes of TWS composition***

While this thesis focused on the seasonal and inter-annual variations of TWS variability and their composition from 2000–2015, it's particular interesting to analyze how these variations and the contributions of different water storages will evolve in future. This is even more pressing in the context of current climate and environmental changes, that will largely alter hydrological processes and water resources. A relatively easy outset to start such investigations would be to force calibrated models of this thesis with forcing data from the CMIP-6 project (Eyring et al. 2016). However, as discussed in previous sections, the interpretability of such future projections of TWS variability based on the here presented approach should be taken with caution, because it is strongly tied to the observational data and thus does not include long-term processes and trends that become relevant on longer time scales.

- ***Tackling data inconsistencies and including further constraints***

Throughout this thesis the need for alternative, and, moreover, additional constraints in order to improve process and parameter identifiability, and to mitigate the effect of possible data inconsistencies, has been stressed several times. Especially data to discriminate between sub-surface water storages (plant-accessible soil moisture in the root zone versus deep soil water and groundwater) and improve the differentiation between short- and long-term water storages that are lumped in the monthly GRACE TWS estimates, are needed. Thus, particular interest lies on surface water and river water observations from the forthcoming SWOT mission, that will help to constrain this important water storage. Estimates of daily GRACE TWS variations might as well be beneficial to improve the decomposition between fast and slow water storages (Eicker et al. 2020). Besides, incorporating spatial pattern of physio-geographic features and indices (such as the Topographic Wetness Index, and soil properties), as well as information on decay parameters – either by transpiration (e.g., soil moisture decay rate by vegetation, Küçük et al. 2022), or baseflow (e.g., baseflow index and other streamflow related indices, Beck et al. 2015) are promising opportunities to improve model performance from a data-driven point of view. In this context, remote sensing data also holds potential beyond their use as an observational constraint per-se and can provide information on identifying and formulating indices and functional relationships across several spatial and temporal scales, which should be addressed in future efforts.

- ***Coupling the water and carbon cycle***

Chapter 3 emphasizes the importance of vegetation and the potential of including related data to improve model performance and hydrologic process identifiability. In this context, a promising opportunity is the synergetic simulation and calibration of the coupled water-carbon cycle, as additional data and constraints of vegetation activity and ecosystem responses are available from the carbon-cycle side, which represent potential complementary constraints for hydrological processes as well.

5.3 Conclusion

This thesis introduced a model-data integration approach that combines GRACE TWS observations and data sets of complementary large-scale observations with simple hydrologic modeling concepts. The approach is used to contribute to the understanding of TWS variability and underlying processes. In particular, this thesis highlights the scale-dependencies of TWS determinants, the key role of vegetation for TWS partitioning among liquid water storages, and the relevance of river and surface water storages for simulating TWS dynamics. While the presented approach can be understood as complementary to global hydrological models, that include a profound process-representation, this thesis proofed its value for diagnostic purposes to understand hydrological processes and to identify exigence and insignificance of hydrological processes for (future) large-scale modeling studies. Besides, the potential of combining multiple data sets to overcome data inconsistencies, infer to un-mapped processes and improve model simulations has been demonstrated. These findings are of general relevance for model development and calibration in future global hydrological studies and encourage to take advantage of the increasing availability of complementary observational data sets.

Bibliography

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, Geophys. J. Int., 192, 557–572, https://doi.org/10.1093/gji/ggs030, 2013.

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Le Moigne, P., Decharme, B., Mahfouf, J.-F., and Calvet, J.-C.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area, Geosci. Model Dev., 10, 3889–3912, https://doi.org/10.5194/gmd-10-3889-2017, 2017.

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

Allen, G. H., and Pavelsky, T. M.: Global extent of rivers and streams, Science, 361, 585-588, doi:10.1126/science.aat0636, 2018.

AMAP: Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017, Arctic Monitoring and Assessment Programme (AMAP), Oslo, Norway, xiv C269, 2017.

Andersen, O. B., Krogh, P. E., Bauer-Gottwein, P., Leiriao, S., Smith, R., and Berry, P.: Terrestrial Water Storage from GRACE and Satellite Altimetry in the Okavango Delta (Botswana), IAG Symp., 135, 521–526, https://doi.org/10.1007/978-3-642-10634-7_69, 2010.

Bai, P., Liu, X., and Liu, C.: Improving hydrological simulations by incorporating GRACE data for model calibration, Journal of Hydrology, 557, 291-304, https://doi.org/10.1016/j.jhydrol.2017.12.025, 2018.

Baldocchi, D., Ma, S., and Verfaillie, J.: On the inter- and intra-annual variability of ecosystem evapotranspiration and water use efficiency of an oak savanna and annual grassland subjected to booms and busts in rainfall, Glob. Change Biol., 27, 359–375, https://doi.org/10.1111/gcb.15414, 2021.

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, J. Hydrometeorol., 10, 623–643, https://doi.org/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 Resour. Res., 43, W02410, https://doi.org/10.1029/2005WR004753, 2007.

Beck, H. E., de Roo, A., and van Dijk, A. I. J. M.: Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments*, Journal of Hydrometeorology, 16, 1478-1501, 10.1175/jhm-d-14-0155.1, 2015.

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

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589-615, 10.5194/hess-21-589-2017, 2017.

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, J. Geophys. Res.-Atmos., 121, 4468–4486, https://doi.org/10.1002/2015jd024546, 2016.

Bergström, S.: Principles and Confidence in Hydrological Modeling, Nord. Hydrol., 22, 123–136, 1991.

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

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, https://doi.org/10.5194/gmd-4-677-2011, 2011.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18-36, 10.1016/j.jhydrol.2005.07.007, 2006.

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.: Hyperresolution global hydrological modeling: what is next?, Hydrol. Process., 29, 310–320, https://doi.org/10.1002/hyp.10391, 2015.

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resources Research, 51, 4923-4947, https://doi.org/10.1002/2015WR017173, 2015.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/a:1010933404324, 2001.

Chen, J., Ban, Y., and Li, S.: Open access to Earth landcover map, Nature, 514, 434–434, https://doi.org/10.1038/514434c, 2014.

Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Wayne Higgins, R., and Janowiak, J. E.: Assessing objective techniques for gauge-based analyses of global daily precipitation, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009132, 2008.

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, J. Geophys. Res.-Solid Ea., 118, 740–747, https://doi.org/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, https://doi.org/10.5194/gmd-4-701-2011, 2011.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, Hydrol. Earth Syst. Sci., 21, 3427-3440, 10.5194/hess-21-3427-2017, 2017.

Contractor, S., Donat, M. G., Alexander, L. V., Ziese, M., Meyer-Christoffer, A., Schneider, U., Rustemeier, E., Becker, A., Durre, I., and Vose, R. S.: Rainfall Estimates on a Gridded Network (REGEN) – a global land-based gridded dataset of daily precipitation from 1950 to 2016, Hydrol. Earth Syst. Sci., 24, 919–943, https://doi.org/10.5194/hess-24-919-2020, 2020.

Cui, Y., Chen, X., Gao, J., Yan, B., Tang, G., and Hong, Y.: Global water cycle and remote sensing big data: overview, challenge, and opportunities, Big Earth Data, 2, 282-297, 10.1080/20964471.2018.1548052, 2018.

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, https://doi.org/10.5194/hess-12-1387-2008, 2008.

DeChant, C. M., and Moradkhani, H.: Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, Water Resources Research, 48, https://doi.org/10.1029/2011WR011011, 2012.

Decharme, B., Martin, E., and Faroux, S.: Reconciling soil thermal and hydrological lower boundary conditions in land surface models, J. Geophys. Res.-Atmos., 118, 7819–7834, https://doi.org/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, https://doi.org/10.1002/2013jd021264, 2014.

Didan, K. and Barreto-Munoz, A.: MODIS Vegetation Index User's Guide (MOD13 Series), The University of Arizona, https://vip.arizona.edu/documents/MODIS/MODIS_ VI_UsersGuide_09_18_2019_C61.pdf (last access: 22 February 2022), 2019.

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

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, https://doi.org/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, 2014.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, J. Hydrol., 270, 105–134, 2002.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M., and Kidd, R.: Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sens. Environ., 162, 380–395, https://doi.org/10.1016/j.rse.2014.07.023, 2015.

Dorigo, W. A., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., de Jeu, R. A. M., and Naeimi, V.: Error characterisation of global active and passive microwave soil moisture datasets, Hydrol. Earth Syst. Sci., 14, 2605-2616, 10.5194/hess-14-2605-2010, 2010.

Dorigo, W., Dietrich, S., Aires, F., Brocca, L., Carter, S., Cretaux, J.-F., Dunkerley, D., Enomoto, H., Forsberg, R., Güntner, A., Hegglin, M. I., Hollmann, R., Hurst, D. F., Johannessen, J. A., Kummerow, C., Lee, T., Luojus, K., Looser, U., Miralles, D. G., Pellet, V., Recknagel, T., Vargas, C. R., Schneider, U., Schoeneich, P., Schröder, M., Tapper, N., Vuglinsky, V., Wagner, W., Yu, L., Zappa, L., Zemp, M., and Aich, V.: Closing the Water Cycle from Observations across Scales: Where Do We Stand?, Bulletin of the American Meteorological Society, 102, E1897-E1935, 10.1175/bams-d-19-0316.1, 2021.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sensing of Environment, 203, 185-215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.

Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., McDowell, W. H., Kortelainen, P., Caraco, N. F., Melack, J. M., and Middelburg, J. J.: The global abundance and size distribution of lakes, ponds, and impoundments, Limnology and Oceanography, 51, 2388-2397, https://doi.org/10.4319/lo.2006.51.5.2388, 2006.

Draper, N. and Smith, H.: Applied Regression Analysis, 2nd edn., Wiley, New York, NY, ISBN 10 0471029955, 1981.

Du, L., Zeng, Y., Ma, L., Qiao, C., Wu, H., Su, Z., and Bao, G.: Effects of anthropogenic revegetation on the water and carbon cycles of a desert steppe ecosystem, Agr. Forest Meteorol., 300, 108339, https://doi.org/10.1016/j.agrformet.2021.108339, 2021.

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

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

Eicker, A., Jensen, L., Wöhnke, V., Dobslaw, H., Kvas, A., Mayer-Gürr, T., and Dill, R.: Daily GRACE satellite data evaluate short-term hydro-meteorological fluxes from global atmospheric reanalyses, Scientific Reports, 10, 4504, 10.1038/s41598-020-61166-0, 2020.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., and Schmied, H. M.: 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, 2014.

Entekhabi, D., Njoku, E. G., Neill, P. E. O., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Zyl, J. V.: The Soil Moisture Active Passive (SMAP) Mission, Proceedings of the IEEE, 98, 704-716, 10.1109/JPROC.2010.2043918, 2010.

ESA: Land Cover CCI Product User Guide Version 2 Tech. Rep., maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf, 2017.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937-1958, 10.5194/gmd-9-1937-2016, 2016.

Famiglietti, J. S., and Rodell, M.: Water in the Balance, Science, 340, 1300-1301, 10.1126/science.1236460, 2013.

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic regulation of plant rooting depth, P. Natl. Acad. Sci. USA, 114, 10572–10577, https://doi.org/10.1073/pnas.1712381114, 2017.

Fekete, B. M., Vörösmarty, C. J., Roads, J. O., and Willmott, C. J.: Uncertainties in Precipitation and Their Impacts on Runoff Estimates, J. Climate, 17, 294–304, https://doi.org/10.1175/1520-0442(2004)017<0294:uipati>2.0.co;2, 2004.

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

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

Fischer, C.: Automatische Kalibrierung hydrologischer Modelle - Entwicklung und Anwendung des Kalibrierungssystems OPTAS, Dr.-Ing., Lehrstuhl für Geoinformatik, Institut für Geographie, Friedrich-Schiller-University, Jena, Germany, 239 pp., 2013.

Fisher, J. B., Huntzinger, D. N., Schwalm, C. R., and Sitch, S.: Modeling the Terrestrial Biosphere, Annual Review of Environment and Resources, 39, 91-123, 10.1146/annurev-environ-012913-093456, 2014.

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 Environ. Change, 23, 144–156, https://doi.org/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 Resour. Res., 48, W01507, https://doi.org/10.1029/2011wr011239, 2012.

Getirana, A., Kumar, S., Girotto, M., and Rodell, M.: Rivers and Floodplains as Key Components of Global Terrestrial Water Storage Variability, Geophysical Research Letters, 44, 10,359-310,368, 10.1002/2017gl074684, 2017.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, Earth System Science Data, 11, 1655-1674, 2019.

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

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

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

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

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling, Journal of Hydrology, 377, 80-91, 10.1016/j.jhydrol.2009.08.003, 2009.

Hagemann, S.: The Hydrological Cycle: How observational data are able to improve climate models, Max-Planck-Institute for Meteorology, Hamburg, 2011.

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, https://doi.org/10.1007/978-3-540-30217-9_29, 2004.

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

Herold, N., Alexander, L. V., Donat, M. G., Contractor, S., and Becker, A.: How much does it rain over land?, Geophysical Research Letters, 43, 341-348, 10.1002/2015GL066615, 2016.

Hicks, F., and Beltaos, S.: River Ice, in: Cold Region Atmospheric and Hydrologic Studies. The Mackenzie GEWEX Experience: Volume 2: Hydrologic Processes, edited by: Woo, M.-k., Springer Berlin Heidelberg, Berlin, Heidelberg, 281-305, 2008.

Huffman, G. J. and Bolvin, D.: Version 1.2 GPCP One-Degree Daily Precipitation Data Set Documentation, NASA, Goddard Space Flight Center, Greenbelt, MD, USA https://rda.ucar.edu/ datasets/ds728.3/#!docs (last access: July 2018), 2013.

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 Multi-satellite Observations, Journal of Hydrometeorology, 2, 36-50, 2000.

Huffman, G. J., Bolvin, D. T., and Adler, R. F.: GPCP Version 1.2 One-Degree Daily Precipitation Data Set, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, https://doi.org/10. 5065/D6D50K46, 2016.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Sorooshian, S., Xie, P., and Yoo, S.-H.: Developing the Integrated Multi-Satellite Retrievals for GPM (IMERG), April 01, 2012, 2012.

Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., Seneviratne, S. I., and Frankenberg, C.: Soil moisture– atmosphere feedback dominates land carbon uptake variability, Nature, 592, 65–69, https://doi.org/10.1038/s41586-021-03325-5, 2021.

Humphrey, V., Gudmundsson, L., and Seneviratne, S. I.: Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes, Surv. Geophys., 37, 357–395, https://doi.org/10.1007/s10712-016-9367-1, 2016.

Humphrey, V., Zscheischler, J., Ciais, P., Gudmundsson, L., Sitch, S., and Seneviratne, S. I.: Sensitivity of atmospheric CO2 growth rate to observed changes in terrestrial water storage, Nature, 560, 628-631, 10.1038/s41586-018-0424-4, 2018.

IPCC: Climate Change 2014: Synthesis Report, in: 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 Resour. Res., 29, 2637–2649, 1993.

Jung, M. and FLUXCOM team: FLUXCOM (RSCMETEO) Global Land Carbon Fluxes using CRUNCEP climate data, FLUXCOM Data Portal, https://www.bgc-jena.mpg.de/geodb/ projects/Home.php, 2016.

Jung, M., Henkel, K., Herold, M., and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, Remote Sens. Environ., 101, 534–553, https://doi.org/10.1016/j.rse.2006.01.020, 2006.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Scientific Data, 6, 74, 10.1038/s41597-019-0076-8, 2019.

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, B. Am. Meteorol. Soc., 77, 437–471, 1996.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Bitar, A. A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., and Delwart, S.: The SMOS Soil Moisture Retrieval Algorithm, IEEE Transactions on Geoscience and Remote Sensing, 50, 1384-1403, 10.1109/TGRS.2012.2184548, 2012.

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

Koirala, S., Yeh, P. J. F., Hirabayashi, Y., Kanae, S., and Oki, T.: Global-scale land surface hydrologic modeling with the representation of water table dynamics, J. Geophys. Res.-Atmos., 119, 75–89, https://doi.org/10.1002/2013JD020398, 2014.

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

Kraft, B., Jung, M., Körner, M., and Reichstein, M.: Hybrid modeling: Fusion of a deep approach and physics-based model for global hydrological modeling, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 1537-1544, 2020.

Kraft, B., Jung, M., Körner, M., Koirala, S., and Reichstein, M.: Towards hybrid modeling of the global hydrological cycle, Hydrol. Earth Syst. Sci. Discuss., 2021, 1-40, 10.5194/hess-2021-211, 2021.

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 Biogeochem. Cy., 19, GB1015, https://doi.org/10.1029/2003GB002199, 2005.

Küçük, Ç., Koirala, S., Carvalhais, N., Miralles, D. G., Reichstein, M., and Jung, M.: Characterising the response of vegetation cover to water limitation in Africa using geostationary satellites, Journal of Advances in Modeling Earth Systems, 14, https://doi.org/10.1029/2021MS002730, 2022.

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, Nat. Geosci., 8, 759–763, https://doi.org/10.1038/ngeo2517, 2015.

Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., Eylander, J. B., and Houser, P. R.: A land surface data assimilation framework using the land information system: Description and applications, Advances in Water Resources, 31, 1419-1432, https://doi.org/10.1016/j.advwatres.2008.01.013, 2008.

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., De Lannoy, G., 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 Resour. Res., 30, 1515–1527, 1994.

Lahoz, W. A., and Schneider, P.: Data assimilation: making sense of Earth Observation, Frontiers in Environmental Science, 2, 10.3389/fenvs.2014.00016, 2014.

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

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 Resour. Res., 51, 7309–7342, https://doi.org/10.1002/2015wr017616, 2015.

Liu, J., Li, Z., Huang, L., and Tian, B.: Hemispheric-scale comparison of monthly passive microwave snow water equivalent products, J. Appl. Remote Sens., 8, 084688, https://doi.org/10.1117/1.JRS.8.084688, 2014.

Livneh, B., and Lettenmaier, D. P.: Multi-criteria parameter estimation for the Unified Land Model, Hydrology and Earth System Sciences, 16, 3029-3048, 10.5194/hess-16-3029-2012, 2012.

Lo, M.-H., Famiglietti, J. S., Yeh, P. J.-F., and Syed, T. H.: Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, Water Resources Research, 46, https://doi.org/10.1029/2009WR007855, 2010.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product, J. Climate, 31, 895–918, https://doi.org/10.1175/JCLI-D-17-0208.1, 2018.

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.

Lu, J., Sun, G., McNulty, S. G., and Amatya, D. M.: A ccomparison of six potential evapotranspiration methods for regional use in the southeastern United States, J. Am. Water Resour. As., 41, 621–633, 2005.

Lu, S., Ouyang, N., Wu, B., Wei, Y., and Tesemma, Z.: Lake water volume calculation with time series remote-sensing images, International Journal of Remote Sensing, 34, 7962-7973, 10.1080/01431161.2013.827814, 2013.

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, European Space Agency study contract report, Finnish Meteorological Institute, Helsinki, http://www.globsnow.info/docs/GlobSnow_2_ Final_Report_release.pdf (last access: July 2018), 2014.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903-1925, 10.5194/gmd-10-1903-2017, 2017.

McColl, K. A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D. J., Lu, H., Pan, M., and Entekhabi, D.: Global characterization of surface soil moisture drydowns, Geophys. Res. Lett., 44, 3682– 3690, https://doi.org/10.1002/2017GL072819, 2017.

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, 2011.

Mizukami, N., Clark, M. P., Gharari, S., Kluzek, E., Pan, M., Lin, P., Beck, H. E., and Yamazaki, D.: A Vector-Based River Routing Model for Earth System Models: Parallelization and Global Applications, Journal of Advances in Modeling Earth Systems, 13, e2020MS002434, https://doi.org/10.1029/2020MS002434, 2021.

Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, American Society of Agricultural and Biological Engineers, 50, 885-900, 2007.

Mostafaie, A., Forootan, E., Safari, A., and Schumacher, M.: Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily GRACE data, Computational Geosciences, 22, 789-814, 10.1007/s10596-018-9726-8, 2018.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781-1800, https://doi.org/10.1016/j.rse.2011.02.019, 2011.

Müller-Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and evaluation, Geosci. Model Dev., 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021, 2021.

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

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

NASA/LARC/SD/ASDC: CERES and GEO-Enhanced TOA, Within-Atmosphere and Surface Fluxes, Clouds and Aerosols Daily Terra-Aqua Edition4A. NASA Langley Atmospheric Science Data Center DAAC [data set], https://doi.org/10.5067/Terra+Aqua/CERES/SYN1degDay_L3.004A, 2017.

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.: epresenting twentieth-century spacetime climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate, J. 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 Resour. Res., 43, W04427, https://doi.org/10.1029/2006WR004941, 2007.

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D., Ashouri, H., and Thorstensen, A. R.: The PERSIANN family of global satellite precipitation data: a review and evaluation of products, Hydrol. Earth Syst. Sci., 22, 5801-5816, 10.5194/hess-22-5801-2018, 2018.

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.

Novák, V.: Evapotranspiration, in: Encyclopedia of Agrophysics, edited by: Gliński, J., Horabik, J., and Lipiec, J., Springer Netherlands, Dordrecht, 280-283, 2011.

O, S., and Orth, R.: Global soil moisture data derived through machine learning trained with in-situ measurements, Scientific Data, 8, 170, 10.1038/s41597-021-00964-1, 2021.

Oki, T., Nishimura, T., and Dirmeyer, P.: Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP), Journal of the Meteorological Society of Japan. Ser. II, 77, 235-255, 1999.

Omlin, M. and Reichert, P.: A comparison of techniques for the estimation of model prediction uncertainty, Ecol. Model., 115, 45–59, https://doi.org/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, J. Hydrometeorol., 14, 1773–1790, https://doi.org/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, https://doi.org/10.1016/j.jhydrol.2015.01.044, 2015.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418-422, 10.1038/nature20584, 2016.

Perlman, H. A., and Evans, J.: The natural water cycle, in, The USGS Water Science School, 2000.

Porporato, A., Daly, E., and Rodriguez-Iturbe, I.: Soil water balance and ecosystem response to climate change, The American Naturalist, 164, 625-632, 2004.

Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using largescale parameters, Mon. Weather Rev., 100, 81–92, 1972.

Quevedo, D. I., and Francés, F.: A conceptual dynamic vegetation-soil model for arid and semiarid zones, Hydrol. Earth Syst. Sci., 12, 1175-1187, 10.5194/hess-12-1175-2008, 2008.

Rakovec, O., Kumar, R., Attinger, S., and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, Water Resources Research, 52, 7779-7792, https://doi.org/10.1002/2016WR019430, 2016.

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 Planet. Change, 53, 198–208, https://doi.org/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.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed, Bulletin of the American Meteorological Society, 93, 811-829, 10.1175/bams-d-11-00052.1, 2012.

Reager, J. T., Thomas, A. C., Sproles, E. A., Rodell, M., Beaudoing, H. K., Li, B., and Famiglietti, J. S.: Assimilation of GRACE Terrestrial Water Storage Observations into a Land Surface Model for the Assessment of Regional Flood Potential, Remote Sensing, 7, 14663-14679, 2015.

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., and De Lannoy, G. J. M.: Assessment of MERRA-2 Land Surface Hydrology Estimates, J. Climate, 30, 2937–2960, https://doi.org/10.1175/jcli-d-16-0720.1, 2017.

Reichle, R. H.: Data assimilation methods in the Earth sciences, Advances in Water Resources, 31, 1411-1418, https://doi.org/10.1016/j.advwatres.2008.01.001, 2008.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 10.1038/s41586-019-0912-1, 2019.

Reyer, C. P. O., Leuzinger, S., Rammig, A., Wolf, A., Bartholomeus, R. P., Bonfante, A., de Lorenzi, F., Dury, M., Gloning, P., Abou Jaoudé, R., Klein, T., Kuster, T. M., Martins, M., Niedrist, G., Riccardi, M., Wohlfahrt, G., de Angelis, P., de Dato, G., François, L., Menzel, A., and Pereira, M.: A plant's perspective of extremes: terrestrial plant responses to changing climatic variability, Global Change Biology, 19, 75-89, https://doi.org/10.1111/gcb.12023, 2013.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24, 3624-3648, 10.1175/jcli-d-11-00015.1, 2011.

Rind, D.: The influence of vegetation on the hydrologic cycle in a global climate model, in: Climate Processes and Climate Sensitivity, edited by: Hansen, J. E. and Takahashi, T., AGU Geophysical Monograph 29, Maurice Ewing American Geophysical Union, 73–91, https://doi.org/10.1029/GM029p0073, 1984.

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.

Rodell, M., Famiglietti, J., Wiese, D., Reager, J., Beaudoing, H., Landerer, F., and Lo, M.-H.: Emerging trends in global freshwater availability, Nature, 2018.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin*, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381-394, 10.1175/bams-85-3-381, 2004.

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.: Basin scale estimates of evapotranspiration using GRACE and other observations, Geophys. Res. Lett., 31, L20504, https://doi.org/10.1029/2004gl020873, 2004.

Rodriguez-Iturbe, I., Porporato, A., Laio, F., and Ridolfi, L.: Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: I. Scope and general outline, Adv. Water Resour., 24, 695–705, https://doi.org/10.1016/S0309-1708(01)00004-5, 2001.

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

Ruiz-Pérez, G., Koch, J., Manfreda, S., Caylor, K., and Francés, F.: Calibration of a parsimonious distributed ecohydrological daily model in a data-scarce basin by exclusively using the spatio-temporal variation of NDVI, Hydrol. Earth Syst. Sci., 21, 6235– 6251, https://doi.org/10.5194/hess-21-6235-2017, 2017.

Sausen, R., Schubert, S., and Dümenil, L.: A model of river runoff for use in coupled atmosphere-ocean models, Journal of Hydrology, 155, 337-352, https://doi.org/10.1016/0022-1694(94)90177-5, 1994.

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

Scanlon, B. R., Zhang, Z. Z., Save, H., Wiese, D. N., Landerer, F. W., Long, D., Longuevergne, L., and Chen, J. l.: Global evaluation of new GRACE mascon products for hydrologic applications, Water Resour. Res., 52, 9412–9429, https://doi.org/10.1002/2016wr019494, 2016.

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller-Schmied, H., van Beek, L. P. H., Wiese, D. N., Wada, Y., Long, D., Reedy, R. C., Longuevergne, L., Döll, P., and Bierkens, M. F. P.: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data, Proceedings of the National Academy of Sciences, 10.1073/pnas.1704665115, 2018.

Schaaf, C., and Wang, Z.: MCD43A1 MODIS/Terra+Aqua BRDF/Albedo Model Parameters Daily L3 Global – 500 m V006, NASA EOSDIS Land Processes DAAC [data set], https://doi.org/10.5067/MODIS/MCD43A1.006, 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.

Schenk, H. J., and Jackson, R. B.: Mapping the global distribution of deep roots in relation to climate and soil characteristics, Geoderma, 126, 129–140, https://doi.org/10.1016/j.geoderma.2004.11.018, 2005.

Schmidt, R., Flechtner, F., Meyer, U., Neumayer, K.-H., Dahle, C., König, R., and Kusche, J.: Hydrological Signals Observed by the GRACE Satellites, Surveys in Geophysics, 29, 319-334, 10.1007/s10712-008-9033-3, 2008a.

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, 2008b.

Seo, D.-J., Cajina, L., Corby, R., and Howieson, T.: Automatic state updating for operational streamflow forecasting via variational data assimilation, Journal of Hydrology, 367, 255-275, https://doi.org/10.1016/j.jhydrol.2009.01.019, 2009.

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, J. Geophys. Res., 115, D20117, https://doi.org/10.1029/2009jd013504, 2010.

Shaad, K.: Evolution of river-routing schemes in macro-scale models and their potential for watershed management, Hydrological Sciences Journal, 63, 1062-1077, 10.1080/02626667.2018.1473871 2018.

Shamir, E., Lee, B.-J., Bae, D.-H., and Georgakakos, K. P.: Flood Forecasting in Regulated Basins Using the Ensemble Extended Kalman Filter with the Storage Function Method, Journal of Hydrologic Engineering, 15, 1030-1044, doi:10.1061/(ASCE)HE.1943-5584.0000282, 2010.

Soltani, S. S., Ataie-Ashtiani, B., and Simmons, C. T.: Review of assimilating GRACE terrestrial water storage data into hydrological models: Advances, challenges and opportunities, Earth-Science Reviews, 213, 103487, https://doi.org/10.1016/j.earscirev.2020.103487, 2021.

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 Resour. Res., 29, 1185–1194, 1993.

Sposób, J.: Water Balance in Terrestrial Ecosystems, in: Encyclopedia of Agrophysics, edited by: Gliński, J., Horabik, J., and Lipiec, J., Springer Netherlands, Dordrecht, 955-959, 2011.

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

Su, Z., Zeng, Y., Romano, N., Manfreda, S., Francés, F., Ben Dor, E., Szabó, B., Vico, G., Nasta, P., Zhuang, R., Francos, N., Mészáros, J., Dal Sasso, S. F., Bassiouni, M., Zhang, L., Rwasoka, D. T., Retsios, B., Yu, L., Blatchford, M. L., and Mannaerts, C.: An Integrative Information Aqueduct to Close the Gaps be-tween Satellite Observation of Water Cycle and Local Sustainable Management of Water Resources, Water, 12, 1495, 2020.

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

Swenson, S., Chambers, D., and Wahr, J.: Estimating geocenter variations from a combination of GRACE and ocean model output, J. Geophys. Res., 113, B08410, https://doi.org/10.1029/2007jb005338, 2008.

Swenson, S.: Assessing High-Latitude Winter Precipitation from Global Precipitation Analyses Using GRACE, J. Hydrometeorol., 11, 405–420, https://doi.org/10.1175/2009jhm1194.1, 2010.

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

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.

Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., Sasgen, I., Famiglietti, J. S., Landerer, F. W., Chambers, D. P., Reager, J. T., Gardner, A. S., Save, H., Ivins, E. R., Swenson, S. C., Boening, C., Dahle, C., Wiese, D. N., Dobslaw, H., Tamisiea, M. E., and Velicogna, I.: Contributions of GRACE to understanding climate change, Nature Climate Change, 9, 358-369, 10.1038/s41558-019-0456-2, 2019.

Telteu, C. E., Müller-Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., Boulange, J. E. S., Andersen, L. S., Grillakis, M., Gosling, S. N., Satoh, Y., Rakovec, O., Stacke, T., Chang, J., Wanders, N., Shah, H. L., Trautmann, T., Mao, G.,

Hanasaki, N., Koutroulis, A., Pokhrel, Y., Samaniego, L., Wada, Y., Mishra, V., Liu, J., Döll, P., Zhao, F., Gädeke, A., Rabin, S. S., and Herz, F.: Understanding each other's models: an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication, Geosci. Model Dev., 14, 3843-3878, 10.5194/gmd-14-3843-2021, 2021.

Telteu, C.-E., Müller-Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., Boulange, J. E. S., Andersen, L. S., Grillakis, M., Gosling, S. N., Satoh, Y., Rakovec, O., Stacke, T., Chang, J., Wanders, N., Shah, H. L., Trautmann, T., Mao, G., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Samaniego, L., Wada, Y., Mishra, V., Liu, J., Döll, P., Zhao, F., Gädeke, A., Rabin, S. S., and Herz, F.: Understanding each other's models: an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication, Geosci. Model Dev., 14, 3843–3878, https://doi.org/10.5194/gmd-14-3843-2021, 2021.

Teuling, A. J., Seneviratne, S. I., Williams, C., and Troch, P. A.: Observed timescales of evapotranspiration response to soil moisture, Geophys. Res. Lett., 33, L23403, https://doi.org/10.1029/2006gl028178, 2006.

Tian, S., Van Dijk, A. I. J. M., Tregoning, P., and Renzullo, L. J.: Forecasting dryland vegetation condition months in advance through satellite data assimilation, Nat. Commun., 10, 469, https://doi.org/10.1038/s41467-019-08403-x, 2019.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., 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, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.

Trautmann, T., Koirala, S., Carvalhais, N., Eicker, A., Fink, M., Niemann, C., and Jung, M.: Understanding terrestrial water storage variations in northern latitudes across scales, Hydrol. Earth Syst. Sci., 22, 4061–4082, https://doi.org/10.5194/hess-22-4061-2018, 2018.

Trautmann, T., Koirala, S., Carvalhais, N., Güntner, A., Jung, M.: The importance of vegetation in understanding terrestrial water storage variations, Hydrology and Earth System Sciences, 26 (4): 1089-1109, doi:10.5194/hess-26-1089-2022, 2022.

Trenberth, K. E., and Asrar, G. R.: Challenges and Opportunities in Water Cycle Research: WCRP Contributions, Surveys in Geophysics, 35, 515-532, 10.1007/s10712-012-9214-y, 2014.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data, J. Hydrometeorol., 8, 758–769, https://doi.org/10.1175/jhm600.1, 2007.

Trischenko, A. P.: Removing Unwanted Fluctuations in the AVHRR Thermal Calibration Data Using Robust Techniques, J. Atmos. Ocean. Tech., 19, 1939–1953, 2002.

Tuozzolo, S., Lind, G., Overstreet, B., Mangano, J., Fonstad, M., Hagemann, M., Frasson, R. P. M., Larnier, K., Garambois, P.-A., Monnier, J., and Durand, M.: Estimating River Discharge With Swath Altimetry: A Proof of Concept Using AirSWOT Observations, Geophysical Research Letters, 46, 1459-1466, https://doi.org/10.1029/2018GL080771, 2019.

United Nations Environment: World atlas of desertification/UNEP, United Nations Environment Programme, Edward Arnold, London, Baltimore, 1992.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47, W07517, https://doi.org/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, https://doi.org/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, Int. J. Geogr. Inform. Sci., 24, 189– 212, https://doi.org/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, Water for a Healthy Country National Research Flagship, Clayton, Australia, 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, Hydrol. Earth Syst. Sci., 18, 2955–2973, https://doi.org/10.5194/hess-18-2955-2014, 2014.

Verpoorter, C., Kutser, T., Seekell, D. A., and Tranvik, L. J.: A global inventory of lakes based on high-resolution satellite imagery, Geophysical Research Letters, 41, 6396-6402, https://doi.org/10.1002/2014GL060641, 2014.

Viovy, N.: CRUNCEP Version 7 - Atmospheric Forcing Data for the Community Land Model. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, https://doi.org/10.5065/PZ8F-F017, 2018.

Viovy, N.: CRUNCEPv6.1 Dataset, http://dods.extra.cea.fr/data/p529viov/cruncep/, last access: 15 September 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, https://doi.org/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.

Wang, F., Polcher, J., Peylin, P., and Bastrikov, V.: Assimilation of river discharge in a land surface model to improve estimates of the continental water cycles, Hydrol. Earth Syst. Sci., 22, 3863– 3882, https://doi.org/10.5194/hess-22-3863-2018, 2018.

Wang, J., Price, K. P., and Rich, P. M.: Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains, International Journal of Remote Sensing, 22, 3827-3844, 10.1080/01431160010007033, 2001.

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J., and Savenije, H. H. G.: Global root zone storage capacity from satellite-based evaporation, Hydrol. Earth Syst. Sci., 20, 1459–1481, https://doi.org/10.5194/hess-20-1459-2016, 2016.

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, J. Geophys. Res., 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 Resour. Res., 50, 7505–7514, https://doi.org/10.1002/2014wr015638, 2014.

Weiss, M., van den Hurk, B., Haarsma, R., and Hazeleger, W.: Impact of vegetation variability on potential predictability and skill of EC-Earth simulations, Climate Dynamics, 39, 2733-2746, 10.1007/s00382-012-1572-0, 2012.

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

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.

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, B. Am. Meteorol. Soc., 77, 853–868, 1996.

Wiese, D. N., Landerer, F. W., and Watkins, M. M.: Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution, Water Resour. Res., 52, 7490–7502, https://doi.org/10.1002/2016wr019344, 2016a.

Wiese, D. N., Yuan, D.-N., Boening, C., Landerer, F. W., and Watkins, M. M.: JPL GRACE Mascon Ocean, Ice, and Hydrology Equivalent Water Height RL05M.1 CRI Filtered Version 2, Ver. 2, PO.DAAC, CA, USA, http://dx.doi.org/10.5067/ TEMSC-2LCR5, 2016b.

Wiese, D. N., Yuan, D.-N., Boening, C., Landerer, F. W., and Watkins, M. M.: JPL GRACE Mascon Ocean, Ice, and Hydrology Equivalent Water Height Release 06 Coastal Resolution Improvement (CRI) Filtered Version 1.0, Ver. 1.0, PO.DAAC, CA, USA, [data set], https://doi.org/10.5067/TEMSC-3MJC6, 2018.

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

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, Hydrol. Earth Syst. Sci., 16, 3083–3099, https://doi.org/10.5194/hess-16-3083-2012, 2012.

Xu, C.-y.: Textbook of Hydrologic Models, Hydrology, Department of Earth Sciences, Uppsala University, Uppsala, Sweden, 2002.

Xu, X., Medvigy, D., Powers, J. S., Becknell, J. M., and Guan, K.: Diversity in plant hydraulic traits explains seasonal and inter-annual variations of vegetation dynamics in seasonally dry tropical forests, New Phytol., 212, 80–95, https://doi.org/10.1111/nph.14009, 2016.

Yamazaki, D., Oki, T., and Kanae, S.: Deriving a global river network map and its sub-grid topographic characteristics from a fine-resolution flow direction map, Hydrol. Earth Syst. Sci., 13, 2241-2251, 10.5194/hess-13-2241-2009, 2009.

Yamazaki, D., Kanae, S., Kim, H., and Oki, T.: A physically based description of floodplain inundation dynamics in a global river routing model, Water Resources Research, 47, https://doi.org/10.1029/2010WR009726, 2011.

Yamazaki, D., Trigg, M. A., and Ikeshima, D.: Development of a global ~90m water body map using multi-temporal Landsat images, Remote Sensing of Environment, 171, 337-351, https://doi.org/10.1016/j.rse.2015.10.014, 2015.

Yang, Y. T., Long, D., Guan, H. D., Scanlon, B. R., Simmons, C. T., Jiang, L., and Xu, X.: GRACE satellite observed hydrological controls on inter-annual and seasonal variability in surface greenness over mainland Australia, J Geophys Res-Biogeo, 119, 2245-2260, 10.1002/2014jg002670, 2014.

Yang, Y., Anderson, M., Gao, F., Hain, C., Noormets, A., Sun, G., Wynne, R., Thomas, V., and Sun, L.: Investigating impacts of drought and disturbance on evapotranspiration over a forested landscape in North Carolina, USA using high spatiotemporal resolution remotely sensed data, Remote Sens. Environ., 238, 111018, https://doi.org/10.1016/j.rse.2018.12.017, 2020.

Yang, Y., Donohue, R. J., and McVicar, T. R.: Global estimation of effective plant rooting depth: Implications for hydrological modeling, Water Resour. Res., 52, 8260–8276, https://doi.org/10.1002/2016WR019392, 2016.

Yu, L., Fatichi, S., Zeng, Y., and Su, Z.: The role of vadose zone physics in the ecohydrological response of a Tibetan meadow to freeze–thaw cycles, The Cryosphere, 14, 4653-4673, 10.5194/tc-14-4653-2020, 2020.

Zeng, Y., Su, Z., Barmpadimos, I., Perrels, A., Poli, P., Boersma, K. F., Frey, A., Ma, X., de Bruin, K., Goosen, H., John, V. O., Roebeling, R., Schulz, J., and Timmermans, W.: Towards a Traceable Climate Service: Assessment of Quality and Usability of Essential Climate Variables, Remote Sens., 11, 1186, 2019.

Zeng, Y., Su, Z., Calvet, J. C., Manninen, T., Swinnen, E., Schulz, J., Roebeling, R., Poli, P., Tan, D., Riihelä, A., Tanis, C. M., Arslan, A. N., Obregon, A., Kaiser-Weiss, A., John, V. O., Timmermans, W., Timmermans, J., Kaspar, F., Gregow, H., Barbu, A. L., Fairbairn, D., Gelati, E., and Meurey, C.: Analysis of current validation practices in Europe for space-based climate data records of essential climate variables, Int. J. Appl. Earth Obs., 42, 150–161, https://doi.org/10.1016/j.jag.2015.06.006, 2015.

Zhang, K., Kimball, J. S., Nemani, R. R., and Running, S. W.: A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006, Water Resources Research, 46, https://doi.org/10.1029/2009WR008800, 2010.

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.

Zhao, F., Veldkamp, T. I. E., Frieler, K., Schewe, J., Ostberg, S., Willner, S., Schauberger, B., Gosling, S. N., Schmied, H. M., Portmann, F. T., Leng, G., Huang, M., Liu, X., Tang, Q., Hanasaki, N., Biemans, H., Gerten, D., Satoh, Y., Pokhrel, Y., Stacke, T., Ciais, P., Chang, J., Ducharne, A., Guimberteau, M., Wada, Y., Kim, H., and Yamazaki, D.: The critical role of the routing scheme in simulating peak river discharge in global hydrological models, Environmental Research Letters, 12, 075003, 10.1088/1748-9326/aa7250, 2017.

Zhuang, R., Zeng, Y., Manfreda, S., and Su, Z.: Quantifying LongTerm Land Surface and Root Zone Soil Moisture over Tibetan Plateau, Remote Sens., 12, 509, 2020.

Appendix A

Supplement of Chapter 2

Supplement material of:

Trautmann, T., Koirala, S., Carvalhais, N., Eicker, A., Fink, M., Niemann, C., Jung, M. (2018): Understanding terrestrial water storage variations in Northern latitudes across scales, Hydrology and Earth System Sciences, 22(7): 4061-4082, doi: 10.5194/hess-22-4061-2018

Available at:

https://doi.org/10.5194/hess-22-4061-2018-supplement

A.1 Detailed Model Description and Formulas

The model consists of three components: (1) a snow component that simulates accumulation and ablation of snow, (2) a soil water component to calculate soil moisture, evapotranspiration and land runoff, and (3) a runoff component that derives total runoff. All modelled fluxes and states correspond to the spatio-temporal resolution of the forcing data, which in this study is a 1°x1° latitude/longitude grid and daily time steps.

The following describes all implemented processes and equations in detail.

A1.1 Snow Component

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

$$\mathrm{SWE_t} = \mathrm{SWE_{t-1}} + \mathrm{SF_t} - \mathrm{ETSub_t} - \mathrm{M_t} \tag{A.1}$$

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

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

$$\mathrm{SF_t} = \mathrm{p_{sf}} \cdot \mathrm{P_t} \quad | \mathrm{T} < 0°\mathrm{C} \tag{A.2}$$

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

$$\mathrm{FSC_t} = \min\left(\frac{\mathrm{SWE_{t-1}}}{\mathrm{sn_c}}, 1\right) \tag{A.3}$$

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

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

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

$$\mathrm{M_t} = (\mathrm{m_t} \cdot \mathrm{T_t} + \mathrm{m_r} \cdot \mathrm{Rn_t}) \cdot \mathrm{FSC_t} \quad | \mathrm{T} > 0°\mathrm{C} \tag{A.4}$$

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

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

$$LE_t = \left(sn_a \cdot \frac{\Delta sn_t}{\Delta sn_t + \gamma sn_t} \cdot (Rn_t - G)\right) \cdot FSC_t \quad (A.5)$$

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

They calculate Δsn_t as a function of T_t [K] (Eq. A.6), and γsn_t as a function of atmospheric pressure Pair [kPa], specific heat of air at constant pressure c_p [MJ kg^{-1} K^{-1}], the ratio molecular weight of water vapor/dry air MW and latent heat of sublimation of ice λsn [MJ kg^{-1}] (Eq. A.6).

$$\Delta sn_t = \left(\frac{5723.265}{T_t^2} + \frac{3.53069}{T_t - 0{,}00728332}\right) \cdot e^{9.550426 - \frac{5723.265}{T_t} + 3.53068 \cdot \ln(T_t) - 0{,}00728332 \cdot T_t} \quad (A.6)$$

$$\gamma sn_t = \frac{Pair \cdot c_p}{MA \cdot \lambda sn_t} \quad (A.7)$$

In Eq. A.7, Pair is assumed to be time- and space-invariant with a uniform value of 101.3 kPa and c_p = 0.001 MJ kg^{-1} K^{-1}. MA is a constant of 0.622 and λsn is defined by Murphy and Koop (2005) as a function of T_t [K]. With a molecular mass of water of 18.01528 g mol^{-1}, λsn can be calculated as:

$$\lambda sn_t = \left(46782.5 + 35.8925 \cdot T_t - 0.07414 \cdot T_t^2 + 541.5 \cdot e^{-\left(\frac{T_t}{123.75}\right)^2}\right) \cdot \frac{0.001}{18.01528} \quad (A.8)$$

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

$$ETSub_t = \frac{LE_t}{\lambda sn_t} \quad (A.9)$$

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

A1.2 Soil Component

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

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

$$\mathrm{SM_t} = \mathrm{SM_{t-1}} + \mathrm{In_t} - \mathrm{ET_t} \tag{A.10}$$

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

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

$$\mathrm{IW_t} = \mathrm{RF_t} + \mathrm{M_t} \tag{A.11}$$

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

$$\mathrm{Qs_t} = \mathrm{IW_t} \cdot \left(\frac{\mathrm{SM_{t-1}}}{\mathrm{s\_max}}\right)^{\mathrm{s_{exp}}} \tag{A.12}$$

In Eq. A.12 Qs_t depends on the inflow IW_t, the runoff coefficient s_{exp} and the actual soil moisture SM_{t-1} compared to its maximum water holding capacity s_{max}. Thus, no land runoff occurs if the soil water storage is empty and all IW is allocated to land runoff if the soil is completely saturated. Between these points, s_{exp} determines the amount of inflow that converts to Qs. While low values of s_{exp} lead to a high amount of Qs even if the soil moisture deficit is low (e.g. low SM/s_{max} ratio), higher values of s_{exp} increase the proportion of IW that infiltrates.

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

$$\mathrm{In_t} = \mathrm{IW_t} - \mathrm{Qs_t} \tag{A.13}$$

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

$$\mathrm{potET_t} = \mathrm{et_a} \cdot \left(\frac{\Delta_t}{\Delta_t + \gamma_t} \cdot \frac{\mathrm{Rn_t}}{\lambda_t}\right) \tag{A.14}$$

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

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

$$\Delta_t = \frac{4098 \cdot 0.611 \cdot e^{\frac{17.27 \cdot T_t}{T_t + 237.3}}}{(T_t \cdot 237.3)^2} \quad \text{(A.15)}$$

$$\lambda_t = 2.501 - (2.361 \cdot 10^{-3}) \cdot T_t \quad \text{(A.16)}$$

Analogue to Eq. A.7, γ_t depends on a constant atmospheric pressure Pair of 101.3 kPa, the specific heat of air at constant pressure c_p [MJ kg^{-1} K^{-1}], the constant MA and the latent heat of vaporization λ_t:

$$\gamma_t = \frac{\text{Pair} \cdot c_p}{\text{MA} \cdot \lambda_t} \quad \text{(A.17}$$

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

$$\text{supET}_t = et_{sup} \cdot (SM_{t-1} + In_t) \quad \text{(A.18)}$$

Finally, actual evapotranspiration ET [mm d^{-1}] at time t is derived by comparing potET_t [mm d^{-1}] and supET_t [mm d^{-1}]:

$$\text{actET}_t = \min(\text{potET}_t, \text{supET}_t) \quad \text{(A.19)}$$

S1.3 Runoff Component

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

$$Q_t = \sum_{i=0}^{60} Qs_{t-i} \cdot \underbrace{\left[e^{-\frac{i}{q\_t}} - e^{-\frac{i+1}{q_t}}\right]}_{\text{delay component}} \quad \text{(A.20)}$$

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

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

$$0 = P_t - actET_t - Q_t + \Delta TWS_t \quad (A.21)$$

with the change of total water storage ΔTWS [mm d^{-1}] resulting from

$$\Delta TWS = (SWE_t - SWE_{t-1}) + (SM_t - SM_{t-1}) + W_t \quad (A.22)$$

so that solving Eq. A.21 and Eq. A.22

$$\Delta RW_t = actET_t + Q_t - P_t - (SWE_t - SWE_{t-1}) - (SM_t - SM_{t-1}) \quad (A.23)$$

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

$$RW_t = RW_{t-1} + \Delta RW_t \quad (A.24)$$

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

$$TWS_t = SWE_t + SM_t + RW_t \quad (A.25)$$

A.2 Uncertainty of the observational Constraints

Maps of the temporal average uncertainties of observed TWS, ET and Q that are used for model calibration are shown in Fig. A.1. For observed SWE a constant average uncertainty of 35 mm is applied.

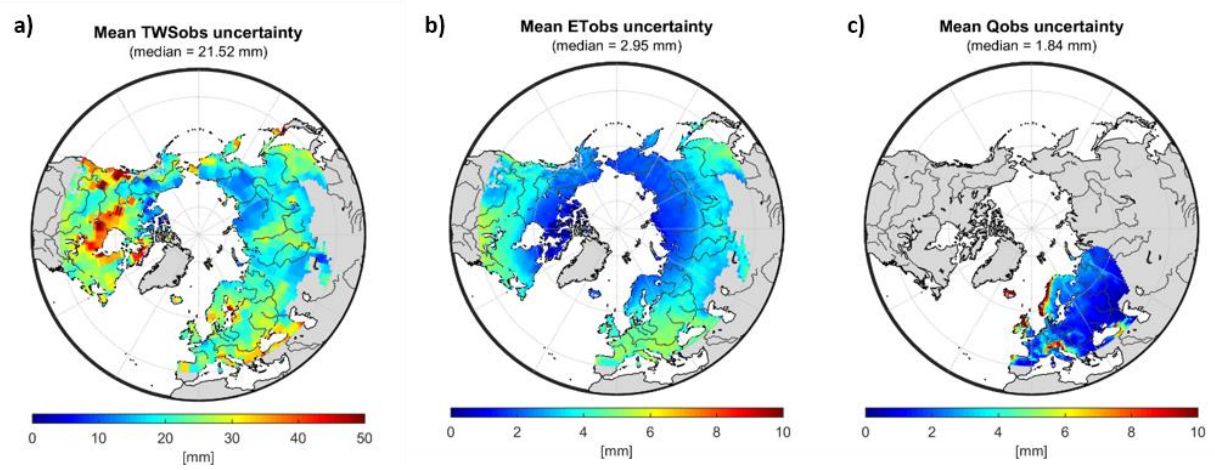

Figure A.1. Mean uncertainty of monthly TWSobs [mm], and of the mean seasonal cycle of ETobs [mm d^{-1}] and Qobs [mm d^{-1}] used for model calibration. Values are truncated to 50 mm resp. 10 mm.

A.3 Cost Terms

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

Table A.1. Cost values obtained with the default and the optimized model parameters using Eq. 2.1.

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

A.4 Model Performance regarding Evapotranspiration and Runoff

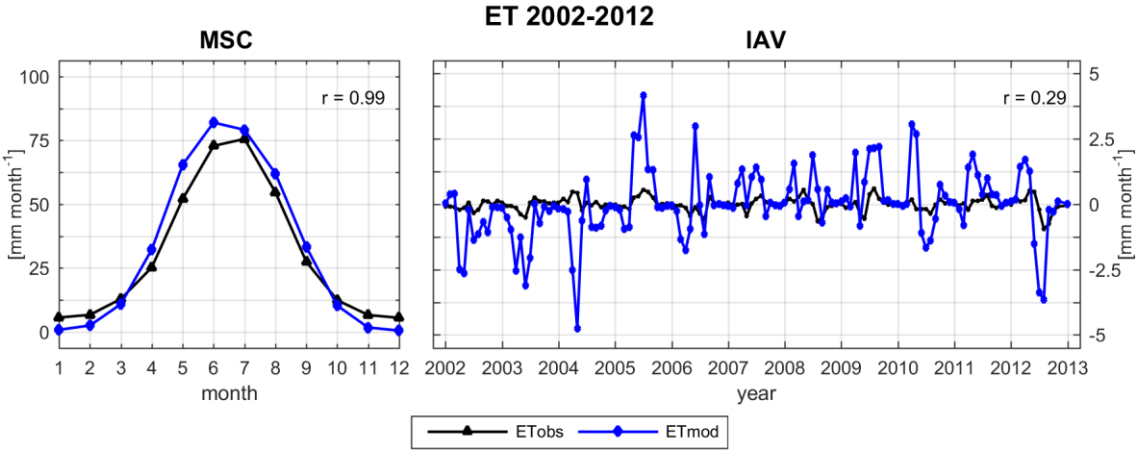

Figure A.2. Spatially averaged mean seasonal cycle (MSC) of the period 2002–2012 and inter-annual variability (IAV, difference between monthly values and the MSC) for ETmod and FLUXCOM based ETobs.

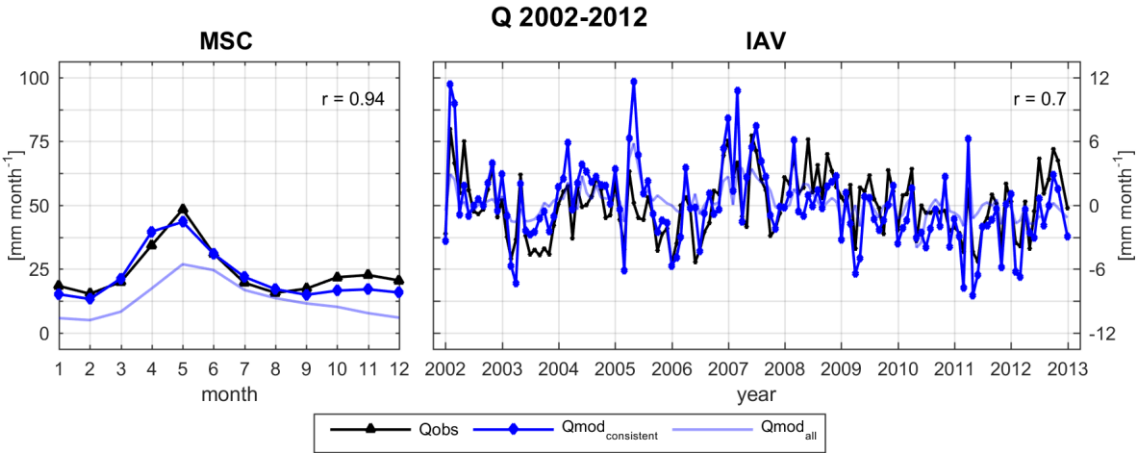

Figure A.3. Spatially averaged mean seasonal cycle (MSC) of the period 2002–2012 and inter-annual variability (IAV, difference between monthly values and the MSC) for Qmod and EU-grid runoff Qobs. $Qmod_{consistent}$ solely considers grid cells that coincide with Qobs, while $Qmod_{all}$ is based on modelled runoff for all grids of the study domain.

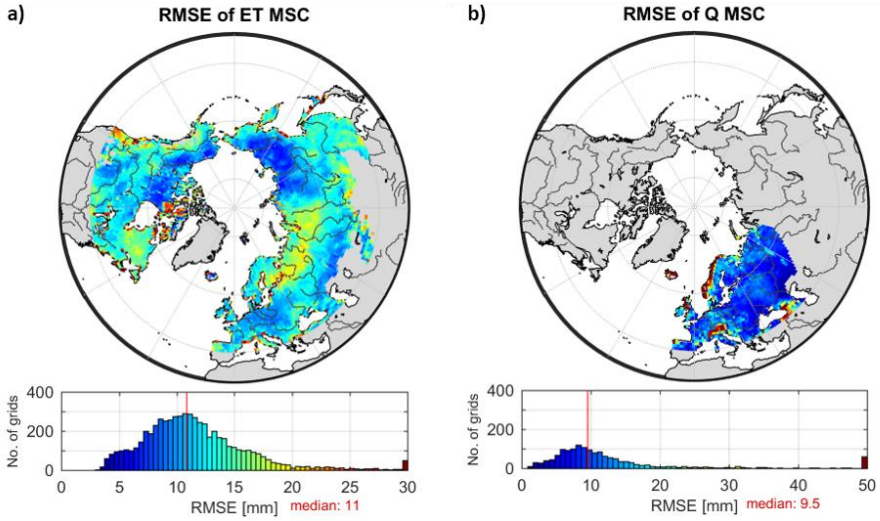

Figure A.4. RMSE of the mean seasonal cycle of simulated and observed a) ET [mm month-1] and b) Q [mm month-1]. RMSE values have been truncated to the range 0–30 (a) resp. 0-50 (b).

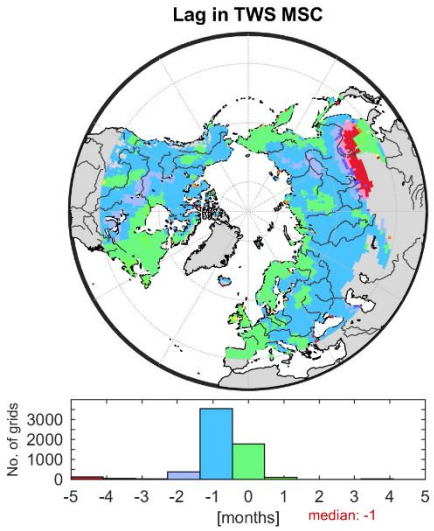

Figure A.5. Grid wise phase lag [months] between mean seasonal TWSobs and TWSmod. Negative values indicate preceding of the model compared to GRACE TWS.

A.6 Comparison with eartH2Observe Models

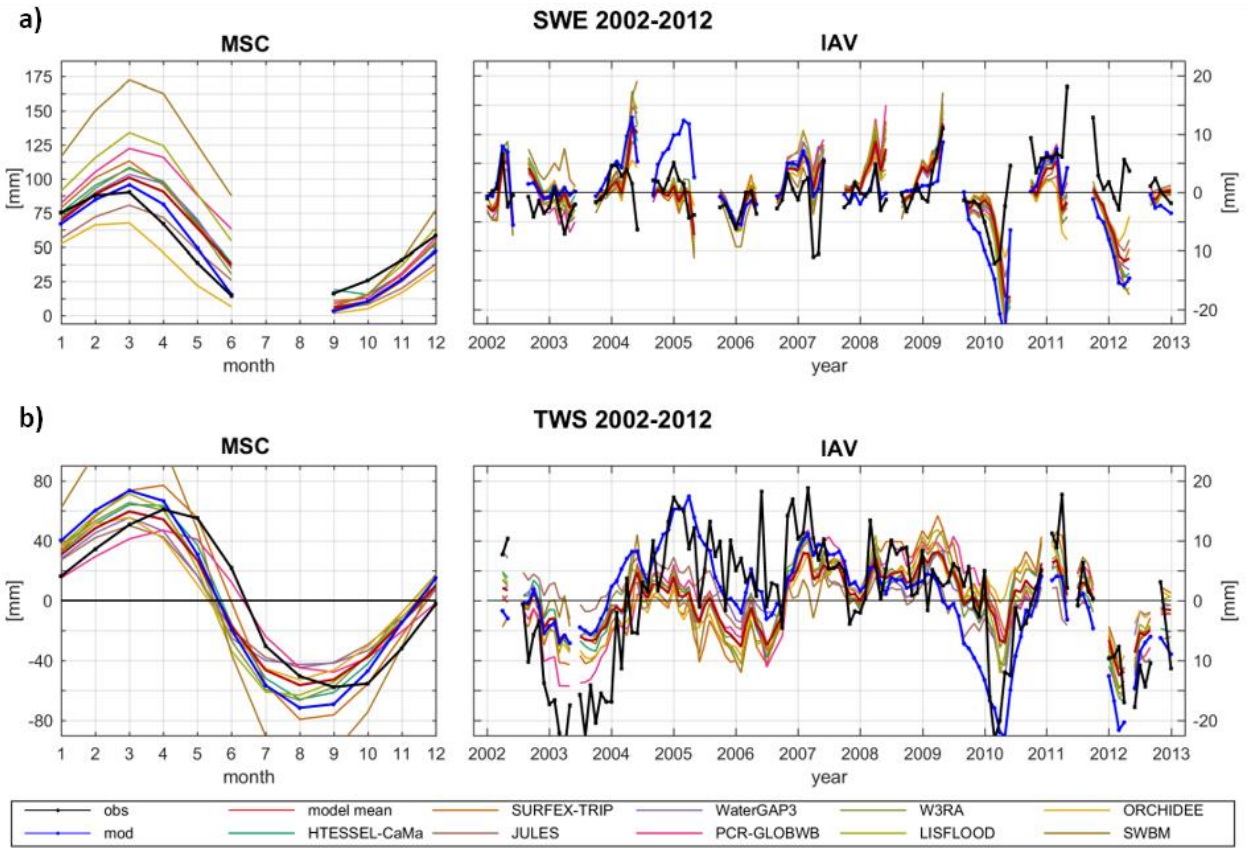

Figure A.6. Comparison of spatially averaged observed (obs) a) SWE (GlobSnow) and b) TWS (GRACE) to simulations of this study (mod) and eartH2Observe models (incl. ensemble mean) in terms of average mean seasonal cycle (MSC) and inter-annual variability (IAV). MSC is calculated for the period 2002–2012, and IAV represents the difference of monthly values from the MSC. Only data points consistent between all models and the respective observational data are considered.

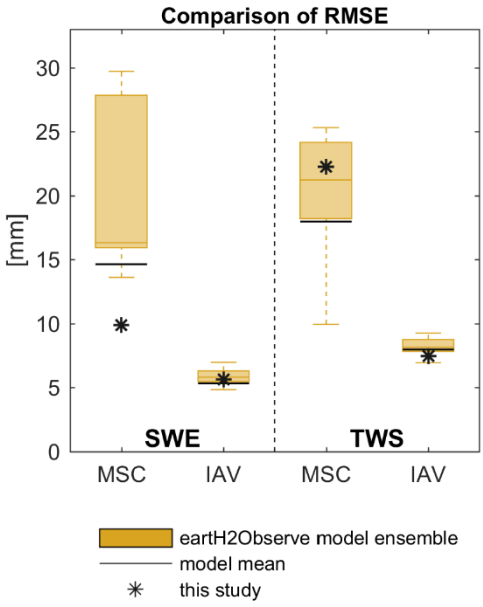

Figure SA.7. RMSE for the spatially averaged SWE (left) and TWS (right) achieved by our model compared to the model ensemble of eartH2Observe models and the ensemble mean across temporal scales.

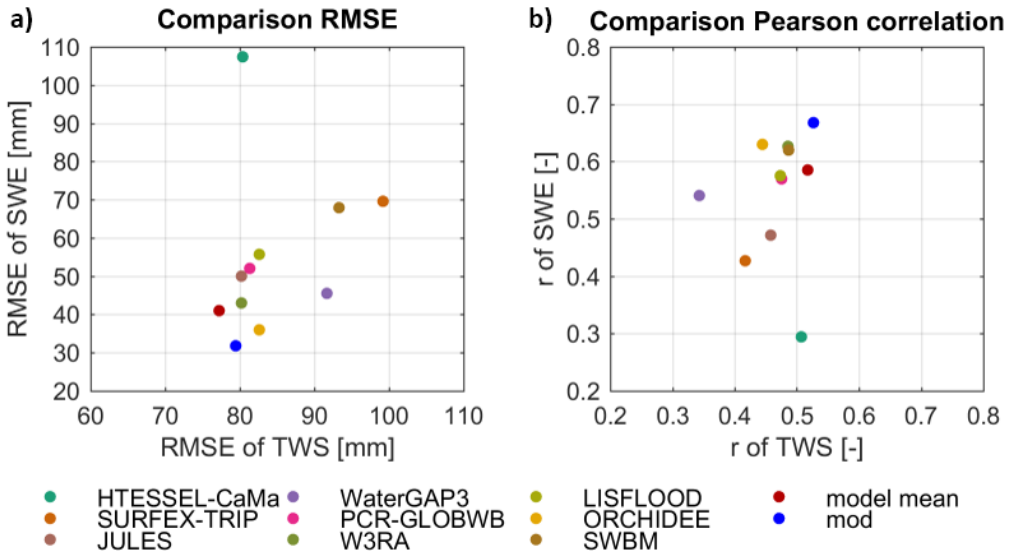

Figure A.8. Comparison of a) RMSE and b) Pearson correlation r for monthly SWE and TWS time series simulated with the eartH2Observe models, the model ensemble mean (model mean) and by our model (mod).

A.7 Uncertainty due to Forcing and Calibration Data

A.7.1 Comparison to WFDEI Precipitation Forcing

To assess the uncertainty in TWSmod and SWEmod that emerges from the choice of precipitation forcing, we calibrated our model in the same manner as before, yet used rain fall and snow fall estimates from the reanalysis based WFDEI product (Weedon et al. 2014) instead of GPCP-1DD precipitation data. Since precipitation is likely the most uncertain input data (Herold et al. 2015, Schellekens et al. 2017), we did not change the temperature and net radiation data sets. The global meteorological WFDEI data for land area is generated by applying the Water and Global Change (WATCH) forcing data methodology to ERA-Interim reanalysis data (Dee et al. 2011). The advantage of the WFDEI product is that it already provides separate values for snow and rain fall, as diagnosed by the reanalysis (Weedon et al. 2014). Therefore, it is not necessary to partition precipitation based on a temperature threshold within the model. We rather applied the provided rain and snow fall estimates directly, and also desisted from scaling snow fall.

Regarding the MSC, we obtained similar model performance in terms of SWE and TWS for both, the spatially averaged dynamics (Fig. A.9, Fig. A.10) and the spatial pattern (not shown). Although the dynamics and thus the correlation coincidence, we obtain a higher amplitude in TWSmod when using WFDEI as forcing compared to the original TWSmod (and TWSobs). This higher amplitude relates to larger seasonal snow accumulation in $SWEmod_{WFDEI}$, because the scaling parameter for snow fall is not calibrated. In terms of IAV, the correlation between observation and WFDEI forced model is comparable for both, TWS and SWE. However, the key findings (Fig. A.11) remain the same as with GPCP precipitation forcing.

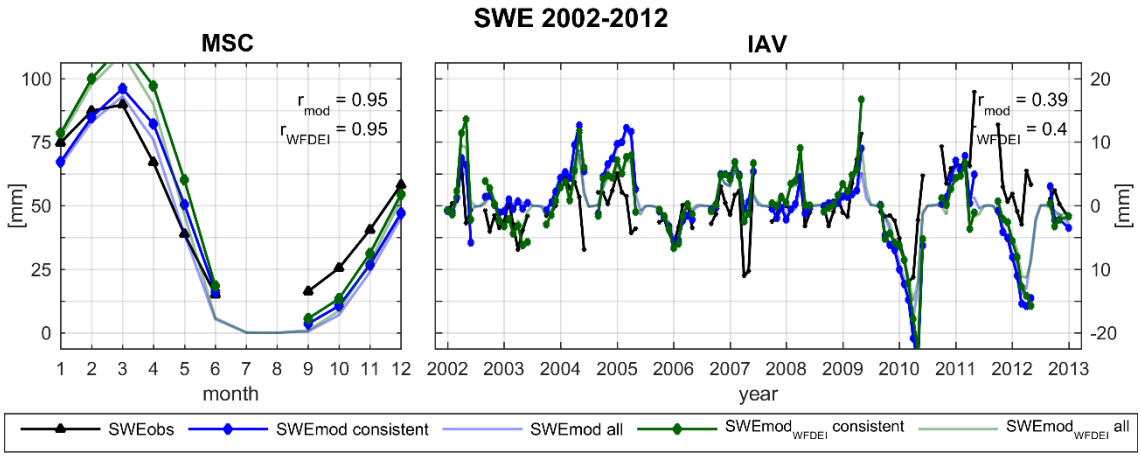

Figure A.9. Comparison of the spatially averaged mean seasonal cycle (MSC) and inter-annual variability (IAV, difference between monthly values and the MSC) of observed SWE (SWEobs), modelled SWE (SWEmod), and modelled SWE based on WFDEI precipitation forcing (SWEmod$_{WFDEI}$). SWEmod consistent and SWEmod$_{WFDEI}$ consistent refers to modelled SWE considering only data points with available SWEobs, while SWEmod all and SWEmod$_{WFDEI}$ all incorporates all time steps for all grids of the study domain.

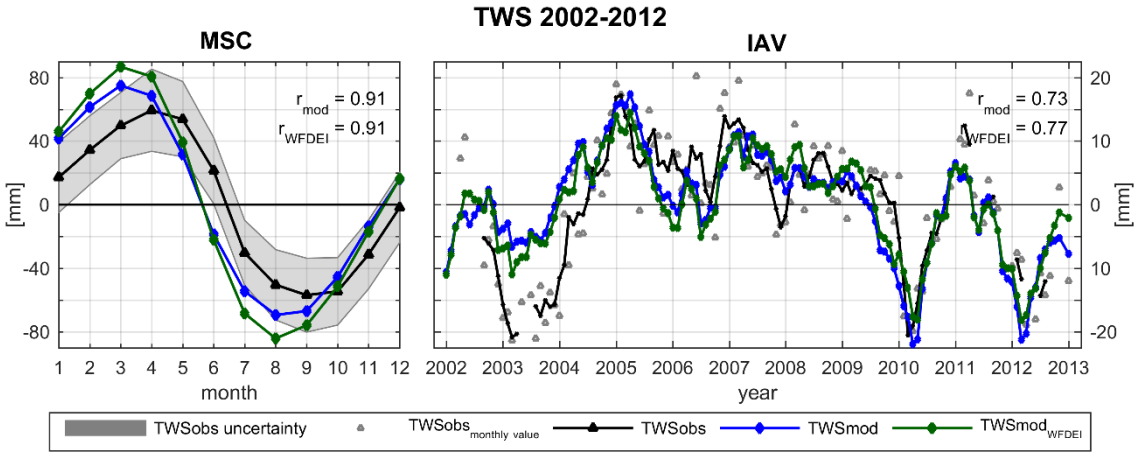

Figure A.10. Comparison of the spatially averaged mean seasonal cycle (MSC) and inter-annual variability (IAV, difference between monthly values and the MSC) of observed TWS (TWSobs), modelled TWS (TWSmod), and modelled TWS based on WFDEI precipitation forcing (TWSmod$_{WFDEI}$). For IAV, TWSobs$_{monthly\ value}$ shows the original IAV of individual TWSobs months, while TWSobs, TWSmod and TWSmod$_{WFDEI}$ are smoothed using a 3-month average moving window filter. Pearson correlation r refers to the smoothed values. For the MSC no smoothing is applied.

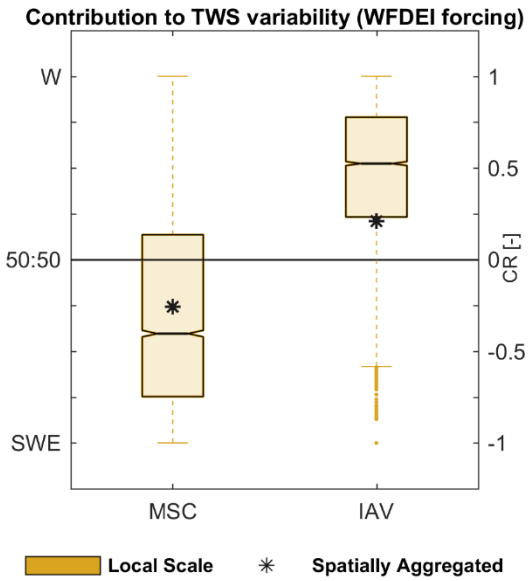

Figure A.11. Relative contribution (based on CR (Eq.2.3)) of snow (SWE) and liquid water (W) to TWS variability on different spatial (local grid scale, spatially integrated) and temporal (mean seasonal MSC, inter-annual IAV) scales when forced with WFDEI rain and snow fall. The boxplots represent the distribution of grid cell CR, with the dashed line marking the corresponding average. The star represents the CR calculated for the spatially integrated values.

A.7.2 Comparison to other GRACE Solutions

In this study we used TWS estimates from the JPL mascon RL05 product for model calibration and evaluation (Watkins et al. 2015;Wiese, 2015). However, various GRACE solutions for TWS from different institutions and using different processing approaches exist. To assess the potential uncertainty resulting from the choice of TWS solution, we compared modelled TWS (mod) and the JPL mascon solution (JPL_{masc}) with other solutions based on different processing approaches. They include the mascon product from the Center of Space Research (CSR at the University of Texas) (CSR_{masc}) (Save et al. 2016), as well as three RL05 solutions based on spherical harmonics provided by JPL, CSR and Geoforschungs Zentrum (GFZ) (Swenson and Wahr 2006, Landerer and Swenson 2012, Swenson 2012). As recommended, we also considered the average of the latter three ($Avg_{JPL/CSR/GFZ}$). All TWS estimates were taken as anomalies to the respective time-mean of 2002–2012, and scaled with the provided gain factors (except for CSR_{masc} that does not require scaling (Save et al. 2016). For comparison, we calculated the spatial average mean seasonal cycle (MSC) and inter-annual variability (IAV) across all grid cells of the study domain (Fig. A.12).

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

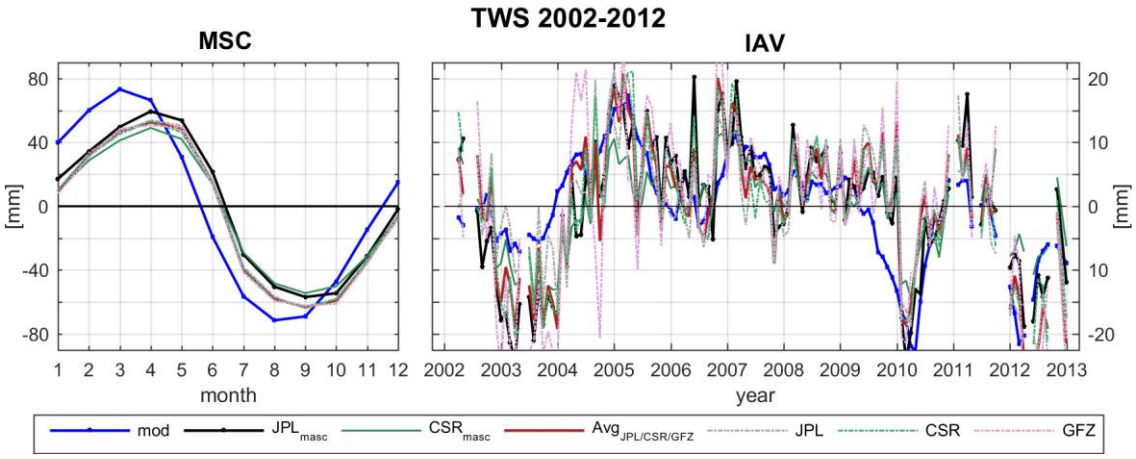

Figure A.12. Comparison of the spatially averaged mean seasonal cycle (MSC) and inter-annual variability (IAV, difference between monthly values and the MSC) of modelled TWS (mod) and observed TWS of different GRACE solutions.

A.8 Covariances between SWE and W

Figure A.13 and Fig. A.14 compare the contribution of the combined SWE and W variances and the covariance of both storages to the total variance of the spatially aggregated TWSmod. On the inter-annual and spatially aggregated scale, 81 % of TWS variability is explained by the variances in SWE and W, suggesting that the covariance between SWE and W only has minor effect. This is underlined by high percentage of SWE and W variance on total TWSmod variance for all grids of the study domain (Fig. A.14). On mean seasonal scales, the majority of spatially aggregated TWS variability is still explained by variances in SWE and W, but the contribution of the covariance increases. This can be expected, as the seasonal variation of snow storage affects the subsequent availability of liquid water storages through the snowmelt process. At the local scale, though, the percentage of SWE and W variance on total TWSmod variance remains high in regions where the dominance of either snow or liquid water components are clear (Fig. 2.7). In regions where covariances of two storage components is larger, the contribution of two storage components to TWS variability are similar resulting in a CR value of around 0. Therefore, we conclude that while the covariances of snow and liquid water can be remarkable on the seasonal scale over a large spatial domain, it does not affect or change the dominant components on the TWS.

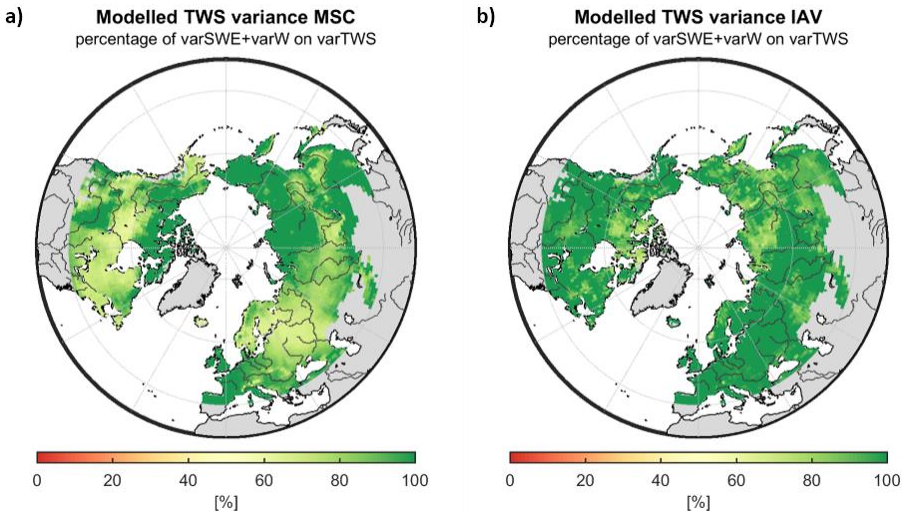

Figure A.13. Percentage of SWE and W variance on total TWSmod variance on mean seasonal (MSC) and inter-annual (IAV) scales.

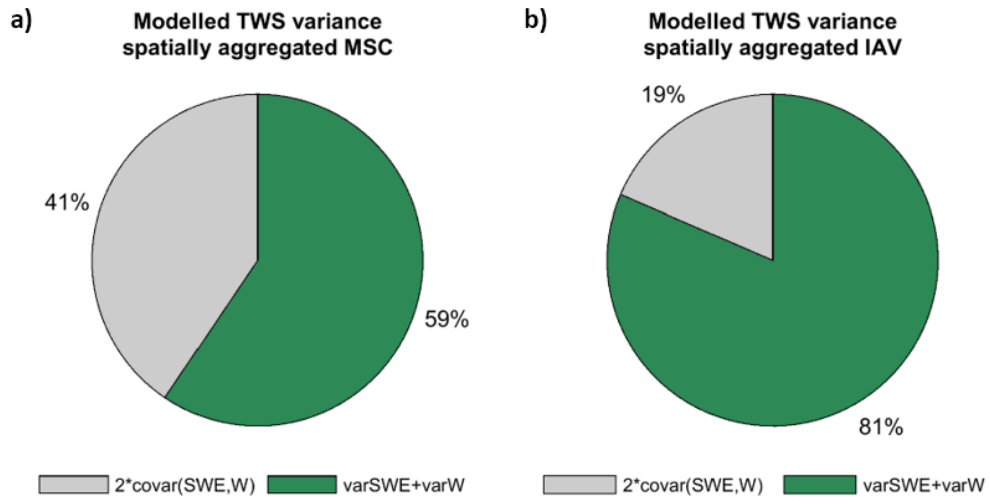

Figure A.14. Percentage composition of spatially aggregated TWSmod variance from the combined variances of SWE and W, and two times the covariance of SWE and W on mean seasonal (MSC) and inter-annual (IAV) scales.

Appendix B

Supplement of Chapter 3

Supplement material of:

Trautmann, T., Koirala, S., Carvalhais, N., Güntner, A., Jung, M. (2022): The importance of vegetation in understanding terrestrial water storage variations, Hydrology and Earth System Sciences, 26 (4): 1089-1109, doi:10.5194/hess-26-1089-2022

Available at:

https://doi.org/10.5194/hess-26-1089-2022-supplement

B.1 Spatial Pattern of scaled Parameters in VEG

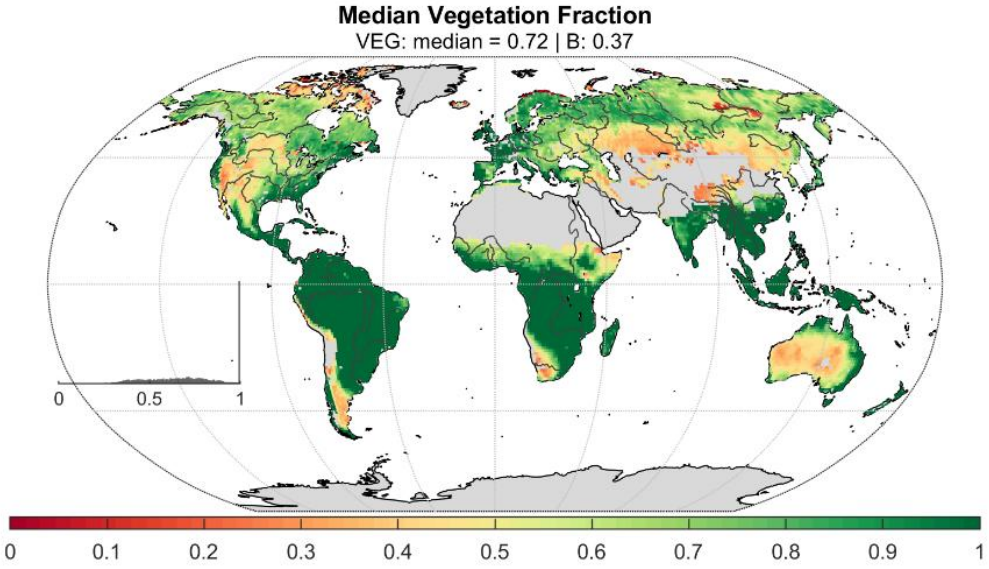

Figure B.1. Global distribution of the median vegetation fraction p_{Veg} after calibration of the VEG experiment.

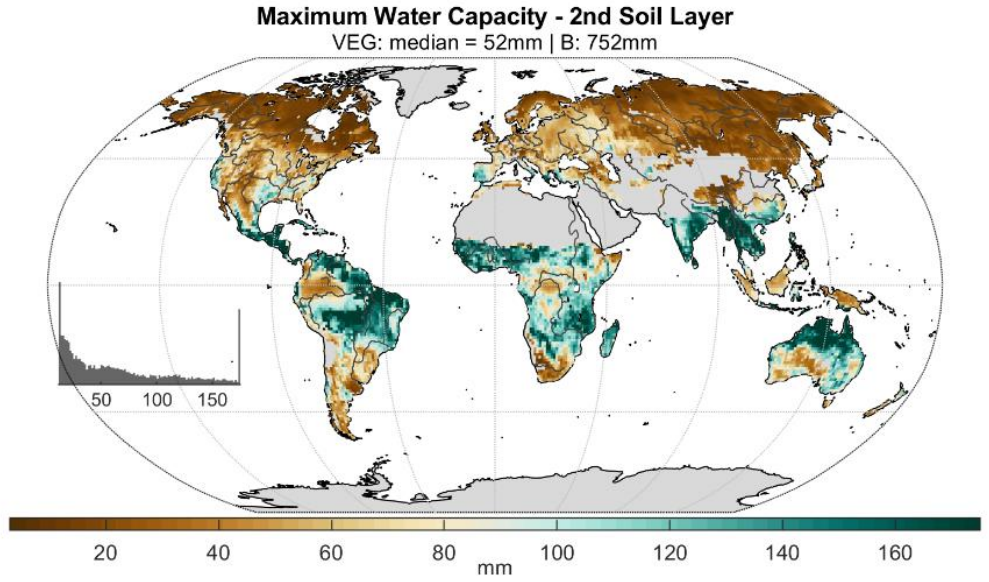

Figure B.2. Global distribution of the maximum water capacity of the 2nd soil layer $wSoil_{max(2)}$ after calibration of the VEG experiment.

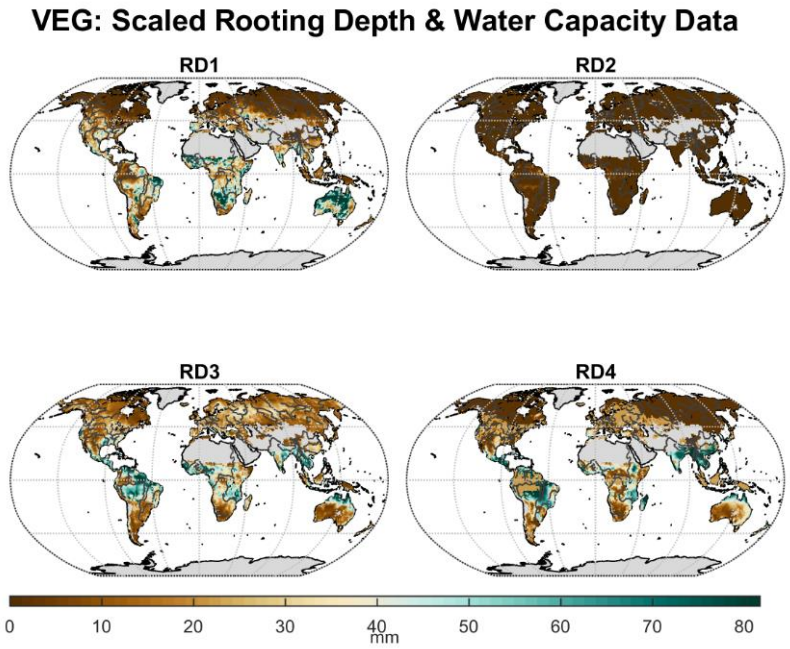

Figure B.3. Global distribution of the maximum water capacity of the 2nd soil layer contributed by each data stream after calibration of their scaling parameters in the VEG experiment. RD1 = maximum rooting depth by Fan et al. 2017; RD2 = effective rooting depth by Yang et al. 2016; RD3 = maximum soil water capacity by Wang-Erlandsson et al. 2016; RD4 = plant available water capacity by Tian et al. 2019.

B.2 Effective alpha Coefficient

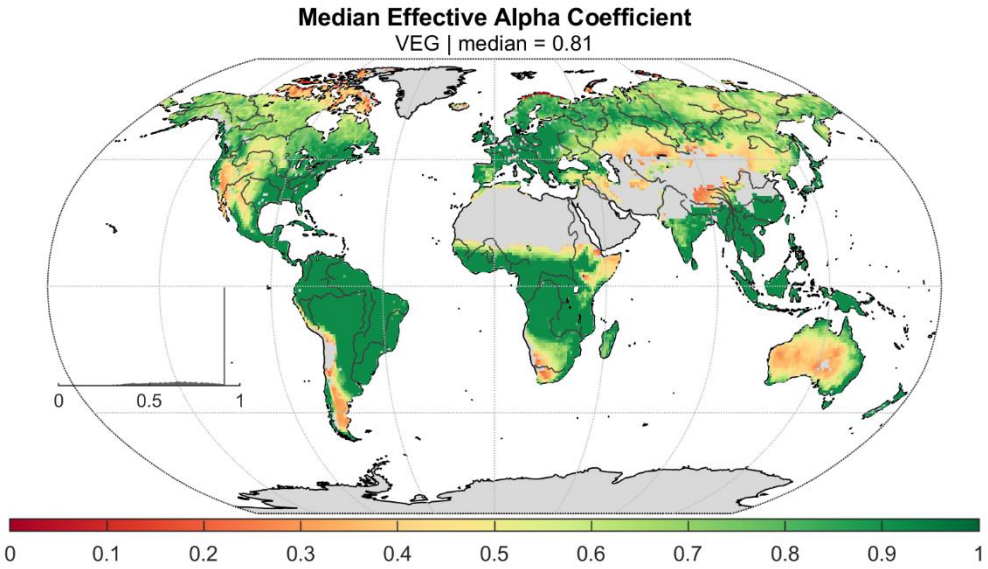

Figure B.4. Global distribution of the median effective alpha coefficient (α_{Veg} * p_{Veg}) in the Priestley-Taylor formula after calibration of the VEG experiment.

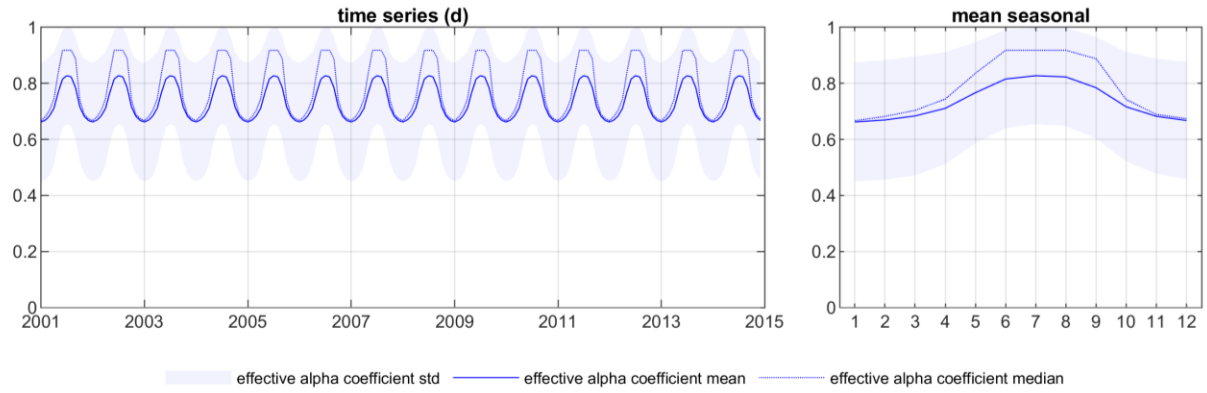

Figure B.5. Daily time series and mean seasonal dynamics of the area weighted average, median and standard deviation of the grid-wise effective alpha coefficient in the Priestley-Taylor formula of the calibrated VEG experiment.

B.3 Parameter Correlation

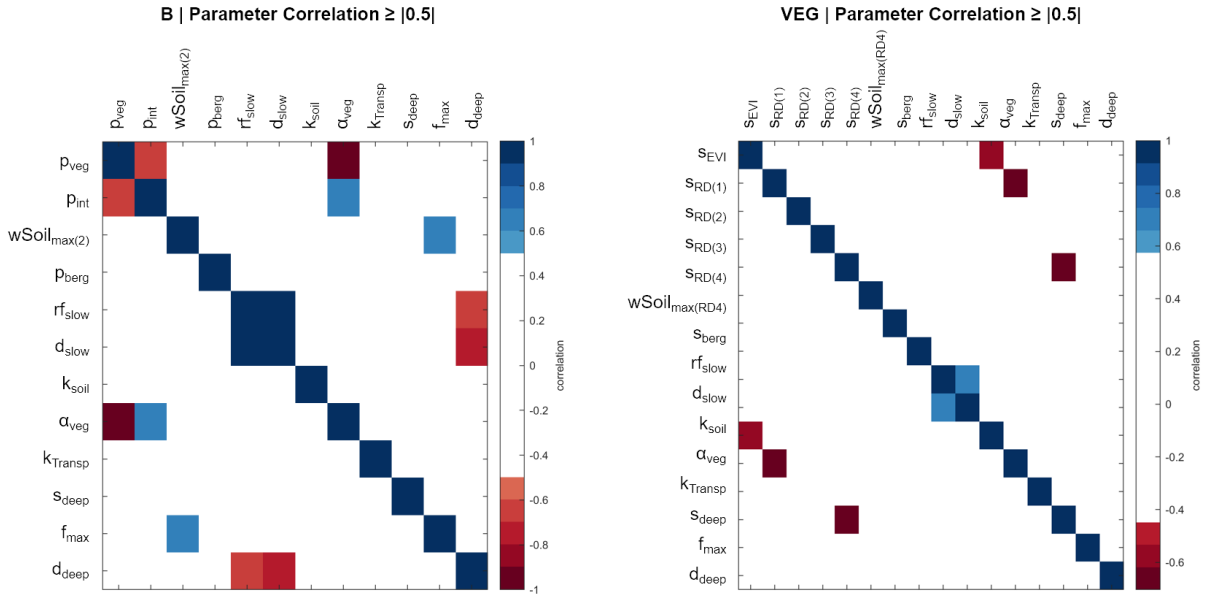

Figure B.6. Correlation (≥ |0.5|) between model parameters for the B and VEG experiment.

B.4 Regional inter-annual TWS Composition

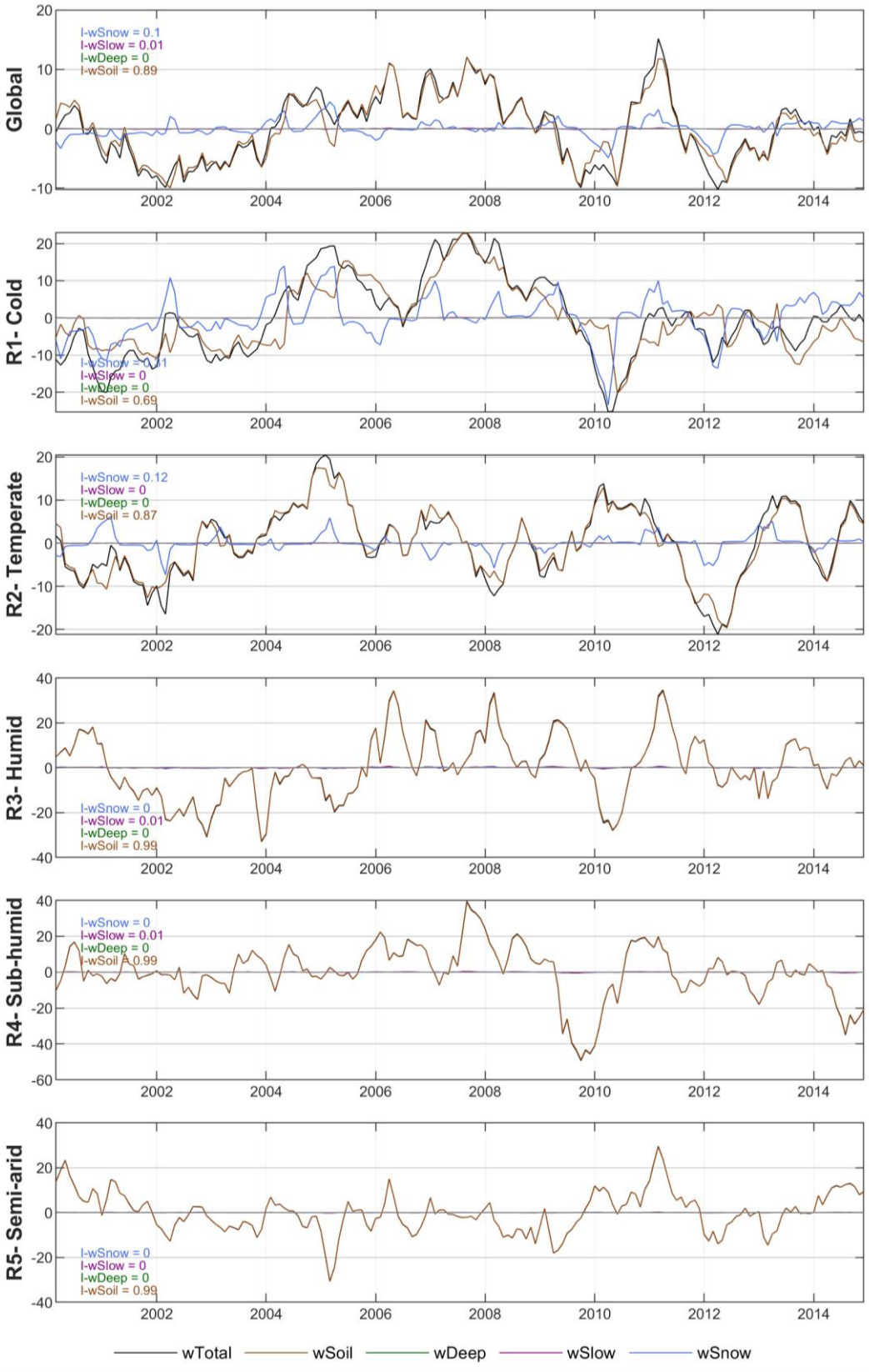

Figure B.7. Global and regional average inter-annual variability of simulated total water storage (wTotal) and its components (wSoil, wDeep, wSlow, wSnow) for B, including the regional Impact Index I for each storage.

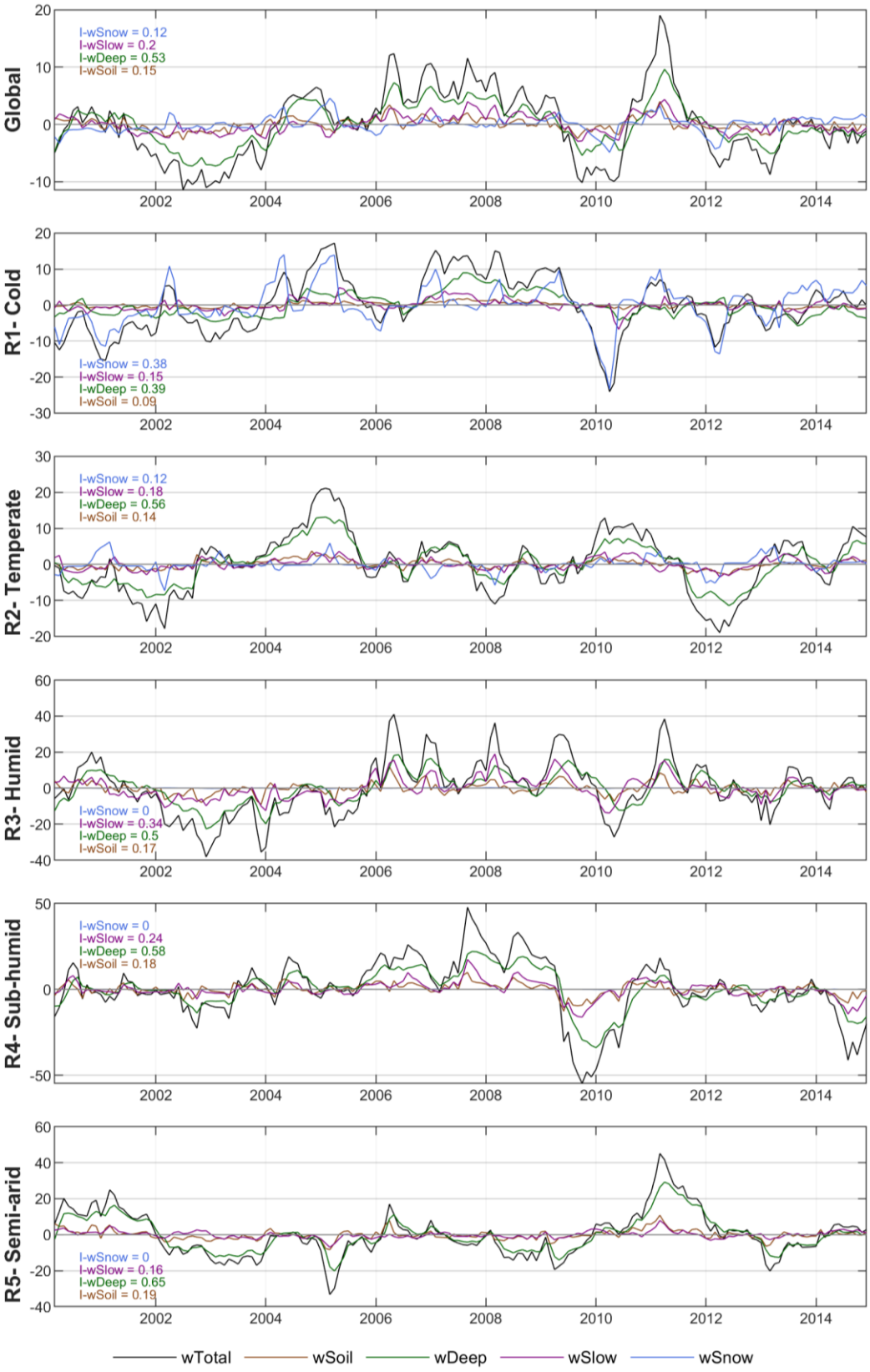

Figure B.8. Global and regional average inter-annual variability of simulated total water storage (wTotal) and its components (wSoil, wDeep, wSlow, wSnow) for VEG, including the regional Impact Index I for each storage.

B.5 Transpiration over Evapotranspiration

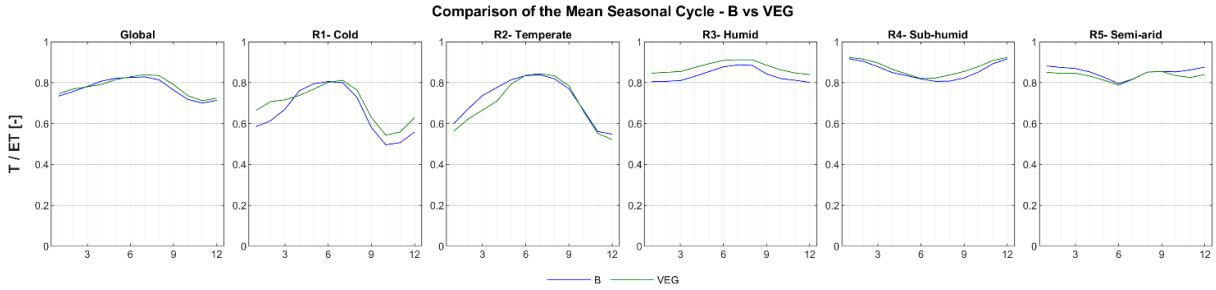

Figure B.9. Global and regional average mean seasonal cycles of modelled transpiration (T) over evapotranspiration (ET) for B and VEG experiments.

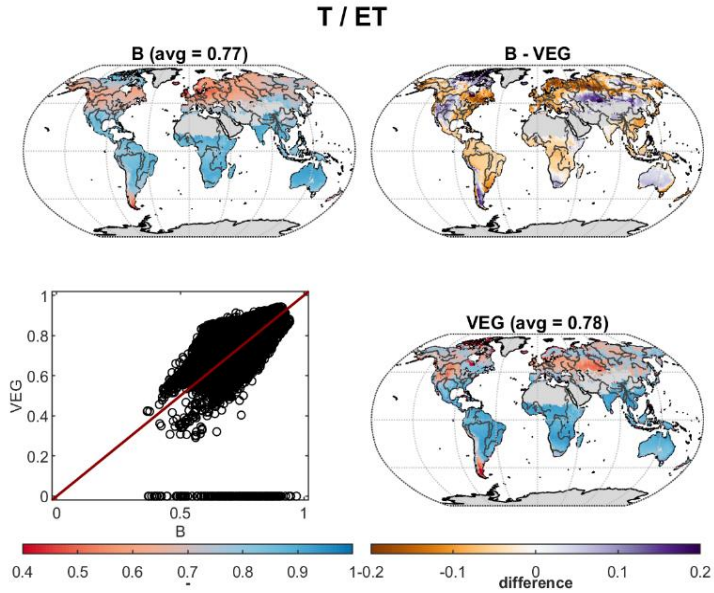

Figure B.10. Global distribution of modelled transpiration (T) over evapotranspiration (ET) for B and VEG experiments, as well as the difference between both (lower right).

B.6 Runoff Components

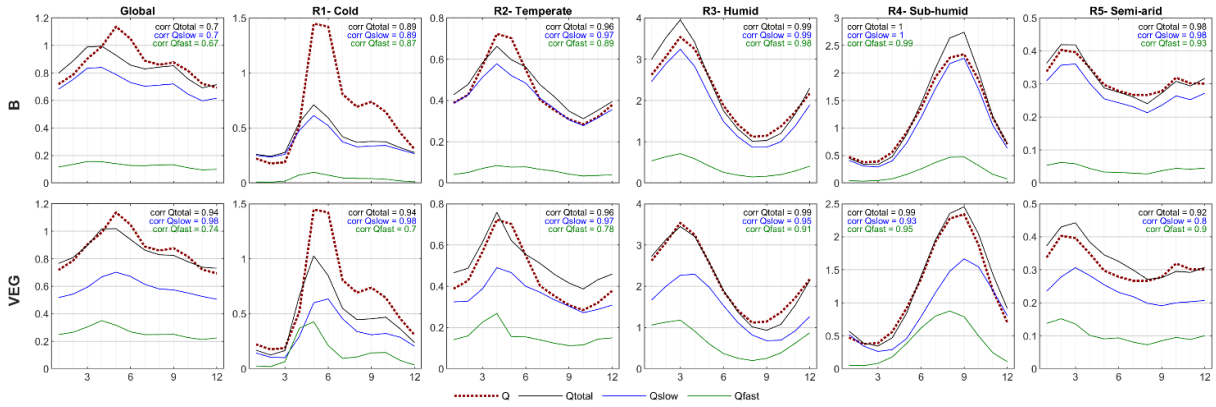

Figure B.11. Global and regional average mean seasonal cycle of observed grid-wise runoff from GRUN (Q) and simulated total runoff (Q_{total}), as well as its components Q_{slow} and Q_{fast}, for the B and VEG experiments. corr is the Pearson correlation coefficient of the respective simulation with observed Q.

B.7 Comparison of VEG & VEG without Capillary Rise

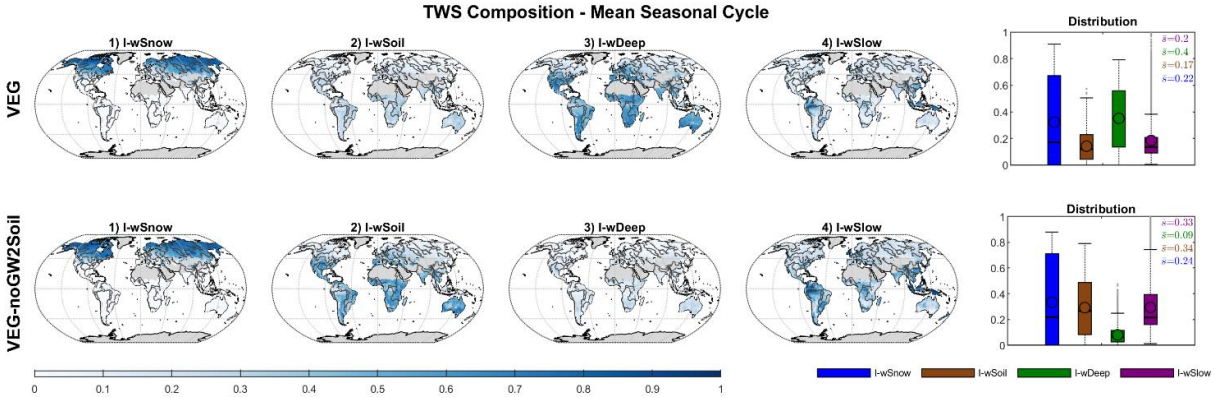

Figure B.12. Global distribution of the Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the mean seasonal cycle of total water storage, for VEG and VEG-noGW2Soil, which is a variant of the VEG experiment, in with the capillary rise from wDeep to wSoil is turned off prior to model calibration.

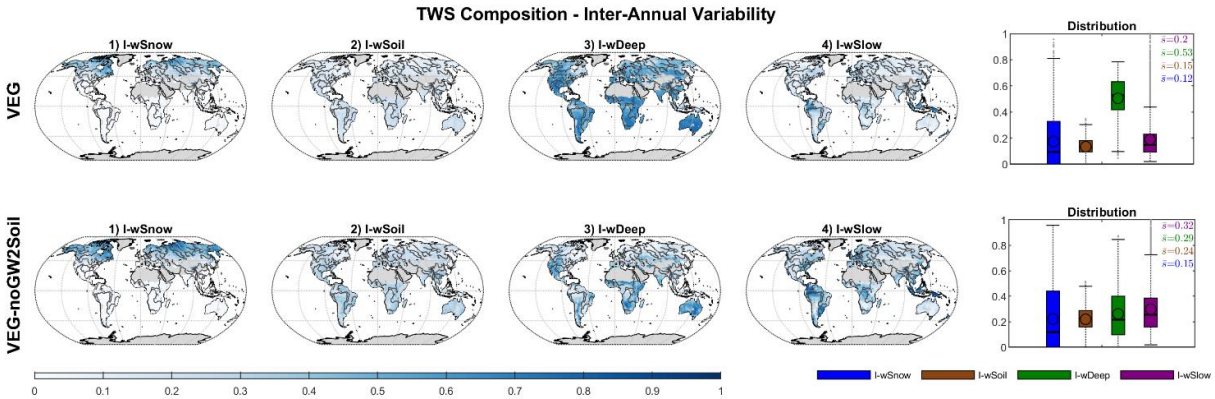

Figure B.13. Global distribution of the Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the inter-annual anomalies of total water storage, for VEG and VEG-noGW2Soil, which is a variant of the VEG experiment, in with the capillary rise from wDeep to wSoil is turned off prior to model calibration.

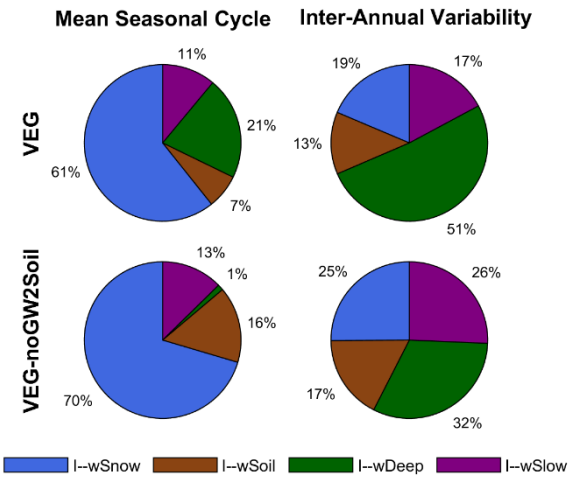

Figure B.14. Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the global average mean seasonal cycle and inter-annual variability of total water storage, for VEG and VEG-noGW2Soil, which is a variant of the VEG experiment, in with the capillary rise from wDeep to wSoil is turned off prior to model calibration.

B.8 Comparison of VEG & VEG with fixed k_{Transp} at 0.05

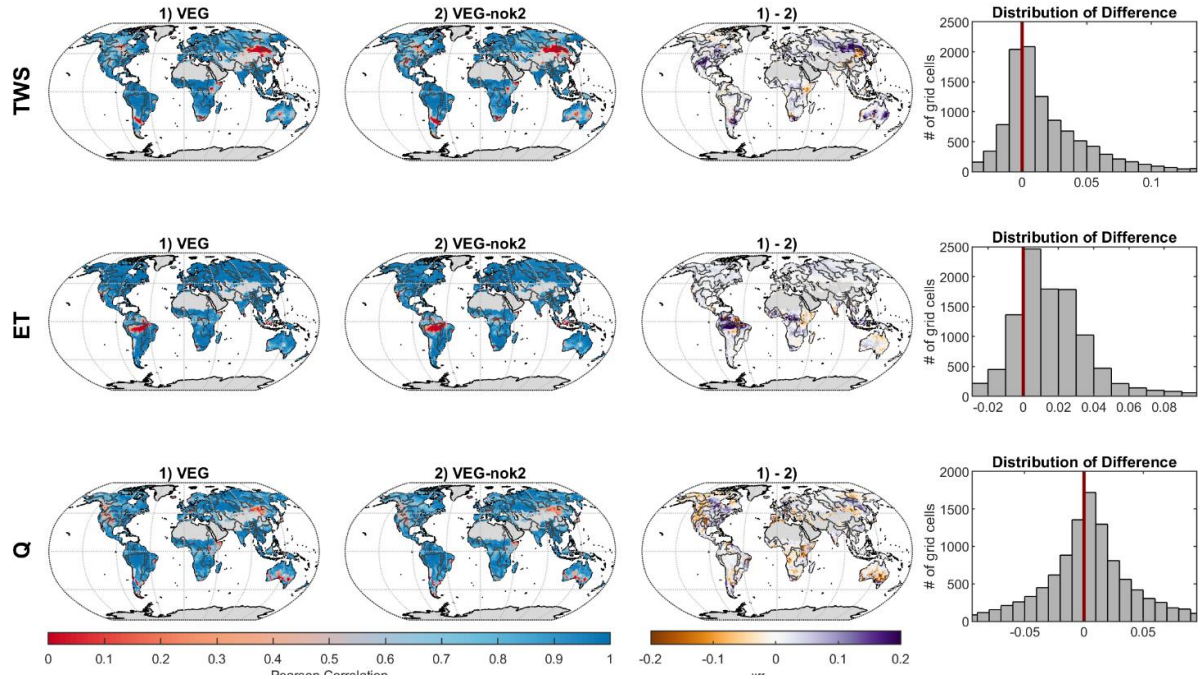

Figure B.15. Grid-wise Pearson's correlation coefficient for total water storage (TWS), evapotranspiration (ET) and runoff (Q) between 1) observations and VEG, and 2) observations and VEG-nok2, as well as differences between 1) and 2) (brown color, i.e., negative values indicate higher correlations for VEG-nok2, while purple color, i.e., positive values indicate better correlation values for VEG). VEG-nok2 is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

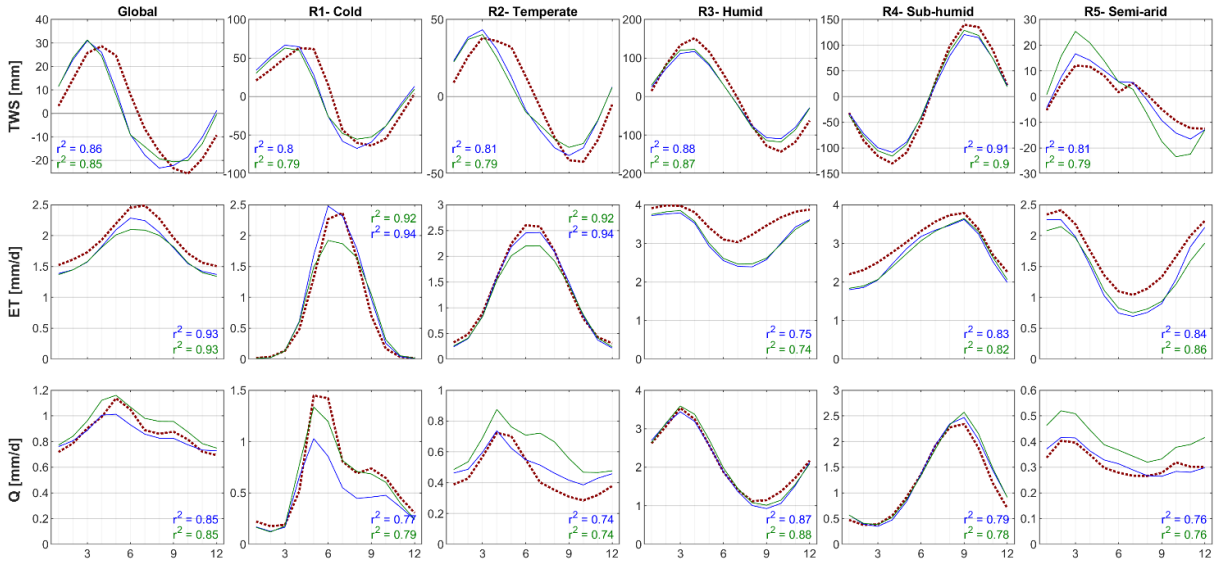

Figure B.16. Global and regional mean seasonal cycles of total water storage (TWS), evapotranspiration (ET) and runoff (Q) for VEG and VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05, compared to the observational constraints by GRACE (TWS), FLUXCOM (ET) and GRUN (Q).

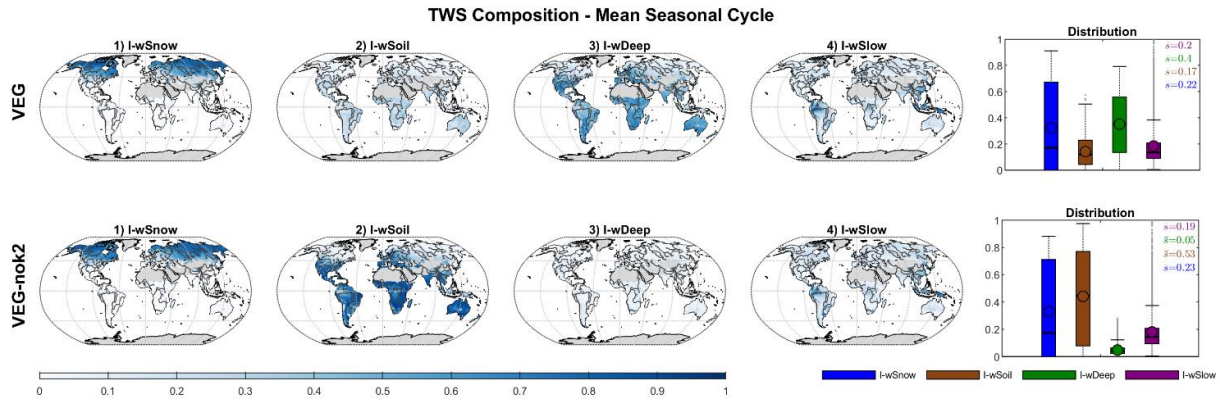

Figure B.17. Global distribution of the Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the mean seasonal cycle of total water storage, for VEG and VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

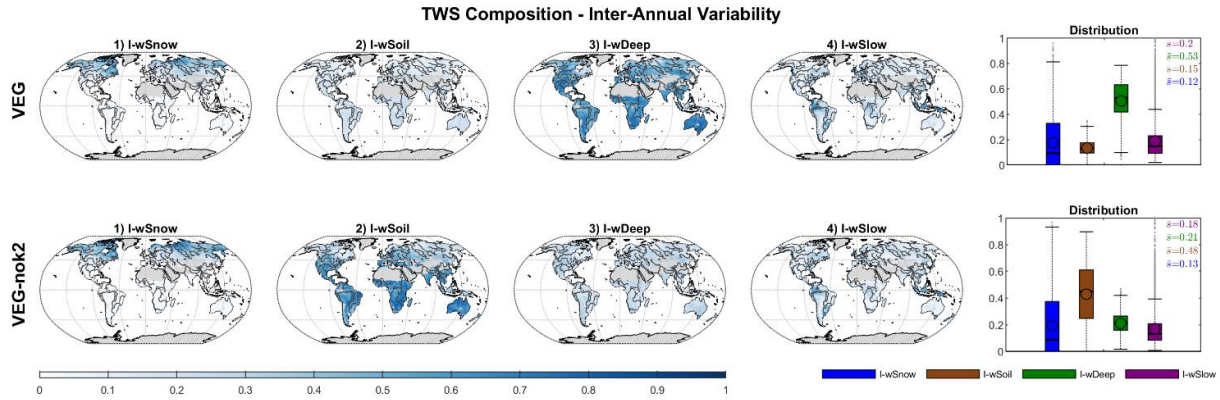

Figure B.18. Global distribution of the Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the inter-annual anomalies of total water storage, for VEG and VEG- VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

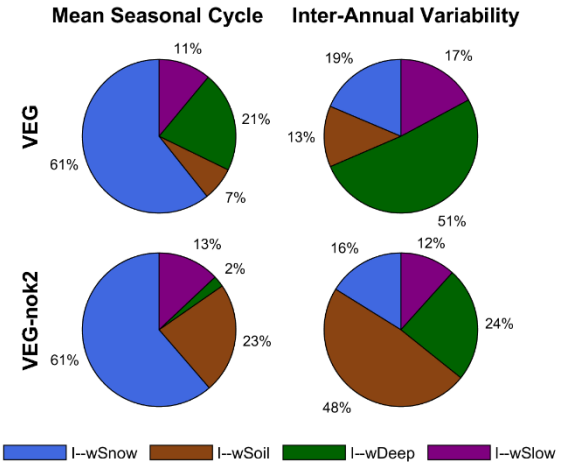

Figure B.19. Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the global average mean seasonal cycle and inter-annual variability of total water storage, for VEG and VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

B.9 PFT Experiment

The following shows an experiment similar to the traditional approach of global hydrological models, in which vegetation-dependent parameters are defined and calibrated for different plant-functional type (PFT) classes separately and then model performance and TWS composition is analyzed in comparison to the B and VEG experiments. The results show that the larger number of parameters (due to different sets for different PFT) does not lead to marked improvements of model performance, but instead increases parameter uncertainty possibly due to overparameterization. In terms of TWS composition, we see substantial differences in the PFT experiment compared to B and VEG, which underlines our conclusions that the representation of vegetation in GHMs is critical for interpreting TWS variations.

Based on the GSWP2 land cover classification (Dirmeyer et al. 2006), we consider 12 PFT classes (Fig. B.20), for which we define individual values of $wSoil_{max(2)}$ (maximum available water capacity of the 2nd soil layer) and s_{berg} (scaling parameter to derive the runoff/infiltration coefficient). Since state-of-the-art global hydrological models (GHMs) usually include seasonal dynamics of leaf area index (LAI) to calculate, e.g., transpiration, we decided to keep the definition of the active vegetation fraction as a function of seasonal EVI data as in the VEG experiment. For the PFT experiment, we focus (i) on $wSoil_{max(2)}$ because GHMs usually apply a PFT specific rooting depth, and (ii) on s_{berg} because this is similar to the runoff coefficient γ which is tuned in some GHMs (e.g., the WaterGAP model (Müller-Schmied et al. 2021)).

When considering these 12 PFT classes, the number of calibration parameters increases from 12 (in B) and 16 (in VEG) to 34 (in PFT). Analysis of parameter uncertainty shows high uncertainties for a set of parameters common with B, while optimized parameter values are between those of B and VEG (Table B.1). Additionally, and unlike B and VEG, PFT has high uncertainty of $wSoil_{max(2)}$ for all PFT classes, and high correlation between each PFT's $wSoil_{max(2)}$ and s_{berg} (Fig. B.21). High uncertainty of $wSoil_{max(2)}$ is an indication that having one $wSoil_{max(2)}$ per PFT may not explain the within-PFT variability. On the other hand, high correlation between each PFT's $wSoil_{max(2)}$ and s_{berg} is systematic, as both parameters are based on the same spatial distribution of PFT classes - and highlights an advantage of the VEG experiment, in which both are based on independent data sets.

In terms of model performance, Fig. B.22 shows a partial improvement for TWS and ET in the PFT experiment. Especially in the Humid and Sub-humid regions, TWS simulation in PFT matches GRACE observations better. These regions include tropical regions, where data for maximum plant available water capacity by Tian et al. 2019 (RD4) are not available. While we filled the missing values for tropical regions with the same $wSoil_{max(RD4)}$ value as in the Northern latitudes, the improved performance in the PFT experiment suggests that at least 2 different $wSoil_{max(RD4)}$ fill values seem necessary for different climate regions. In contrast to TWS and ET, PFT performance of Q is poorer than in B and VEG, with a clear underestimation of the seasonal variability. To consider model performance in relation to the number of calibration parameters, we calculated the Akaike information criterion (AIC). Since low values of AIC indicate better performance compared to the other experiments, PFT only performs superior regarding ET, while the increased number of model parameters isn't advantageous regarding TWS and Q simulations. Also, note that the increased number of model parameters comes at an additional computational cost.

Furthermore, the results of the PFT experiment confirm that changing the representation of vegetation has a marked impact on the simulated TWS composition (Figs. B.23-B.25). In PFT, among the liquid water storages wSoil contributes most to mean seasonal TWS variability, with Impact Index values between those of B and VEG (Fig. B.23, Fig. B.25). Compared to VEG, wSlow is in general less important in PFT, while wDeep has a less impact on mean seasonal TWS, but its contribution to inter-annual TWS variability increases.

Overall, this analysis underlines that including continuous fields of vegetation parameters is preferable than the 'traditional' PFT-based approaches of defining parameters for distinct PFT classes (and their calibration) - in terms of model calibration and the uncertainty of calibrated model parameters, but also regarding model performance in relation to the number of model parameters. Furthermore, we could

highlight that the representation of vegetation in hydrological models is crucial for the partitioning of simulated TWS.

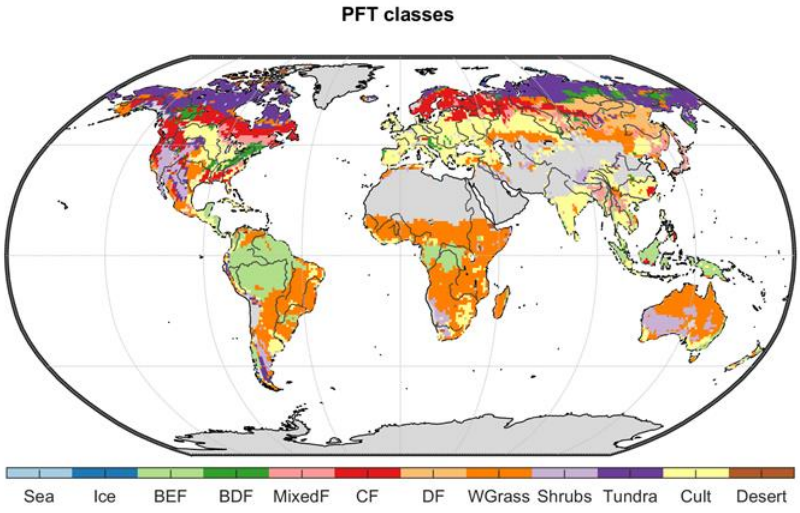

Figure B.20. Classes of plant functional type used in the PFT experiment. (Sea (PFT0); Ice=Continental Ice (PFT1); BEF=Broadleaf Evergreen Forest (PFT2); BDF=Broadleaf Deciduous Forest & Woodland (PFT3); MixedF=Mixed Coniferous & Broadleaf Deciduous Forest & Woodland (PFT4); CF= Coniferous Forest & Woodland (PFT5); DF=High Latitude Deciduous Forest & Woodland (PFT6); WGrass= Wooded C4 Grassland (PFT7); Shrubs=Shrubs & Bare Ground (PFT8); Tundra (PFT9); Cult=Cultivation (PFT10); Desert (PFT11)).

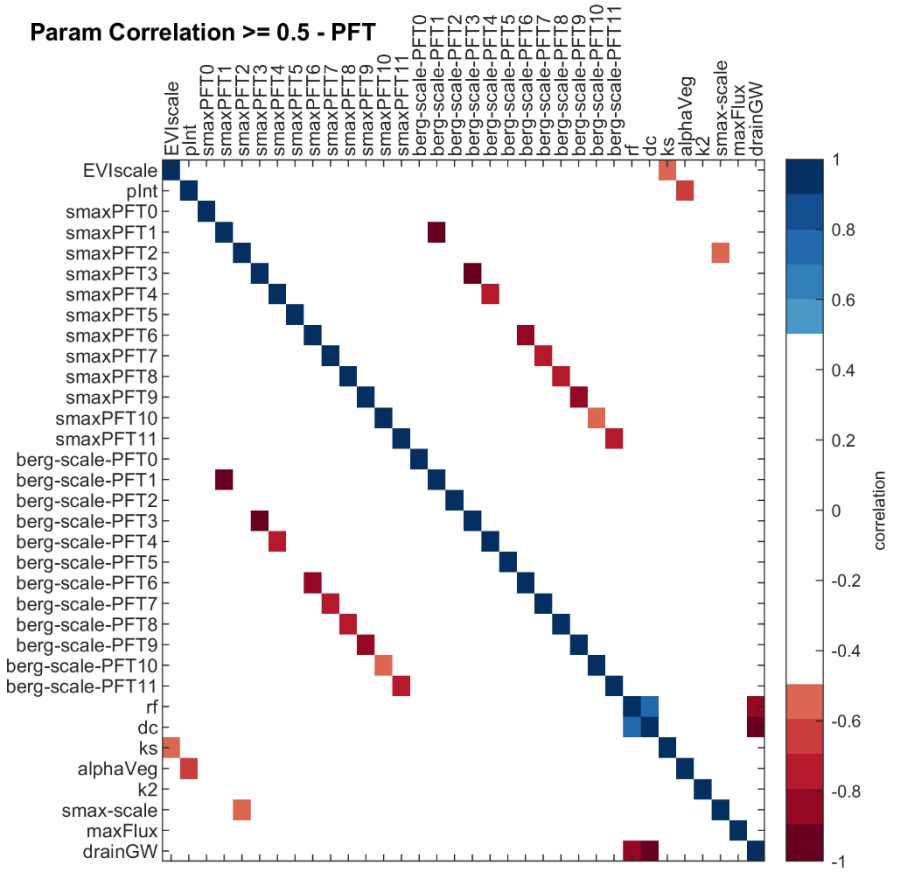

Figure B.21. Correlation of calibrated parameters for the PFT experiment. Shown are only correlation coefficients $|r| \geq 0.5$.

Table B.1. Calibrated parameter values and their uncertainty for B, VEG and PFT. Italic font indicates a calibrated parameter that hits the parameter bounds, and underlined numbers indicate parameter uncertainty ≥ 20%.

parameter	calibrated values ± uncertainty					
	B		VEG		PFT	
vegetation fraction						
p_{veg}	0.37	± 0.05				
s_{EVI}			3.89	± 0.05	3.75	± 0.03
evapotranspiration						
p_{int}	1	± 0.08	0.6	± 0.02	0.71	± 0.02
k_{Soil}	*0.1*	± 0.01	0.4	± 0.08	0.27	± 0.04
α_{veg}	2.25	± 0.15	0.92	± 0.00	0.87	± 0
k_{Transp}	0.12	± 0.32	0.48	± 1.76	0.5	± 4.32
deep soil						
s_{Deep}	9.1	± 461317	5.6	± 0.21	8.48	± 0.24
f_{max}	1.5	± 0.00	5.1	± 0.01	11.77	± 0.02
d_{Deep}	*1*	± 5.61	0.01	± 0.00	0.03	± 0
delayed water storage						
rf_{Slow}	0.78	± 1.72	0.68	± 0.01	0.62	± 0.05
d_{Slow}	*1*	± 2329	0.02	± 0.03	0.03	± 0.19
infiltration/runoff						
p_{berg}	1.32	± 0.02				
s_{berg}			3.08	± 0.02		
s_{berg_PFT0}					3.7	± 0.45
s_{berg_PFT1}					3.11	± 0.32
s_{berg_PFT2}					1.87	± 0.01
s_{berg_PFT3}					2.57	± 0.09
s_{berg_PFT4}					2.04	± 0.03
s_{berg_PFT5}					4.31	± 0.05
s_{berg_PFT6}					0.5	± 0.01
s_{berg_PFT7}					2.9	± 0.03
s_{berg_PFT8}					0.48	± 0.01
s_{berg_PFT9}					0.69	± 0.01
s_{berg_PFT10}					1.36	±0.01
s_{berg_PFT11}					2.5	± 0.11
soil moisture						
$wSoil_{max(2)}$	752	± 0.02				
$s_{RD(1)}$			0.01	± 0.00		
$s_{RD(2)}$			*0*	± 0.00		
$s_{RD(3)}$			0.15	± 0.06		
$s_{RD(4)}$			0.15	± 0.07		
$wSoil_{max(RD4)}$			145	± 0.08		
$wSoil_{max\_PFT0}$					1.57	± 8.94
$wSoil_{max\_PFT1}$					0.78	± 10.23
$wSoil_{max\_PFT2}$					1.01	± 0.41
$wSoil_{max\_PFT3}$					1.27	± 1.42
$wSoil_{max\_PFT4}$					0.5	± 0.5
$wSoil_{max\_PFT5}$					0.54	± 0.32
$wSoil_{max\_PFT6}$					0.85	± 2.53
$wSoil_{max\_PFT7}$					01.01	± 0.57
$wSoil_{max\_PFT8}$					1.45	± 2.72
$wSoil_{max\_PFT9}$					0.56	± 1.07
$wSoil_{max\_PFT10}$					0.39	± 0.2
$wSoil_{max\_PFT11}$					0.7	± 3.23

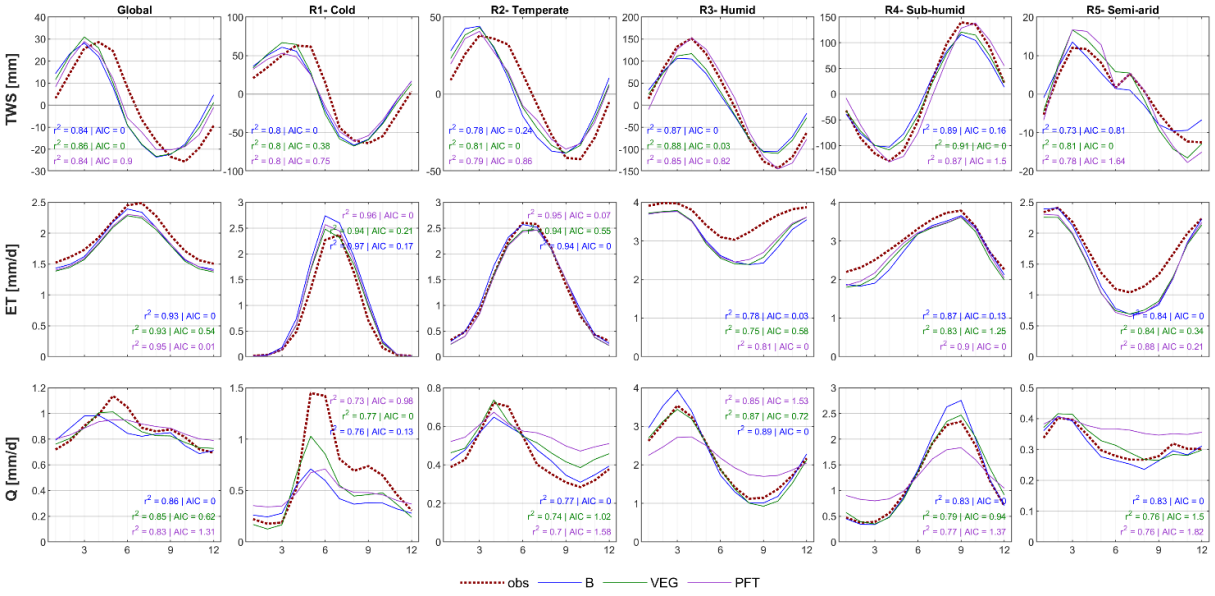

Figure B.22. Global and regional mean seasonal cycles of total water storage (TWS), evapotranspiration (ET) and runoff (Q) for the B, VEG and PFT experiments compared to the observational constraints by GRACE (TWS), FLUXCOM (ET) and GRUN (Q). For each, the Pearson correlation (r^2) and Akaike information criterion (AIC) are calculated to compare model performance in terms of seasonal dynamics and of mean standard error in relation to the number of calibration parameters.

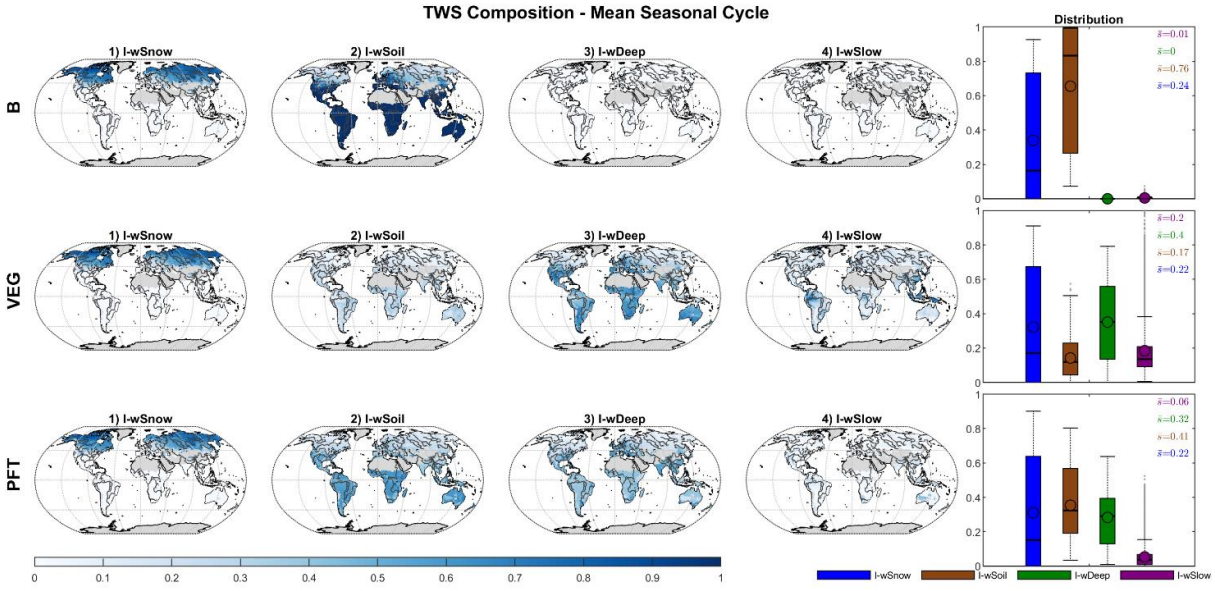

Figure B.23. Global distribution of the Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the mean seasonal cycle of total water storage, for B, VEG and PFT.

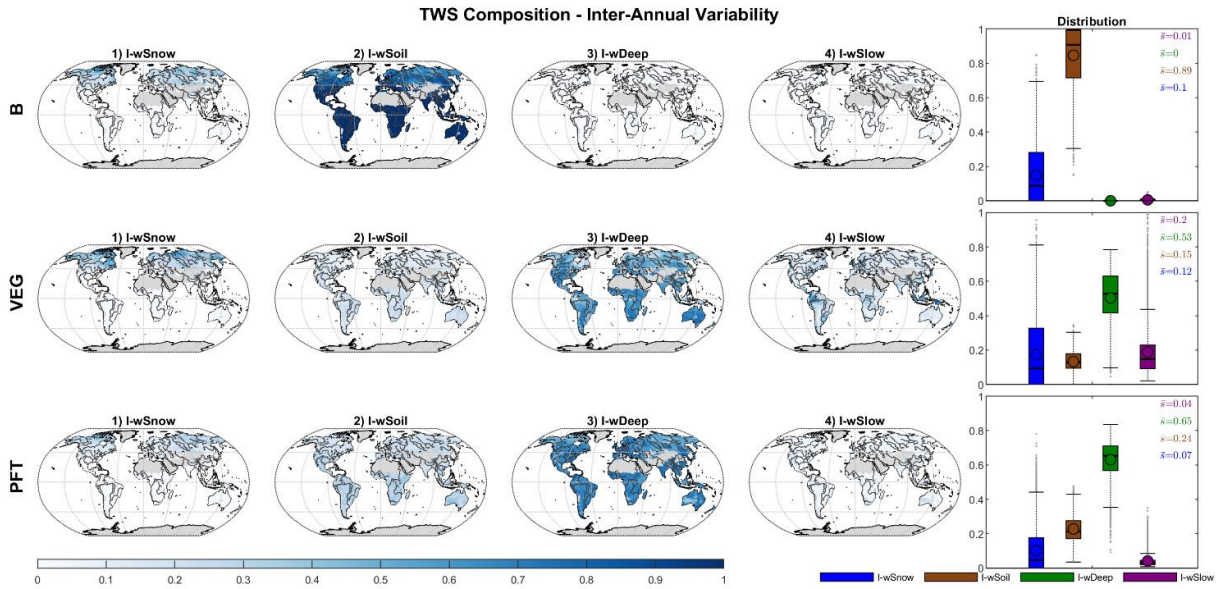

Figure B.24. Global distribution of the Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the inter-annual variability of total water storage, for B, VEG and PFT.

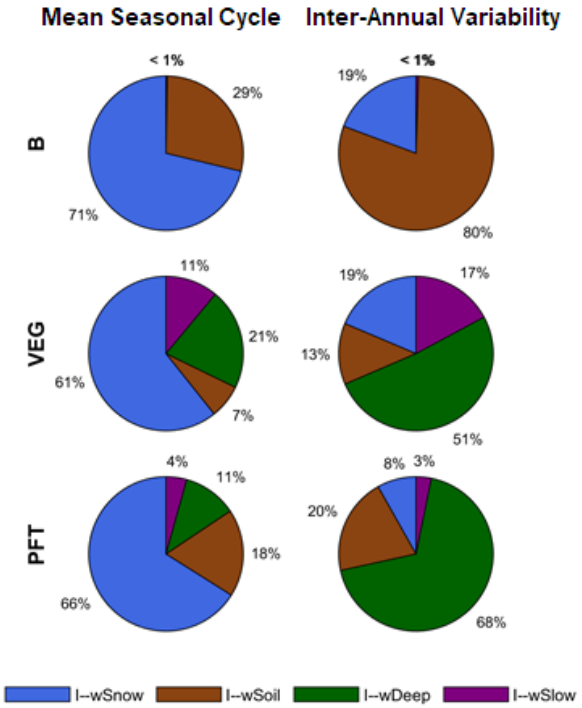

Figure B.25. Impact Index I for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the global average mean seasonal cycle and inter-annual variability of total water storage, for B, VEG and PFT.

B.10 Consistency Check of observational Data

In the following, we check for possible inconsistencies between the different observational data products. Similar to Rodell et al. 2015, we calculate the monthly water (im)balance, WB, from the observations for the period 01/2004-11/2010 (the time period in which none of the observation data has missing monthly values):

$$WB = P_{GPCP1DD} - ET_{FLUXCOM} - Q_{GRUN} - dS_{GRACE} \quad (B.1)$$

with ideally WB = 0.

Fig. B.26 shows the average monthly water imbalance scaled by each grid's average monthly precipitation $P_{GPCP1DD}$. While regionally large differences exist, the global mean and median are around 0. The global mean value of - 0.05 corresponds to a water balance residual of ~ 5% of precipitation - which is similar to the global residual of 4.3 % of precipitation reported in Rodell et al. 2014. Also temporally, the global average (Fig. B.27) varies around 0, suggesting no major systematic inconsistency at the global scale, yet with a small imbalance with a tendency to negative values. This suggests that more water leaves the system than comes in when looking at the observational data. In comparison, there is obviously no imbalance for the simulations from B and VEG as they close the water balance by definition of the model - which represents the major advantage of using models instead of observational based data from different sources.

We also calculated each variable in Eq. B.1 by solving the water balance with the other observed components and we compared the resulting water-balance-derived variable with the actual observed one. Differences between both indicate inconsistencies between a particular observed variable and the remaining observational variables. For ET, Q and TWS, we additionally plot the modelled fluxes and storage changes from B and VEG to evaluate the effect of observational inconsistencies on model simulations (Fig. B.27). The modelled fluxes are smoother and closer to the observations than the same estimate of the variable from the water balance. Therefore, we find that the model allows to potentially bridge the inconsistencies between the different data products. However, for dS, B and VEG show a time shift compared to the observed storage change, that is not reflected in dS calculated from P, ET and Q observations. Accordingly, this underlines that the phase lag between observed and modelled TWS variations is not caused by data inconsistencies, but rather related to the potential deficiencies in the model structure, as already discussed in the main text of the manuscript.

Fig. B.28 compares the residuals of the simulated and observed ET, Q and dS (mod-obs), and the ones of the water-balance derived and the observed variables (WB-obs). Large residuals of WB-obs point again to data inconsistencies among the observed variables. When the residuals WB-obs and mod-obs in a region agree, it implies that the multi-criteria calibration approach prevents overfitting of the model(s) to an observed variable that is inconsistent with the remaining observed variables. Therefore, the model performance in these regions might be relatively poor in view of the inconsistent data streams, which is in fact a desirable behavior in the model calibration (e.g., ET in the Semi-arid region and dS in Temperate and Humid region).

When the residuals of mod-obs are considerably smaller than WB-obs, the model fits an observed variable well although it is inconsistent with the remaining observed variables (e.g., Q and dS in the Semi-arid region). Further, when the residuals of mod-obs are large but WB-obs does not indicate data inconsistencies, it points to issues related to model structure and parameter identifiability (e.g., Q in the Cold region, where the model(s) lacks the representation of permafrost, freeze/thaw dynamics and ice jam in rivers).

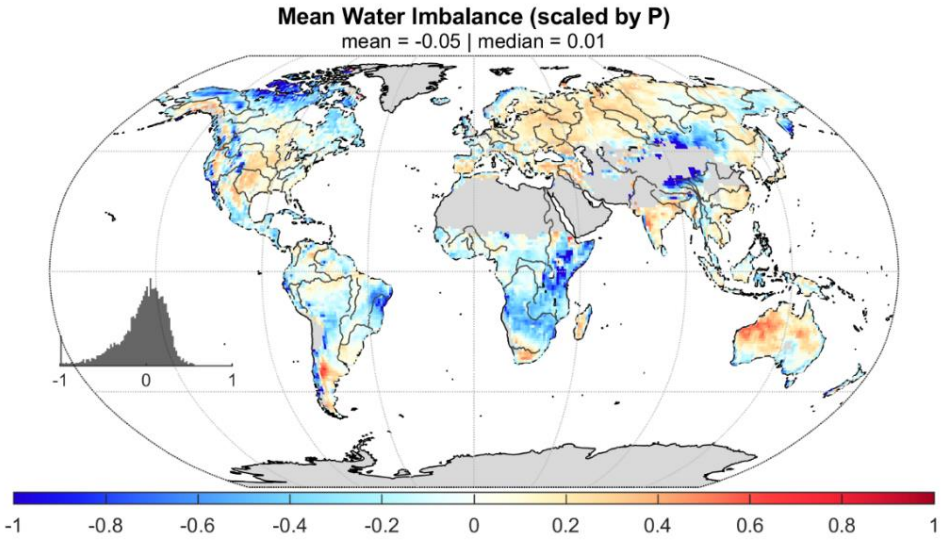

Figure B.26. Mean water imbalance scaled by mean precipitation.

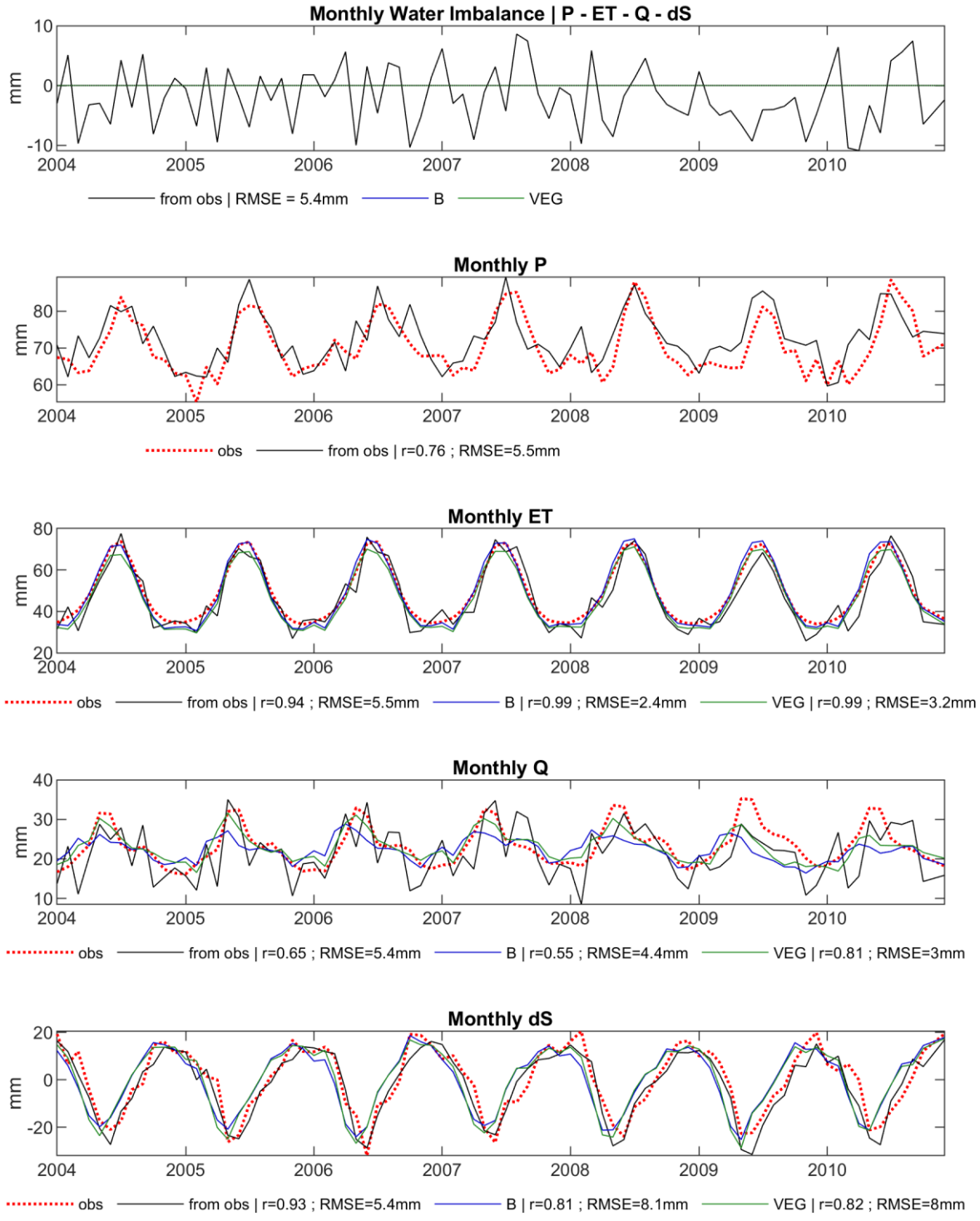

Figure B.27. Global average time series of the water imbalance calculated from the observations (top row), and of water balance variables calculated from the other observations by resolving the water balance equations (from obs) vs the observed variable (obs) vs the simulated variable of the B and VEG simulations.

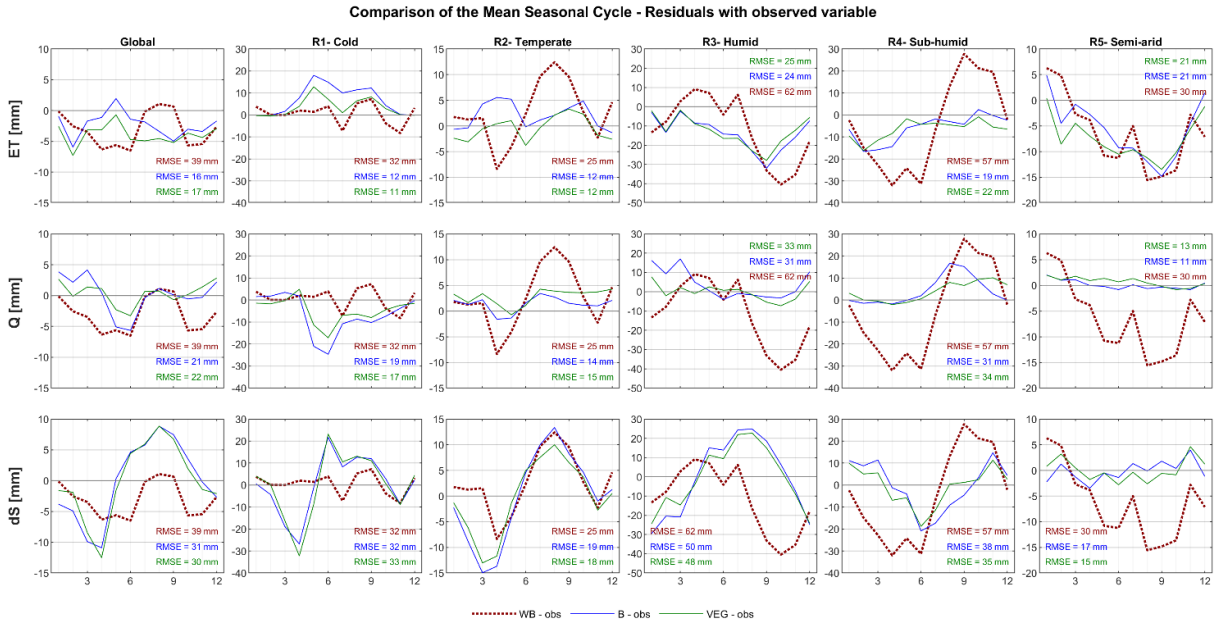

Figure B.28. Global and regional mean seasonal cycles of the difference between the simulations of ET, Q and dS with B and VEG and the respective observations (B-obs, VEG-obs), as well as difference between observed variables and the same variables calculated via the water balance from the other observations (WB-obs).

B.11 Analysis for Koeppen-Geiger Zones

Additionally, to the hydroclimatic cluster analysis shown in the main text, we performed a regional analysis for Koeppen-Geiger climate zones. To do so, we aggregated Koeppen-Geiger subgroups considering the main climate group and distinguishing between humid and semi-arid conditions. The resulting zones are shown in Fig. B.29. Fig. B.30 evaluates model performance for the Koeppen-Geiger regions and Fig. B.31 shows the composition of seasonal TWS variations therein. Note that most parts of the Polar and Boreal Koeppen-Geiger (KG) zone are included in the Cold region (R1) of the hydroclimatic cluster classification. We find that the regional averages are very similar for both classification schemes in terms of model performance and composition of seasonal TWS variations.

The Northern Hemisphere Temp and Boreal-sa KG zones are both included in the Temperate hydroclimatic region (R2). Temp KG and the Temperate region (R2) agree well regarding model performance and seasonal cycles, although we see a slightly better performance for the Temp KG regarding TWS and Q. In the Boreal-sa KG, B and VEG do not reproduce the spring peak of Q and precede the observed TWS significantly, decreasing model performance slightly when combining the Temp and Boreal-sa KG zones in one hydroclimatic region. Therefore, it would make sense to further split up the Temperate hydroclimatic cluster region. However, Boreal-sa KG spans Northern China, where poorer model performance is also evident from the performance maps in Fig. 4 of the main manuscript.

However, as mentioned in the main text, the advantage of the hydroclimatic cluster regionalization becomes obvious when interpreting results of the Arid and Temp-sa KG zones. This is because these climate zones are distributed across the Southern and Northern Hemisphere, causing 2 peaks in the regional seasonal cycles for TWS, ET and Q, due to opposing seasonal dynamics. The Arid KG zone includes the Semi-arid cluster regions (R5) in the Southern Hemisphere, as well as parts of the Temperate region (R2) (mainly in North America). The Temp-sa KG zone covers a rather small fraction of the study area, that is spread over the Temperate region (R2) in the Northern Hemisphere and the Semi-arid region (R5) of the Southern Hemisphere.

The effect of opposing seasonal cycles also exists in the Tropic KG zone, although less pronounced due to the proximity to the equator where the climate is more homogeneous and seasonality is low. The Tropic KG corresponds to the Humid cluster region (R3) on the Southern Hemisphere, and parts of the Sub-humid region (R4) on the Northern Hemisphere. Compared to the hydroclimatic cluster regions, the Tropic KG has less seasonal variation (a smaller amplitude) of TWS, ET and Q, due to its larger area North and South

of the equator. Both, B and VEG underestimate the ongoing depletion of TWS from September to December in Tropic KG, which is likely related to the opposing seasonal cycles of TWS in the Humid (R3) and the Sub-humid (R4) cluster regions. In the Tropic KG, Q peaks in March (as in Humid (R3)) and has a second, smaller peak in September (when Q peaks in the Sub-humid region (R4)). However, model performance is very similar for Tropic KG and the Humid and Sub-humid cluster regions.

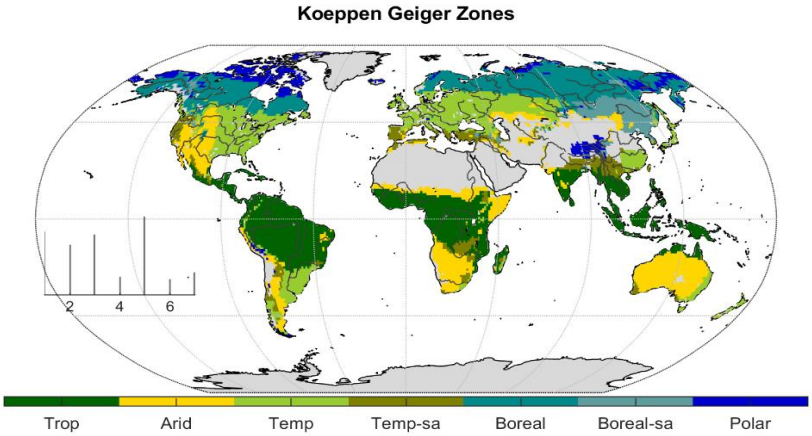

Figure B.29. Regions based on Koeppen-Geiger climate zones (Trop = Af, Am, As, Aw; Arid = BSh, BSk, BWh, BWk; Temp = Cfa, Cfb, Cfc, Dfa, Dfb; Temp-sa = Csa, Csb, Csc, Cwa, Cwb, Cwc; Boreal = Dfc, Dfd; Boreal-sa = Dsa, Dsb, Dsc, Dwa, Dwb, Dwc, Dwd; Polar = EF, ET).

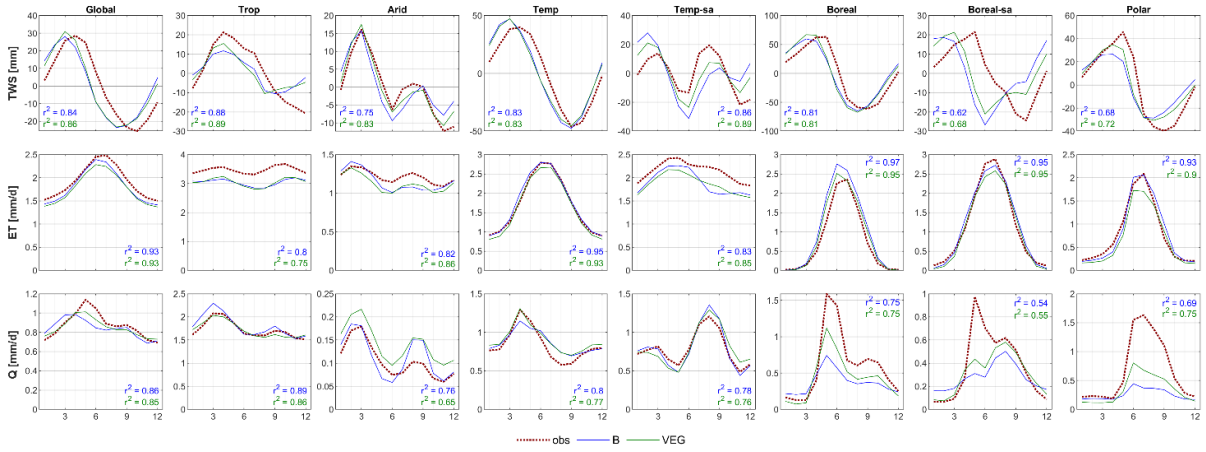

Figure B.30. Global and regional mean seasonal cycles of total water storage (TWS), evapotranspiration (ET) and runoff (Q) for the B and VEG experiments compared to the observational constraints by GRACE (TWS), FLUXCOM (ET) and GRUN (Q).

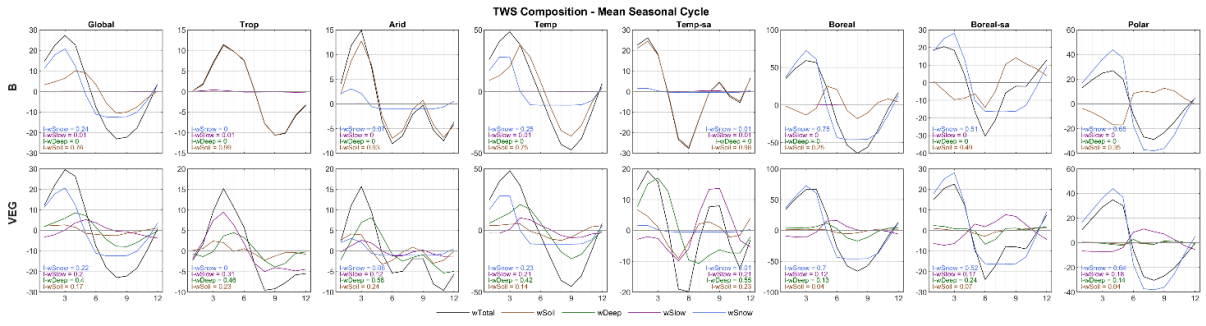

Figure B.31. Global and regional mean seasonal cycles of simulated total water storage and its components for B and VEG, including the regional Impact Index I for each storage.

Appendix C

Supplement of Chapter 4

Supplement material of:

Trautmann, T., Koirala, S., Güntner, A., Kim, H., Jung, M. (2022): Implications of river storage for integrating GRACE TWS observations into a global hydrological model, *submitted to Environmental Research Communications*.

C.1 Comparison of Calibration Runs in terms of calibrated Parameters and final Costs

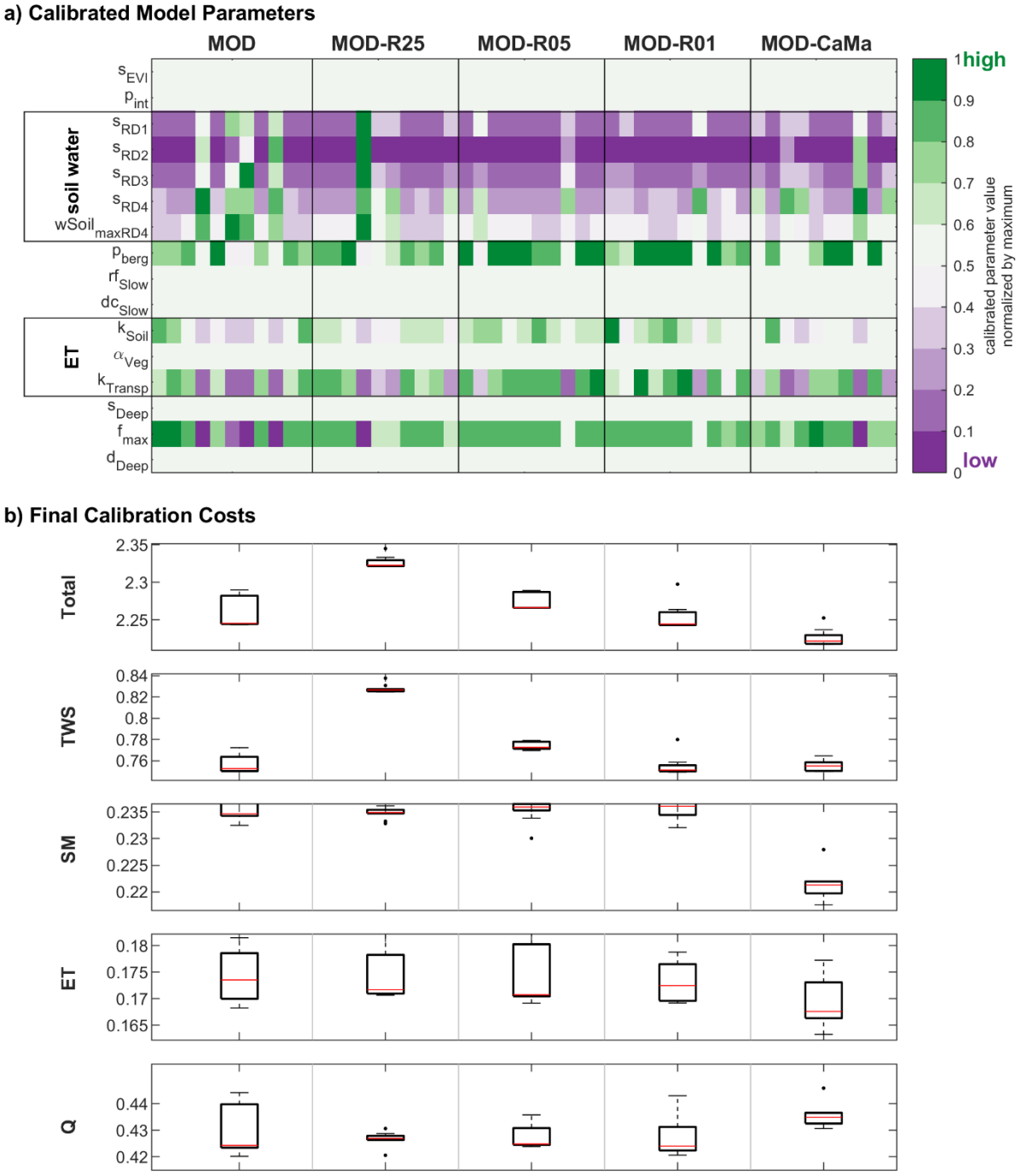

Figure C1. Calibrated parameters **(a)** and distribution of obtained total cost and cost components **(b)** for the 10 calibration runs of MOD and MOD-R▲. In **a)** calibrated parameter values are normalized by the maximum value of each parameter and those parameters with a low variance (<0.025) are set to 0.5 to emphasizes the 2 different parameter sets with either higher (green color) or lower (purple color) parameters that define maximum soil water storage and evapotranspiration (ET) processes. Along with the soil water and ET parameters, also the runoff-infiltration coefficient (p_{berg}) and the maximum flux rate to the deeper soil water storage (f_{max}) have either higher or lower values. Calibration runs with higher soil moisture and lower ET parameters have relatively higher costs than the other calibration runs of the respective experiment. In **b)**, the distribution of model costs of 10 calibration runs is shown for each experiment. Please note the different range of the y-axis for each cost component.

C.2 Influence of the Selection of calibration Grid Cells on Model Calibration

In the calibration approach presented in the main manuscript, we used the same grid cells for model calibration as in Trautmann et al. 2022, that are selected by a stratified random sampling among Koeppen-Geiger climate zones. This calibration subset is representative for the global study area, as both have the same distribution regarding all observational constraints, and also regarding the fraction of grid cells with large river water storage (Fig C.2). Therefore, calibration does not specifically focus on grid cells in which river storage is relevant. Instead, the same distribution as the global study area allows calibration of parameter values that are globally applicable and is thus preferable for a global study.

To assess the impact of the selection of calibration grid cells, we excluded grid cells with an average river storage below the median average of all grid cells (based on routed GRUN Q_R with an effective flow velocity of 0.5 ms^{-1}) from the stratified random sampling and select a new calibration subset *opti_relRiv*. By that, model calibration focuses on grid cells with relevant river water storage, i.e., we impose larger changes to the TWS constraint when removing river storage. This should lead to pronounced differences in calibrated parameter values among the experiments. The experiment calibrated against the original GRACE TWS adjusts its parameters at least partly to represent the river storage that is included in GRACE TWS and in *opti_relRiv*. Therefore, poorer model performance can be expected when these parameters are applied globally, where river storage is less relevant. On the contrary, the experiments calibrated against TWS without river storage do not need to implicitly account for river storage and thus their parameters should lead to good model performance globally.

Compared to the global study area and the calibration subset of the main manuscript, the new calibration subset *opti_relRiv* has significantly different distributions (Fig. C.2). Grid cells of *opti_relRiv* have higher ET and Q_R, and the original GRACE TWS has larger anomalies, i.e., more variations, than globally.

These differences lead to partly different calibrated parameter values for *opti_relRiv* calibrations than obtained by the calibration runs presented in the main manuscript. In detail, vegetation fraction and soil water storage are larger, counteracted by smaller transpiration parameters, which approximately keeps simulated ET comparable. Besides, less water infiltrates into soil and the delayed water storage depletes faster, which increases (fast) runoff. Additionally, the deep water pool that interacts with soil water and slow storage doesn't play a role.

These differences are due to the different hydrological characteristics of *opti_relRiv* and the previous calibration subset, not only in terms of TWS, but also regarding ET and Q_R.

Unlike in the main manuscript we do see some tendencies when contrasting the *opti_relRiv* experiments that are calibrated against GRACE TWS (MOD-relRiv) and those calibrated against TWS without river storage (MOD-R01-relRiv, MOD-R05-relRiv, MOD-R25-relRiv).

However, differences in parameter values among calibration experiments for *opti_relRiv* are by far smaller than the differences between calibrations done for different calibration subsets. This underlines the rather low impact of river storage on model calibration and suggests that the impact from changes in the other constraints is more influential.

Similar as shown in the main text (section 3.1.2), resulting mean seasonal dynamics do not differ between different experiments calibrated for *opti_relRiv* (Fig. C.3). In contrast, differences between *opti_relRiv* and the MOD_{best} experiment from the main manuscript are evident for Q_R in various regions, but also for TWS in the Semi-arid region, indicating again a larger effect of differences in the ET and Q_R constraint than from river water storages.

When adding simulated river storage to modeled TWS for model validation against GRACE TWS, calibrations of *opti_relRiv* outperform MOD_{best} on global average and in most regions, except the Semi-arid region. MOD-R01-relRiv and MOD-R05-relRiv achieve nearly perfect agreement of seasonal TWS variations in the Humid and Sub-humid region. Locally, differences in MEF are significant, showing improved performance in Northern low and wetland, tropics and some semi-arid regions, whereas MEF of other grid cells is worse than with MOD_{best}.

Altogether, calibrating for a spatial subset of grid cells where river water storage is relevant shows qualitatively the same pattern regarding the effect on model calibration (just a little systematic impact) and model validation (regional improvement) as with the previous calibration subset. Instead, we found that the choice of calibration grid cells has a larger impact than removing river water storage from GRACE TWS, due to changes in other constraints.

This underlines and supports the findings presented in the main text.

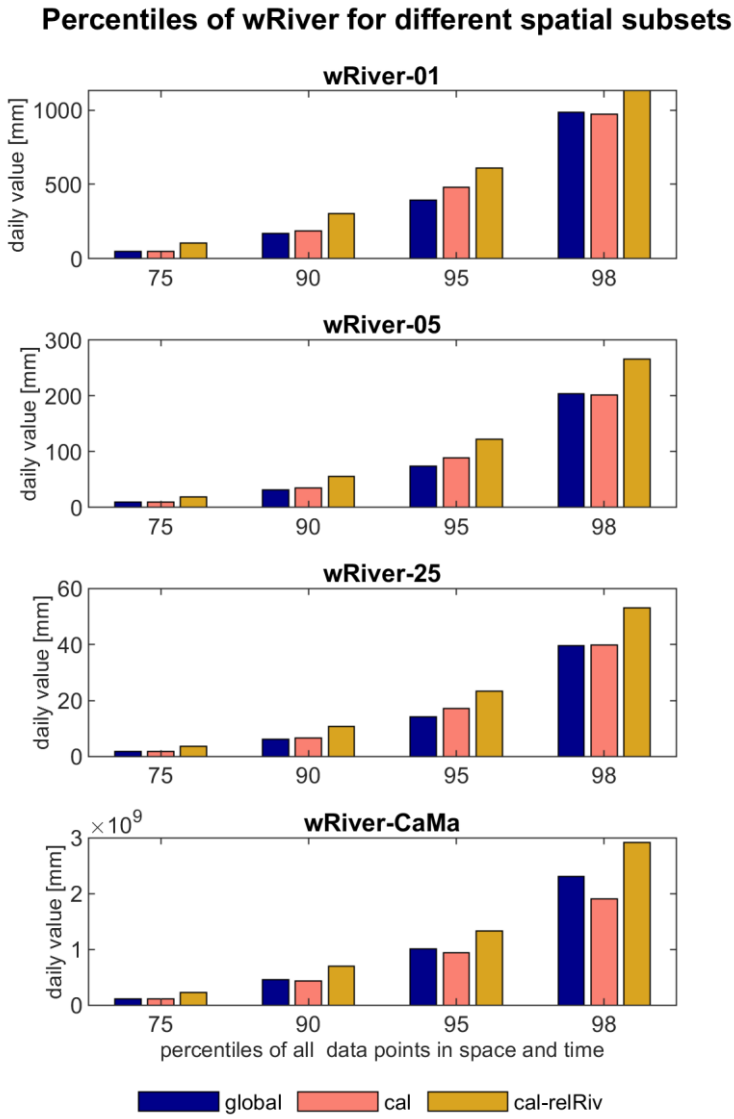

Figure C.2. Distributions of river storage based on CaMa-Flood and TRIPy with different effective flow velocities for different spatial subsets. 75th, 90th, 95th, and 98th percentiles of river storage of the global study area (global), the calibration subset of the main study (cal), and a calibration subset with relevant river storage (cal-relRiv) for different flow velocities (0.1; 0.5; and 2.5 ms^{-1}).

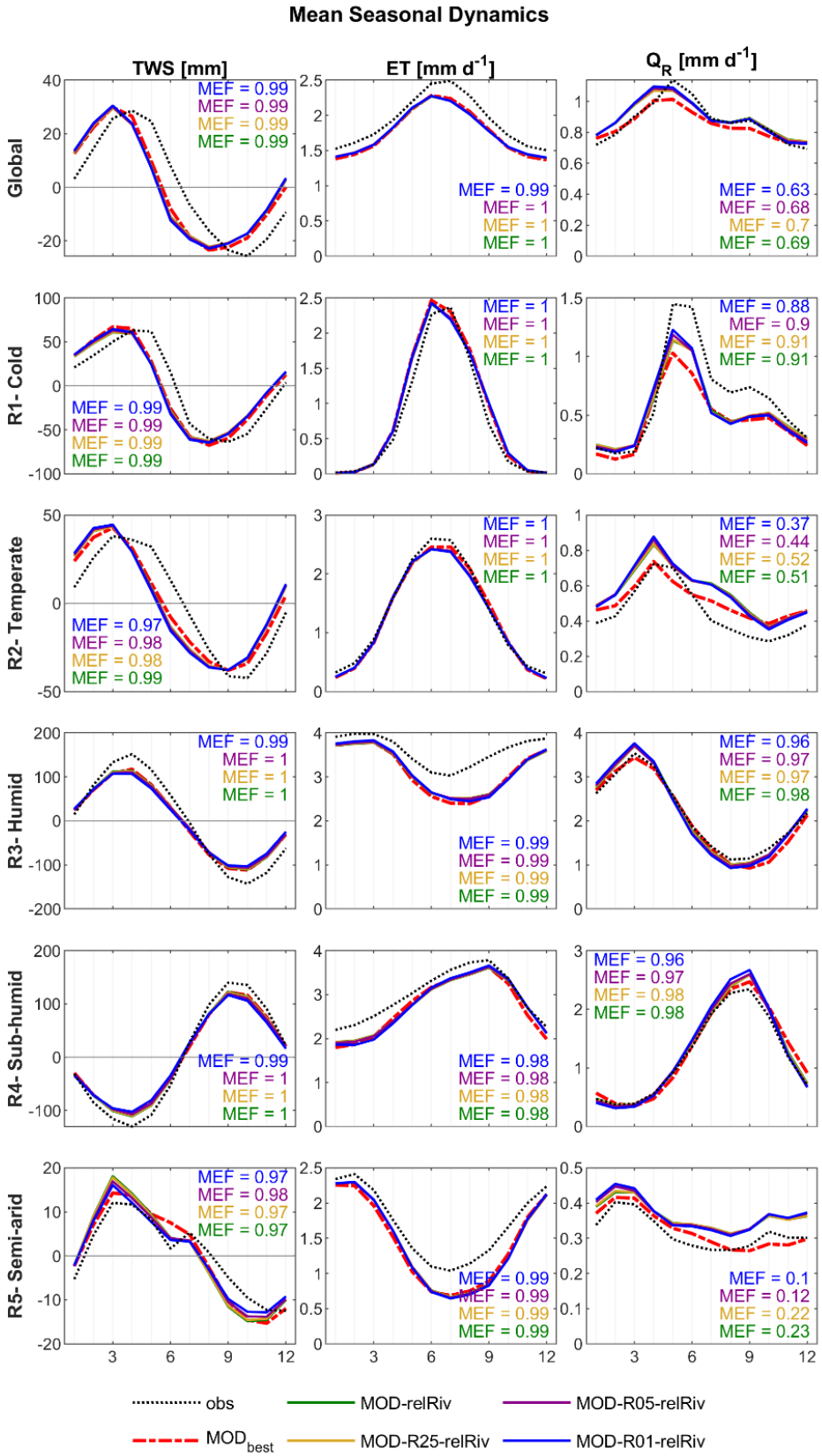

Figure C.3. Mean seasonal dynamics of simulated and observational TWS, ET, and Q_R summarized globally and for different regions. Solid lines show simulations by different experiments calibrated for the *opti_relRiv* subset. Red dashed lines denote MOD_{best} that was calibrated for the calibration subset of the main study (and against the original GRACE TWS). Black dotted lines denote observational GRACE TWS, FLUXCOM ET, and GRUN Q_R, respectively. Listed MEF compares the respective calibrated experiment against MOD_{best}.

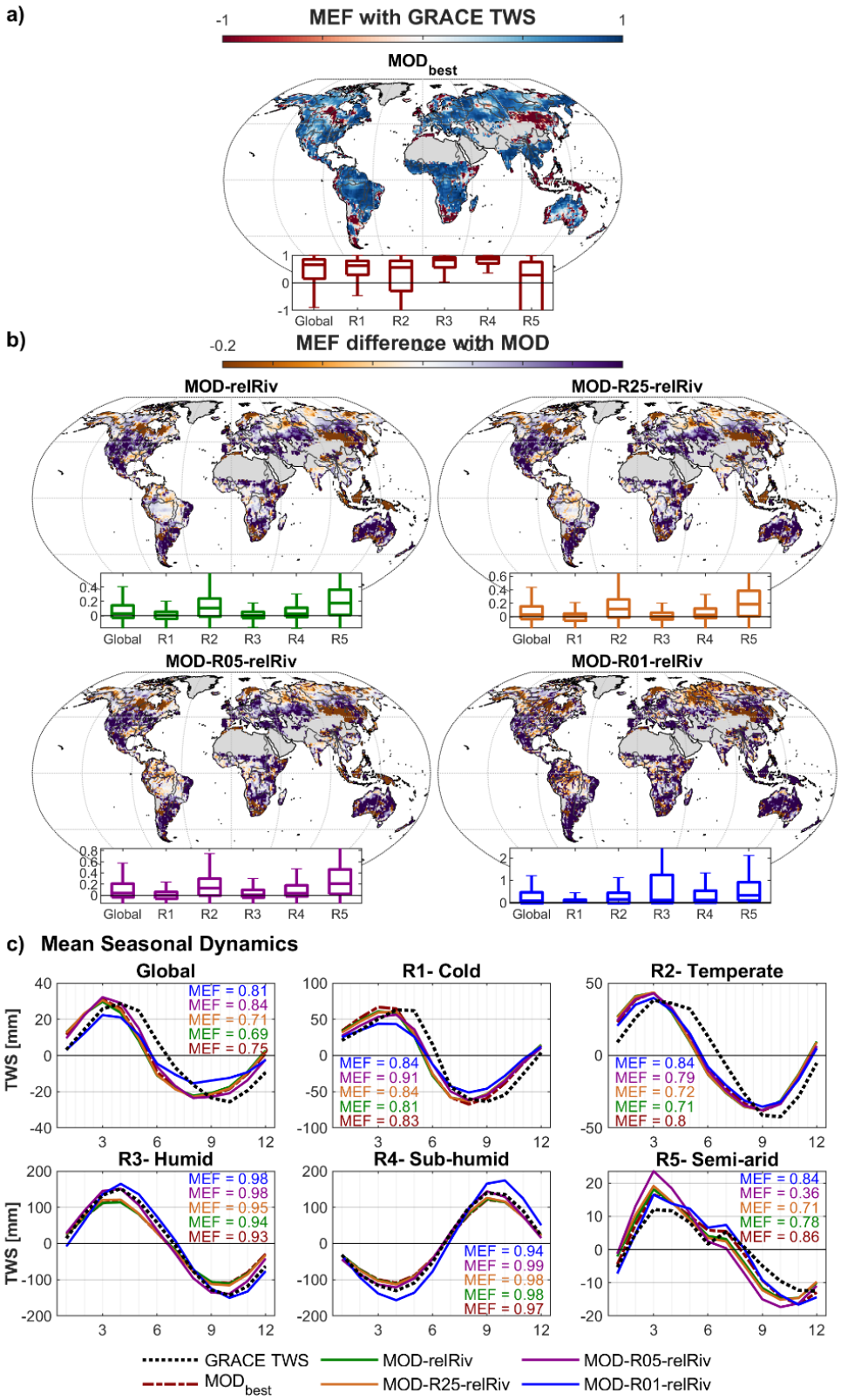

Figure C.4. Comparison of the original GRACE TWS, TWS simulated by MOD_{best} (from the main manuscript), and TWS (MOD-relRiv) resp. TWS+wRiver (MOD-R▲-relRiv) from experiments calibrated for the *opti_relRiv* subset. **a)** grid-wise Nash-Sutcliff efficiency (MEF) between GRACE TWS and MOD_{best} with distribution of grid-wise MEF per regions; **b)** grid-wise differences in MEF between MOD_{best} and MOD-relRiv resp. MOD-R▲-relRiv with distribution of grid-wise MEF differences per regions; and **c)** mean seasonal dynamics averaged globally and for different regions.

C.3 Comparison with eartH2Observe Models

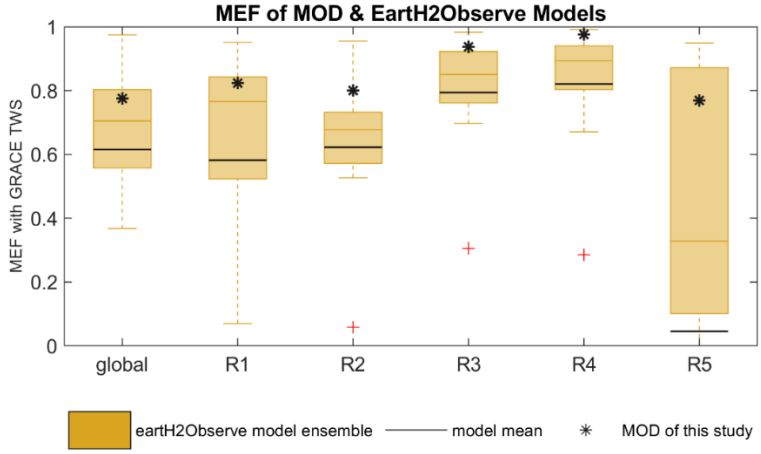

Figure C.5. Comparison of regional Nash-Sutcliff efficiency (MEF) obtained with the eartH2Observe model ensemble (box plots, Schellekens et al. 2017) resp. MOD_{best} of this study (asterisk) and GRACE TWS. Regions R1-R5 refer to hydro-climatic cluster regions shown in Fig. 4.2 of the main manuscript.

C.4 Evaluation of Discharge at GRDC Stations

Table C.1. GRDC stations used for evaluation of simulated discharge (Q_{Dis}) in Fig. C.6.

River	GRDC station name	Location (lat/lon in 1° grid)	Catchment size [km²]	Overlapping time period	Region
Lena	Stolb	72.5/126.5	2 460 000	03/2000-01/2003	R1-Cold
Yenisey	Igakra	67.5/86.5	2 440 000	03/2000-12/2011	R1-Cold
Ob	Salekhard	66.5/66.5	2 949 998	03/2000-12/2010	R1-Cold
Danube	Ceatal Izmail	45.5/28.5	807 000	03/2000-12/2010	R2-Temperate
Mississippi	Vicksburg, MS	32.5/-91.5	2 964 255	03/2000-12/2014	R2-Temperate
Amazonas	Obidos-Linigrafo	-2.5/-55.5	4 680 000	03/2000-01/2008	R3-Humid
Congo	Kinshasa	-4.5/15.5	3 475 000	03/2000-12/2010	R3-Humid
Zambezi	Katima Mulilo	-17.5/24.5	334 000	03/2000-12/2014	R5-Semi-arid

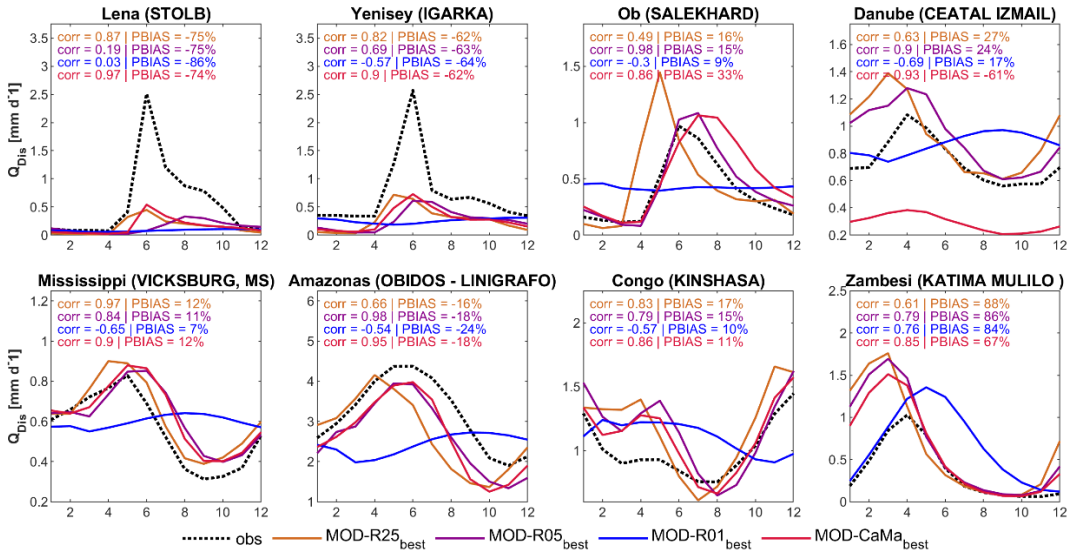

Figure C.6. Mean seasonal observed and simulated discharge Q_{Dis} from MOD-R01, MOD-R05, MOD-R25 and MOD-CaMa at 8 GRDC stations (see Table C.1 for station specifics).

Declaration of Authorship

I hereby declare that I prepared this dissertation entitled ‘Understanding global Water Storage Variations using Model-Data Integration’ on my own without any illegal assistance and that I did not use any other sources and means than specified.

This dissertation has not been submitted to any other university for examination, neither in Germany nor in any other country.

Jena, November 2022 Tina Trautmann