INSTITUT FÜR GEODÄSIE UND GEOINFORMATION

A water storage reanalysis over the European continent: assimilation of GRACE data into a high-resolution hydrological model and validation

Dissertation

zur Erlangung des akademischen Grades

Doktorin der Ingenieurwissenschaften

(Dr.-Ing.)

der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Anne Springer

aus Siegburg

Bonn 2019

Referent: - Prof. Dr. Jürgen Kusche
Korreferent: - Dr. Laurent Longuevergne
Korreferent: - Prof. Dr. Stefan Kollet
Tag der mündlichen Prüfung: - 08.03.2019
Erscheinungsjahr: - 2019
URN: - urn:nbn:de:hbz:5n-53930

A water storage reanalysis over the European continent: assimilation of GRACE data into a high-resolution hydrological model and validation

Summary

Continental water storage and redistribution within the Earth's system are key variables of the terrestrial water cycle. Changes in water storage and fluxes may affect resources for drinking water and irrigation, lead to drought or flood conditions, or cause severe changes of ecosystems e.g., through salinification. Hydrological models, which map water storages and fluxes, are being continuously improved and deepen our understanding of geophysical processes related to the water cycle. However, models are built on a simplified representation of reality, which leads to limited predicting skills of the simulation results. Assimilating remotely sensed total water storage variability from the Gravity Recovery and Climate Experiment (GRACE) mission has become a valuable tool for reducing uncertainties of hydrological model simulations. Simultaneously, coarse GRACE observations are disaggregated spatially and temporally through data assimilation.

In this thesis, GRACE data are assimilated into the Community Land Model version 3.5 (CLM3.5) yielding a unique daily 12.5 km reanalysis of total water storage evolution over Europe (2003 to 2010). Independent observations are evaluated to identify model deficits and to validate the performance of data assimilation. For the first time, the effect of data assimilation on modeled total water storage is also shown on the level of GRACE K-band observations. Optimal strategies for assimilating GRACE data into a high-resolution hydrological model are investigated through synthetic experiments. These experiments address the choice of the assimilation algorithm, localization, inflation of the ensemble of model states, ensemble size, error model of the observations, and spatial resolution of the observation grid.

As the assimilation of GRACE data into CLM3.5 is realized within the Terrestrial Systems Modeling Platform (TerrSysMP), future assimilation experiments can be extended for the groundwater and atmosphere components included in TerrSysMP.

Eine Reanalyse des europäischen Wasserspeichers: Assimilierung von GRACE Daten in ein hochaufgelöstes hydrologisches Modell und Validierung

Zusammenfassung

Änderungen im kontinentalen Wasserspeicher und im Transport von Wasser durch das Erdsystem sind wichtige Einflussgrößen für die Verfügbarkeit von Frischwasserresourcen, die Entstehung von Dürren und Überschwemmungen, sowie für die Erhaltung von Ökosystemen, welche z.B. durch Versalzung gefährdet werden. Hydrologische Modelle, die die Speicherung und den Transport von Wassermassen abbilden, werden stetig verbessert und helfen unser Verständnis von hydrologischen Prozessen zu vertiefen. Allerdings ermöglichen hydrologische Modelle nur eine vereinfachte Abbildung der Realität, sodass die Aussagekraft der Simulationsergebnisse beschränkt ist. Die Assimilierung von Wasserspeicheränderungen, gemessen von den GRACE (Gravity Recovery and Climate Experiment) Satelliten, kann hydrologische Simulationen verbessern und erlaubt gleichzeitig eine räumliche und zeitliche Differenzierung der grobaufgelösten GRACE Beobachtungen.

In dieser Doktorarbeit werden GRACE Daten in das Land-Oberflächen-Modell CLM3.5 (Community Land Model Version 3.5) assimiliert, um eine neuartige Reanalyse täglicher Wasserspeicheränderungen (2003 bis 2010) für Europa mit 12.5 km Auflösung zu generieren. Durch unabhängige Beobachtungen werden Defizite des Modells identifiziert und das Ergebnis der Datenassimilierung beurteilt. Zum ersten Mal wird auch die Auswirkung der Assimilierung direkt auf Basis der GRACE K-Band Beobachtungen untersucht. Mit Hilfe synthetischer Experimente wird die beste Strategie zur Assimilierung von GRACE Daten in ein hochaufgelöstes hydrologisches Modell ermittelt. Dabei wird der Einfluss unterschiedlicher Assimilierungsstrategien untersucht, unter anderem die Wahl des Assimilierungsalgorithmus, die Lokalisierung des Einflussbereichs von Beobachtungen, die Erhöhung der Spannweite der Ensemblemitglieder des Modells, die Ensemblegröße, das Fehlermodell der Beobachtung und die räumliche Auflösung des Beobachtungsgitters.

Da die Assimilierung von GRACE in das CLM3.5 Modell unter Verwendung von TerrSysMP (Terrestrial Systems Modeling Platform) geschieht, können die Assimilierungsexperimente in Zukunft auf die zusätzliche Verwendung des in TerrSysMP enthaltenen Grundwasser- und des Atmosphärenmodells erweitert werden.

Acknowledgements

First of all, I would like to thank Jürgen Kusche for supervising my PhD thesis. He guided me through the whole process and gave me the possibility to work on several different aspects of physical and satellite geodesy and to gain insights also into neighboring disciplines. I always appreciated our talks and I have learned a lot from your explanations and from your always further reaching questions. For me, it was also essential that I could develop my own ideas and that I had the possibility to discuss them at several international conferences.

I would like to thank Annette Eicker for our productive collaborations and for sharing her knowledge with me. I always enjoyed a lot to solve problems jointly with you.

Furthermore, I would like to thank Jessica Keune for providing the initial set-up of CLM3.5 and for answering a number of hydrology-related questions with a lot of patience. I am also grateful to Wolfgang Kurtz who provided the interfaces routines between PDAF and CLM3.5 to me. I really enjoyed our talks and your interest in my work. My understanding of interrelations between geodesy and hydrology were also furthered by discussions with Bibi Naz, Stefan Kollet, Laurent Longuevergne, Thomas Poméon, and Bernd Diekkrüger. Thank you for opening new and interesting perspectives to me. In particular, I would like to thank Laurent Longuevergne for helping me to gain experiences with a first very simple hydrological model. A meeting with Lars Nerger, the developer of the parallel data assimilation framework PDAF, was particular important for me to understand that the results from data assimilation are not always straightforward. This was very encouraging for the next steps of my work.

I acknowledge a DAAD grant for spending three months in Rennes (France) with Laurent Longuevergne. This was a very fruitful time for me and I really enjoyed the working environment. I would also like to acknowledge funding by the Centre for High-Performance Scientific Computing in Terrestrial Systems (HPSC-TerrSys) and funding by the COAST (Studying changes of sea leveland water storage for coastal regions in West-Africa using satellite and terrestrial datasets) project, supported by the Deutsche Forschungsgemeinschaft under Grant No. KU1207/20-1.

Proof reading made this thesis a great deal better and, therefore, I would like to thank my colleagues Bernd Uebbing, Christina Lück, Alexander Harker, Kerstin Schulze, Joël Köhler, Helena Gerdener, Kristin Vielberg, Sophie Stolzenberger, and Anno Löcher for reading parts of my thesis and for making valuable suggestions. Many thanks go to Katrin Almon who spend a lot of time reading the whole thesis. I am really grateful for this. Whenever my computer or our compute cluster reacted not as expected I got help from Judith Schall, Roelof Rietbroek, Jan Martin Brockmann, Christian Neyer, Michael Plümer, and Bernd Uebbing. Thank you for spending time with these issues.

Finally, I would like to thank all my past and present colleagues of APMG and TG for a fruitful and friendly working environment, which makes it easy to work efficiently. I am particularly grateful for my wonderful office mates Judith and Helena. I really enjoyed our talks.

A completely different contribution to this work was made by my parents, Carola and Bernhard Springer, through baby-sitting, cooking, and baking. Your support was indispensable for finishing this thesis. Thank you for your efforts! Andreas Ernst provided me with continuous support, interest in my work, and trust in my abilities during the process that lead to this thesis. I thank you with all my heart. Finally, my dear son Fabian (unwittingly) helped me to finish this thesis by distracting me from work (but not too much) and by opening to me a completely different view on the world.

Contents

8 Data Assimilation Experiments 83

Chapter 1

Introduction

1.1 Motivation

Terrestrial water storage is a key variable of the global water cycle. The total water storage (TWS) on the continents includes all water components on and underneath the Earth's surface, i.e. groundwater, soil moisture, surface waters (wetlands, rivers, lakes), snow water, and canopy water. Changes in TWS directly affect freshwater availability (Döll et al., 2016; Solander et al., 2017; Rodell et al., 2018), and can be related to natural disasters such as droughts and floods (Leblanc et al., 2009; Chew and Small, 2014; Sun et al., 2017). In addition, TWS changes are an important indicator for climate change (Green et al., 2011; Teutschbein et al., 2011; Yang et al., 2015; Kusche et al., 2016). Given the strong interconnection between human life and the terrestrial water cycle, millions of people are affected by natural disasters that can be expressed through TWS change. Well-documented examples are the California drought (Nelson and Burchfield, 2017), and the increasing intensity and duration of monsoon floods in South Eastern Asia (Dewan, 2015).

Changes in TWS and its components are associated with changes in hydrological fluxes, such as infiltration rates, runoff, groundwater recharge, and evapotranspiration. Indirectly, changes in TWS components also affect the atmospheric part of the water cycle and the Earth's energy cycle. Soil moisture feedbacks on the atmosphere change variables such as air temperature and wind speed. Snow cover affects surface albedo and, thereby, induces changes in atmospheric circulation. Due to its memory effect, groundwater has a huge impact on long-term climate variability. Groundwater recharges soil moisture and, thus influences near-surface processes such as land-use change. Clearly, TWS is linked to a number of variables that were defined as Essential Climate Variables (ECV) by the Global Observing System for Climate (GCOS) under the auspices of United Nations organizations (Bojinski et al., 2014).

Accurate knowledge of TWS and TWS changes is indispensable for sustainable land and water management. Projections of TWS evolution are a basis for developing strategies that guarantee food and water supply, protect human health, preserve ecosystems, control energy generation and prevent migration streams. Finally, monitoring TWS directly targets the United Nation's sustainable development goal 6, i.e. "ensure availability and sustainable management of water and sanitation for all" [1].

Despite the importance of TWS, its monitoring is challenging. Hydrological models, which map the terrestrial water cycle, are only a simplified representation of reality and suffer from

---

[1] `https://unstats.un.org/sdgs/report/2018/goal-06/`

the limited quality of input data. In-situ observations of TWS are sparse and represent pointwise measurements. TWS observations from remote sensing are available from the Gravity Recovery and Climate Experiment (GRACE) mission, but have a limited spatial and temporal resolution and, thus, do not resolve fast hydrological processes at small spatial scales.

Based on these constraints, the motivation of this thesis is to enhance our knowledge about the evolution of TWS and the understanding of hydrological processes, combining both observations and models.

1.2 Terrestrial Water Cycle and Space Gravimetry

Our knowledge of the terrestrial water cycle is based on two complementary elements: (i) observations from in-situ stations or remote sensing and (ii) physical laws represented by numerical models.

Hydrological models map the individual components of TWS and simulate water and energy fluxes. Conceptual hydrological models are based on empirical equations and parameters, and were developed for water resources management, as such, they often also model human water use. Hydrological land-surface models aim to represent physical processes of the real world by using mathematical equations. Hydrological models are run at different spatial scales (grid cells of a few meters to a hundred kilometers) and different temporal scales (hourly to monthly time steps). Uncertainties exist due to insufficient realism of model equations and structure, imperfect model parameters, and imperfect forcing and surface data sets. These uncertainties lead to deficiencies in representing trends and temporal variability of TWS. As a consequence, simulations from individual hydrological models differ largely (Schewe et al., 2013; Scanlon et al., 2018).

The GRACE mission (2002 – 2017) allowed the monitoring of large scale changes in the integral sum of TWS (hydrosphere, biosphere, atmosphere, and oceans) from space. GRACE was a geodetic mission that observed the Earth's time-variable gravity field by measuring changes in the distance between two satellites. These changes can be converted to TWS variability (Wahr et al., 1998). The temporal resolution of GRACE solutions is typically one month with a spatial resolution of a few hundred kilometers. Depending on the analysis strategy, daily GRACE solutions can be provided, but with a loss in spatial resolution. Decomposition of GRACE observations into individual components of TWS and to smaller spatial and temporal scales requires complementary information from models, e.g. via statistical decomposition techniques (Rietbroek, 2014; Forootan et al., 2014).

The joint evaluation of hydrological models and GRACE-derived TWS variability has been beneficial for our knowledge about the terrestrial water cycle. Comparing modeled and observed TWS variability provides information on model deficiencies (Niu and Yang, 2006; Alkama et al., 2010; Zhang et al., 2017). Calibrating hydrological model parameters against GRACE observations (usually model parameters have been calibrated against discharge observation) helps to improve the representation of the terrestrial water cycle in the models (Werth et al., 2009; Chen et al., 2017).

Merging hydrological models and GRACE observations via data assimilation, i.e. modifying the model state towards the observations in an optimal sense via a dynamical model, implies (i) disaggregation of GRACE observations vertically, horizontally, and temporally, and (ii) a more realistic representation of TWS compartments in the model with the potential to improve also modeled water fluxes (Zaitchik et al., 2008; Girotto et al., 2016; Schumacher et al., 2016). Therefore, GRACE-assimilating hydrological models provide unique information on the distribution and redistribution of water.

1.3 Scientific Context and Objectives of the Thesis

Assimilating GRACE data into hydrological models presents several challenges. The temporal and spatial (horizontal and vertical) resolution mismatch between simulated model states and observed TWS variability requires the use of sophisticated strategies to connect them (Girotto et al., 2016). Furthermore, data assimilation requires information on uncertainties of model and observations. The model error is difficult to quantify as it depends on uncertainties in model structure, atmospheric forcings, and soil data sets. GRACE-derived TWS maps are contaminated with correlated noise, which requires careful post-processing of the GRACE solutions. Currently, no standard way exists for the assimilation of GRACE data into hydrological models.

So far, only few groups have assimilated GRACE data into a hydrological model. Typically, the applied hydrological models have a spatial resolution between 0.5° ($\sim$50 km) to 1° ($\sim$100 km) and run at daily time steps, whereas the resolution of climate data might be restricted (e.g., number of cloud-free days per month, precipitation averaging). At our institute, GRACE data was assimilated into the conceptual WaterGAP Global Hydrology Model (WGHM) at 0.5° resolution (Schumacher, 2016). One exception for assimilation experiments at higher spatial resolution is the assimilation of GRACE data into a lumped rainfall-runoff model set up for the Rhine catchment at 1 km resolution, which however uses atmospheric forcing data resolved at 0.25° to 0.5° (Tangdamrongsub et al., 2015). Another example is the catchment-based land surface model running at a 1/8° degree grid with 15 minute time steps and daily atmospheric forcing data (Kumar et al., 2016).

This thesis aims at assimilating GRACE-derived TWS variability into the Community Land Model version 3.5 (CLM3.5) at 12.5 km spatial resolution over the entire of Europe using 3-hourly atmospheric forcing data. CLM3.5 is a physics-based land-surface model, which has a more complex structure than conceptual models like WGHM, as several hydrological, biogeophysical, and biogeochemical processes are represented and water is stored in ten soil layers and up to five snow layers. Physical relationships between model variables cause particular challenges regarding a physically consistent update of the model states during assimilation.

CLM3.5 is part of the Terrestrial Systems Modeling Platform (TerrSysMP), which also includes a groundwater component and an atmospheric model (Gasper et al., 2014; Shrestha et al., 2014). Including the assimilation of GRACE data into CLM3.5 using TerrSysMP will allow for extending the experiments to the groundwater component, or even to simulations of the whole terrestrial water cycle, i.e. simulations with a fully coupled model that includes atmospheric, land-surface, and groundwater components.

Previous studies investigated different strategies of assimilating GRACE data by varying governing parameters such as, the assimilation algorithm (Khaki et al., 2017a), correlated versus white observation noise (Schumacher et al., 2016), and the observation grid (Khaki et al., 2017b). This thesis goes further and provides a systematic study of the most important assimilation parameters, which results in a ranking of their influences on the assimilation results.

This thesis addresses two main issues:

1. What is the optimal way of assimilating GRACE data into a high-resolution (12.5 km) continental scale land-surface model in terms of an improved description of water storage variability?

2. What is the impact of the assimilation of GRACE data on the performance of CLM3.5 in terms of the realism of simulated water storage compartments and water fluxes?

For answering the first question, a detailed review on the findings from previous GRACE assimilation studies is performed and, subsequently, a number of synthetic experiments are carried out. The second question is answered by validating water storage compartments and water fluxes of CLM3.5 against independent observation-based data sets, both before and after data assimilation.

Finally, this thesis provides an unprecedented reanalysis of TWS and its components at 12.5 km resolution and (sub-) daily time steps over Europe for the time span 2003 to 2010.

1.4 Organization of the Thesis

The thesis starts with an introduction into the characteristics of the GRACE mission (Chapter 2). The GRACE measurement principle is explained and different concepts of gravity field processing are introduced. The chapter concludes with an overview of recent GRACE solutions. Chapter 3 discusses studies related to the validation of hydrological models using GRACE observations and aims at a complete survey of studies that assimilate GRACE data into hydrological models. In Chapter 4, the structure and setup of CLM3.5 are explained and different global models are introduced, followed by an overview of the observation-based data sets that are used for model validation.

In Chapter 5 the processing of GRACE data, as performed in this thesis, is described in two parts: First, the computation of gridded TWS anomalies from GRACE spherical harmonics and corresponding error information is explained. Second, the processing of GRACE level 1B data for hydrological model validation is outlined. Chapter 6 gives an overview on individual data assimilation algorithms and tuning approaches. The strategy for assimilating GRACE data into CLM3.5, which was realized within this thesis, is explained in Chapter 7.

The results from data assimilation experiments are provided in Chapter 8. First, synthetic experiments addressing the influence of different assimilation strategies are evaluated. Then, modeled trends and phase shifts with respect to GRACE-derived TWSA are investigated before and after data assimilation. Finally, the real-case scenario, an 8-year (2003 to 2010) assimilation run, is validated against independent observations. Additionally, the representation of extreme events is assessed for the assimilated model. In Chapter 9 hydrological signals are investigated for their skill in explaining level 1B GRACE K-band observations. First, global hydrological models are assessed and then the CLM3.5 over Europe is evaluated with respect to the impact of data assimilation. The thesis closes with a summary, followed by final conclusions and an outlook on future possible extensions of this work (Chapter 10).

Chapter 2

Gravity Recovery and Climate Experiment GRACE

2.1 GRACE Satellite Mission

The Gravity Recovery and Climate Experiment (GRACE) mission measured spatial and temporal variations of the Earth's gravity field from March 2002 to June 2017. Variations in gravity are due to mass changes in hydrosphere, atmosphere, biosphere, oceans, and due to mass variations inside of the Earth. Monthly high-resolution gravity field solutions contribute to better knowledge about climate-relevant parameters and processes:

- Groundwater depletion and droughts threaten sustainable water supply and food production in many regions of the world, like e.g. California, Southern Europe, the Middle East, Africa, India, parts of China, and Australia (Mekonnen and Hoekstra, 2016). GRACE is recognized as a unique tool for measuring changes of **water resources on the continents** with global coverage. The satellite mission measured the total sum of groundwater, soil moisture, snow, canopy water, and surface waters. An outstanding example for the potential of the GRACE mission was the combination of GRACE and ancillary observations with hydrological modeling in order to quantify groundwater depletion in the Northwest India Aquifer and to relate it to human activities (Rodell et al., 2009; Long et al., 2016). Furthermore, the joint evaluation of GRACE observations and hydrological models enhanced the reliability of drought indicators and flood potential (Houborg et al., 2012; Reager et al., 2014). This highlights that both society and economy benefit from the measurements of the GRACE mission.
- The **atmospheric-terrestrial flux**, precipitation (P) minus evapotranspiration (E), is a key component for investigating the interaction between atmosphere and land surface, and for understanding human influences on climate change, e.g., via land use change. However, in particular evapotranspiration is still not well known from observations and often poorly represented by Numerical Weather Prediction (NWP) model simulations. By combining GRACE data with $P - E$ and river discharge R the closure of the water budget equation can be assessed (Lorenz et al., 2014; Springer et al., 2014, 2017; Lv et al., 2017). Furthermore, trends in P, E, and R indicate that fluxes move away from an equilibrium state. This might lead to an intensification of the water cycle. Accelerations in total water storage can be derived from GRACE observations and allow the validation of such trends in water fluxes (Eicker et al., 2016).

- **River discharge** is an important indicator for freshwater availability along rivers and of major importance for water management. As **runoff from the continents** occurs mainly at the river mouths, it acts as a driver for regional oceanic circulations (Carton, 1991), which are also of major importance for forcing climate models. Solving the water budget equation for R using GRACE data is one possibility for overcoming the problem of the decreasing number of gauging stations (Syed et al., 2009).
- Ocean currents transport heat and mass from the equator toward the poles, and, thus, regulate global and regional climatic conditions. GRACE emerged as valuable tool for quantifying changes in **strength and direction of ocean currents** and for validating ocean circulation models (Landerer et al., 2015).
- **Sea level rise** contaminates aquifers and agricultural soils and reshapes the world's shorelines, which causes each year the displacement of thousands of people living in coastal areas. Ocean mass changes derived from GRACE observations allow us to monitor global and regional mass induced sea level variations (Lombard et al., 2007; Johnson and Chambers, 2013). One-half to two-thirds of sea level rise is attributed to **mass loss from glacier and ice sheets** (Gardner et al., 2013). GRACE observes glacier and ice sheet melting and, thus, also contributes to the decomposition of the sea level budget (Rietbroek et al., 2016).

GRACE was a joint mission of National Aeronautics and Space Administration (NASA) and Deutsches Zentrum für Luft- und Raumfahrt (DLR). The University of Texas Center for Space Research (CSR) was responsible for the overall GRACE mission. The project management was carried out by the Jet Propulsion Laboratory (JPL) and the German mission contributions were supplied by GeoForschungsZentrum (GFZ) in Potsdam. CSR, JPL, and GFZ are part of the GRACE Science Data System (GSDS), which is responsible for system development, data processing, and data archival (Bettadpur, 2012b).

GRACE consisted of two identical satellites, chasing each other with a distance of about 220 km on the same near circular orbit (Figure 2.1). The 89° orbit inclination ensured near-global coverage (Tapley et al., 2004a). The initial height of 500 km decreased down to 330 km by mid-2017 due to atmospheric drag. The low orbit permitted detailed mapping of anomalies of the gravity field down to spatial scales of a few hundred kilometers. The orbital period of about 94 minutes ensured dense spatial coverage after 30 days of continuous observations.

GRACE realized the low-low Satellite-to-Satellite Tracking (SST) concept. In an idealized case (i.e. no atmospheric drag, no solar radiation pressure, etc.), the change in distance between two co-orbiting satellites is nearly proportional to the difference in the Earth's gravitational potential at the respective locations (Wolff, 1969). When the leading satellite approaches a positive gravity anomaly, it is attracted towards the anomaly and the distance to the trailing satellite increases. The distance decreases, when the trailing satellite is also attracted towards the gravity anomaly.

The key instrument of the GRACE satellites was a dual one-way K-band microwave ranging (KBR) system, which measured the inter-satellite range and its derivatives (Figure 2.2). The KBR system of each satellite was equipped by a horn antenna that transmitted carrier phase signals on two different frequencies. The phase measurements at each frequency are the basis for deriving inter-satellite biased ranges with an accuracy of 1 µm. Subsequently, biased ranges from both frequencies are combined for removing ionospheric effects. After compression, biased ranges, range-rates and range accelerations are available at 5 second sampling. Non-gravitational forces acting on the satellites, such as air drag and solar radiation pressure, are accounted for by measurements from accelerometers, which were located at the center of mass of each satellite. Besides that, each satellite was equipped with two simultaneously

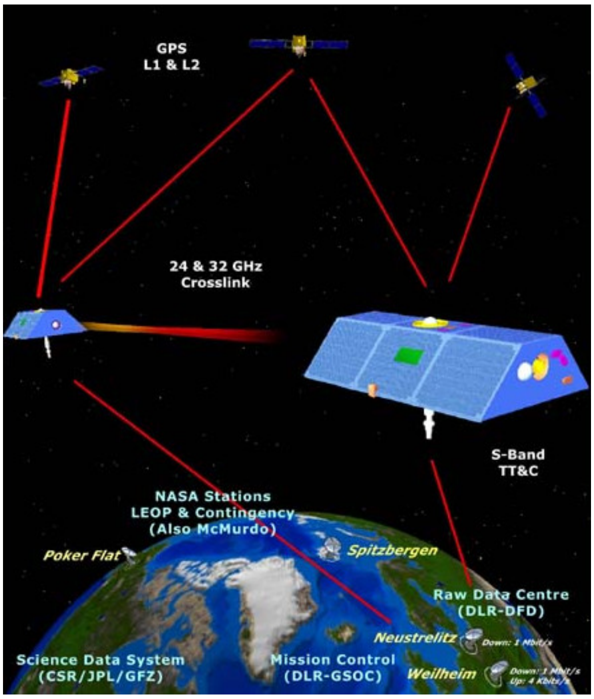


Figure 2.1: Measurement principle of the GRACE mission (Source: http://www2.csr.utexas.edu).

operating star cameras, which provided the internal orientation of the satellites with respect to their line of sight. Additionally, precise orbit determination and time tagging of all sensors was achieved with Global Positioning System (GPS) space receivers. In the end, the Earth's gravity field can be computed from measured ranges and from the satellite's accelerations, orientations, and positions (see Section 2.2).

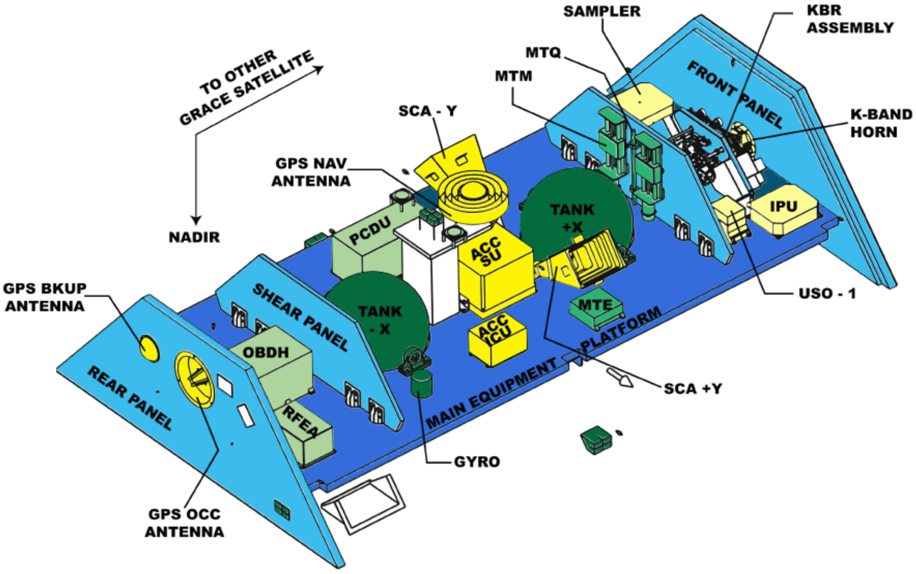


Figure 2.2: Design of a GRACE satellite (Source: http://op.gfz-potsdam.de).

Initially, the GRACE mission was designed for a 5 year life period (Tapley et al., 2004b, 2016). Since 2011, battery issues lead to missing solutions approximately every 6 months. Finally, GRACE was operative until June 2017 providing 15 years of continuous records. An overlap period with the GRACE Follow-On (GRACE-FO) mission, launched in May 2018, was not achieved. Strategies for closing the gap include the computation of gravity field solutions from kinematic orbits of the Swarm Earth Explorer satellites (Lück et al., 2018). GRACE-FO now carries on the observation strategy of the GRACE mission, at the same time testing a new laser ranging system, which is more precise than the K-band instrument (Sheard et al., 2012; Tapley et al., 2016). First GRACE-FO gravity solutions are expected to become available beginning of 2019.

2.2 GRACE Data Processing

Before the data collected on board of the GRACE satellites can be used for gravity field investigations, several correction and filtering steps are required (Bettadpur, 2012b). Depending on the processing state, the data products are divided into five different categories:

- Level 0: The level 0 data products are raw data, which were received by the GRACE Raw Data Center (RDC) at DLR in Neustrelitz (Germany) via telemetry at each pass of the satellites. The down-link from each satellite includes one file with science instrument data and one file with spacecraft housekeeping data.
- Level 1A: The binary encoded measurements from the level 0 data products are converted to engineering units and time tagged to Global Positioning System (GPS) time. Furthermore, quality control flags are added and the data are reformatted for further processing.
- Level 1B: Level 0 and level 1A data products are transformed into a uniform reference system and the sampling of the data products is reduced by temporal filtering. Besides, ancillary data sets like preliminary orbit information belong to the level 1B data products.
- Level 2: Level 2 data products provide gravity field solutions derived from level 1B data on the basis of spherical harmonic coefficients or other base functions.
- Level 3: Gridded time series of total water storage (TWS) variations are obtained from level 2 data products by applying post-processing steps, like e.g. spatial filtering. These products typically have a spatial resolution of a few hundred kilometers and a temporal resolution of one month.

The different categories of GRACE products address the needs of different user groups, which vary from the improvement of gravity field estimation techniques based on level 1B data to mere applications of level 3 data for e.g. hydrological studies. In this thesis, level 1B data (see Section 5.2) and of level 2 data (see Section 5.1) are used.

Different strategies exist for the computation of gravity field solutions from KBR, GPS, star camera, and accelerometer observations. All approaches are based on Newton's equation of motion, which relates the orbit of a satellite of mass m to the forces acting on the satellite for any given time t,

$$\ddot{\mathbf{r}} = \frac{1}{m} f(t, \mathbf{r}, \dot{\mathbf{r}}), \tag{2.1}$$

with $\mathbf{r}$, $\dot{\mathbf{r}}$, and $\ddot{\mathbf{r}}$ denoting the satellite's position, velocity and acceleration, respectively. The force function f includes gravitational (conservative) accelerations and non-gravitational (non-conservative) accelerations. In general, a two-step approach is applied to derive gravity field solutions (Bettadpur, 2012a):

1. The orbit of each satellite is derived by numerical integration of the equation of motion, considering an apriori mean static gravity field model in combination with so-called background force models that represent forces from tides and from non-tidal oceanic and atmospheric mass variations. Generally, the following background models are taken into account (see e.g.; Dahle et al., 2012; Bettadpur, 2012c; Watkins and Yuan, 2014):
    - **N-Body Perturbations**: The satellites' orbits are influenced by gravitational forces from third bodies, i.e. the sun, the moon, and the planets. Tidal forces also act on the Earth and cause mass variations, which, in turn, influence the satellite. Direct and indirect accelerations acting on the GRACE satellites are computed from planetary ephemerides (Agnew, 2010).
    - **Solid Earth Tides**: Tidal forces induce deformations of the solid Earth, which can be modeled using Love numbers (Wahr, 1981). Subsequently, the influence on the gravitational potential is computed.
    - **Ocean Tides**: Oceanic mass variations due to tidal forces are taken into account by combining models of the different tidal constituents.
    - **Pole Tides**: Changes in the Earth's rotation result in changes of the centrifugal force. The influence on the gravitational potential is computed from the polar motion and (i) an Earth model for the contribution of the solid Earth and (ii) an equilibrium model for the oceanic contribution (Desai, 2002).
    - **De-aliasing**: High-frequency ($<$ 30 days) non-tidal variations of atmosphere and ocean are modeled and removed using the Atmosphere and Ocean De-aliasing Level-1B (AOD1B) product (Flechtner et al., 2015; Dobslaw et al., 2017a). By this means, aliasing of high frequencies into monthly gravity field solutions is avoided.
    - **Non-gravitational Forces**: Non-gravitational accelerations, i.e. atmospheric drag and radiation pressure from Sun and Earth, were measured by the accelerometers on-board of the satellites. A few-parameter model relating linear acceleration observations to non-gravitational forces is set up. During gravity field estimation, biases and scale factors of the accelerometer measurements are co-estimated.
    - **Relativistic Correction**: Instead of a fully consistent relativistic formulation of the equation of motion, general relativistic corrections are applied as described in Chapter 10 of the IERS 2010 conventions (Petit and Luzum, 2010).
2. Residuals are computed between the predicted orbit and the GPS and KBR observations. Observation equations that relate gravity field parameters and additional parameters to the residuals are linearized and solved within an iterative least squares adjustment.

Popular approaches for setting up the observation equations are (i) the classical variational or dynamic approach (Riley et al., 1967; Bettadpur and McCullough, 2017), (ii) the acceleration approach (Ditmar and van der Sluijs, 2004; Weigelt, 2017), and (iii) the energy balance approach (Jekeli, 1999, 2017). For the classical variational approach, a so-called state transition matrix is set up that relates the residuals to variations in the gravity field parameters and to initial conditions for the satellites' positions and velocities. Then, corrections to the gravity field parameters are computed iteratively by integrating the variational equations. Hereby, initial conditions are estimated for individual arcs, which have typically a length between 6 and 24 hours as a trade-off between linearization errors and the number of unknowns. Instead of solving the variational equations, the acceleration approach directly links range accelerations to gravity gradients. In doing so, linearization errors are avoided and the computational costs are limited as no iterations are necessary. However, the double differentiation amplifies the noise, making more sophisticated data processing necessary. The energy balance approach is based on the conservation of potential and kinetic energy of a satellite constellation. The main idea is to represent the Earth's gravitational potential as work acting on the satellites. Again, integration of the variational equations and iterations are redundant, but the noise situation is challenging in the inversion process.

Variants of the variational approach are the short arc approach (Mayer-Gürr, 2006) and the modified short arc approach (Chen et al., 2015), which minimize model errors using very short arcs between 30 minutes and 2 hours. The acceleration apporach was extended by Liu (2008) to the average acceleration approach that includes smoothing of the satellites' orbits and considers a noise model for data weighting. Finally, the celestial mechanics approach (Beutler et al., 2010a,b) solves for the gravity field parameters in a generalized orbit determination problem and is based on the Bernese GPS Software (Dach et al., 2007).

All of the above approaches yield at estimating spherical harmonic coefficients for representing the gravity field. A particular challenge of the GRACE mission is the measurement configuration, which leads to stripes and requires careful post-processing and spatial filtering of level 2 data (see Chapter 5). An alternative approach to the estimation of spherical harmonic coefficients are so-called mass concentration blocks (mascons), which are regularized solutions of the gravity field. Measured range-rates or range-accelerations are directly fit to specified grid locations while applying geophysical or data-driven constraints (Watkins et al., 2015; Wiese et al., 2016; Luthcke et al., 2014; Save et al., 2016).

After more than 15 years of GRACE data evaluation, the processing strategies and background models still evolve. In particular the AOD1B product for high-frequency non-tidal mass transport in the atmosphere and oceans is developed further (Dobslaw et al., 2017a) with first studies taking also hydrological variability into account (Zenner et al., 2014). Moreover, strategies for handling non-conservative forces become more and more elaborate (Chen et al., 2016; Klinger and Mayer-Gürr, 2016). Another example for the improvement of the processing strategy is the refinement of orbit integration, which is necessary for linearization of the observation equations (Ellmer and Mayer-Gürr, 2017).

2.3 GRACE Analysis Centers and GRACE Solutions

The three official processing centers within the GSDS are GFZ, JPL, and CSR. They published the reprocessed GRACE RL06 in 2018. Improvements with respect to RL05 were achieved in particular through the improved de-aliasing product AOD1B-RL06 (Dobslaw et al., 2017b). All of the official processing centers use the classical variational approach (or dynamic approach) for deriving monthly sets of spherical harmonic coefficients. However, the individual solutions differ regarding the processing strategy, the background models and the spectral resolution. Besides, several other groups compute their own gravity field models with alternative approaches and with different spatial and temporal resolution.

Table 2.1 gives an overview on selected gravity field solutions. The temporal resolution of the gravity field solutions vary between one month (1M) and one day (1D) (column 3). The spatial resolution is between degree and order (d/o) 30 to 120 for solutions provided as spherical harmonic (SH) coefficients (column 4). The corresponding wavelengths λ can be computed with the Earth's radius R and and the maximum degree n_{max} according to $\lambda = (\pi R / n_{max})$. Solutions up to n_{max}=30 cover wavelengths down to 670 km and solutions up to n_{max}=120 cover wavelengths down to 170 km, respectively. Some of the solutions provided as spherical harmonic coefficients are available after filtering in spectral domain, which is indicated by the index ^{f} in column 1. The mascon solutions are provided on 1° and 3° grids. Please note that the actual spatial resolution of GRACE data is difficult to quantify due to the spatial correlation structure and furthermore depends on the latitude. In general, at the equator a resolution of about 300 km is assumed.

The dynamic approach is used by the three official processing centers and by the Graz University of Technology (ITSG-Grace2014, ITSG-Grace2016, ITSG-Grace2018; Mayer-Gürr et al., 2016, 2018), the GNSS Research Center of Wuhan University (WHU-Grace01s; Zhou et al., 2015), the Huazhong University of Science and Technology (HUST-Grace2016; Zhou et al., 2016), the Institute of Geodesy and Geophysics at the Chinese Academy of Sciences (IGG-RL01; Wang et al., 2015), the Leibniz University Hannover (LUH-GRACE2018; Naeimi et al., 2018), the Faculty of Geoscience and Environment Engineering at Southwest Jiaotong University (SWJTU-Grace-RL01), and the Space Geodesy Research Group (CNES/GRGS; Lemoine et al., 2007; Bruinsma et al., 2010). ITSG and WHU-Grace2016 provide monthly solutions for d/o 60, 90, and 120. ITSG solutions also include daily solutions, which are estimated within an Kalman smoother framework using temporal correlation patterns derived from geophyical models (Kurtenbach et al., 2012). Additionally, ITSG-Grace2016 is provided with full covariance information. Besides the three official processing centers, ITSG-Grace2018 is currently the only publicly available solution that uses the AOD1B-RL06 product instead of AOD1B-RL05. Unlike other solutions, the CNES/GRGS-RL03 gravity fields are constrained towards a mean field, which stabilizes the solutions and makes filtering unnecessary (indicated by the ** in Table 2.1).

Bonn University (ITG-Grace2010) and Tongji University (Tongji-RL02; Chen et al., 2015, 2016) use the short-arc and modified short-arc approaches. Furthermore, ITG-Grace2010 provides daily solutions, which are based on the same method as ITSG-Grace2016 (Kurtenbach, 2011). Gravity field solutions based on the acceleration approach are published by Delft University of Technology (DMT-1b; Liu, 2008; Liu et al., 2010) and solutions based on the celestial mechanics approach are computed at Bern University (AIUB-RL02; Meyer et al., 2016).

Sakumura et al. (2014) found that the combination of several gravity field solutions using the ensemble mean reduces the noise. The European Gravity Service for Improved Emergency Management (EGSIEM) combined for the first time gravity fields of different analysis centers (AIUB-RL02, GFZ-RL05, CNES/GRGS-RL03, ITSG-Grace2016) on normal equation level, while optimally weighting the individual solutions.

While mascon solutions of JPL (Watkins et al., 2015; Wiese et al., 2016) are constrained with information from geophysical models, CSR (Save et al., 2012, 2016) and NASA's Goddard Space Flight Center (NASA-GSFC) (Luthcke et al., 2013, 2014) apply constraints that are exclusively based on GRACE data. CSR also computed a daily mascon product based on a sliding window approach, which is not yet publicly available (Sakumura et al., 2016).

Obviously, the question arises, which of the GRACE solutions should be used. This depends mainly on the application requirements. Several studies compare and validate GRACE solutions (e.g. Sakumura et al., 2014; Scanlon et al., 2016). Generally, a combined solution as derived by the EGSIEM project reduces noise and is more reliable. All products based on spherical harmonics require careful post-processing including filtering and signal restoration due to leakage effects. In contrast, filtering is not necessary when using Mascon solutions. However, these solutions are drawn towards constraints during the estimation process, which is an additional error source. Most of the current global gravity field solutions can be downloaded from the ICGEM website (`http://icgem.gfz-potsdam.de/home`). ICGEM also provides an online tool to convert monthly spherical harmonic coefficients from individual solutions to gridded maps of gravity anomaly, TWS, etc. Furthermore, interactive tools for displaying and comparing different GRACE solutions are available (e.g. `http://www.thegraceplotter.com/`). Besides global gravity models, a number of regional solutions exist.

The different gravity field solutions allow some kind of quality control, but as all of them are based on the same level 1B data, they contain the same instrument errors and the same errors from atmospheric and oceanic de-aliasing products. Therefore, differences between the solutions cannot be used as error estimate. The error estimate of the GRACE solutions is indeed challenging and represents a specific field of research, which will be discussed in Section 5.1.5.

In this work, I use the solutions from ITSG-Grace2016 as they (i) cover the whole study period, (ii) are unconstrained solutions, (iii) provide full error covariance matrices, and (iv) also include a daily product. The full error covariance matrices are interesting for studying the influence of correlations in the context of data assimilation.

Table 2.1: This table provides an overview on the currently publicly available GRACE solutions (as of December 2018). The responsible institutions and references are given in the text. The index ^{f} in Column 1 indicates solutions of spherical harmonic (SH) coefficients that are also available after spectral filtering. Column 2 gives the approaches that are used for estimating the gravity field solutions (see Section 2.2). Constrained solutions are indicated by **. The temporal resolution of the products is provided in Column 3, where 1M indicates one month and 1D one day. For products provided as SH coefficients, the spatial resolution is indicated by the maximum degree of the SH coefficients, whereas for mascon products the grid size is given (Column 4).

Solution	**Approach**	T	**Resolution**	**Time span**
CSR-RL05^f	SH: dynamic	1M	96	2002-04 to 2017-06
CSR-RL06	SH: dynamic	1M	60	2002-04 to 2016-08
JPL-RL05^f	SH: dynamic	1M	90	2002-04 to 2017-06
JPL-RL06	SH: dynamic	1M	96	2002-04 to 2016-08
GFZ-RL05^f	SH: dynamic	1M / 7D	90/30	2002-04 to 2017-06
GFZ-RL06	SH: dynamic	1M	96	2002-04 to 2014-12
ITSG-Grace2014^f	SH: dynamic	1M / 1D	60,90,120/40	2003-02 to 2014-06
ITSG-Grace2016^f	SH: dynamic	1M / 1D	60,90,120/40	2002-04 to 2017-06
ITSG-Grace2018	SH: dynamic	1M / 1D	60,90,120/40	2002-04 to 2016-08
WHU-Grace01s	SH: dynamic	1M	60,90,120	2002-04 to 2016-07
HUST-Grace2016^f	SH: dynamic	1M	60	2003-01 to 2016-03
IGG-RL01	SH: dynamic	1M	60	2002-04 to 2016-07
LUH-Grace2018	SH: dynamic	1M	80	2003-01 to 2009-12
SWJTU-Grace-RL01	SH: dynamic	1M	60	2003-03 to 2011-10
CNES/GRGS-RL03	SH: dynamic**	1M / 10D	80/80	2002-08 to 2014-06
ITG-Grace2010	SH: short arc	1M / 1D	120/40	2002-08 to 2009-08
Tongji-RL02^f	SH: mod. short arc	1M	60	2003-01 to 2015-09
DMT-1b	SH: acceleration	1M	120	2003-02 to 2009-12
AIUB-RL02^f	SH: cel. mechanics	1M	60,90	2003-03 to 2014-03
EGSIEM	SH (combined)	1M	90	2006-01 to 2007-12
JPL-RL05	mascons**	1M	3°	2002-04 to 2017-06
JPL-RL06	mascons**	1M	3°	2002-04 to 2017-06
CSR-RL05	mascons**	1M	1°	2002-04 to 2017-06
NASA-GSFC	mascons**	1M	1°	2003-01 to 2016-07

Chapter 3

Related Work

Over the past decade, the Gravity Recovery and Climate Experiment (GRACE) mission has contributed significantly to the understanding of hydrological processes (Tapley et al., 2004a). In particular, a large number of studies has been using GRACE data as a validation tool for hydrological models (see Section 3.1). These studies are usually based on gridded or basin-averaged total water storage anomalies (TWSA, Section 3.1.1), and suffer from the downward continuation and filtering steps. To overcome these limitations, Eicker et al. (2016) evaluated global hydrological models in-orbit on the basis of level 1B data by computing K-band range-rate (KBRR) residuals. This approach was extended in this thesis to examining different model versions and signals on smaller spatial scales. Section 3.1.2 introduces studies which assess different geophysical signals using KBRR residuals. With Zaitchik et al. (2008), a new application of GRACE data emerged, the assimilation of total water storage anomalies (TWSA) into a hydrological model, as also realized within this thesis. During the last few years, important progress has been made regarding the assimilation of GRACE data, but still only few studies have been published (Section 3.2).

3.1 Validation of Hydrological Models Using GRACE Data

Hydrological models aim at representing water, energy, and biogeochemical fluxes at different spatial and temporal scales, and thereby contribute to a better understanding of the Earth's system. Total water storage (TWS) is one key variable of hydrological models as the distribution of water within the model controls all major hydrological processes. However, uncertainties in (i) model structure or representation of physical processes, (ii) model parameters, and (iii) atmospheric forcing data lead to limited predicting skills. Before the start of the GRACE mission, validation of TWS simulated by hydrological models was restricted to few ground-based measurements.

Indeed, the start of the GRACE mission in 2002 has opened up new possibilities for assessing the quality of hydrological models and for determining model deficiencies. Nevertheless, it should be remembered that GRACE observed the spatio-temporal variability of the total sum of the water column, whereas hydrological models simulate a number of individual storage compartments. Furthermore, in contrast to hydrological models, which usually run at daily to hourly time steps on grids of 1° or smaller, GRACE-derived TWSA are monthly data sets with a spatial resolution of about 200 km. The resulting challenges are described in detail in Section 5.1. Below, the records of hydrological model validation with GRACE data on the basis of TWS time series are presented (Section 3.1.1), followed by an introduction into studies relevant for the validation of hydrological signals based on GRACE level 1B data (Section 3.1.2).

3.1.1 Total Water Storage Time Series

The very first studies comparing hydrological models and GRACE primarily explored the fundamental information content of the new type of observations provided by GRACE. In this scope, Andersen (2005) evaluated 15 months of gravity data and compared them to four hydrological models on an inter-annual scale. At that time, observations and models diverged widely, with only one out of four models being correlated with GRACE data at the global scale. In general, studies of this period were limited to comparing spatial maps of the seasonal cycle and basin-averaged TWSA time series of large river catchments, thereby verifying continental scale patterns and general seasonal dynamics (Tapley et al., 2004a; Wahr et al., 2004; Ramillien et al., 2004, 2005). Research tended to focus on basic GRACE processing strategies, e.g. filtering and treatment of the low-degree coefficients (Chen et al., 2005; Chen, 2005), rather than on hydrological phenomena.

A step forward was the assessment of specific hydrological extreme events in GRACE-derived TWSA and hydrological models. In an early study by Andersen et al. (2005) the manifestation of the 2003 European heat wave in GRACE data was confirmed by three independent data sets, which outlined the potential of GRACE for investigating extreme climate events. Chen et al. (2009) reported the 2005 Amazon drought being distinct in the GRACE time series, whereas, in this case, hydrological models failed in representing the drought related decline in TWS. Similarly, Chen et al. (2010) found a detailed picture of the 2008/2009 La Plata drought in the GRACE data, while this event was only partly captured by the investigated land surface model. A more recent study by Jin and Feng (2013) surveyed the general difficulties of global hydrological models in matching GRACE-derived trends and accelerations for regions affected by extreme events.

Several studies investigated the general performance of different model versions and/or different hydrological models. In this scope, Niu and Yang (2006) used GRACE data to demonstrated the better performance of a modified version of the Community Land Model (CLM) compared to the standard version. The value of GRACE for detecting missing processes in hydrological models or improving existing ones was confirmed by Alkama et al. (2010), who carried out a statistical comparison between GRACE and a hydrological model for the 183 largest river basins of the world. Grippa et al. (2011) evaluated six GRACE solutions and soil moisture from nine land surface models over West Africa, and identified processes to be improved in the land surface modeling. Pokhrel et al. (2013) included an interactive groundwater store in a hydrological model over the Amazon basin and validated both model versions against GRACE, achieving significant improvement especially with respect to annual amplitudes and phases. Recently, Getirana et al. (2017) used GRACE data to prove the influence of rivers and floodplanes on TWS. General strengths and weaknesses of current global hydrological models were evaluated in a comprehensive study by Zhang et al. (2017), who compared the performance of four different models in 31 catchments situated within different climate zones. Scanlon et al. (2018) showed that global hydrological models underestimate TWS trends relative to GRACE observations.

Overall, a considerable number of studies suggested that hydrological models generally tend to underestimate the amplitude of TWS compared to GRACE and that they often have a time shift of about one month.

In recent years, decreasing errors in the GRACE solutions enabled the validation of hydrological models also at smaller spatial scales. This concurrently stimulated research regarding processing strategies for ensuring the consistency of GRACE data and output from hydrological models (Longuevergne et al., 2010).

Progress in the quality of the GRACE solutions also enabled the evaluation of individual hydrological compartments. For instance, Jin and Feng (2013) computed ground water variability from the difference of GRACE-derived TWSA and surface water, soil moisture, snow, ice and canopy water from two different hydrological models. They found one of the models representing a relatively reliable data set of groundwater variability compared to in-situ observations. Similarly, Joodaki et al. (2014) estimated the human contribution to groundwater depletion in the Middle East. Döll et al. (2014b) contributed to a better understanding of groundwater depletion and its causes by jointly interpreting TWSA from GRACE and an improved global hydrological model, which considers groundwater recharge from surface water bodies and major aquifers.

Another step forward was the separation of GRACE-derived TWSA into different components using statistical decomposition techniques. Temporal and spatial decomposition of TWSA time series enables a more detailed evaluation of hydrological processes including interannual and non-periodic phenomena. For the first time, Rangelova et al. (2007) applied Principle Component Analysis (PCA) to GRACE data and to three hydrological models over North America. They found a strong annual signal in the first spatial mode due to snow accumulation and inter-annual mass changes in the second mode. For the hydrological models similar patterns were revealed, although they rather underestimated mass changes.

In later studies, the application of statistical decomposition techniques were extended from pure validation of hydrological models to joint evaluation of TWS observations and models in order to gain maximum information on hydrological dynamics in different storage compartments. Awange et al. (2014) applied PCA to TWSA from GRACE, to soil moisture from a hydrological model, and to remotely sensed precipitation. Hence, they concluded about the relationship between rainfall and ground water variations in Ethiopia. Forootan et al. (2014) went even further and separated water storage compartments over Iran by (i) decomposing data sets from hydrological modeling and altimetry into statistically independent components, and (ii) fitting them to TWSA from GRACE. Andrew et al. (2017) used a wavelet approach in order to partition TWSA from GRACE into its compartments. Recently, Felfelani et al. (2017) isolated anthropogenic water storage changes from multiple GRACE solutions and two hydrological models, in which human factors were excluded. In particular, their results demonstrated different performance of the two models in highly-managed catchments and in snow-dominated regions.

Lately, there has been growing interest in the quantification of uncertainty of TWS products from GRACE and from hydrological models, respectively (Long et al., 2017). A further focus is put on deriving one 'best' GRACE product and to make it available to a wider user community, as attempted by the European Gravity Service for Improved Emergency Management (EGSIEM) project.

In this review, I highlighted a subjective selection of studies that represent the progress made in validating hydrological models since the start of the GRACE mission. An overview on all GRACE related publications including applications related to ocean and atmosphere is provided by GeoForschungsZentrum (GFZ) at `http://www-app2.gfz-potsdam.de/pb1/op/grace/references/sort_date.html`.

3.1.2 K-Band Range Rate Residuals

Most of the above discussed studies assess monthly gravity field solutions provided either as sets of spherical harmonic coefficients or as gridded values of equivalent water heights. Gravity field processing requires downward continuation from the satellites' altitude, which amplifies high-frequency measurement errors. Consequently, the recovered gravity fields are noisy and require a tailored filtering procedure. Filtering results into attenuated fields of monthly mass variations at a coarse spatial resolution of a few hundred kilometers, which complicates comparison with high-resolution geophysical or hydrological models. The limitations due to downward continuation and filtering can be overcome by comparing model outputs and GRACE observations directly on the basis of GRACE level 1B K-band observations. In this scope, K-band range-rate (KBRR) or K-band range-acceleration (KBRA) observations are simulated from modeled mass variations and subtracted from original GRACE K-band observations leading to KBRR or KBRA residuals.

Since the early years of the GRACE mission, KBRR residuals were evaluated to assess the quality of GRACE solutions, ocean models, atmospheric models, and of the GRACE de-aliasing products (Bosch et al., 2009; Zenner et al., 2012; Dobslaw et al., 2013). Therefore, the ability of different products in reducing KBRR residuals was compared. For instance, Kusche et al. (2009a) computed KBRR residuals for GRACE solutions filtered with different de-correlation filters. Filtering reduced the KBRR residuals, which was however inconclusive (Kusche et al., 2009a). More recently, Dobslaw et al. (2017b) used KBRA residuals to localize regions that are problematic for de-aliasing. However, KBRA are more noisy than KBRR due to differentiation, which is challenging regarding the evaluation of small signals.

Han et al. (2008) explained the computation of KBRR in detail and investigated hydrological signatures and signatures from glacial isostatic adjustment over South America, Africa, and North America. Moreover, Han et al. (2010a) detected signatures of the 2010 Maule (Chile) earthquake in the KBRR residuals. They investigated range-rate data close to the epicenter before and after the earthquake and were able to infer a negative gravity anomaly after the earthquake.

Different hydrological models were first evaluated using GRACE level 1B data in the studies of Han et al. (2009) and Han et al. (2010b). They assessed individual arcs of range-rate residuals for the Amazon basin regarding the impact from surface waters and soil moisture. Later on, Zenner et al. (2014) computed KBRR residuals for hydrological signals from two global hydrological models used in addition to the atmosphere and ocean de-aliasing product and compared them to KBRR residuals derived from the daily Kalman-based GRACE solutions of ITG-Grace2010 (Kurtenbach, 2011) used in addition to the atmosphere and ocean de-aliasing product. GRACE-derived ITG daily Kalman solutions turned out to better fit to GRACE KBRR observations than daily hydrological model output. At the moment it is not clear, whether daily hydrological model output could improve GRACE solutions through de-aliasing.

3.2 Assimilation of GRACE Data into Hydrological Models

Traditionally, hydrological models have been calibrated against discharge observations (Duan et al., 2006). Discharge reflects the response to all hydrological processes within a catchment. However, models calibrated only with respect to discharge often fail in providing reliable estimates of other relevant hydrological variables (Guo et al., 2017). Therefore, multi-objective calibration with complementary observations became more and more relevant. A decade ago, studies have started to investigate the integration of GRACE observations into the calibration process of hydrological models (e.g., Werth et al., 2009; Lo et al., 2010; Livneh and Lettenmaier, 2012; Rakovec et al., 2016; Chen et al., 2017; and Bai et al., 2017), which generally results in an improved representation of the continental water cycle.

A step further in tuning models towards observations is taken by data assimilation, i.e. the hydrological model states are updated towards the observations. Assimilation of GRACE observations allows for improving model estimates of TWS while simultaneously disaggregating the coarse TWS observations spatially and temporally. This section provides a literature review on assimilation or joint calibration/assimilation experiments that integrate GRACE observations into hydrological models (Table 3.1).

The following topics and key challenges regarding the assimilation of GRACE data into hydrological models are addressed in literature (see last column of Table 3.1):

1. GRACE solutions are usually provided as spherical harmonic coefficients, which are converted to gridded data or catchment averages. This raises the question about the optimal representation of GRACE observations for data assimilation purposes.
2. GRACE data has a temporal resolution of one month and a spatial resolution of about 200 km, while models are usually resolved on finer temporal and spatial scales. Thus,
   (a) the formulation of the observation operator, which transforms the model into observation space, and
   (b) the computation and application of the analysis increment, which represents the correction of the model state towards the observations,

   are challenging research areas.
3. The uncertainties of model states and observations govern the analysis increment.
   (a) In the case of ensemble-based filter algorithms the model uncertainty is represented by the ensemble spread, which is generated by disturbing (i) forcing data, (ii) model parameters, and (iii) model states.
   (b) Due to the challenging measurement configuration the derivation of realistic error estimates for gridded or basin averaged GRACE observations remains a matter of research.
4. A basic choice in data assimilation has to be made regarding the filter algorithm. Today, a huge number of different algorithms exists with individual strengths and weaknesses, especially with respect to computational costs, the representation of the ensemble of model states, and the disturbance of the observations. Furthermore, filter algorithms may be tuned by applying localization (i.e., limiting the influence of the individual observation to a certain radius) and/or by ensemble inflation (i.e. increasing the ensemble spread).
5. The validation of assimilation results is performed using observational data sets, which need to be comparable to the model output. The choice of the validation data sets and metrics depends on the study region.
6. Some GRACE-related assimilation experiments address specific hydrological applications, e.g.
   (a) the improvement of snow representation,
   (b) the monitoring of droughts,
   (c) the assessment of regional flood potential, and
   (d) the quantification of groundwater depletion.
7. Usually, the quality of the representation of long-term and interannual variability by hydrological models is limited. Thus, the improvement of this aspect within data assimilation frameworks is an interesting research question.

In their influential paper Zaitchik et al. (2008), for the first time, introduced GRACE data into a hydrological model through data assimilation. Zaitchik et al. (2008) developed a smoother algorithm that mimics the GRACE observation frequency and distributes the monthly analysis increment to each day of the month by reiterating the model. In their analysis, Zaitchik et al. (2008) demonstrated that the assimilation of GRACE data increases the temporal correlation with gauged streamflow and to some extent with groundwater well observations, ans also impacts evapotranspiration and root zone soil moisture. Furthermore, Zaitchik et al. (2008) identified limitations of the Catchment Land Surface Model (CLSM in representing the annual minimum in TWS and suggested to modify the model accordingly.

Based on the framework developed by Zaitchik et al. (2008), Forman et al. (2012) showed that the error of snow simulation in the Mackenzie river basin declines due to the assimilation of GRACE observations. Forman et al. (2012) also pointed out short-comings of Glacial Isostatic Adjustment (GIA) models which significantly affect the quality of the assimilated GRACE data, and limitations due to lacking validation data sets. Another study focusing on the improvement of snow data through assimilation was performed by Su et al. (2010), who complemented the assimilation of Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction into the CLM over North America by additionally assimilating GRACE data. Su et al. (2010) attributed improvement of modeled snow water equivalent with respect to an observation-based data set to the unique information provided by GRACE. Su et al. (2010) applied the same model (CLM) as used in this thesis. However, validation is exclusively limited to snow water equivalent and the developed assimilation framework leaves room for improvement.

The framework of Zaitchik et al. (2008) was adjusted in Li et al. (2012) in order to overcome the deficiencies in the model physics identified by Zaitchik et al. (2008). Li et al. (2012) underlined the increase of biases in fluxes (evapotranspiration, discharge) as a result of assimilation and pronounced the need for simultaneously assimilating complementary observations. Furthermore, the framework was used for assessing the impact of GRACE data assimilation on drought monitoring (here: agricultural and hydrological droughts). Indeed, Li et al. (2012) found a positive contribution of the assimilated model to drought monitoring due to its higher spatial resolution compared to pure GRACE observations and due to a better representation of deep soil water. The experiment was carried out over Europe, which suffered from a serious drought in 2003. Up to today, this study is the only data assimilation experiment performed over larger parts of Europe that includes GRACE observations.

In a later study, Houborg et al. (2012) extended the experiments of Li et al. (2012) and developed drought indicators for North America based on the assimilated model. Previously, drought indicators were mainly restricted to the use of precipitation indices, whereas less quickly varying components of the water cycle like deep soil moisture and groundwater were not included. Now, Houborg et al. (2012) integrated the spatially and temporally downscaled GRACE observations into the U.S. and North American Drought Monitors yielding a statistically significant improvement of drought detection. Similarly, Reager et al. (2015) reported on the value of the assimilated model for characterizing flood potential in the Missouri basin, and pointed out that the assimilated model allowed to draw a more detailed picture of flood conditions. Besides, Houborg et al. (2012) put some effort into improving the data assimilation system by generating more reliable model errors through the consideration of correlations between the model forcing fields, and by rescaling the GRACE observations in order to take into account signal attenuation due to filtering.

A stronger focus on the assimilation framework itself was placed by Forman and Reichle (2013), who performed a number of synthetic twin experiments to investigate the optimal way of assimilating GRACE data into the CLSM. On the one hand, they investigated the impact of the correlation length used for perturbing the forcing data, and, on the other hand, they analyzed the impact of introducing GRACE data with different spatial resolution by dividing the Mackenzie River Basin in one to six basins. Although an increasing error correlation length of the perturbed forcing data was found to the model performance with respect to TWS and snow water, a larger impact was obtained from the spatial resolution of the observations. In their conclusions, Forman and Reichle (2013) recommended that GRACE data should be assimilated at the smallest spatial scale, at which the observations can be assumed to be uncorrelated.

So far, assimilation experiments with the CLSM were limited to catchment-averaged GRACE-derived time series. In line with studies from other groups (e.g., Eicker et al., 2014), Kumar et al. (2016) moved on to incorporate gridded GRACE observations (on a $1^\circ\times1^\circ$ grid) with corresponding distributed error estimates instead of uniform ones. Interestingly, they found only small differences compared to the results for the catchment-averaged observations, which is in contrast to Schumacher et al. (2016). Additionally, Kumar et al. (2016) assessed influences from distributed/uniform measurement errors and from different sets of scaling factors, with ambiguous results that need more detailed investigations. All in all, Kumar et al. (2016) confirmed through an integrated and systematic validation environment for land model analysis that data assimilation leads to improvement, even when model structure and parameters are optimized and the best available forcing data are used.

Girotto et al. (2016) revisited the assimilation algorithm developed by Zaitchik et al. (2008) and suggested a new approach (similar to Eicker et al., 2014), which yielded better results than the previous approaches from Zaitchik et al. (2008) and Su et al. (2010) (for more details see Section 7.2.4). In contrast to other studies, Girotto et al. (2016) rescaled the GRACE observations to match the climatology of the model in order to avoid impacts from systematic differences. This procedure ensures consistency of observed and modeled dynamic range of TWS, but on the other hand certain model deficiencies cannot be revealed. Additionally, Girotto et al. (2016) limited the influence of each observation grid cell to a certain radius by localization. Girotto et al. (2017) used the modified algorithm to force the model to represent the Indian groundwater depletion, which is not simulated by the open-loop model due to missing anthropogenic processes. Admittedly, TWS from the assimilated model declines significantly, but, as deep aquifers are not modeled, the assimilation of GRACE introduced unrealistic trends in the shallow groundwater and an unrealistic reduction of evapotranspiration. This again, emphasizes the need for representing all relevant processes of the water cycle.

The value of GRACE data for introducing missing trends into a hydrological model was confirmed by Tangdamrongsub et al. (2017) for the Hexi Corridor in Northern China, that suffers from huge human-induced water abstraction. Tangdamrongsub et al. (2017) assimilated gridded GRACE data into the global distributed hydrological model PCR-GLOBWB, with and without considering observation error correlation up to a certain localization radius using the filter approach of Forman et al. (2012). In line with Schumacher et al. (2016), Tangdamrongsub et al. (2017) found significant impact of error correlation on the assimilation result, and concluded that white-noise GRACE errors severely overestimate the information content of the observations.

Tangdamrongsub et al. (2015) investigated the effect of GRACE data assimilation in the presence of different uncertainty assumptions for model forcings and parameters. In this scope, Tangdamrongsub et al. (2015) set up an assimilation framework for the Rhine river basin using the rainfall-runoff model wflow_hbv, which is based on the distributed HBV-96 model. GRACE observations were interpolated to 5-day time series and assimilated on the 1 km model grid assuming uniform and uncorrelated observation errors. All assimilation experiments significantly improved the model with respect to groundwater well data, whereas only slight improvements for streamflow data were determined.

In her dissertation, Schumacher (2016) established a joint calibration and data assimilation framework for integrating GRACE observations into the WaterGAP Global Hydrology Model (WGHM). The thesis provides a comprehensive and thorough discussion of several aspects discussed at the beginning of this Section and resulted into different publications. In Eicker et al. (2014) the calibration and data assimilation framework was introduced with a model state vector augmented by sensitive parameters that were previously analyzed by Schumacher et al. (2012). The contribution of GRACE data to individual storage compartments was assessed within a synthetic experiment. To my knowledge, Eicker et al. (2014) for the first time assimilated gridded ($5° \times 5°$) GRACE observations using full error covariance matrices propagated from the spherical harmonic coefficients, and contrasted the results to the application of basin-averaged GRACE observations for the Mississippi river basin. They found a smaller root mean squared deviation (RMSD) between GRACE and the assimilation run with gridded observations than with basin-averaged observations. Another interesting aspect was that the assimilated model followed GRACE more closely after the calibration and data assimilation period due to the calibrated parameters. In Schumacher et al. (2016) the calibration and data assimilation framework was extended for different Kalman filter algorithms and was tuned by variance inflation. Within a synthetic twin experiment, Schumacher et al. (2016) demonstrated spatial error correlation of GRACE data having significant impact on the filter update, with ambiguous results. In contrast, the effect from different filter algorithms was found to be marginal. Schumacher et al. (2018) transferred the calibration and data assimilation framework to the Murray-Darling basin and included a discussion on the parameter equifinality problem. The study showed that the negative trend of TWS related to the millenium drought could be introduced to the groundwater compartment of WGHM by assimilating GRACE. Besides, Schumacher et al. (2018) demonstrated that using different GRACE solutions has only a small impact on the assimilated model.

The first joint assimilation experiment of remotely sensed TWS and soil moisture was performed by Tian et al. (2017) with the World Wide Water Resources Assessment (W3RA) tool, a hybrid between a land surface and a catchment model. Here, for the first time, a GRACE mascon product was used in the assimilation process. Main challenges arised from the different uncertainties of the two products, which were rescaled correspondingly. Tian et al. (2017) showed that the assimilation of GRACE data mainly impacts groundwater and TWS, whereas the assimilation of soil moisture has large effects on top and shallow soil water. Indeed, the joint assimilation of TWS and soil moisture lead to better results than the individual assimilation of TWS and soil moisture, also in comparison to other independent observation-based data sets.

Khaki et al. (2017b) investigated the same model as Tian et al. (2017) over Australia, but with focus on the assimilation algorithm. Their comprehensive experiments addressed the choice of the observation grid taking into account full error correlations during the assimilation process. For smaller observation grids localization was applied, which improved the results significantly with respect to groundwater observations. Khaki et al. (2017a) extended the framework for a number of different filter algorithms, including the classical Ensemble Kalman Filter (EnKF), the Ensemble Transform Kalman Filter (ETKF), the SQuare Root Analysis (SQRA) scheme, the Deterministic Ensemble Kalman Filter (DEnKF), the Ensemble Square Root Filter (EnSRF), the Ensemble Optimal Interpolation (EnOI) scheme, and two particle filter variants (PF). Best results were obtained using the SQRA, the EnSRF, or the EnKF. In Khaki et al. (2018) the EnSRF was applied to improve the W3RA over Iran. One focus was the analysis of the effect of GRACE data assimilation on different water storage compartments. The study of Khaki et al. (2018) also resulted into more reliable quantification of groundwater depletion.

In contrast to all previous studies, van Dijk et al. (2014) proposed an innovative approach for combining off-line model output from different LSMs with ancillary data sets (e.g. groundwater, river flow, lake water storage) and with GRACE-derived TWS estimates within an ensemble Kalman filter. Observation-based TWS was derived from three different GRACE solutions and model-based TWS was computed using five land surface models. The respective error estimates were obtained through a triple collocation technique (Stoffelen, 1998), which, however showed to be problematic as it requires uncorrelated time series. The results were validated against several remote sensing products, which indicated only minor improvements of modeled river and snow storage. Finally, van Dijk et al. (2014) concluded that the main contribution of the assimilation procedure is a better agreement with the GRACE observations.

All in all, the presented studies furthered the search for an optimal assimilation strategy for GRACE-type TWS observations, yet, they do not finally answer the research questions formulated at the beginning of this section. However, implications can be drawn also for determining the optimal assimilation strategy of GRACE into the CLM3.5 model over Europe, which has a higher spatial and temporal resolution than the presented models.

Table 3.1: This table gives an overview on studies dealing with the assimilation of GRACE data into hydrological models. The first column names the study, the second to fourth columns provide the study area and the applied hydrological model together with its resolution. The fifth and sixth columns inform about the resolution of the GRACE solutions and the structure of the error covariance matrix of gridded TWSA, with the possible cases of (i) a diagonal covariance matrix (diag.), (ii) correlations represented by assuming a certain spatial error correlation length (corr), and (iii) a full covariance matrix (full). In case (i) and (ii), the errors are either uniform for all grid cells / catchment averages (unif.) or distributed (distr.), n.r. means not reported. The filter algorithm is reported in column eight and subscript C marks simultaneous calibration of model parameters. The study focus is indicated by the last column and relates to main challenges and questions formulated above. (*LSMs used in the study: GLDAS(CLM, MOS, NOAH, VIC), W3RA)

Reference	**Study Area**	**Model**	**Model Resolution**	**GRACE Resolution**	**GRACE Error [mm]**	**Assimilation Technique**	**Focus of Study**
Zaitchik et al. (2008)	Mississippi	CLSM	4000 km^2	4 subbasins	diag. & 20 unif.	EnKS	2,4
Forman et al. (2012)	Mackenzie R.	CLSM	4000 km^2	4 subbasins	diag. & 16-18 unif.	EnKS	3b,6a
Li et al. (2012)	Europe	CLSM	1500 km^2	9 subbasins	diag. & 15 unif.	EnKS	6b
Houborg et al. (2012)	N. America	CLSM	3640 km^2	26 subbasins	diag. & 10, 20 unif.	EnKS	6b
Forman and Reichle (2013)	Mackenzie R.	CLSM	3000 km^2	1 to 6 subb.	diag. & 8-20 unif.	EnKS	1,3a
Reager et al. (2015)	Missouri	CLSM	4000 km^2	1 subbasin	diag. & n.r.	EnKS	7c
Kumar et al. (2016)	USA	CLSM	1/8°x1/8°	1°x1°, 26 subb.	diag. & 20 distr.	EnKS	1,3b,5,6b
Girotto et al. (2016)	USA	CLSM	36 km	1°x1°	corr. & >=15 distr.	EnKS	1,2b
Girotto et al. (2017)	India	CLSM	36 km	1°x1°	corr. & >=15 distr.	EnKS	7
Su et al. (2010)	North America	CLM	n.r.	4°x4°	diag. & 20 unif.	EnKS	6a
van Dijk et al. (2014)	global	5 LSMs*	1°x1°	1°x1°	diag. & 11-12 distr.	EnKF	3,4
Eicker et al. (2014)	Mississippi	WGHM	0.5°x0.5°	5°x5°, 1 subb.	full	EnKF^C	1,2a,3a,3b
Schumacher et al. (2016)	Mississippi	WGHM	0.5°x0.5°	5°x5°, 1, 4, 11, subb.	diag & 9-25 distr., full	EnKF^C, SQRA^C, SEIK^C	1,3b,4
Schumacher (2016)	Mississippi, Murray Darling	WGHM	0.5°x0.5°	5°x5°, 1, 4, 11, subb.	diag & 9-25 distr., full	ENKF^C, SQRA^C, SEIK^C	1,3b,4,6b,7
Schumacher et al. (2018)	Murray Darling	WGHM	0.5°x0.5°	4 subbasins	full	EnKF^C	1,6b,7
Tangdamrongsub et al. (2015)	Rhine	wflow_hbv	1 km x 1 km	1 km x 1 km	diag. & 20 unif.	EnKF	3a
Tangdamrongsub et al. (2017)	Hexi Corridor (China)	PCR-GLOBWB	0.3°x0.3°	0.5°x0.5°	full	EnKS	3b,7
Tian et al. (2017)	Australia	W3RA	n.r.	0.5°x0.5°	diag. & 20 distr.	EnKF	4,5
Khaki et al. (2017b)	Australia	W3RA	1°x1°	1° to 5°, 12 subb.	full	SQRA	1,3b,5
Khaki et al. (2017a)	Australia	W3RA	1°x1°	1°	full	(D)EnKF, ETKF, SQRA, EnSRF, EnOI, PF	4
Khaki et al. (2018)	Iran	W3RA	1°x1°	1°	full	EnSRF	6d

Chapter 4

Modeling Terrestrial Water Storage

A large number of hydrological models has been developed to improve our understanding of individual components of the terrestrial water cycle. Hydrological models can be classified in three categories: empirical models, conceptual models, and physics-based models (Devia et al., 2015). Empirical models establish a functional relationship (not based on physical processes) between model input and output. In contrast, conceptual models describe different components of the water cycle considering physical laws through empirical equations with model parameters that need to be calibrated. Finally, physics-based models aim at mapping the processes of the real world through mathematical equations as realistically as possible, using state variables that vary in time and space. Simulations with physics-based hydrological models are often computationally expensive and produce a huge amount of data. This chapter gives an overview of the hydrological models used in this thesis. The regional physics-based Community Land Model version 3.5 (CLM3.5, Section 4.1) is used for data assimilation experiments, while global models (Section 4.2) serve for deriving rescaling factors of gridded GRACE data (Section 5.1.3.3) and for investigating short-term hydrological signals on the level of GRACE range-rate and range-acceleration residuals (Chapter 9). Supplementary observation-based data sets, used for validation purposes, are introduced in Section 4.3.

4.1 Community Land Model version 3.5

The land-surface model CLM consists of multiple modules that represent biogeophysical and biogeochemical processes, dynamic vegetation composition and structure, plant phenology, and the hydrological cycle. Energy, water, and carbon fluxes are simulated in response to atmospheric forcings (Section 4.1.3.1) and land surface characteristics (Section 4.1.3.2). For this thesis, the simulation of the water cycle is of particular importance (Section 4.1.2).

CLM was first introduced by Dai and Trenberth (2003) and is still improved and refined. Version 3.5 includes parametrizations of canopy interception, frozen soil, soil water availability and soil evapotranspiration, surface and subsurface runoff, and a simple groundwater model, which lead to an improved annual cycle of total water storage (TWS; Oleson et al., 2008). With version 4.0 more sophisticated representations of soil hydrology and snow processes were included. CLM version 4.5 (Oleson et al., 2013) incorporates modifications regarding the hydraulic properties of frozen soils and a new surface water store. In the latest CLM version 5.0, important modifications have been applied to soil and plant hydrology, snow density, and river modeling. Furthermore, later CLM versions realize a more detailed hierarchical grid structure.

Although more recent versions exist, in this thesis CLM3.5 was used, as it is the land-surface component of the fully coupled soil-vegetation-atmosphere model realized within the Terrestrial Systems Modeling Platform (TerrSysMP; Gasper et al., 2014; Shrestha et al., 2014). TerrSysMP consists of three individual models, the surface-subsurface model Parallel Flow (ParFlow), the one dimensional land-surface model CLM3.5, and the COnsortium for Small Scale MOdelling (COSMO) numerical weather prediction model. After gaining experience with GRACE assimilation into CLM3.5 stand-alone, it will be possible to extend the simulations for the other two components of TerrSysMP. Coupling of the three models is realized with the Ocean Atmosphere Sea Ice Soil coupling tool (OASIS3-MCT; Craig et al., 2017), and allows for running individual models or two or three coupled models. A fully coupled experiment for the European COordinated Regional Downscaling EXperiment CORDEX) area was set up by Keune et al. (2016) at 0.11° (∼ 12.5 km) resolution. Ensemble runs and data assimilation are realized within the integrated framework of TerrSysMP and the Parallel Data Assimilation Framework (PDAF; Nerger et al., 2005; Nerger and Hiller, 2013) developed by Kurtz et al. (2016).

4.1.1 Model Structure

Spatial land surface heterogeneity is realized in CLM3.5 by dividing each grid cell into multiple nested subgrid levels (Oleson et al., 2004). More precisely, one grid cell can have up to five landunits, namely glacier, wetland, vegetated, lake, and urban. Parallelization is realized by grouping grid cells into ‘clumps’. Each ‘clump’ is then assigned to different processors. Each vegetated landunit is divided into up to four Plant Functional Types (PFT). The PFTs capture ecological and hydrological characteristics of 15 categories of plants (Bonan et al., 2002) and influence the fluxes to and from the land surface. Furthermore, each land unit and each PFT has a separate column for state variables of water and energy and related fluxes, with each column consisting of 10 soil layers and up to five snow layers. The soil layers are unevenly spaced with increasing thickness from the surface down to 3 m depth.

When starting a model run, the current model time step is determined and corresponding atmospheric forcing data are read. Then, two driver modules are called. The first driver module determines the ‘clump’ boundaries and the surface albedo and, subsequently, calls several components of the model, which are summarized roughly in the following:

1. The water balance at the beginning of the time step is computed from all storage compartments. Carbon and nitrogen balances are computed if carbon and nitrogen cycle dynamics are coupled to CLM.

2. The first hydrology module computes water storage changes due to precipitation. This includes processes such as interception, throughfall, and canopy drainage. Furthermore, the snow layers are initialized.

3. The absorbance of surface radiation by vegetation and land surface is computed.

4. The first biogeophysics module uses ground temperature to determine leaf temperature and surface fluxes.

5. Surface fluxes and canopy fluxes are computed including sensible and latent heat fluxes, leaf fluxes, transpiration, dew accumulation, and photosynthesis.

6. The second biogeophysics module updates soil, snow, and ground temperatures as well as related surface fluxes.

7. The second hydrology module determines changes in TWS-related variables caused by processes such as snow melting and compaction, surface runoff, infiltration into the surface soil layer, and subsurface drainage. Furthermore, the water balance at the end of the model time step is computed.

8. Finally, balances of water, energy, carbon, and nitrogen are checked by contrasting states at the beginning and at the end of the time step against simulated fluxes.

Then, the second driver module writes the diagnostics and model output, and updates accumulated output fields. If selected in the model governing files, initial files and restart files are written. In this thesis, processes simulated within the two hydrology modules are of particular interest and described in the following.

4.1.2 Water Balance

CLM3.5 models several components of the water cycle including interception, throughfall, canopy drip, snow accumulation and melt, water transfer between snow layers, infiltration, surface and sub-surface runoff, water redistribution within the soil, and groundwater recharge. These processes constrain TWS, which is composed by canopy water W_{can} [mm], snow water W_{sno} [mm], soil water $w_{liq,i}$ [kg/m^2] and soil ice $w_{ice,i}$ [kg/m^2] at layers $i = 1...10$, and water in the unconfined aquifer W_{aqu} [mm]. This means, total water storage W can be computed from

$$W = W_{can} + W_{sno} + W_{aqu} + \sum_{i=1}^{10} \left(w_{liq,i} \frac{\rho_{h2o}}{1000} + w_{ice,i} \frac{\rho_{ice}}{1000} \right), \tag{4.1}$$

where ρ_{h2o}=1000 kg/m^3 and ρ_{ice}=917 kg/m^3 are the densities of fresh water and ice, respectively.

At each model step, balance checks ensure the conservation of energy and mass. The water balance is checked by contrasting TWS change during the current time step against the sum of water fluxes arising during the time step, i.e. liquid and solid precipitation, evapotranspiration from vegetation and ground, surface runoff, and sub-surface drainage (Figure 4.1). The following descriptions are based on Oleson et al. (2004) and Oleson et al. (2008).

The advanced snow model of CLM3.5 consists of up to five layers $i = [0, ..., -4]$. The number of snow layers varies according to snow depth (Figure 4.2). The central variables, snow water W_{sno} [mm] and snow depth ΔZ [m], are computed from the layer-dependent variables liquid water $w_{liq,i}$, ice lenses $w_{ice,i}$, and layer thickness Δz_i according to

$$W_{sno} = \sum_{i=0}^{snl+1} \left(w_{liq,i} + w_{ice,i} \right), \tag{4.2}$$

and

$$\Delta Z = \sum_{i=0}^{snl+1} \Delta z_i, \tag{4.3}$$

where *snl* is the negative of the number of snow layers. Ice and water content within each snow layer change due to melting and water flow between the layers. Additionally, in the top layer, snow accumulates due to precipitation and shrinks due to sublimation. Snow layers are subdivided when a prescribed maximum thickness is exceeded. Conversely, layers are

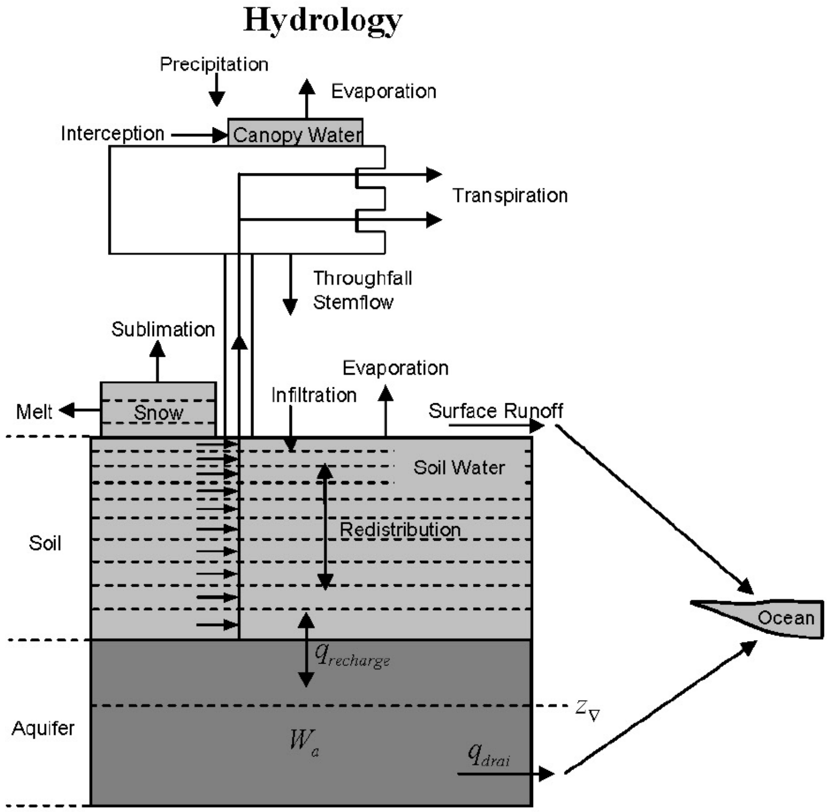


Figure 4.1: Hydrology module of the Community Land Model version 3.5 (CLM3.5; taken from Oleson et al., 2008).

combined if a layer reaches a prescribed minimum thickness after snow compaction, due to crystal breakdown, pressure, or melting. W_{sno} is capped at a maximum of 1 m, exceeding snow water is assigned to runoff. In case of less than 0.01 m depth, snow exists without explicit snow layers. In the case of no existing snow layers but snow depth $\Delta Z \geq 0.01\,\mathrm{m}$, a snow layer is initialized according to:

$$\begin{aligned}
snl &= -1 \\
\Delta z_0 &= \Delta Z \\
z_0 &= -0.5\Delta z_0 \\
z_{h,-1} &= -\Delta z_0 \\
\tau_{sno} &= 0 \\
T_0 &= min(T_f, T_{atm}) \\
w_{ice,0} &= W_{sno} \\
w_{liq,0} &= 0 \\
f_{ice} &= 1,
\end{aligned} \tag{4.4}$$

where τ_{sno} is the snow age. The snow temperature T_0 is the minimum of freezing temperature T_f and atmospheric temperature T_{atm}. The fraction of ice relative to total water f_{ice} equals 1 for a newly initialized snow layer.

Canopy water is controlled by intercepted precipitation relative to the exposed leaf and stem area index, drips off the vegetation, evapotranspiration, and the maximum storage capacity. Precipitation that is not intercepted by canopy as well as melting water from the snow pack are distributed to surface runoff and infiltration into the surface soil layer.

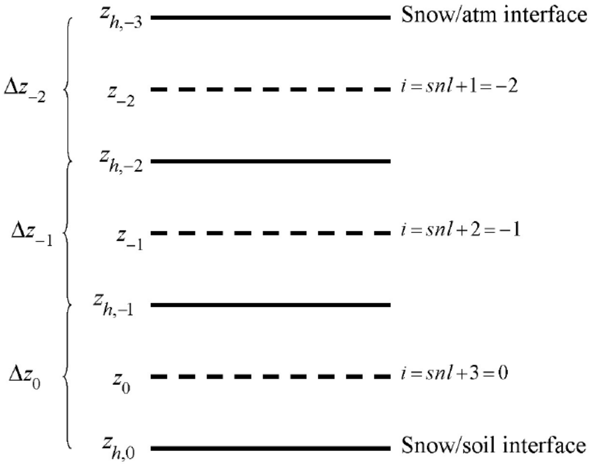


Figure 4.2: The total snow pack is represented in CLM3.5 by up to five layers, here, the three layer case (snl=-3) is shown with i=-2 being the top layer, i=0 being the bottom layer, and snl being the negative of the number of snow layers. For each layer, the node depth z_i and the layer interface depth $z_{h,i}$ depend on the layer thickness Δz_i. The sum of all Δz_i gives the depth of the snow pack ΔZ. (Figure taken from Oleson et al., 2004).

The vertical distribution of soil moisture is governed by infiltration, runoff, gradient diffusion, gravity, and root extraction through canopy transpiration. Soil liquid water $w_{liq,i}$ and soil ice $w_{ice,i}$ are predicted for a 10 layer soil scheme.

Starting from Darcy's law (Darcy, 1856; Whitaker, 1986) and assuming mass conservation, vertical water flow at time t and depth z is computed from a modified Richard's equation (Richards, 1931; Ross, 1990)

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[k\left(\frac{\partial \theta}{\partial z}\frac{\partial \psi}{\partial \theta} + 1\right)\right], \tag{4.5}$$

which is solved numerically by integrating over each layer with boundary conditions of infiltration and drainage at the top and bottom layers, respectively. Hydraulic conductivity k [mm/s] and soil matrix potential ψ [mm] are highly dependent on soil texture, i.e. the percentage of clay and sand (Section 4.1.3.2). The volumetric soil water content θ_i [mm^3 of water / mm^{-3} of soil] is composed by soil liquid water $w_{liq,i}$ and soil ice $w_{ice,i}$ according to

$$\theta_i = \frac{w_{liq,i}}{\Delta z_i \rho_{liq}} + \frac{w_{ice,i}}{\Delta z_i \rho_{ice}}, \tag{4.6}$$

where Δz_i is the thickness of soil layer i. In CLM3.5 a frozen soil scheme was added, which aims at a more accurate representation of runoff and soil water storage (Niu and Yang, 2006). Soil ice $w_{ice,i}$ is computed among other parameters from soil temperature and freezing temperature according to a freezing point depression equation. The volumetric soil water content θ_i and subsequently $w_{liq,i}$ and $w_{ice,i}$ are capped by saturation. A minimum water content for $w_{liq,i}$ of 0.01 mm is ensured by water from lower layers and ultimately from aquifer water and subsurface runoff.

In the case that the water table is below the soil column, water from the bottom soil layer seeps into the unconfined aquifer W_{aqu} following Darcy's law. This process depends on the hydraulic conductivity of the aquifer, on the water table depth, and on the matric potential of the bottom soil layer. The initial amount of water in the unconfined aquifer is assumed to be 4.8 m. Considering an average specific yield of 0.2, a corresponding water table depth of 1 m below the soil column is obtained. When the water table reaches to the bottom of the soil column, the aquifer water is at its maximum of 5 m. In the case that the water table is within the soil column, there is no water flux between the soil column and the underlying aquifer. The liquid water content of soil layers within the water table is then kept between assigned minimum and maximum values by subsurface drainage, i.e. the lateral flow of water out of the soil columns. If the water table is below the soil column, subsurface drainage is removed from the aquifer.

4.1.3 Model Setup

CLM3.5 was set up for the European CORDEX EUR-11 domain illustrated in Figure 4.3. The model is based on a 0.11° artificial non-rotated regular grid, which was transformed to the CORDEX grid for evaluations. The model domain has $436 \times 424 = 184864$ grid cells, from which 99395 are located over land. For all model simulations performed in this thesis, a temporal resolution of one hour was chosen.

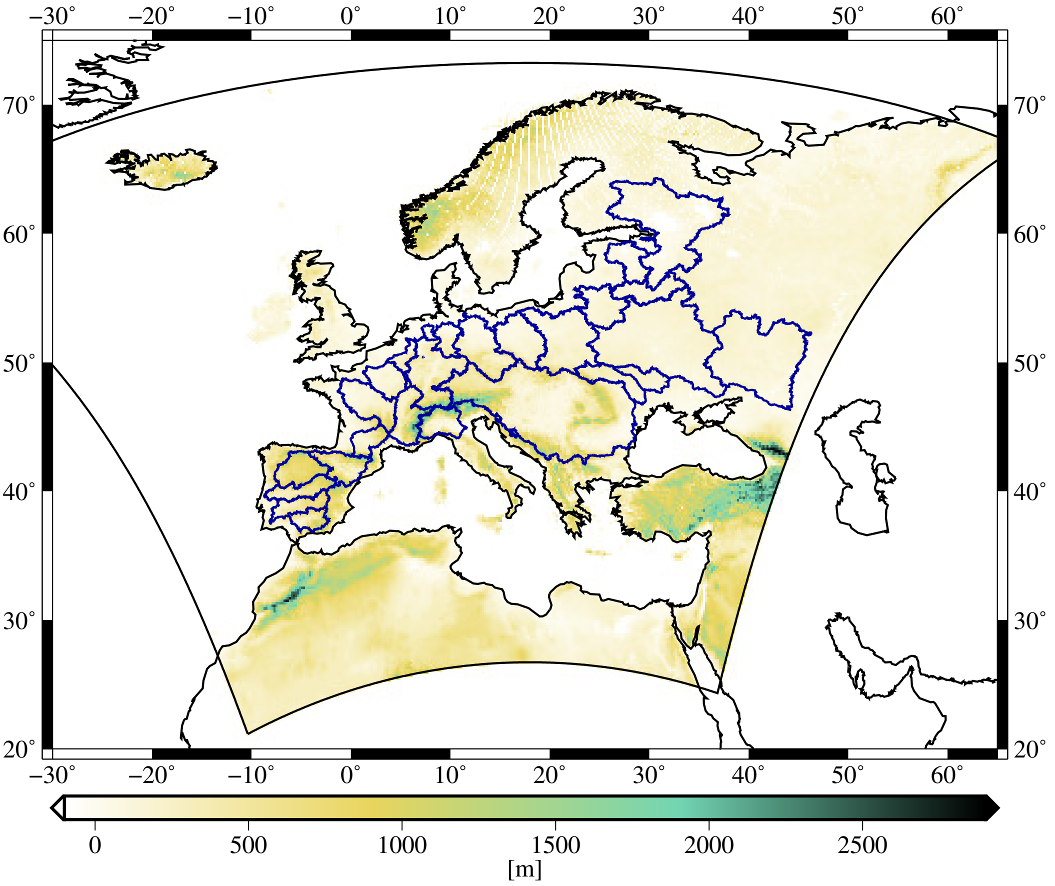


Figure 4.3: Topography over the European COordinated Regional Downscaling EXperiment (CORDEX domain). Major European river catchment are drawn in blue.

4.1.3.1 Atmospheric Forcings

CLM3.5 is forced by temperature, wind speed, specific humidity, surface precipitation, surface pressure, and incident solar and long-wave radiation. High-resolution atmospheric precipitation data are essential as they lead to a more rapid response of the model with more runoff and less infiltration, while coarse resolution precipitation smoothes the model in space and time.

In this thesis, two different atmospheric forcing data sets were used. Weather Research and Forecasting (WRF; Skamarock et al., 2008) version V3.3.1 model output generated by the Centre de Recherche Public - Gabriel Lippmann (CRP-GL) for the European CORDEX area was made available by Jessica Keune[12] for the time span 2002 to 2010 at 3-hourly resolution. The meso-scale non-hydrostatic WRF model has 50 vertical layers and 0.11° horizontal resolution (Vautard et al., 2013; Kotlarski et al., 2013). Data assimilation is realized for the WRF model using the three-dimensional variational (3D-Var) algorithm (Barker et al., 2004). It should be noted, that the WRF model was found to considerably overestimate summer precipitation over Southern Europe and to underestimate summer temperatures for most regions (Kotlarski et al., 2013).

The COnsortium for Small Scale MOdelling REAnalysis at 6 km horizonal resolution (COSMO-REA6) of the CORDEX EUR-11 domain is based on the meso-scale non-hydrostatic COSMO model and has 40 vertical layers and a horizontal resolution of 0.055° ($\sim$6 km) Bollmeyer et al. (2015). A continuous nudging scheme is used for assimilating observations from radiosondes, air planes, vertically pointing wind profilers, surface stations, buoys, and ships. Precipitation is not assimilated but derived internally from the model state variables. Bollmeyer et al. (2015) and Springer et al. (2017) found no significant precipitation biases with respect to observation-based data sets. COSMO-REA6 model output was provided by Bibi Naz[3] for the time span 2002 to 2006 at 6-hourly resolution. As short-wave radiation was not yet available from COSMO-REA6 at the time of computation, the data set was complemented by Princeton Meteorological Forcings at 0.25×0.25 (Sheffield et al., 2006). All data was then regridded to the CLM3.5 resolution. Long-wave radiation was not provided to the model in this case.

The WRF forcings were used for data assimilation experiments with real GRACE data. The COSMO forcings were applied to run a reference model for synthetic data assimilation experiments.

4.1.3.2 Surface Data Sets

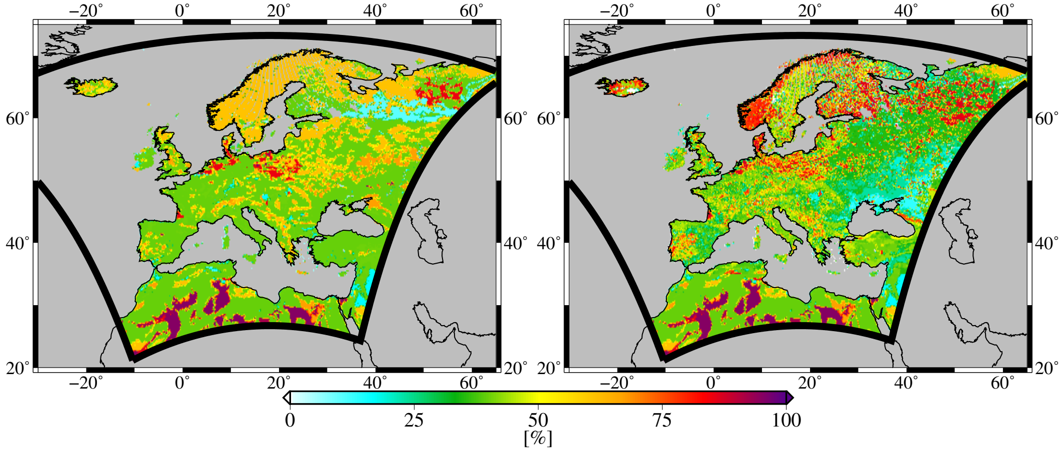
Surface data sets of CLM3.5 are static and include topography, soil properties, PFTs, and physiological vegetation parameters. Elevation data was derived from the 1 km Global Multiresolution Terrain Elevation Data (GMTED2010) data set (Danielson and Gesch, 2011). Two soil texture data sets were applied. Jessica Keune provided percentages of sand and clay based on the Food and Agriculture Organization / United Nations Educational, Scientific and Cultural Organization (FAO/UNESCO) Digital Soil Map of the World (Figure 4.4 a, c). This data set was used for assimilation experiments with real GRACE data. Additionally, maps of clay and sand of higher spatial resolution were derived from the European Soil Data Centre (ESDAC; Hiederer et al., 2013, Figure 4.4 b, d) for a reference run needed for synthetic data assimilation experiments. Soil properties derived from ESDAC have a higher resolution than those from FAO/UNESCO. Differences between the two data sets also exist on larger scales. PFTs are based on land use data from the Moderate Resolution Imaging Spectroradiometer (MODIS; Friedl et al., 2002) as provided by the global CLM3.5 surface data set of Oleson et al. (2008). This surface data set was also used for properties of vegetation, such as leaf and stem area index and top and bottom canopy height.

---

[1] Meteorological Institute, Bonn University, Bonn, Germany

[2] Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium

[3] Institute of Bio- and Geosciences (Agrosphere, IBG-3), Research Centre Jülich, Jülich, Germany


(a) Sand FAO (b) Sand ESDAC

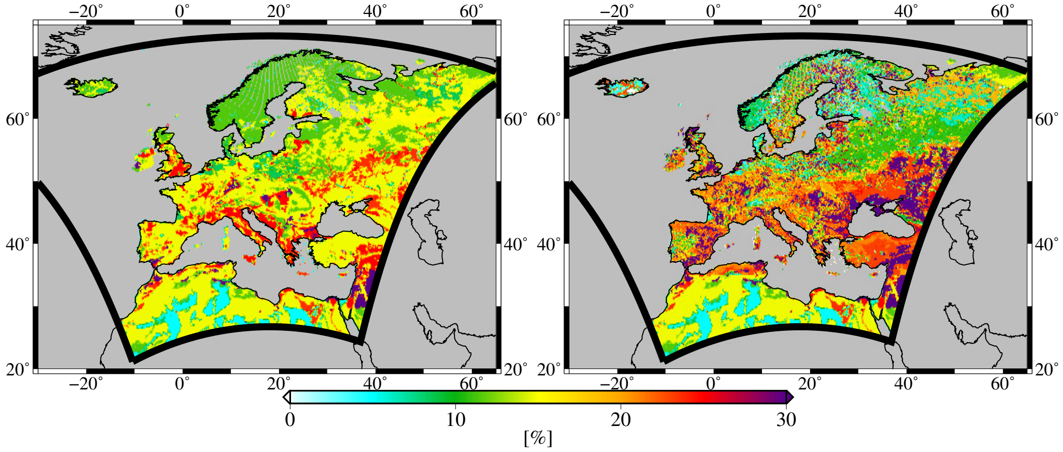

(c) Clay FAO (d) Clay ESDAC

Figure 4.4: Percentage of sand (a, b) and clay (c, d) from the Food and Agriculture Organization / United Nations Educational, Scientific and Cultural Organization (FAO/UNESCO) Digital Soil Map of the World and from the European Soil Data Centre (ESDAC), respectively.

4.2 Global Hydrological Models

4.2.1 WGHM

The conceptual WaterGAP Global Hydrology Model (WGHM; Döll et al., 2003) simulates water flows between all continental water storage compartments except glaciers. Water storage compartments include canopy, snow, soil, groundwater, lakes, man-made reservoirs, wetlands and rivers. Interestingly, WGHM also accounts for human-induced water abstraction from rivers, lakes, reservoirs, and groundwater, which are modeled by an additional groundwater and surface water use model (Döll et al., 2012, 2014a). The model is calibrated against mean annual river discharge from more than thousand gauging stations. Here, daily output and monthly averages at $0.5^\circ \times 0.5^\circ$ are obtained from the version STANDARD of WGHM 2.2 described in Müller Schmied et al. (2014).

4.2.2 LSDM

The Land Surface Discharge Model (LSDM; Dill, 2008) was originally developed for investigating hydrospheric-induced Earth orientation and gravity field parameters. The model was created by combining and extending a simplified land surface scheme and a hydrological discharge model. LSDM enables operational simulations of vertical and lateral water transports and storages on the continents. The model captures all major hydrological processes and models water stored as soil moisture, snow, ice, groundwater, and in wetlands, rivers, lakes, and reservoirs. The model output is provided on a $0.5^\circ \times 0.5^\circ$ grid with daily resolution.

4.2.3 GLDAS Land Surface Models

Four different land surface models from the Global Land Data Assimilation System version 2.0 (GLDAS; Fang et al., 2009) were used in this thesis. GLDAS drives offline land surface models while integrating several observation-based data sets. Data can be downloaded from the Goddard Earth Science Data and Information Services Center (GES DISC). For this thesis, I downloaded monthly averages of (1) the CLM4.0, (2) the MOSAIC land surface model, (3) the Variable Infiltration Capacity (VIC) Macroscale Hydrological Model, and (4) the NOAH land surface model. Additionally, 3-hourly output from NOAH was downloaded and averaged to daily data. All models are available on a $1^\circ \times 1^\circ$ grid. TWS was derived by adding water from all soil levels, snow water equivalent, and canopy water. Groundwater and surface water storages are not part of the models.

CLM4.0 (Lawrence et al., 2011) is the follow-up version of CLM3.5 introduced in Section 4.1. The MOSAIC (Koster and Suarez, 1992, 1996) land surface model is named after the implemented 'mosaic' strategy, which accounts for subgrid heterogeneity of the land surface characteristics. An electrical resistance network analog is applied for calculating energy and water fluxes. The VIC model (Gao et al., 2010) is a large-scale, semi-distributed land surface model with typically three soil layers, where infiltration into the upper layer is controlled by a parameterization of the variable infiltration capacity. The NOAH land surface model (Ek, 2003) is based on diurnal computation of Penman potential evaporation, a multilayer soil model, a canopy model, surface hydrology, and snow and sea ice parameterizations. The four GLDAS land surface models differ significantly with respect to their structure and their performance in simulating the hydrological cycle (e.g.; Zaitchik et al., 2010; Yang et al., 2013; Cai et al., 2014; Scanlon et al., 2018).

4.3 Validation Data Sets

Model states and fluxes from CLM simulations are validated against independent observation-based data sets. In-situ and remotely sensed validation data sets include observations of soil moisture, discharge, and evapotranspiration.

4.3.1 Soil Moisture

Remotely sensed soil moisture achieves near-global coverage, but it is only sensitive to the first few centimeters of the soil (Dorigo et al., 2011). In contrast, most in-situ soil moisture sensors measure at least down to 30 cm depth and sometimes even reach a depth of 1.5 m. However, only few in-situ soil moisture stations are available over Europe during the study period. Furthermore, in-situ measurements are only representative for one specific location, which makes it difficult to validate models with grid cells of several kilometers. Thus, in-situ observations and remotely sensed soil moisture complement each other to a certain extent.

4.3.1.1 In-Situ Measurements from the International Soil Moisture Network

Harmonized in-situ soil moisture measurements are available via the International Soil Moisture Network (ISMN; Dorigo et al., 2011, 2013). The ISMN provides quality checked volumetric soil moisture observations, measured at different depths by 59 soil moisture networks world wide. The data base is continuously being extended. Volumetric soil moisture is expressed as the volumetric fraction of water within a given soil layer (m^3 water per m^3 soil). Many stations provide time series at hourly resolution, yet, some stations only provide few observations per month. Measurements are performed at depths between 2 cm and 1.5 m. Furthermore, for some networks measurements at more than 10 depths are available whereas other networks only measure at one single depth.

Here, data from 67 stations organized within 17 European networks were evaluated. For validation, the time series were aggregated to monthly time steps. Stations with less than two years of data within the study period from 2003 to 2010 were excluded. About 90 % of the stations were available for more than 3 years and 40 % of the stations were available for more than 5 years. Only stations that provide measurements down to a depth of at least 5 cm were considered in order to validate soil moisture variability beyond surface soil moisture. Largest networks used here are the networks REMEDHUS (orange) with 20 suitable stations, SMOSMANIA (red) with 11 suitable stations, and WEGENERNET (yellow) with 12 stations (Figure 4.5). REMEDHUS is located about 300 km west of Madrid and operated by the University of Salamanca. Soil moisture is measured between 0 cm and 5 cm depth. SMOSMANIA is located in South-Western France close to the border to Spain and stations measure down to a depth of 30 cm. Finally, WEGENERNET is located in South-Eastern Austria and stations measure at depths of 20 cm and 30 cm.

4.3.1.2 Remote Sensing from ESA CCI

Daily surface soil moisture at 0.25° resolution is provided by the European Space Agency (ESA) Climate Change Initiative (CCI). The product v03.2 was retrieved from multiple remotely operating sensors (Dorigo et al., 2017). Between 2003 and 2010, passive microwave observations were obtained from DMSP SSM/I, TRMM TMI, AQUA AMSR-E, Coriolis WindSat, and SMOS MIRAS, and active observations from C-band scatterometers of ERS1/2 and

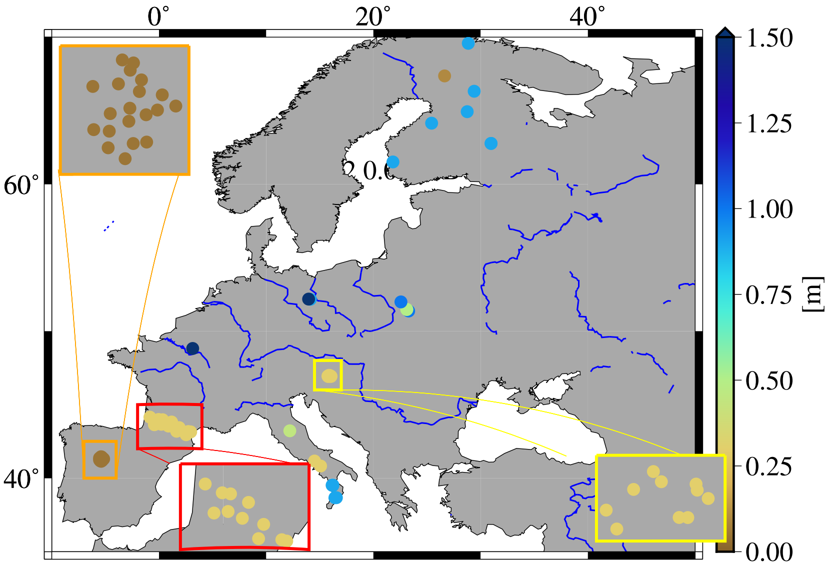


Figure 4.5: Maximum depth [m] of soil moisture stations of the International Soil Moisture Network (ISMN).

MetOp-A/B. Absolute soil moisture was derived by rescaling the retrievals against modeled soil moisture from GLDAS-NOAH using cumulative density function matching. Here, the merged product from active and passive sensors was used as it performs better than each of the individual data sets (Liu et al., 2011). The data was resampled to the CLM grid by nearest neighbor interpolation. The data coverage declines in northern latitudes and for mountainous areas due to snow coverage during winter and spring (Figure 4.6).

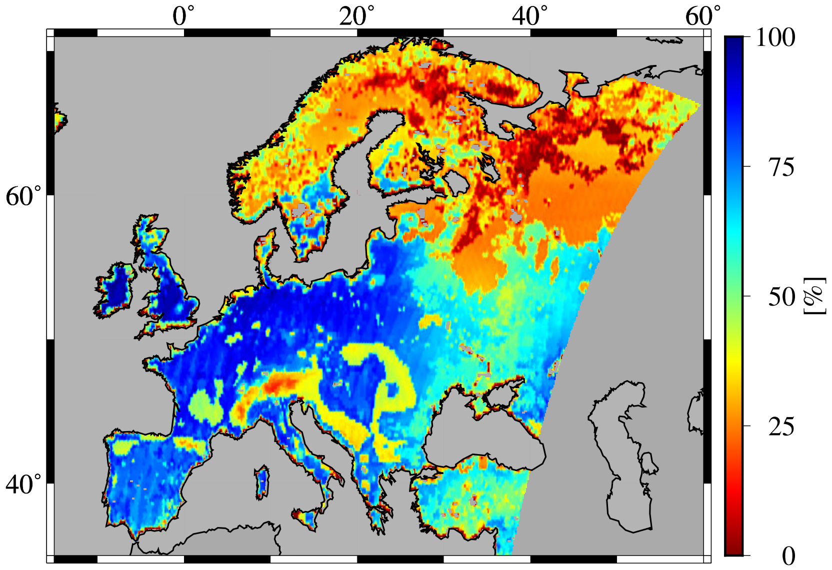


Figure 4.6: Availability of daily soil moisture data from the European Space Agency (ESA) Climate Change Initiative (CCI).

4.3.2 Discharge

Monthly discharge data from the Global Runoff Data Centre (GRDC) is available for restricted time periods only (Figure 4.7). Here, monthly discharge of 26 major European rivers was evaluated at the most downstream gauging station of each catchment. Mereley 10 of them provide data within the study period 2003 to 2010.

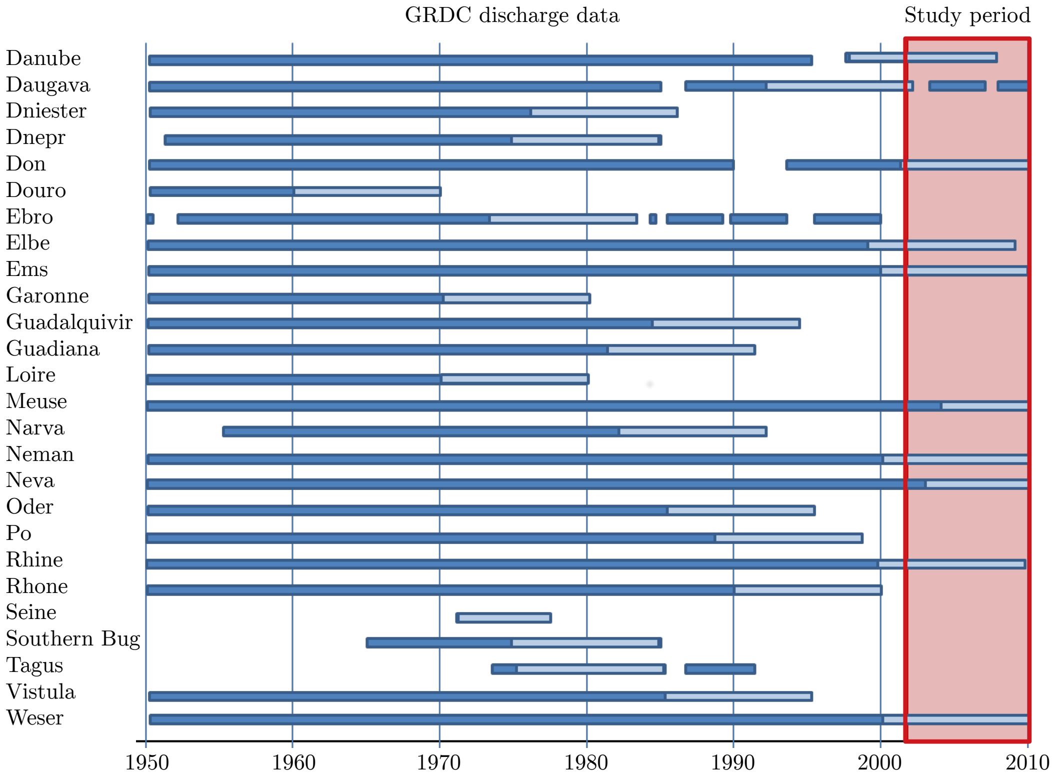


Figure 4.7: The availability of discharge observations from the Global Runoff Data Centre (GRDC) for the most downstream gauging station of each river is indicated in dark blue (modified from Springer et al., 2017). Light blue indicates the calibration period of the lumped rainfall-runoff model GR2M-snow.

Approaches for extending discharge time series in time include satellite altimetry, runoff-precipitation ratios, and runoff-storage relationships. Remote sensing is no option over Europe so far, as rivers are too small. Here, discharge time series were extended to the study period 2003 to 2010 by calibrating the lumped rainfall-runoff model GR2M (Mouelhi, 2003). In a previous study, this model was successfully employed for closing the terrestrial water budget equation over Europe (Springer et al., 2017). Here, the same settings were applied.

The GR2M is a two-parameter model that requires monthly precipitation and potential evapotranspiration as input. Two stores are included, a production store of variable size and a routing store with a fixed capacity of 60 mm. Calibration parameters are the capacity of the production storage and an exchange coefficient, which accounts for the exchange of water with the outside of the basin. Here, GR2M was extended by a distributed Hydrologiska Byrans Vattenavdelning (HBV) type snow model, which requires gridded temperature as input and adds three more calibration parameters (melting temperature, melting coefficient, and temperature separating rain and snow). The precipitation and temperature data sets were obtained from the European daily high-resolution gridded dataset (Haylock et al., 2008), and potential evapotranspiration was obtained from the Climate Research Unit (CRU) data set TS3.22 (Harris et al., 2014).

GR2M-snow was calibrated against GRDC data using the 10 most recent continuous years available (marked in light blue in Figure 4.7) applying the least squares approach. GRDC data was found to be erroneous for Neman and Vistula, which were instead calibrated against the observational gridded runoff estimates for Europe (E-RUN; Gudmundsson and Seneviratne, 2015, 2016). After calibration, the model was used to simulate discharge for the period 1959 to 2010. Limitation of this approach include changing river dynamics e.g. due to construction.

4.3.3 Evapotranspiration

Acquisition of observation-based evapotranspiration data is challenging (Wang and Dickinson, 2012). Evapotranspiration is directly related to latent heat flux Q_L via $E = \lambda^{-1} Q_L$, with the latent heat of vaporization for water $\lambda = 2.5 \times 10^3 kJkg^{-1}$. Over Europe about 200 FLUXNET towers adopt the eddy covariance method to assess variables such as latent heat flux. Here, a global data set of upscaled FLUXNET observations was used. The data set is provided by Jung et al. (2011) at the Max Planck Institute (MPI) for Biogeochemistry in Jena. They derived monthly gridded latent heat flux at a resolution of 0.5°×0.5° from observations of FLUXNET towers together with meteorological and remote sensing observations applying a machine learning approach.

Chapter 5

Processing of GRACE Data

Gravity Recovery and Climate Experiment (GRACE) level 2 and level 1B data were processed for two different purposes. (1) ITSG-Grace2016 (Section 2.3, Mayer-Gürr et al., 2016) level 2 solutions solved up to degree and order 90 served as observational input for the data assimilation framework that was set up within this thesis. Therefore, global gravity field solutions provided as spherical harmonic coefficients were converted into gridded total water storage anomalies (TWSA, i.e. the deviation of total water storage TWS from its temporal mean), as described in Section 5.1. The derivation of TWSA was accompanied by a thorough error assessment, which is fundamental in the context of data assimilation. (2) GRACE level 1B data was used for computing K-band range-rate (KBRR) and K-band range-acceleration (KBRA) residuals defined as the difference between measured and modeled range-rates or range-accelerations (Section 5.2). K-band residuals bypass the downward continuation and filtering steps required for validating hydrological models, which are based on gridded maps. Here, K-band residuals were computed (i) for assessing temporally high-frequent hydrological mass variations from global hydrological models, and (ii) for validating the performance of GRACE-assimilating hydrological model output in comparison to output from an open-loop run.

5.1 Computing Gridded Total Water Storage Anomalies from GRACE Level-2 Data

GRACE level 2 data products are generally provided as sets of spherical harmonic (SH) coefficients, so-called gravity coefficients or Stokes coefficients (see Chapter 2). The mathematical concept of gravity potential and its representation using spherical harmonics is introduced in Section 5.1.1.1 and Section 5.1.1.2, respectively. Gridded TWSA were derived from Stokes coefficients using the approach of Wahr et al. (1998) (Section 5.1.1.3).

Due to the measurement configuration of the GRACE mission an elaborate post-processing strategy has to be applied to the SH coefficients (Figure 5.1). Low degree SH coefficients cannot or only with limited quality be measured by the GRACE satellites (Section 5.1.2). Therefore, the degree-1 coefficients and the c_{20} coefficient were replaced by external data sets. As most processing centers provide unconstrained solutions, decorrelation and filtering is necessary in order to remove correlated noise (Section 5.1.3.2). As a consequence, signal is attenuated and needs to be restored (Section 5.1.3.3), which was realized here by applying rescaling factors derived from an ensemble of hydrological models. Furthermore, the impact from glacial isostatic adjustment (GIA) needs to be removed for hydrological applications

(Section 5.1.4).

In this thesis, the uncertainty characteristics of the GRACE mission are of major importance, as they control the weight of the observations in the data assimilation process. In Section 5.1.5 error estimates from different processing centers are discussed and the error propagation from SH coefficients to gridded TWSA is described.

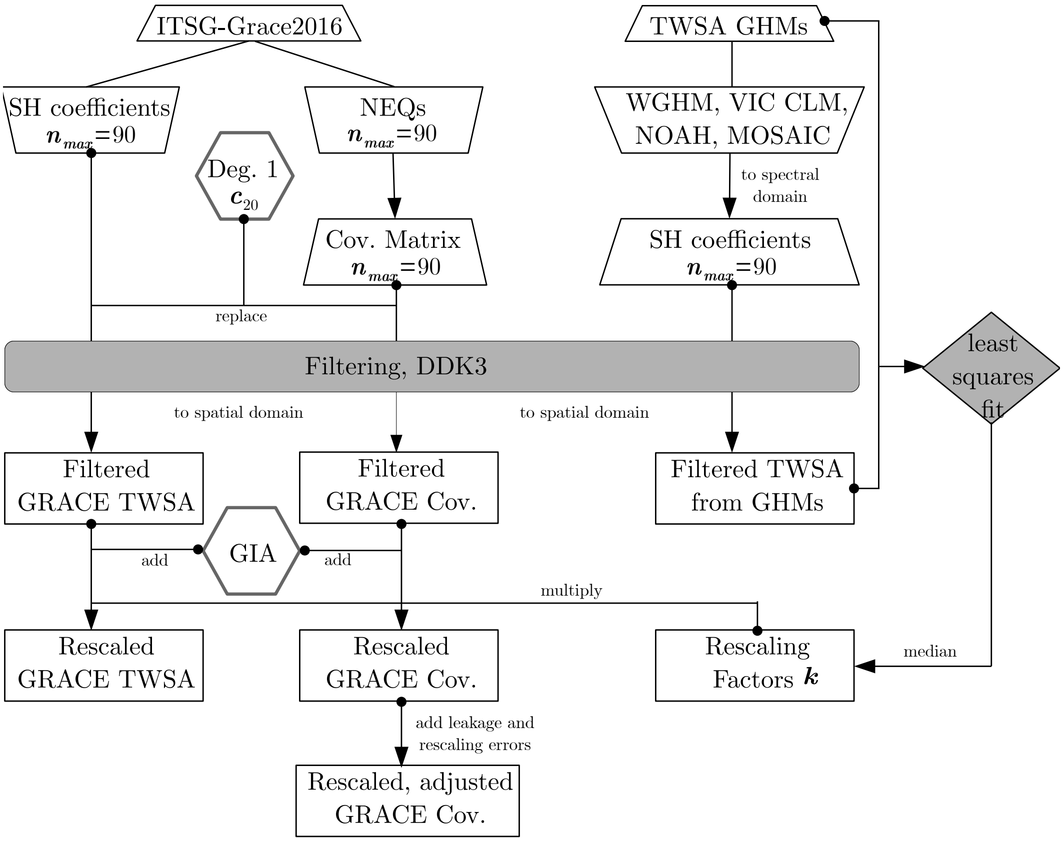


Figure 5.1: This flow chart illustrates the processing steps implemented for deriving gridded total water storage anomalies (TWSA) from spherical harmonic (SH) coeffcients including a thorough error propagation.

5.1.1 From Gravity Potential to Total Water Storage Anomalies

The mathematical description of the Earth's gravity potential and its representation using SH functions follows Hofmann-Wellenhof and Moritz (2006). The formulation of the conversion from Stokes coefficients to gridded mass estimates is adopted from Wahr et al. (1998).

5.1.1.1 The Earth's Gravity Field

Newton's law of gravitation states that two point masses M and M' attract each other along their connection line with the force

$$\mathbf{F} = GMM'\frac{\mathbf{r}-\mathbf{r}'}{r^3}, \tag{5.1}$$

where $r = |\mathbf{r} - \mathbf{r}'|$ is the distance between the two masses and $G = 6.67428 \times 10^{-11} m^3\, kg\, s^{-2}$ is Newton's gravitational constant. As gravity is a conservative force field, the force vector $\mathbf{F}$ can be derived as the gradient of the gravitational potential V,

$$\mathbf{F} = \nabla V. \tag{5.2}$$

When defining M as attracting mass and M' as an attracted uniform mass, the gravitational potential induced by mass M can be written as

$$V = \frac{GM}{r}. \tag{5.3}$$

According to the superposition principle, the potential of several point masses is the sum of the individual contributions. Consequently, the gravitational potential of a solid body can be described by Newton's integral

$$V = G \iiint\limits_{v} \frac{dm}{r}, \tag{5.4}$$

where v is the volume of the solid body and $dm = \rho\, dv$ is a mass element with density ρ. The gravitational potential is continuous and differentiable outside of the solid body and vanishes for $r \to \infty$. Outside of the attracting masses, the Earth's gravitational potential V satisfies Laplace's equation

$$\Delta V = 0. \tag{5.5}$$

This means, the gravitational potential is harmonic in the exterior space, i.e. it is source free.

5.1.1.2 Spherical Harmonics

The solutions of Laplace's equation (Eq. (5.5)) are harmonic functions. Classically, the gravitational potential is expressed as series of spherical harmonics

$$V(\lambda, \theta, r) = \frac{GM}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{(n+1)} Y_n(\lambda, \theta), \tag{5.6}$$

with the Earth's mean equatorial radius R and the Earth's mass M. The spherical geocentric position is denoted by longitude λ, co-latitude θ, and radius r, which is defined as the distance from the origin of an Earth-fixed coordinate system (with $r >= R$ as the solution is only defined in the exterior space). The factor $(R/r)^{(n+1)}$ is 1 at the Earth's surface and smaller than 1 in the outer space, and models the upward continuation of the gravitational potential, which attenuates with increasing distance from the Earth's origin. The surface spherical harmonics $Y_n(\lambda, \theta)$ of degree n define a complete set of orthogonal functions in a system of spherical coordinates and have the form

$$Y_n(\lambda, \theta) = \sum_{m=0}^{n} \left(c_{nm} \cos(m\lambda) C_{nm}(\cos\theta) + s_{nm} \sin(m\lambda) S_{nm}(\cos\theta)\right). \tag{5.7}$$

In Eq. (5.7), C_{nm} and S_{nm} denote the surface spherical harmonics of degree n and order m. The corresponding unit less spherical harmonic (SH) coefficients c_{nm} and s_{nm} are the so-called Stokes' coefficients. The surface spherical harmonics are derived according to

$$\left\{ \begin{array}{c} C_{nm}(\lambda, \theta) \\ S_{nm}(\lambda, \theta) \end{array} \right\} = P_{nm}(\cos\theta) \left\{ \begin{array}{c} \cos(m\lambda) \\ \sin(m\lambda) \end{array} \right\}, \tag{5.8}$$

where $P_{nm}(\cos\theta)$ are the fully normalized Legendre functions computed using a stable recursion formula.

5.1.1.3 Total Water Storage Anomalies

Variations of the gravitational potential are due to mass redistribution. The derivation of mass redistribution is an ambiguous problem as different mass distributions can generate the same gravitational potential. However, mass redistribution can be assumed to happen within a thin layer at the Earth's surface. Thus, Wahr et al. (1998) defined surface density changes $\Delta\sigma$ as the radial integral of the density redistribution within this thin layer. The integral is related to the gravitational potential by an expansion into spherical harmonics, according to

$$\Delta\sigma(\lambda,\theta) = \frac{M}{4\pi R^2} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{2n+1}{1+k_n'} \left(\Delta c_{nm}\cos(m\lambda)C_{nm}(\cos\theta) + \Delta s_{nm}\sin(m\lambda)S_{nm}(\cos\theta)\right). \tag{5.9}$$

The coefficients Δc_{nm} and Δs_{nm} represent the anomalies of the Stokes coefficients c_{nm} and s_{nm} with respect to a static reference field. Besides direct gravitational changes, surface mass redistribution also induces deformation of the solid Earth, which indirectly causes changes in gravity. This indirect effect is taken into account via the dimensionless degree-dependent Love numbers k_n'. As all kind of Love numbers, the gravitational load Love numbers are derived from Earth models, which represent the average Earth properties (Farrell, 1972). For hydrological applications the change in surface density is commonly expressed in equivalent water height E

$$\Delta E(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{2n+1}{1+k_n'} \left(\Delta c_{nm}\cos(m\lambda)C_{nm}(\cos\theta) + \Delta s_{nm}\sin(m\lambda)S_{nm}(\cos\theta)\right), \tag{5.10}$$

where $\rho_w = 1025 kg/m^3$ is the average density of sea water. Eq. (5.10) represents the central formula for deriving gridded total water storage anomalies (TWSA) from Stokes coefficients. However, beforehand a thorough post-processing of the GRACE-derived Stokes coefficients is required, which is described below.

5.1.2 Coefficients of Lower Degree

Stokes coefficients of degree 0, 1, and 2 are directly related to the shape of the Earth (Hofmann-Wellenhof and Moritz, 2006). The coefficient c_{00} is connected to the mass of the Earth's system. Thus, c_{00} is assumed to be a constant, which means that Δc_{00} is zero. In contrast, the degree 1 and 2 coefficients need special attention when evaluating GRACE derived gravity fields (Chen et al., 2015).

5.1.2.1 Δc_{10}, Δc_{11}, Δs_{11}

The center of mass (CM) of the whole Earth system is defined as the geocenter, and moves with respect to the center of figure (CF) of the solid Earth. The geocenter motion is linked to changes in the degree-1 coefficients Δc_{10}, Δc_{11}, Δs_{11}. The two GRACE satellites orbit the CM and, furthermore, the origin of the reference frame of the gravity field solutions is the CM.

Hence, the degree-1 Stokes coefficients derived from GRACE are zero by definition. This is problematic when using GRACE together with hydrological models, as these models simulate changes of water with respect to the CF. Significant terrestrial water storage changes related to degree-1 coefficients were found by Meyrath et al. (2017).

From the above follows that degree-1 coefficients have to be added in the GRACE gravity solutions using external time series. Different approaches exist for deriving time-variable degree-1 coefficients. Satellite laser ranging (SLR) measures unambiguous ranges to geodetic satellites, from which variations of the vector between the CM and the CF can be obtained (Cheng et al., 2013b). Determining geocenter motion from Global Positioning System (GPS) observations suffers from the inhomogeneous distribution of GPS stations with lacks in the oceans and remote land locations (Zhang and Jin, 2014). Another approach is the combination of GRACE with ocean model output, which leads to estimates of the degree-1 coeffcients that are consistent with the other coefficients (Swenson et al., 2008; Sun et al., 2017). Finally, degree-1 coeffcients can also be estimated from an inversion approach using GPS, GRACE, and ocean bottom pressure, as implemented by Rietbroek et al. (2012).

Amplitude and phase of degree-1 time series from different approaches were compared e.g., by Sun et al. (2017). Accordingly, c_{10} from Swenson et al. (2008) has the smallest annual amplitude with 1.9 mm. The optimized combination of GRACE and ocean bottom pressure by Sun et al. (2017) yields a larger amplitude of about 3 mm. This is in good agreement with the inversion approach of Rietbroek et al. (2012), whcih leads to amplitudes between 2.2 mm and 3.5 mm. Largest annual amplitudes are obtained from SLR with 4.2 mm (Cheng et al., 2013b). Annual amplitudes of c_{11} vary between 1.9 mm (Swenson et al., 2008) and 2.9 mm (Cheng et al., 2013b). The annual amplitude of s_{11} is estimated to values between 2.5 mm (Swenson et al., 2008) to 3.4 mm (Rietbroek et al., 2012). In this thesis, the time series of geocenter motion and corresponding error information from Rietbroek et al. (2012) was used.

5.1.2.2 Δc_{20}

Degree 2 coefficients are related to the Earth's inertia tensor, and the zonal c_{20} coefficient is directly related to the Earth's dynamic oblateness. Due to the low orbits and the short baseline between the two GRACE satellites, Δc_{20} is corrupted by aliasing effects, e.g., from the 161-day-period ocean tide (Cheng et al., 2013a). Therefore, for hydrological applications c_{20} is replaced by a time series of c_{20} coefficients from SLR. In this thesis the data set from Cheng et al. (2013a) was applied.

5.1.3 Removing Correlated Errors

GRACE level 2 Stokes coeffcients contain errors due to (i) instrument noise, (ii) non-isotropic spatial sampling of the orbit, and (iii) temporal aliasing caused by imperfect background models for short-term mass variations (Schmidt et al., 2007). Instrument noise includes noise from the K-band ranging system, accelerometer errors, and orbit errors. The K-band measurements are the primary observable and sensitive in along-track direction. Thus, due to the near-polar orbit, GRACE measurements are highly sensitive in north-south direction and have limited sensitivity in east-west direction. This property, together with the temporal sampling and mismodeled short-term mass variations, leads to temporal aliasing, which manifests itself by north-south stripes in the gridded TWSA. The three mentioned types of errors are summarized as measurement errors. In the following, the characteristics of measurement errors and their handling are discussed.

5.1.3.1 Uncertainty Characteristics of the SH Coefficients

The Earth's gravitational potential decreases with increasing distance from its source, which is expressed by the factor $(R/r)^{(n+1)}$ in Eq. (5.10). This factor is degree-dependent and induces stronger damping for high-degree coefficients, which are related to the shorter wavelengths. Correspondingly, measurement errors are amplified and cause large errors for the high-degree coefficients. For this reason, the gravity field recovery is limited to degree and order (d/o) n_{max}, which is typically between 40 and 120, and leads to

$$\Delta E(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w} \sum_{n=0}^{n_{max}} \sum_{m=0}^{n} \frac{2n+1}{1+k'_n} \left(\Delta c_{nm} \cos(m\lambda) C_{nm}(\cos\theta) + \Delta s_{nm} \sin(m\lambda) S_{nm}(\cos\theta)\right). \tag{5.11}$$

The truncation of Eq. (5.11) neglects signal from degrees $n > n_{max}$, which results into omission errors smaller than 5 mm for $n_{max} = 90$ (Gunter et al., 2006). The contribution of each degree is described by the squared sum of all coefficients of degree n, the so-called degree variances

$$\sigma_n^2 = \sum_{m=0}^{n} (c_{nm}^2 + s_{nm}^2). \tag{5.12}$$

The square root of degree variances of the high-resolution static global gravity field model GOCO05s (Mayer-Guerr, 2015) is displayed in terms of equivalent water height in mm by the gray line in Figure 5.2. The signal content decreases with increasing d/o. The signal of the time-varying gravity field of the ITSG solution with respect to the static field (GOCO05s) is shown by the black line for January 2007. The difference in signal content with respect to the static field increases for shorter wavelengths. Eq. (5.12) can also be applied to the error spectrum. Each processing center provides together with the Stokes coefficients formal errors. These are over-optimistic as they result from error propagation of observation and background models errors, which are both not well-known. For a more realistic error estimate, Schmidt et al. (2007) applied a degree-dependent scaling factor in order to match certain characteristics of the gravity field. The resulting calibrated errors differ from the formal errors by a factor of about 3.5 for the GFZ-RL05 solution (Figure 5.2, green lines).

Formal errors derived for the ITSG solutions are more realistic due to an empirical noise model based on K-Band residual analysis. They differ from the calibrated errors of GFZ-RL05 only by a factor of 1.5 for solutions up to d/o 60 (yellow line). For higher degree coefficients, the difference between calibrated errors and the ITSG solutions is larger. Furthermore, Figure 5.2 shows that the errors of the ITSG solutions for d/o 60, 90, and 120 differ for the high degree coefficients. These differences decrease when the gravity field solutions are smoothed as described in Section 5.1.3.2. It should be kept in mind that the degree variances only assess the main diagonal of the covariance matrix of the SH coefficients. In Section 5.1.5 the error analysis is extended to the spatial domain.

5.1.3.2 Filtering and Decorrelation

Filters, which aim at the removal of the correlated error patterns, have become more and more sophisticated. Generally, the filters are described by a filter matrix $\mathbf{W}$, which contains weighting factors for each Stokes coefficient. Isotropic filters (e.g., the Gaussian filter) are only degree-dependent and, thus, independent of direction in spatial domain. In contrast, anisotropic filters, which depend on degree and order, are location dependent in spatial domain

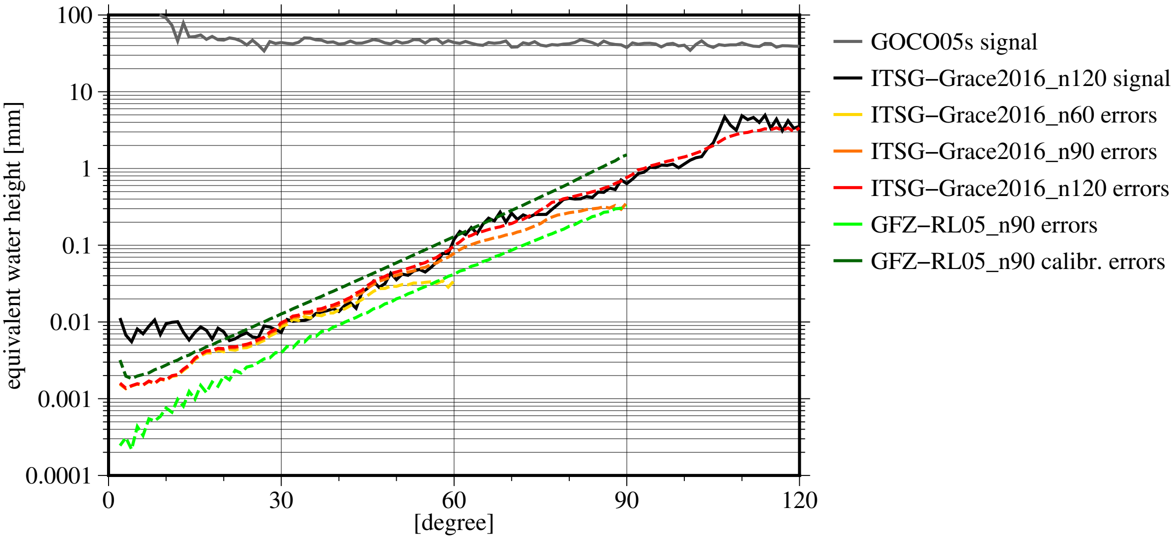


Figure 5.2: Square root of degree variances of the static field (gray) and of the ITSG-Grace2016 solution for January 2007 (black) with respect to the static field. Additionally, the square root of degree variances of different error models is shown (dashed). The degree variances are shown in terms of equivalent water height given in mm.

and can be designed specifically for removing the north-south striping patterns. The filter matrix $\mathbf{W}$ is directly applied to the SH coefficients arranged in the vector $\mathbf{x}$ according to

$$\mathbf{x}^f = \mathbf{W}\mathbf{x} = \mathbf{W}\begin{pmatrix} c_{nm} \\ s_{nm} \end{pmatrix}, \tag{5.13}$$

with $\mathbf{x}^f$ denoting the vector of filtered SH coefficients. Accordingly, Eq. (5.11) is modified to

$$\Delta E(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w} \sum_{n=0}^{n_{max}} \sum_{m=0}^{n} \frac{2n+1}{1+k'_n} \left(\Delta c^f_{nm} \cos(m\lambda) C_{nm}(\cos\theta) + \Delta s^f_{nm} \sin(m\lambda) S_{nm}(\cos\theta)\right), \tag{5.14}$$

with c^f_{nm} and s^f_{nm} denoting the filtered SH coefficients.

An example for isotropic filters is the Gaussian filter proposed by Jekeli (1981). The gravity signal (Figure 5.3 a) is convoluted with a normalized bell-shaped Gaussian weighting function. The filter width is usually chosen between 300 km and 1000 km and defines the radius, at which the weighting function fades to half of its maximum. As the Gaussian filter is independent of orientation the removal of north-south striping patterns often requires very large filter radii, which causes strong signal attenuation (Figure 5.3 b, Chen et al., 2005).

Swenson and Wahr (2006) introduced an empirical decorrelation step before applying the Gaussian filter. They analyzed the correlation patterns of the SH coefficients and found apparent correlations between coefficients of the same order for even and odd degrees, separately. Subtracting a quadratic polynomial fitted to the coefficients of same parity significantly reduces the stripes in the gravity field solutions. Persisting stripes at the equator were removed by applying a Gaussian filter after the polynomial filter. The empirical decorrelation method was modified and refined by ,e.g., Duan et al. (2009), Chambers and Bonin (2012), and Bonin et al. (2012).

Alternatively, correlated noise can be removed by separating signal and noise using statistical decomposition techniques like principle component analysis (Schrama et al., 2007) or independent component analysis (Frappart et al., 2011). Recently, Wang et al. (2016) developed a new stochastical decorrelation method which simultaneously estimates signal and correlated noise in a Bayesian framework.

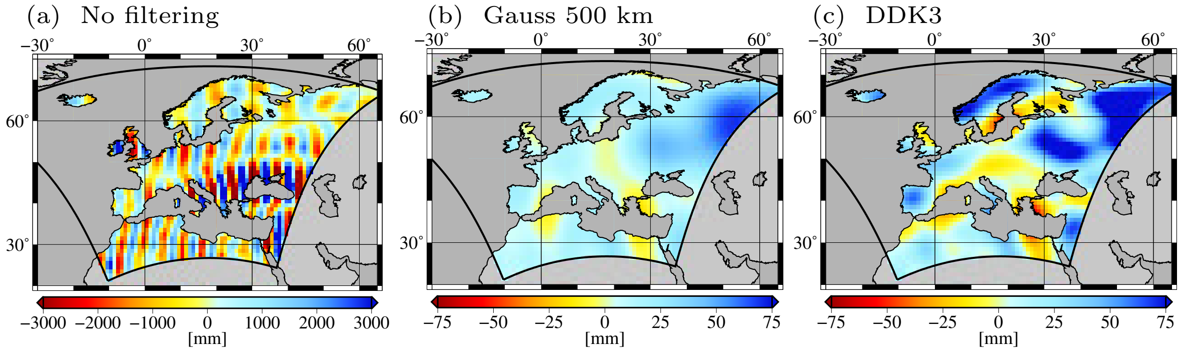


Figure 5.3: TWSA of ITSG-Grace2016 (solution up to degree n_{max}=90) in January 2007 (a) for the original solution, (b) after applying a Gaussian filter with 500 km radius, (c) after applying the DDK3 filter.

In this study, the anisotropic DDK filter, suggested by Kusche (2007), was applied. The DDK filter combines decorrelation and filtering (Figure 5.3 c). The filter is based on a simplified GRACE error covariance matrix and imitates the regularization of a set of monthly GRACE normal equations. The filter strength is related to the regularization parameter, which defines the filters DDK1 to DDK5, where DDK1 implies strongest smoothing and DDK5 weakest smoothing. In spatial domain the filter is latitude-dependent with stronger impact in east-west direction than in north-south direction. Decorrelation is achieved through negative side lobes. In Figure 5.3 the performance of the Gaussian filter with 500 km radius (Figure 5.3 b) is compared to the DDK3 filter (Figure 5.3 c), which is used in this study. Both filters successfully remove the stripe,s which dominate the gravity field solution (Figure 5.3 a). However, the chosen Gaussian filter causes stronger attenuation of the gravity signal than the DDK3 filter. The impact of different filters on GRACE solutions was assessed by, e.g., Klees et al. (2007), Werth et al. (2009), and Kusche et al. (2009b).

5.1.3.3 Compensation of Signal Attenuation

Both, truncation of the spherical harmonics to degree n_{max} (Section 5.1.3.1) and filtering (Section 5.1.3.2) distort the gravity signal (Klees et al., 2007). Depending on the mass distribution and the filter strength, signal is transported out of the target area (leakage-out) or leaks into the target area (leakage-in). The leakage effect is particularly large for (i) strong filters, (ii) small target areas, (iii) differing mass distribution outside and inside of the target region, and (iv) near the coast.

The evaluation of GRACE data in conjunction with hydrological models requires consistent time series of TWSA. One straightforward way to harmonize GRACE and model data is to filter both data sets in the same way. This approach is applied in several studies that validate hydrological models against measured TWSA, e.g., Schmidt et al. (2006), Güntner (2008), Xie et al. (2012), and Felfelani et al. (2017). However, in the context of data assimilation, it is beneficial to process GRACE data in a way that it can directly be related to the model outputs. Thus, signal loss due to truncation and filtering is restored by rescaling either basin averaged TWSA, which is the case occuring more often, or as in this study by rescaling gridded TWSA.

Long et al. (2015) assessed the skill and uncertainties of three different rescaling approaches.

The first approach can be applied to both, individual grid cells and river basins, whereas the second and third approach require basin-averaged TWSA. (1) The scaling factor approach computes multiplicative scaling factors by fitting filtered and unfiltered TWSA from a hydrological model in a least squares sense. It is the most widely used approach (Fenoglio-Marc et al., 2006; Swenson and Wahr, 2007; Famiglietti et al., 2011; Landerer and Swenson, 2012; Long et al., 2015; Zhang et al., 2016). (2) For the additive correction approach leakage-in and leakage-out is computed separately from the filtered and the unfiltered basin function and from a hydrological model. Subsequently, signal loss due to leakage-out is added to the filtered GRACE solutions and additional signal due to leakage-in is subtracted (Klees et al., 2007; Longuevergne et al., 2010). (3) The multiplicative correction approach applies a multiplicative factor to the difference between filtered GRACE signal and the signal leaking into the basin. The factor is derived from the amplitude of the filtered and unfiltered basin function assuming a uniform distribution of TWS within the basin. Long et al. (2015) summarized that the three approaches show major differences in arid and semiarid regions and for relatively small basins.

In this thesis, the scaling factor approach was applied as it is the only approach suited for gridded TWSA. Landerer and Swenson (2012) used TWS estimates from the Global Land Data Assimilation System (GLDAS) Noah Land Surface Model (NOAH) to compute scaling factors for GRACE data truncated at degree 60 and smoothed with a Gaussian filter of 300 km radius for a 1°×1° grid. Here, the rescaling strategy of Landerer and Swenson (2012) was adopted for computing rescaling factors for the ITSG solutions of degree 90. Landerer and Swenson (2012) used outputs from one single hydrological model and later experiments by Long et al. (2015) and Zhang et al. (2016) showed that over Europe the impact from different hydrological models is rather small (see also Section 5.1.5, Figure 5.8 a). Nevertheless, in this study the rescaling factors were computed from five different global hydrological models in order to guarantee transferability to other study regions and for obtaining a robust estimate. The following five models were evaluated at monthly resolution: (1) the GLDAS Community Land Model version 4.0 (CLM4.0), (2) the GLDAS MOSAIC land surface model, (3) the GLDAS Variable Infiltration Capacity (VIC) Macroscale Hydrological Model, (4) the GLDAS NOAH land surface model, and (5) the WaterGAP Global Hydrology Model (WGHM) (see Section 4.2 for details on the models). The GLDAS models were downloaded from the Goddard Earth Science Data and Information Services Center (GES DISC).

First, TWSA were computed for each hydrological model from the individual storage compartments for the study period 2003 to 2010. Second, modeled TWSA were converted to SH coefficients up to degree 90. Third, the SH coefficients were filtered with the DDK3 filter and converted back to gridded TWSA. The rescaling factors were then derived for each hydrological model and for each grid cell by minimizing the difference between original TWS S^o and filtered TWS S^f in a least squares sense according to

$$M = \sum_t (\Delta S_t^o - k \Delta S_t^f)^2, \tag{5.15}$$

where the summation was performed for all months t from 2003 to 2010. Long et al. (2015) found only little temporal variability of rescaling factors in Europe. Therefore, and in order to limit influence from hydrological model errors, a temporally constant rescaling factor was applied in this study. In line with Long et al. (2015), negative rescaling factors were set to zero and rescaling factors greater than three were set to three. Finally, for each grid cell the median of the rescaling factors from all five models was computed. Figure 5.4 shows rescaling factors close to one for most inland points, which indicates only weak signal attenuation. In contrast, along coastlines large rescaling factors indicate leakage from the weaker ocean signal. In Northern Africa rescaling factors are about 0 as TWSA are very small in this region and signal leaking from the surrounding regions needs to be removed. The estimated rescaling factors are consistent with those derived by Landerer and Swenson (2012), Long et al. (2015), and Zhang et al. (2016). Applying rescaling factors introduces an additional source of uncertainty into the derivation of TWSA. Therefore, error estimates of the rescaling procedure are derived in Section 5.1.5.

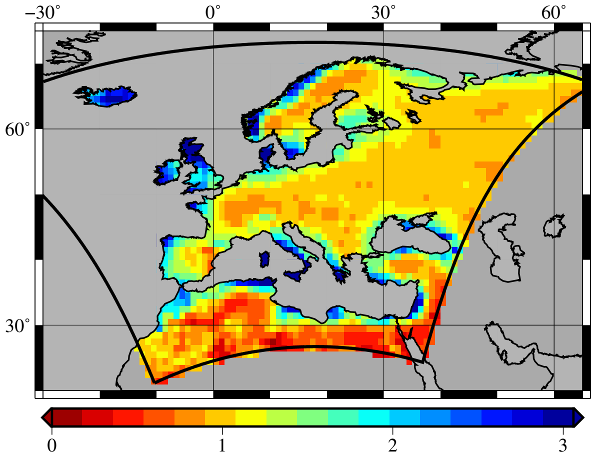


Figure 5.4: Rescaling factors for the ITSG-Grace2016 solution ($n_{max} = 90$) on a 1° grid.

Kumar et al. (2016) came to the conclusion that the role of rescaling factors in the context of GRACE assimilation into hydrological models is rather marginal. For data assimilation purposes, Girotto et al. (2016) suggested as alternative approach to rescale GRACE observations to match the long-term mean and standard deviation of the hydrological model in order to guarantee climatological consistency (Draper et al., 2015). In this study, TWS anomalies and corresponding errors shall be represented as realistically as possible, since this is a necessary prerequisite for including also other observation-based data sets into the developed data assimilation framework.

5.1.4 Glacial Isostatic Adjustment

The Earth's delayed viscoelastic response to the retreat of the ice sheets, which covered the Earth during the last glacial maxima, is called glacial isostatic adjustment (GIA). The isostatic equilibrium is adjusted by the influx of mantle material to the previously glaciated regions. As a consequence, long-term gravity changes arise. In Europe, for example, Scandinavia experiences uplifts of about 1cm/year (Kierulf et al., 2014).

Literature on GIA modeling used to focus on providing the location and strength of the signal. One frequently applied approach is to simulate the viscoelastic response of a compressible Earth to surface load (A et al., 2013). Uncertainty estimates were based on inter-comparison studies between GPS and global GIA models suggesting an uncertainty of 20 % for the GIA models being reasonable (Kierulf et al., 2014). Recently, research moved on to provide uncertainty estimates together with the individual models. Simon et al. (2018) developed an inversion approach for computing GIA and corresponding uncertainties from GRACE gravity data and vertical land motion data from GPS over Northern Europe (Figure 5.5 a, e). Over Scandinavia trends of up to 30 mm/year with corresponding standard deviations of up to 1.2 mm/year were found.

The results from Simon et al. (2018) were compared to three global GIA models from A et al. (2013), Klemann and Martinec (2009), and Wang et al. (2008) (Figure 5.5 b–d). Differences

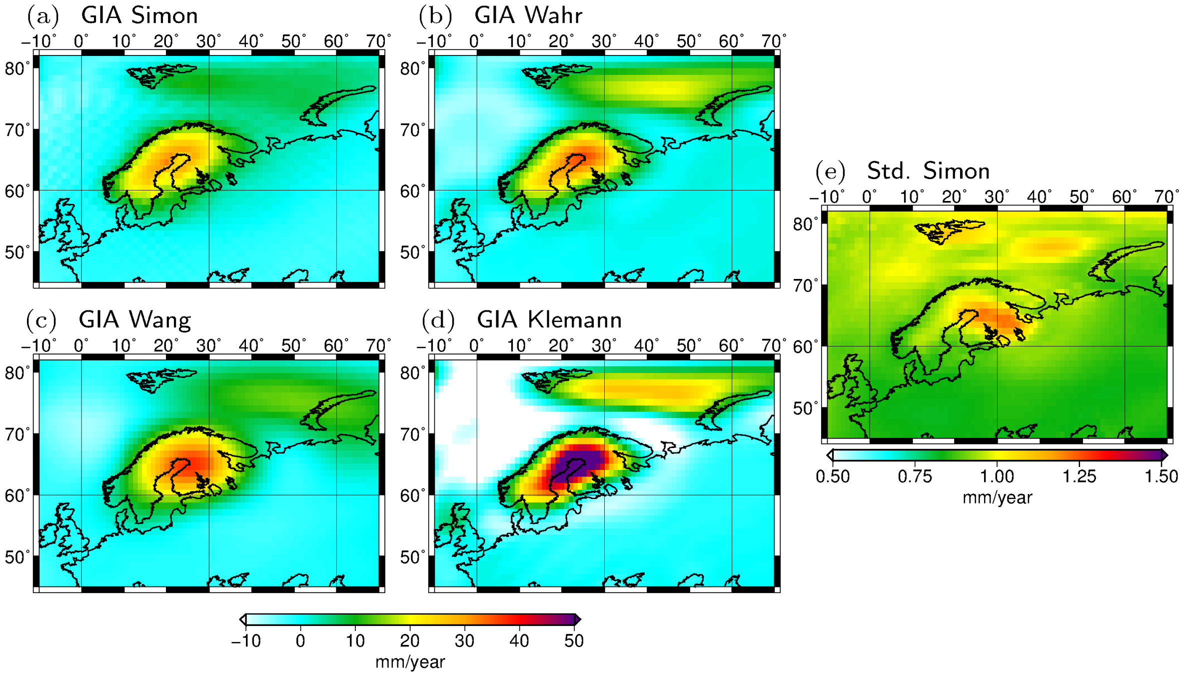


Figure 5.5: Trend induced by glacial isostatic adjustment (GIA) over Northern Europe given in equivalent water height (EWH) according to (a) Simon et al. (2018), (b) A et al. (2013), (c) Wang et al. (2008), and (d) Klemann and Martinec (2009), and (e) the corresponding uncertainty estimate from Simon et al. (2018).

arise with respect to the shape and the magnitude of the signal.

In the context of hydrological studies, GIA is removed from the GRACE solutions by applying the GIA model either directly to the Stokes coefficients or to the gridded maps of TWSA. Here, the data set of Simon et al. (2018) was applied to the gridded maps of TWSA. Error estimates of the GIA model were not considered further in this study.

5.1.5 Error Propagation

Estimating TWSA from observations damped at satellite altitude is an ill-posed problem and methods for constraining solutions may lead to biased estimates (Kusche and Springer, 2017). The following considerations aim at deriving the representation of the error covariance matrix of gridded, filtered, and rescaled TWSA estimates.

The expectation of gridded TWSA is denoted as $E(\hat{\mathbf{x}}) \neq \mathbf{x}$, where $\hat{\mathbf{x}}$ is the vector of estimated TWSA and $\mathbf{x}$ is the vector of unknown true TWSA. The corresponding covariance matrix of gridded TWSA is denoted as $\boldsymbol{\Sigma}(\hat{\mathbf{x}}) = E\left((\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))(\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))^T\right)$. Filtering of the TWSA with some filter matrix $\mathbf{W}$ (this step is usually performed in spectral domain) leads to the expectation and the covariance matrix of filtered TWSA $\hat{\mathbf{x}}^f$,

$$
\begin{aligned}
E(\hat{\mathbf{x}}^f) &= E(\mathbf{W}\hat{\mathbf{x}}) \neq \mathbf{x}, && (5.16)\\
\boldsymbol{\Sigma}(\hat{\mathbf{x}}^f) &= E\left((\mathbf{W}\hat{\mathbf{x}} - E(\mathbf{W}\hat{\mathbf{x}}))(\mathbf{W}\hat{\mathbf{x}} - E(\mathbf{W}\hat{\mathbf{x}}))^T\right) && (5.17)\\
&= \mathbf{W}(\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))(\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))^T \mathbf{W}^T && (5.18)\\
&= \mathbf{W}\boldsymbol{\Sigma}(\hat{\mathbf{x}})\mathbf{W}^T. && (5.19)
\end{aligned}
$$

Applying the vector $\mathbf{k}$ that contains the rescaling factors for each grid cell leads to

$$E(\hat{\mathbf{x}}^{+}) = E\left(\mathbf{kW}\hat{\mathbf{x}}\right) = \mathbf{kW}\hat{\mathbf{x}} \neq \mathbf{x} \tag{5.20}$$

$$\boldsymbol{\Sigma}\left(\hat{\mathbf{x}}^{+}\right) = \mathbf{KW}\boldsymbol{\Sigma}(\hat{\mathbf{x}})\mathbf{W}^{T}\mathbf{K}, \tag{5.21}$$

where $\mathbf{K}$ is a diagonal matrix containing the rescaling factors $\mathbf{k}$ on its main diagonal and $\hat{\mathbf{x}}^{+}$ is the vector of filtered and rescaled TWSA. In total, this gives rise to the bias $\mathbf{b}_{\hat{\mathbf{x}}^{+}}$ of filtered and rescaled TWSA:

$$\mathbf{b}_{\hat{\mathbf{x}}^{+}} = E\left(\hat{\mathbf{x}}^{+} - \mathbf{x}\right). \tag{5.22}$$

Obviously, the covariance matrix $\boldsymbol{\Sigma}\left(\hat{\mathbf{x}}^{+}\right)$ does not account for spreading of the estimate $\hat{\mathbf{x}}^{+}$ with respect to the truth $\mathbf{x}$. Therefore, the mean square error (MSE) matrix $\mathbf{M}(\hat{\mathbf{x}}^{+})$ is considered as

$$\mathbf{M}(\hat{\mathbf{x}}^{+}) = \boldsymbol{\Sigma}\left(\hat{\mathbf{x}}^{+}\right) + \mathbf{b}_{\hat{\mathbf{x}}^{+}}\mathbf{b}_{\hat{\mathbf{x}}^{+}}^{T}. \tag{5.23}$$

Unfortunately, the bias $\mathbf{b}_{\hat{\mathbf{x}}^{+}}$ is not known. Here, I chose to approximate $\mathbf{b}_{\hat{\mathbf{x}}^{+}}\mathbf{b}_{\hat{\mathbf{x}}^{+}}^{T}$ by leakage and rescaling errors that can be derived according to Landerer and Swenson (2012). Please note that the derivation of unbiased TWSA with consistent error estimates from GRACE is a current matter of research that cannot be entirely solved in this thesis.

In the following, first the error propagation from the full error covariance matrix of SH coefficients to the covariance matrix $\boldsymbol{\Sigma}\left(\hat{\mathbf{x}}^{+}\right)$ of gridded, filtered, and rescaled TWSA is described (Section 5.1.5.1). Then, the computation of the mean square error matrix $\mathbf{M}(\hat{\mathbf{x}}^{+})$ is derived (Section 5.1.5.2).

5.1.5.1 Covariance Matrix of Gridded, Filtered, and Rescaled TWSA

For this study, thorough error propagation was performed starting from the SH coefficients along with full error covariance information and resulting into errors of gridded and rescaled TWSA including full error covariance matrices.

The ITSG-Grace2016 formal error estimates are closer to calibrated errors of GFZ than to formal errors of other processing centers (Section 5.1.3.1). Filtered (DDK3) and gridded error estimates of GFZ formal errors amount to about 11 mm equivalent water height (EWH) over the European CORDEX area, while calibrated errors are three times larger with about 35 mm EWH. On the contrary, formal errors propagated from the ITSG-GRACE2016 solution are on average about 21 mm EWH, when starting from the diagonal covariance matrix of SH coefficients, and 18 mm EWH, when starting from the full covariance matrix of SH coefficients, respectively. Errors propagated from the main diagonal of the SH covariance matrix are nearly longitude-independent, with decreasing values toward the poles due to better spatial sampling (Figure 5.6 a). In contrast, errors propagated from the full error covariance matrix show spatial patterns beyond changes with latitude (Figure 5.6 b). Zhang et al. (2016) concluded that in particular in higher latitudes errors are underestimated when neglecting error correlations. However, over large parts of Central Europe errors are overestimated when neglecting error correlations.

It is challenging to compare the error estimates of Figure 5.6 to previous studies due to different GRACE releases and filtering. Wahr et al. (2006) constructed measurement errors for GRACE mass estimates depending on latitude and smoothing radius. They obtained errors of about 16 mm EWH in Europe for CSR-RL02 gravity fields, which were smoothed with a Gaussian filter of 750 km radius. These errors matched well the calibrated errors for

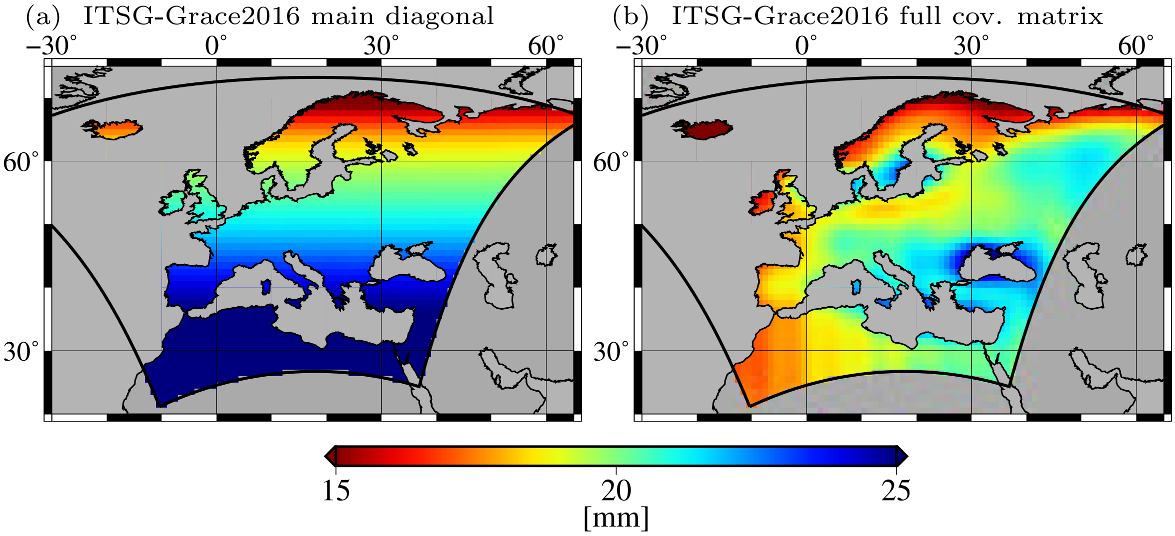


Figure 5.6: Gridded standard deviations of TWSA propagated from (a) the diagonal and (b) the full error covariance matrix of SH coefficients from ITSG-Grace2016 (n_{max}=90) for January 2007.

the same solution. Landerer and Swenson (2012) applied the destriping filter of Swenson and Wahr (2006) followed by a Gaussian filter of 300 km radius to CSR-RL04 and estimated measurement errors of about 10 mm EWH according to the method of Wahr et al. (2006). For the newest GRACE releases errors decreased significantly. However, Zhang et al. (2016) derived measurement errors from calibrated errors of GFZ-RL05a between 20 and 40 mm EWH applying the DDK2 filter and also using the method of Wahr et al. (2006). All in all, I would conclude, that error propagation from formal ITSG errors still lead to a rather optimistic error estimate.

Neighboring grid cells of GRACE solutions are highly correlated due to the limited spatial resolution, and also remote grid cells can be correlated due to the special error structure of GRACE data. Figure 5.7 shows correlations of 0.7 and 1 between Bonn ($\lambda = 7.1°$, $\phi = 50.7°$, indicated by the dot) and grid cells within a radius of about 200 km in east-west direction and 300 km in north-south direction. For remote cells small positive and negative correlations of up to 0.4 are obtained with large-scale spatial patterns.

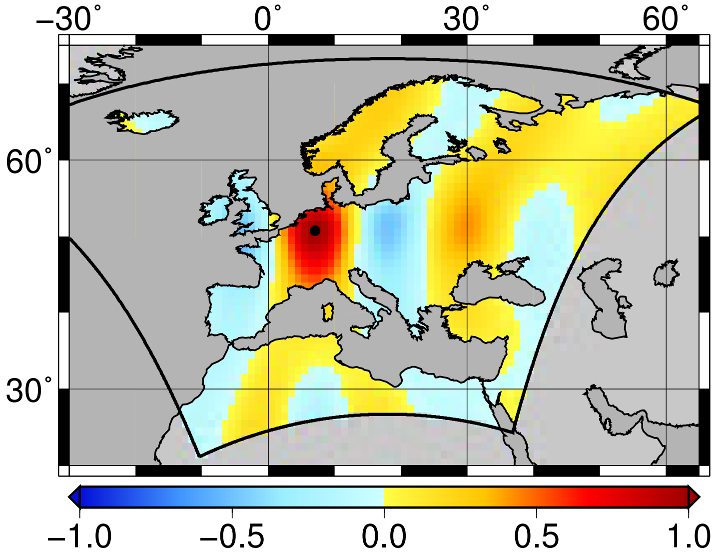


Figure 5.7: Correlation between Bonn ($\lambda = 7.1°$, $\phi = 50.7°$, indicated by the dot) and the grid points of the European CORDEX region for January 2007.

The ITSG Graz provides normal equations for the monthly GRACE solutions, which were solved here up to d/o 90. Error propagation was performed using the filter matrix of the DDK3 filter. Then, errors for degree 1 coefficients and c_{20}, which come along with the applied time series, were replaced on the main diagonal of the filtered covariance matrices, assuming zero correlation with the other coefficients. Finally, the covariance matrix $\mathbf{\Sigma}(\hat{\mathbf{x}}^f)$ of gridded and filtered TWSA was obtained from

$$\mathbf{\Sigma}(\hat{\mathbf{x}}^f) = \mathbf{F}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{W}^T\mathbf{F}^T, \tag{5.24}$$

where $\mathbf{\Sigma}_{\mathbf{xx}}$ is the covariance matrix of the SH coefficients with degree 1 and c_{20} replaced, $\mathbf{W}$ is the filter matrix for the DDK3 filter, and the Jacobian matrix $\mathbf{F}$ contains the partial derivatives of Eq. (5.11) for each SH coefficient.

Filtering induces leakage errors and therefore gridded TWSA were rescaled according to Eq. (5.15) using static rescaling factors (Figure 5.4). Dependence of rescaling factors from the choice of the hydrological models and from the seasonal cycle were assessed by computing coefficients of variation (CV, Figure 5.8). The CVs are defined as the standard deviation of a data set normalized by its mean. Largest model-dependent variations occur in Northafrica where TWS variability is close to zero. In contrast, over Europe, 70 % of the CVs are below 0.15 indicating consistency of TWSA from different models for most regions in Europe. Yet, over mountainous areas, CVs reach values of up to 0.3 (Figure 5.8 a). Overall, the analysis of CVs confirmed findings of Long et al. (2015) and Zhang et al. (2016). Additionally, for each month, rescaling factors were estimated separately and show only small variability (Figure 5.8 b). In fact, 60 % of the corresponding CVs were below 0.1 and 94 % below 0.2. Largest seasonal dependence of rescaling factors was found over Skandinavia and over the Alps. An analysis of the seasonal cycle indicated slightly smaller rescaling factors in summer and winter than in spring and autumn.

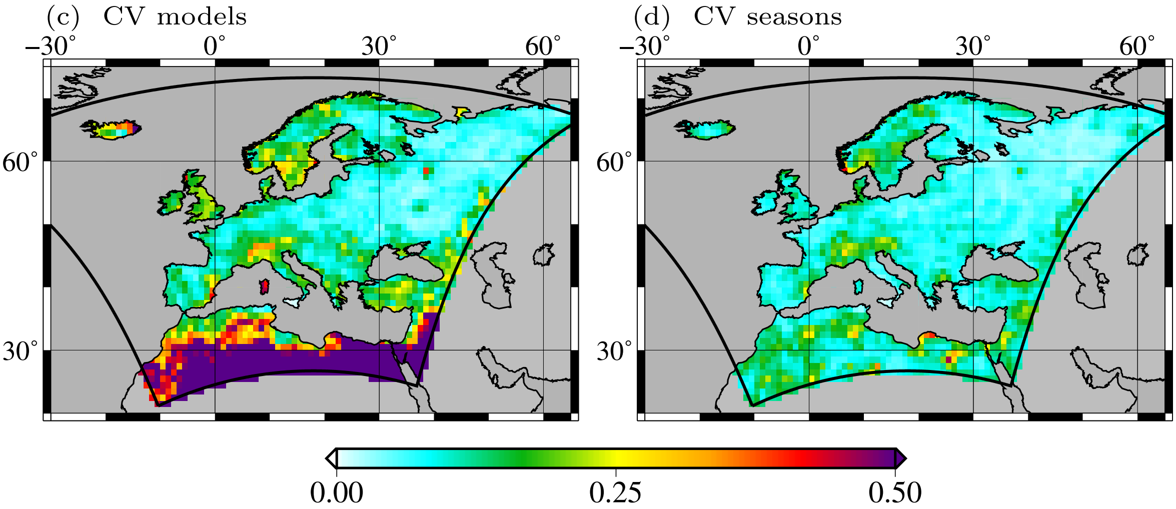


Figure 5.8: Coefficients of variation (CVs, defined as the ratio of standard deviation and mean) for (a) rescaling factors derived from five different hydrological models and (b) rescaling factors estimated for each month individually.

The final measurement error covariance matrix was obtained by applying the static rescaling factors $\mathbf{k}$ to the main diagonal of $\mathbf{\Sigma}(\hat{\mathbf{x}}^f)$, resulting into

$$\mathbf{\Sigma}(\hat{\mathbf{x}}^+) = \mathbf{K}\mathbf{F}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{W}^T\mathbf{F}^T\mathbf{K}. \tag{5.25}$$

Inner parts of the continents then show errors of about 20 mm, while errors of more than 30 mm are obtained towards the coast (Figure 5.9 a).

5.1.5.2 Mean Square Error Matrix

In order to approximate the spreading of gridded, filtered, and rescaled TWSA with respect to true TWSA, the mean square error matrix $\mathbf{M}$ was set up. This means, the matrix $\mathbf{b}_{\hat{\mathbf{x}}^+}\mathbf{b}_{\hat{\mathbf{x}}^+}^T$ (Eq. (5.23)) of rank 1 was approximated by leakage and rescaling errors derived following Landerer and Swenson (2012) and Zhang et al. (2016).

Leakage errors e_{leak} account for differences in the variability of TWSA from GRACE and TWSA from hydrological models, i.e. the part of leakage that is not modeled when contrasting filtered and unfiltered TWSA estimates from the models. Here, leakage errors were computed for each model and each grid cell individually by weighting the leakage effect by the ratio of the RMS variability of the filtered GRACE time series and the filtered model time series according to

$$e_{leak} = RMS\left(\Delta\mathbf{S}^o - k\Delta\mathbf{S}^f\right)\frac{RMS\left(GRACE^f\right)}{RMS\left(\Delta\mathbf{S}^f\right)}, \tag{5.26}$$

where $\mathbf{S}^o$ is the time series of original TWSA of the hydrological model and $\mathbf{S}^f$ is the time series of filtered TWSA of the hydrological model. The median of the leakage errors of all models is shown in Figure 5.9 b. The distribution of leakage errors is similar to the distribution of rescaled measurement errors. It is striking that leakage errors are with about 15 mm over inner Europe only slightly smaller than the measurement errors and have about the same size near the coasts. This underlines the importance of taking the contribution of leakage errors into account for the total error estimate.

Rescaling errors e_{resc} account for differences in simulated TWSA from different hydrological models, i.e. the uncertainty of hydrological model estimates. Rescaling errors were computed for each grid cell from the deviation of the rescaling factors of the individual models from their median, and from the RMS variability of the filtered GRACE time series $RMS\left(GRACE^f\right)$ according to

$$e_{resc} = RMS([k_1, ..., k_M] - median([k_1, ..., k_M]))RMS\left(GRACE^f\right), \tag{5.27}$$

where $[k_1, ..., k_M]$ represents the rescaling factors of M hydrological models for one grid cell. The rescaling errors have only a small contribution to the overall error with about 5 mm (Figure 5.9 c). As stated before, hydrological models agree well over Europe with slightly larger differences only over the Alps, Skandinavia, Great Britain, and Turkey.

As rescaling factors are spatially correlated, the matrices $\mathbf{E}_{leak}$ and $\mathbf{E}_{resc}$ for leakage and rescaling errors were set up using the squared exponential correlation function proposed by Landerer and Swenson (2012) according to

$$\mathbf{E}_{leak}(x_i, x_j) = e_{i,leak}e_{j,leak}exp\left(\frac{-d_{ij}^2}{2d_0^2}\right), \tag{5.28}$$

$$\mathbf{E}_{resc}(x_i, x_j) = e_{i,resc}e_{j,resc}exp\left(\frac{-d_{ij}^2}{2d_0^2}\right),, \tag{5.29}$$

where e_i and e_j are the errors for grid points i and j, d_{ij} is the distance between the two points, and d_0 represents the decorrelation length. The decorrelation length was chosen as

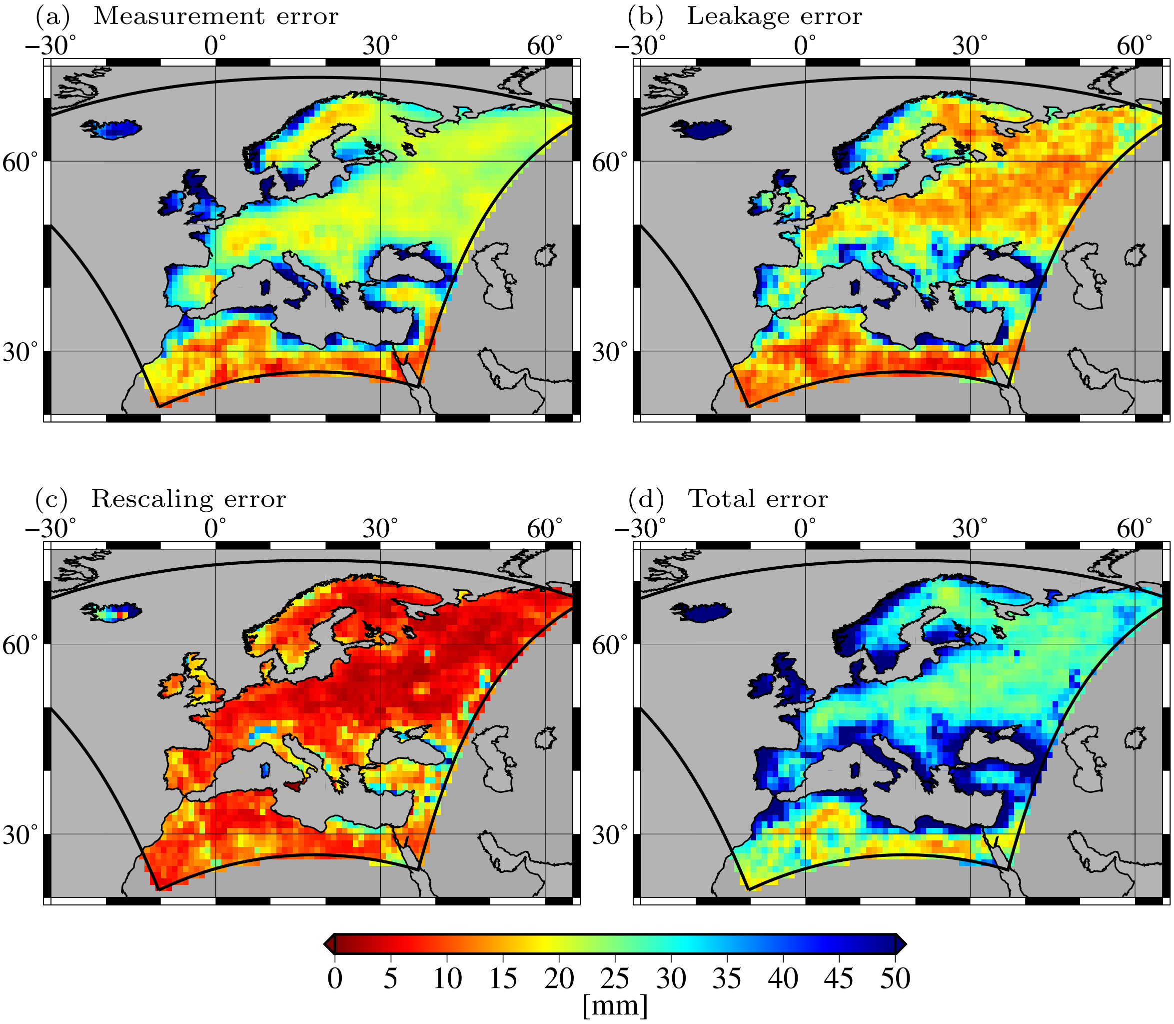


Figure 5.9: Errors of the DDK3 filtered and rescaled TWSA from ITSG-Grace2016 for January 2007 on a 1 ° grid: (a) rescaled measurement error, (b) leakage error, c) rescaling error, and d) total error.

100 km and 10 km for leakage error and rescaling error, respectively (Zhang et al., 2016).

In total, the mean square error matrix $\mathbf{M}(\hat{\mathbf{x}}^{+})$ of a monthly solution of gridded and rescaled TWSA (Figure 5.9 d) is computed as the sum of measurement error covariance matrix, leakage error matrix, and rescaling error matrix according to

$$\mathbf{M}(\hat{\mathbf{x}}^{+}) = \mathbf{K}\mathbf{F}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{W}^{T}\mathbf{F}^{T}\mathbf{K} + \mathbf{E}_{leak} + \mathbf{E}_{resc}, \tag{5.30}$$

where $\mathbf{E}_{leak} + \mathbf{E}_{resc}$ approximate $\mathbf{b}_{\hat{\mathbf{x}}^{+}}\mathbf{b}_{\hat{\mathbf{x}}^{+}}^{T}$ of Eq. (5.23). The total error of TWSA over inner parts of Europe amounts to about 30 mm and increases up to 50 mm and more towards the coasts. Overall, 10 % of the errors are even larger than 70 mm. A more extensive discussion on the three types of errors can be found in Zhang et al. (2016).

5.2 Computing K-Band Residuals from GRACE Level 1B Data

GRACE K-band range-rate (KBRR) and range-acceleration (KBRA) residuals were computed to investigate hydrological mass variability directly on the basis of level 1B observations. Both types of K-band observations were simulated from modeled water mass variations and subtracted from the original measurements to obtain residuals. This procedure has two advantages: (i) models can be validated without the downward continuation and filtering steps generally required for validation using gridded gravity field maps, and (ii) the content of high-frequent (daily) signals in hydrological models can be assessed. Furthermore, investigating the K-band residuals of the GRACE assimilating CLM3.5 model implies a step towards using the original GRACE in-orbit observations in the assimilation process.

K-band residuals were computed using the short-arc approach of Mayer-Gürr (2006), which is implemented within the GRavity field Object Oriented Programming System (GROOPS) gravity field analysis software. First, level 1B raw data from the GRACE satellite mission was downloaded and prepared for processing (Section 5.2.1). K-band observations were corrected for known effects beyond hydrology. Then, KBRR and KBRA were simulated from hydrological models and subtracted from the K-band measurement to obtain KBRR residuals and KBRA residuals (Section 5.2.2). Finally, the along-track residuals were converted to gridded time series (Section 5.2.3).

5.2.1 Preparation of GRACE Level 1B Data

GRACE level 1B release 02 (RL02) data from JPL include for each of the two satellites dynamic orbits, accelerometer data, star camera data, GPS observations, and furthermore K-band observations between the two satellites together with light-time corrections and antenna centre corrections. GPS orbits and clocks are provided by the Astronomical Institute of the University of Bern (AIUB). Additionally, Earth Orientation Parameters (EOP) were obtained from the International Earth Rotation and Reference Systems Service (IERS). Finally, AOD1B RL05 data of high-frequency non-tidal mass transport in the atmosphere and oceans (Flechtner et al., 2015) was downloaded from JPL. All of the mentioned data sets were converted into a GROOPS specific format before further processing.

Data from the instruments on board of the GRACE satellites were adjusted. This includes interpolation of small gaps (< 30 seconds) of the star camera observation and synchronization with orbit data. Furthermore, bias and scale corrections were applied to the accelerometer data using an empirical model. Then, the K-band observations were adjusted (i) for changes introduced by the time of flight of the K-band signal, so-called light-time corrections, and (ii) for geometric effects arising as the position of the K-band antenna phase center differs from its nominal position, so-called antenna phase center corrections.

5.2.2 Computation of K-band Residuals

A number of background models describing effects beyond hydrology-induced gravity changes were removed from the K-band observations (see also Section 2.2):

- Tidal effects from the Sun, the Moon, and the planets were computed from the JPL DE405 ephemerides (Standish, 1998). Deformations of the solid Earth arising due to tidal forces from third bodies were modeled according to the IERS 2003 conventions (McCarthy and Petit, 2004). Ocean tides were removed using the EOT11a model of Savcenko and Bosch (2012). The atmospheric tides S1 and S2 were modeled according to Biancale and Bode (2006).
- The GFZ de-aliasing product AOD1B-RL05 provides non-tidal, high-frequency mass redistribution of atmosphere and ocean. The S1 and S2 atmospheric tides were removed to avoid double book-keeping (Flechtner et al., 2010).
- As static reference field served the ITG-Grace2014s up to degree $n_{max} = 200$. However, experiments with different static reference fields showed that the choice of the static field in not relevant for the results achieved in this thesis.

Additionally, different models representing mainly hydrological mass variations were considered, such as GRACE solutions, output from hydrological models, or mass variations of large reservoirs. These data sets were converted into spherical harmonic coefficients by discretizing an integral according to

$$\left\{ \begin{array}{c} c_{nm} \\ s_{nm} \end{array} \right\} = \frac{1}{4\pi} \sum_{i}^{I} ewh\,(\lambda_i, \theta_i)\,\Delta A_i \left\{ \begin{array}{c} \cos(m\lambda_i) \\ \sin(m\lambda_i) \end{array} \right\} P_{nm}(\cos(\theta_i)), \tag{5.31}$$

where the area element ΔA_i is given for each grid point with the coordinates λ_i and θ_i.

For all of the above data sets, the forces along each of the satellite's orbit were computed. Then, observation equations were set up for 60-min arcs of the satellites' orbit according to Section 4.2.4.4 of Mayer-Gürr (2006) using the above models as background models. The observation equations were solved for KBRR and KBRA residuals at a sampling rate of 5 seconds co-estimating boundary positions of the two satellites and constant accelerometer bias parameters (Chapter 5 of Mayer-Gürr, 2006).

For future investigations, the version of the JPL ephemerides should be updated, e.g., to version DE430. Furthermore, the latest release of the GRACE de-aliasing product, AOD1B-RL06, should be considered.

5.2.3 Evaluation of K-band Residuals

KBRR and KBRA residuals of each arc were tested for outliers. Each contaminated arc with an RMS of the residuals larger than $1{\times}10^{-6}$m/s for KBRR residuals and $1{\times}10^{-7}$m/s^2 for KBRA residuals was removed. As the ratio of deleted arcs can reach up to 10 % to 20 % it would be interesting to remove outliers in the observations prior to the estimation of the residuals. From the arcs remaining after outlier detection, daily, monthly, and yearly RMS values were computed for individual grid cells and accumulated over specific study areas. In this scope, each residual was assigned to the mid-point of the orbit positions of the two satellites.

Please note that mass changes affect the KBRR residuals not at the location of the mass change, but northern and southern. In contrast, KBRA residuals are located directly above a mass change, but evaluation is difficult as they are more noise than KBRR residuals.

Chapter 6

Concepts of Sequential Data Assimilation

Data assimilation is a mathematical framework combining model simulations that seek to represent physical processes with information provided by measurements, in order to obtain a better representation of reality. There are two classes of data assimilation concepts: variational methods and sequential methods. Both approaches can be applied to all kind of data assimilation problems and have different strengths and weaknesses, which have frequently been discussed in literature (Rabier et al., 1992; Robert et al., 2006; Goodliff et al., 2015; Abaza et al., 2015). Various open issues exist regarding the optimal data assimilation algorithm for a given application.

Variational data assimilation approaches minimize a cost function that describes the misfit between model and observations over the whole assimilation period (Dimet and Talagrand, 1986; Evensen, 2009). For this, the model must be integrated forward and backward in time. In its strong formulation, the result from variational approaches must be a valid trajectory of the model, while the weak constraint problem allows for deviations from the model trajectory by taking into account model errors (Sasaki, 1970). As variational methods provide an improved estimate of the initial conditions, they are of particular interest for atmospheric modeling. In contrast, sequential data assimilation approaches correct the model state whenever observations become available and then propagate from the updated state. This procedure assumes that (i) the model depends only on the previous time step, and that (ii) the observations are independent in time. Sequential data assimilation is often used in the context of hydrological models as these models are largely dependent on climate forcing fields whereas the initial conditions are less relevant.

Two classes of sequential data assimilation approaches exist: ensemble Kalman filters and particle filters. While ensemble Kalman filters are based on Gaussian assumptions, particle filters can also represent non-Gaussian probability density functions of model states. However, classical particle filter methods suffer from divergence problems for large-scale systems, which can only be overcome by localization or by applying future observations. Therefore, this thesis focuses on ensemble Kalman filter methods. Reviews on different data assimilation algorithms are provided e.g. by Houtekamer and Zhang (2016) with a discussion of ensemble Kalman filters and tuning algorithms, and by Bannister (2017) with a review on variational and ensemble-variational data assimilation methods. Vetra-Carvalho et al. (2018) provide a coherent mathematical description of the principle data assimilation approaches and also discuss computational aspects.

Typically, ensemble based sequential data assimilation algorithms are derived from the

Kalman Filter (KF) (Evensen, 2009) and include variants that can deal with non-linear high-dimensional models efficiently. Furthermore, smoothing algorithms allow to modify model states in the past according to newly available observations. In Section 6.1, the classical KF and its non-linear variant, the Extended Kalman Filter (EKF), are introduced. Then, the Ensemble Kalman Filter (EnKF) for large-scale problems (Section 6.1.2) follows together with computationally more efficient modifications, such as the Ensemble Transform Kalman Filter (ETKF, Section 6.1.3), the Singular Evolutive Interpolated Kalman (SEIK, Section 6.1.4) filter, and the Error Subspace Transform Kalman Filter (ESTKF, Section 6.1.5).

In large scale problems, the number of state variables is several magnitudes larger than the number of ensemble members, but the ensemble covariance matrix has a rank smaller or equal to the ensemble size minus one. This so called undersampling of the ensemble can cause problems such as (i) underestimation of the ensemble spread, (ii) filter divergence, and (iii) error in estimated correlations (Vetra-Carvalho et al., 2018). The effect of undersampling and problems arising due to model errors and inappropriate ensemble generation can be minimized through localization and ensemble covariance inflation (Section 6.2).

6.1 Ensemble Kalman Filter Approaches

The common ground of all data assimilation algorithms is the Bayes theorem (van Leeuwen and Evensen, 1996; Koch, 2007)

$$p\left(\mathbf{x}|\mathbf{y}\right) \propto p\left(\mathbf{y}|\mathbf{x}\right) p\left(\mathbf{x}\right). \tag{6.1}$$

The Bayes theorem relates the posterior or analysis probability density function (PDF) $p\left(\mathbf{x}|\mathbf{y}\right)$ of the model states $\mathbf{x}$ depending on the observations $\mathbf{y}$ to the observation or likelihood PDF $p\left(\mathbf{y}|\mathbf{x}\right)$, which depends on the model states, and to the model forecast PDF $p\left(\mathbf{x}\right)$. Both prior and likelihood PDF are assumed to be Gaussian and, thus, lead to a Gaussian posterior PDF. This also holds for non-linear models where higher moments are ignored.

All data assimilation algorithms aim at deriving a posterior PDF of the model state, $\mathbf{x}^a$, by combining the model forecast $\mathbf{x}^f$ with the observations $\mathbf{y}$. The forecasted model state $\mathbf{x}_k^f$ at time t_k is obtained by evolving the model state at the previous time step with the model operator $\mathcal{M}$ according to

$$\mathbf{x}_k^f = \mathcal{M}\mathbf{x}_{k-1}^f + \mathbf{q}_k, \tag{6.2}$$

where $\mathbf{q}$ denotes the model error assuming white Gaussian noise with zero mean. The corresponding model error covariance matrix is denoted as $\mathbf{C}_{xx}^f(k)$. Furthermore, a set of observations $\mathbf{y}_k$ at time t_k is linked to the model states via the linear or linearized observation equation

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k^f + \epsilon_k, \tag{6.3}$$

where $\mathbf{H}_k$ denotes the linearized form of the observation operator $\mathcal{H}$, and ϵ_k is the observation error vector. The observation error covariance matrix is denoted as $\mathbf{C}_{yy}(k)$.

All considered data assimilation algorithms are composed by the forecast step (Eq. 6.2), and the filter specific analysis step, in which the updated model state is computed. The basic formulation of the traditional KF is presented in Section 6.1.1 and describes the assimilation of observations into linear models assuming Gaussian errors. Non-linear models are treated by the EKF, which uses the linearized model equations to approximate the model covariance matrix. In the case of large-scale non-linear geophysical models the EKF is no option as linearization leads to unstable evolution of model errors for high-dimensional models and, in fact, errors cannot be explicitly represented by an error covariance matrix due to limitations in computation time and storage. Therefore, it was necessary to develop computationally efficient algorithms with reduced memory requirements, which lead to the EnKF (Section 6.1.2) and variants like the ETKF (Section 6.1.3), the SEIK filter (Section 6.1.4), and the ESTKF (Section 6.1.5).

6.1.1 The Extended Kalman Filter

The Kalman Filter (KF) is the optimal sequential data assimilation algorithm for linear models, where model errors and observation errors are Gaussian, uncorrelated, and unbiased. The derivation of the following basic equations is explained extensively by Evensen (2009).

In case of the linear Kalman filter the model operator $\mathcal{M}$ introduced in Eq. (6.2) can be expressed by the matrix $\mathbf{M}$, leading to the forecast step

$$\mathbf{x}_k^f = \mathbf{M}\mathbf{x}_{k-1}^f + \mathbf{q}_k. \tag{6.4}$$

Correspondingly, the propagation of the model covariance matrix from time step $k-1$ to time step k is obtained through formal error propagation according to

$$\mathbf{C}_{xx}^f(k) = \mathbf{M}^T\mathbf{C}_{xx}^f(k-1)\mathbf{M} + \mathbf{Q}_k, \tag{6.5}$$

where $\mathbf{Q}_k$ represents the contribution of the model error $\mathbf{q}_k$. At each time step where observations are available the best linear unbiased estimator (BLUE, Koch (1988)) for the analyzed model state $\mathbf{x}_k^a$ is obtained by minimizing the cost function

$$J(\mathbf{x}_k^a) = \left(\mathbf{x}_k^f - \mathbf{x}_k^a\right)^T \mathbf{C}_{xx}^f(k)^{-1} \left(\mathbf{x}_k^f - \mathbf{x}_k^a\right) + \left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^a\right)^T \mathbf{C}_{yy}(k)^{-1} \left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^a\right) \tag{6.6}$$

with respect to $\mathbf{x}_k^a$. The update equation can be derived according to Evensen (2009) as

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k\mathbf{d}_k, \tag{6.7}$$

where $\mathbf{K}_k$ denotes the Kalman gain and $\mathbf{d}_k$ the innovation defined as

$$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^f, \tag{6.8}$$

$$\mathbf{K}_k = \mathbf{C}_{xx}^f(k)\mathbf{H}_k^T \left(\mathbf{H}_k\mathbf{C}_{xx}^f(k)\mathbf{H}_k^T + \mathbf{C}_{yy}(k)\right)^{-1}. \tag{6.9}$$

The innovation $\mathbf{d}_k$ describes the difference between observations and model prediction. This difference is weighted by the Kalman gain according to the respective covariance matrices of model and observations. The weighted difference is then added to the model prediction. The covariance matrix $\mathbf{C}_{xx}^a(k)$ of the analyzed model state is calculated by formal error propagation via

$$\mathbf{C}_{xx}^a(k) = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{C}_{xx}^f(k). \tag{6.10}$$

Eqs. (6.7) and (6.10) represent the analysis step of the data assimilation algorithm.

In the non-linear case, the Extended Kalman Filter (EKF) forecast equations are derived from Eq. (6.2) by Taylor expansion. Assuming all higher order terms to be negligible yields the EKF forecast equations analogous to Eq. (6.4) and (6.5). However, neglecting higher order terms can lead to an unrealistic representation of the model error covariance due to unbounded error growth. This can result into instabilities of the filter algorithm. The relevance of higher order approximations for the model error evolution is discussed by (Miller, 1994). The analysis steps of the EKF are identical to those of the KF and thus given by Eqs. (6.7) to (6.10). However, for non-linear systems, the EKF analysis equations represent only an approximation to the optimal estimate and the linearization of the forward integration of the error covariance matrix is computationally expensive.

6.1.2 The Ensemble Kalman Filter

An alternative to the EKF is the Ensemble Kalman Filter (EnKF) developed by Evensen (1994). The EnKF resolves the problems arising in large scale systems regarding the evolution of the error covariance matrix. Instead of computing and storing a full error covariance matrix, the error statistics are represented by an ensemble of N_e model states. The ensemble members usually differ by the start values and input data of the model. Each ensemble member $\mathbf{x}_k^{f(i)}$ with $i = 1, ..., N_e$ evolves according to the nonlinear model of Eq. (6.2). As the true model state is not known, the ensemble covariance matrix is defined around the ensemble mean $\overline{\mathbf{x}}_k^f$ as

$$\tilde{\mathbf{C}}_{xx}^f(k) = \frac{1}{N_{e-1}} \sum_{i=1}^{N_e} \left(\mathbf{x}_k^{f(i)} - \overline{\mathbf{x}}_k^f\right) \left(\mathbf{x}_k^{f(i)} - \overline{\mathbf{x}}_k^f\right)^T . \tag{6.11}$$

During the analysis step, each ensemble member is updated according to Eq. (6.7) using an observation vector of size N_y from a random ensemble of observations, which is generated from the observation error covariance matrix (Burgers et al., 1998; Houtekamer and Mitchell, 1998). Defining the ensemble of model forecasts as $\mathbf{X}_k^f = \left[\mathbf{x}_k^{f(1)}, \mathbf{x}_k^{f(2)}, ..., \mathbf{x}_k^{f(N_e)}\right]$, Eq. (6.7) can be rewritten (dropping the time index k) as

$$\mathbf{X}^a = \mathbf{X}^f + \mathbf{KD} = \mathbf{X}^f + \tilde{\mathbf{C}}_{xx}^f \mathbf{H}^T \left(\mathbf{H}\tilde{\mathbf{C}}_{xx}^f \mathbf{H}^T + \mathbf{C}_{yy}\right)^{-1} \left(\mathbf{Y} - \mathbf{H}\mathbf{X}^f\right), \tag{6.12}$$

where $\mathbf{Y} = \left[\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(N_e)}\right]$ is the ensemble of perturbed observations. Since the direct computation of the model error covariance matrix $\tilde{\mathbf{C}}_{xx}^f$ is computationally inefficient for large scale problems, it is common to use a so-called square-root formulation of the EnKF. To this end, the ensemble perturbation matrix is defined as $\mathbf{X}'^f = \mathbf{X}^f - \overline{\mathbf{X}^f}$, where the mean of the ensemble is stored in each column of $\overline{\mathbf{X}^f}$. In fact, the ensemble perturbations are a scaled matrix square root of the model error covariance matrix

$$\tilde{\mathbf{C}}_{xx}^f = \frac{\mathbf{X}'^f \left(\mathbf{X}'^f\right)^T}{N_e - 1} \tag{6.13}$$

and replace $\tilde{\mathbf{C}}_{xx}^f$ in Eq. 6.12. The analysis equation (6.12) can then be expressed as

$$\mathbf{X}^a = \mathbf{X}^f + \frac{1}{N_e - 1}\mathbf{X}'^f \left(\mathbf{X}'^f\right)^T \mathbf{H}^T \left(\frac{1}{N_e - 1}\mathbf{H}\mathbf{X}'^f \left(\mathbf{X}'^f\right)^T \mathbf{H}^T + \mathbf{C}_{yy}\right)^{-1} \left(\mathbf{Y} - \mathbf{H}\mathbf{X}^f\right). \tag{6.14}$$

With the ensemble perturbation matrix in observation space, $\mathbf{S} = \mathbf{H}\mathbf{X}'^f$, and the innovation covariance matrix, $\mathbf{F} = 1/(N_e - 1)\mathbf{S}\mathbf{S}^T + \mathbf{C}_{yy}$, Eq. (6.14) can be rewritten as

$$\mathbf{X}^a = \mathbf{X}^f + \frac{1}{N_e - 1}\mathbf{X}'^f \mathbf{S}^T \mathbf{F}^{-1} \left(\mathbf{Y} - \mathbf{H}\mathbf{X}^f\right). \tag{6.15}$$

Eq. (6.15) represents an optimized formulation of the EnKF in the case $N_y >> 2N_e$, as it is possible to implement the EnKF defining only one matrix of size $N_y \times N_y$, while all other required matrices are of size $N_y \times N_e$. In the case of few observations ($N_e > 2N_y$), a different formulation is applied (Evensen, 1994).

The covariance matrix of the analyzed model state is then defined by the general square root form of the Ensemble Kalman filters, which can be derived from Eq. (6.10) and transforms the forecasted ensemble perturbations to the analyzed ensemble perturbations according to

$$\mathbf{X}'^{a} \left(\mathbf{X}'^{a}\right)^{T} = \mathbf{X}'^{f} \left(\mathbf{T}_{EnKF}\mathbf{T}_{EnKF}^{T}\right) \left(\mathbf{X}'^{f}\right)^{T}. \tag{6.16}$$

The transform matrix $\mathbf{T}_{EnKF}$ is defined as

$$\mathbf{T}_{EnKF}\mathbf{T}_{EnKF}^{T} = \mathbb{1} - \mathbf{S}^{T}\mathbf{F}^{-1}\mathbf{S}. \tag{6.17}$$

As indicated by Eq. (6.15), in the case of the EnKF the transform matrix is not directly computed. However, for other filter algorithms different ways for the computation of $\mathbf{T}$ were developed (Section 6.1.3, 6.1.4, 6.1.5).

In order to unify the formulation of several Kalman filter algorithms, Vetra-Carvalho et al. (2018) wrote the analysis update equation as linear transformations of ensemble mean $\overline{\mathbf{x}}^{a}$ and ensemble perturbations $\mathbf{X'^{a}}$ using the filter specific weight vector $\overline{\mathbf{w}}$ and the weight matrix $\mathbf{W}'$

$$\overline{\mathbf{x}}^{a} = \overline{\mathbf{x}}^{f} + \mathbf{X}'^{f}\overline{\mathbf{w}}, \tag{6.18}$$

$$\mathbf{X}'^{a} = \mathbf{X}'^{f}\mathbf{W}', \tag{6.19}$$

$$\mathbf{X}^{a} = \mathbf{X}'^{a} + \overline{\mathbf{X}}^{a}. \tag{6.20}$$

In the case of the EnKF, the transformation of the ensemble mean is not necessary, so that only the weight matrix $\mathbf{W}'$ has to be set up according to

$$\mathbf{W}'_{EnKF} = \frac{1}{N_e - 1}\mathbf{S}^{T}\mathbf{F}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X}^{f}). \tag{6.21}$$

In the following, alternative more efficient ensemble Kalman filters are presented based on the notation of Vetra-Carvalho et al. (2018).

6.1.3 The Ensemble Transform Kalman Filter

Perturbation of observation errors as realized by the EnKF introduce extra sampling errors and spurious correlations between the model state ensemble and the observations (Nerger, 2004). Among others, Bishop et al. (2001) developed a square root analysis scheme that avoids the perturbation of observation errors by explicitly transforming the forecast ensemble perturbations $\mathbf{X}'^{f}$ to the updated analysis perturbations according to Eq. (6.16). In the formulation of the resulting Ensemble Transform Kalman Filter (ETKF) the matrix $\mathbf{T}\mathbf{T}^{T}$ is rewritten as

$$\mathbf{T}_{ETKF}\mathbf{T}_{ETKF}^{T} = \left(\mathbb{1} + \frac{1}{N_e - 1}\mathbf{S}^{T}\mathbf{C}_{yy}^{-1}\mathbf{S}\right)^{-1} \tag{6.22}$$

using the Sherman-Morrison-Woodbury identity (Golub and Loan, 2012; Vetra-Carvalho et al., 2018). With the eigenvalue decomposition $\left(\mathbf{T}_{ETKF}\mathbf{T}_{ETKF}^{T}\right)^{-1} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{T}$ and its symmetric square root $\mathbf{T}_{ETKF} = \mathbf{U}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{U}^T$ the weight matrices defined in the previous section can be written as

$$\mathbf{W}'_{ETKF} = \mathbf{U}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{U}^T, \tag{6.23}$$

$$\overline{\mathbf{w}}_{ETKF} = \frac{1}{\sqrt{N_e - 1}}\mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T \left(\mathbf{X}'^f\right)^T \mathbf{H}^T\mathbf{C}_{yy}^{-1}\left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^f\right). \tag{6.24}$$

Obviously, the observation covariance matrix $\mathbf{C}_{yy}^{-1}$ is used directly (Eq. 6.24) and does not need to be sampled. However, the performance of the filter depends on the efficient computation of the inverse observation covariance matrix $\mathbf{C}_{yy}^{-1}$.

6.1.4 The Singular Evolutive Interpolated Kalman Filter

The Singular Evolutive Interpolated Kalman (SEIK) filter (Pham et al., 1998; Pham, 2001) was the first square root filter and computes the analysis step in the ensemble error subspace. Hence, the computation time is significantly reduced with respect to the EnKF (Nerger et al., 2007). In the SEIK filter, the ensemble mean and the last column of $\mathbf{X}'^f$ are removed by defining the matrices $\mathbf{A}$ and $\mathbf{L}$ as

$$\mathbf{L} = \mathbf{X}^f\mathbf{A}_{SEIK}, \tag{6.25}$$

$$\mathbf{A}_{SEIK} = \begin{bmatrix} \mathbf{1}_{N_e-1\times N_e-1} \\ \mathbf{0}_{1\times N_e-1} \end{bmatrix} - \frac{1}{N_e}\left[\mathbf{1}_{N_e\times N_e-1}\right], \tag{6.26}$$

where $\mathbf{0}$ and $\mathbf{1}$ represent matrices whose elements all equal to 0 and 1, respectively. The product of the transform matrix $\mathbf{T}$ (Eq. 6.22) leads in the ensemble error subspace to

$$\mathbf{T}_{SEIK}\mathbf{T}_{SEIK}^T = \left(\mathbf{A}^T\mathbf{A} + \frac{1}{N_e - 1}\left(\mathbf{HL}\right)^T\mathbf{C}_{yy}^{-1}\left(\mathbf{HL}\right)\right)^{-1} \tag{6.27}$$

In contrast to $\mathbf{T}_{ETKF}\mathbf{T}_{ETKF}^T$ from Eq. (6.22), $\mathbf{T}_{SEIK}\mathbf{T}_{SEIK}^T$ has the size $N_e - 1 \times N_e - 1$. The weight matrices of Eqs. (6.18) and (6.19) become

$$\mathbf{W}'_{SEIK} = \mathbf{A}\mathbf{T}_{SEIK}\boldsymbol{\Omega}, \tag{6.28}$$

$$\overline{\mathbf{w}}_{ETKF} = \frac{1}{\sqrt{N_e - 1}}\mathbf{A}\mathbf{T}_{SEIK}\mathbf{T}_{SEIK}^T\left(\mathbf{HL}\right)^T\mathbf{C}_{yy}\left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^f\right), \tag{6.29}$$

where $\mathbf{T}_{SEIK}$ is obtained from Cholesky decomposition of $(\mathbf{T}_{SEIK}\mathbf{T}_{SEIK}^T)^{-1}$, and $\boldsymbol{\Omega}$ is an orthonormal rotation matrix that can be either random or deterministic. In case of deterministic $\boldsymbol{\Omega}$, a stable ensemble is obtained when using a symmetric square root of $\mathbf{T}_{SEIK}\mathbf{T}_{SEIK}^T$ (Nerger et al., 2012).

6.1.5 The Error Subspace Transform Kalman Filter

The difference between the updated ensemble of ETKF and SEIK is small, and the ensemble transformation becomes identical when choosing the projection matrix $\mathbf{A}_{SEIK}$ according to Nerger et al. (2012). The resulting Error Subspace Transform Kalman Filter (ESTKF) is a favorable combination of the ETKF (Section 6.1.3) and the **S**ingular **E**volutive **I**nterpolated **K**alman filter SEIK (Section 6.1.4). Like the SEIK algorithm, the ESTKF computes the ensemble transformation in the error subspace, and likewise to the ETKF a minimum ensemble transformation is performed (Yang et al., 2009). The projection matrix $\mathbf{A}_{ESTKF}$ of size $N_e \times N_e - 1$ is defined as

$$\mathbf{A}_{ESTKF\{i,j\}} = \begin{cases} 1 - \frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i = j, i < N_e \\ -\frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i \neq j, i < N_e \\ -\frac{1}{\sqrt{N_e}} & \text{for } i = N_e \end{cases} . \tag{6.30}$$

While the SEIK filter drops the last column of $\mathbf{X}'^f$ through the multiplication with $\mathbf{A}_{SEIK}$, $\mathbf{A}_{ESTKF}$ redistributes the information from the last column to the other columns. The forecasted model ensemble is then transformed into the error subspace using equation (6.26). Similar to Eq. (6.27) the product of the transform matrix is computed according to

$$\mathbf{T}_{ESTKF}\mathbf{T}_{ESTKF}^T = \left(\mathbb{1} + \frac{1}{N_e - 1}(\mathbf{HL})^T \mathbf{C}_{yy}^{-1} (\mathbf{HL})\right)^{-1}, \tag{6.31}$$

where $\mathbf{T}_{ESTKF} = \mathbf{U\Sigma}^{-\frac{1}{2}}\mathbf{U}^T$ is obtained from the symmetric square root of $(\mathbf{T}_{ESTKF}\mathbf{T}_{ESTKF}^T)^{-1}$. The weight matrices applied in Eq. (6.18) and (6.19) are given by

$$\mathbf{W}'_{ESTKF} = \mathbf{A}\mathbf{T}_{ESTKF}\mathbf{A}^T, \tag{6.32}$$

$$\overline{\mathbf{w}}_{ESTKF} = \frac{1}{\sqrt{N_e - 1}}\mathbf{AU\Sigma}^{-1}\mathbf{U}^T (\mathbf{HL})^T \mathbf{C}_{yy} \left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^f\right). \tag{6.33}$$

Applying the ESTKF leads to the same results as obtained by the ETKF at slightly lower computational costs (Nerger et al., 2012).

6.1.6 Smoother Extensions

Ensemble Kalman smoothers process measurements sequentially in time just like the Ensemble Kalman filters, but additionally correct model states backwards in time (Evensen and van Leeuwen, 2000). Hence, the optimal model state is obtained from all (including future) observations. This is usually realized by smoothing the analysis ensemble at previous time steps via the filter transform matrix $\mathbf{TT}^T$. Consequently, arbitrary Ensemble Kalman filters can easily been extended to Ensemble Kalman smoothers (Kirchgessner et al., 2017). However, due to the Gaussian assumptions Ensemble Kalman smoothers do not behave optimally for non-linear systems (Nerger and Hiller, 2013), which makes careful tuning through localization and inflation necessary (Section 6.2).

6.2 Tuning of the Filter Algorithms

Restrictions of the ensemble size due to computational issues lead to sampling errors. More precisely, the model covariance matrix obtained from the state ensemble has a maximum rank of $N_e - 1$, while the state dimension is much larger for large-scale models. Thus, problems like underestimation of the ensemble spread, filter divergence, and spurious correlations arise. Localization and inflation techniques are used to minimize the effect of undersampling.

6.2.1 Localization

Localization limits spatial correlations to certain distances and thus suppresses long range correlations. However, localization also introduces imbalances into the ensemble state vector

(Mitchell et al., 2002). Mainly two types of localization are used in ensemble data assimilation algorithms: (i) domain localization, i.e. only observations within a certain distance are considered, and (ii) covariance localization, i.e. the observation covariance matrix $\mathbf{\Sigma}_{yy}$ is modified in a way that the influence of remote observations is down weighted or even equal to zero. Both approaches are described below following Vetra-Carvalho et al. (2018).

The performance of both approaches depends on the choice of the localization radius. The optimal localization radius is usually derived from several data assimilation experiments with different localization radii. Over the last decade several adaptive localization methods have been developed in order to avoid numerical tuning, e.g. (Anderson, 2007; Bishop and Hodyss, 2009; Flowerdew, 2015). However, the success of adaptive localization methods depends on the properties of the data assimilation system and requires careful testing before application.

6.2.1.1 Domain Localization

Domain localization implies that the analysis step is applied not to the whole state vector, but to specified model subdomains using only observations within a certain localization radius. Therefore, the model state perturbations $\mathbf{x'}^f_i$ of each ensemble member $i = 1, ..., N_e$ are transformed to $\gamma = 1, ..., \Gamma$ subdomains using the linear transformation matrix $\mathbf{D}_\gamma$. Likewise, the observation vector $\mathbf{y}$, the observation error covariance matrix $\mathbf{C}_{yy}$, and the observation operator $\mathbf{H}$ are transformed using the linear transformation matrix $\hat{\mathbf{D}}_\gamma$ that defines all observations within a certain distance from subdomain $\mathbf{D}_\gamma$, resulting into

$$\mathbf{x'}_{i,\gamma} = \mathbf{D}_\gamma \mathbf{x'}^f_i, \tag{6.34}$$

$$\mathbf{y}_\gamma = \hat{\mathbf{D}}_\gamma \mathbf{y}, \tag{6.35}$$

$$\mathbf{H}_\gamma = \hat{\mathbf{D}}_\gamma \mathbf{H}, \tag{6.36}$$

$$\mathbf{C}_{yy,\gamma} = \hat{\mathbf{D}}_\gamma \mathbf{C}_{yy} \hat{\mathbf{D}}^T_\gamma \tag{6.37}$$

The updated local state ensemble is then obtained from

$$\overline{\mathbf{x}}^a_\gamma = \overline{\mathbf{x}}^f_\gamma + \mathbf{X}'^f_\gamma \overline{\mathbf{w}}_\gamma, \tag{6.38}$$

$$\mathbf{X'}^a_\gamma = \mathbf{X}'^f_\gamma \mathbf{W}'_\gamma, \tag{6.39}$$

where $\overline{\mathbf{w}}_\gamma$ and $\mathbf{W}'_\gamma$ are computed from the local model forecasts and from the localized observation vector. To avoid unrealistic small scale patterns observation domains of neighboring model domains must have sufficient overlap. As each model domain is updated independently from the others, the analysis step can easily being parallelized, which reduces computation time. Domain localization is typically applied to the ETKF, SEIK, and ESTKF algorithms.

6.2.1.2 Covariance Localization

Covariance localization is typically used for the classical EnKF. A weight or localization matrix $\mathbf{L}$ is directly applied to the model covariance matrix $\tilde{\mathbf{C}}^f_{xx}$. The matrix $\mathbf{L}$ is a correlation matrix set up from an auto-correlation function with compact or space-limited support, similar to the shape of a Gaussian function (Gaspari and Cohn, 2006). In practice, covariance localization is realized by computing the Schur product (Houtekamer and Mitchell, 2001), i.e. Eq. (6.12) is modified to

$$\mathbf{X}^a = \mathbf{X}^f + \mathbf{L} \circ \left(\tilde{\mathbf{C}}^f_{xx} \mathbf{H}^T \right) \left(\mathbf{L} \circ \left(\mathbf{H} \tilde{\mathbf{C}}^f_{xx} \mathbf{H}^T \right) + \mathbf{C}_{yy} \right)^{-1} \left(\mathbf{Y} - \mathbf{H}\mathbf{X}^f \right), \tag{6.40}$$

where $\circ$ denotes the element-wise matrix multiplication (Schur product).

6.2.2 Covariance Inflation

All Ensemble Kalman filter algorithms have to deal with filter divergence caused by underestimation of the model state errors due to undersampling. In this case, during the analysis step, the forecasted model state obtains too much weight while the observations have increasingly less impact. In large-scale systems localization alone is not sufficient to reduce the impact from undersampling, thus, covariance inflation is applied to increase the ensemble spread artificially. Most commonly a constant multiplicative inflation factor (Anderson and Anderson, 1999) r is applied to the ensemble via

$$\mathbf{x}_i^f = r\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right) + \overline{\mathbf{x}}^f, \tag{6.41}$$

where $r \geq 1$. Alternatively, the so-called forgetting factor $\rho = r^{-2} \leq 1$ as introduced by Pham et al. (1998) can be applied efficiently during the computation of the transform matrix and modifies e.g. Eq. (6.22) to

$$\mathbf{T}\mathbf{T}^T = \left(\rho\mathbb{1} + \frac{1}{N_e - 1}\mathbf{S}^T\mathbf{C}_{yy}^{-1}\mathbf{S}\right)^{-1}. \tag{6.42}$$

The optimal inflation strength is usually derived by tuning experiments. In order to avoid empirical tuning, adaptive inflation algorithms were developed, which allow for spatially and temporally varying forgetting factors. Anderson (2009), for instance, estimated individual inflation factors for each state vector entry using a Bayesian estimator. However, adaptive inflation algorithms come along with extra computational burdens. Instead of applying ensemble inflation to the model states, observation errors can be inflated (Minamide and Zhang, 2016), which is also interesting in the case of different observation types. In this thesis, data assimilation experiments are performed applying a fixed multiplicative forgetting factor, derived from tuning experiments, to the model states.

Chapter 7

Implementing the Assimilation of GRACE Data into CLM3.5

Total water storage anomalies (TWSA) from the Gravity Recovery and Climate Experiment (GRACE) mission are assimilated into the Community Land Model version 3.5 (CLM3.5) within the modular high-performance data assimilation framework TerrSysMP-PDAF version 1.0 (Kurtz et al., 2016). The framework consists of the Parallel Data Assimilation Framework (PDAF), which implements the data assimilation algorithms introduced in Chapter 6 in an optimal way for parallel applications, and the Terrestrial Systems Modeling Platform (TerrSysMP), which was extended by calls to PDAF filter routines. As a result, one single executable is obtained for the integration of the model ensemble and the assimilation step.

TerrSysMP-PDAF (Section 7.1) requires as input (i) parameters defining the filter algorithm, (ii) observation files, and (iii) instructions for generating an ensemble of model runs (Figure 7.1). Observed TWSA and corresponding error information were computed from GRACE level 2 data according to Section 5.1 and model input data were generated during pre-processing according to Section 7.3. Post-processing includes the evaluation of CLM3.5 output using independent data sets.

Both components of the data assimilation framework TerrSysMP-PDAF are implemented in Fortran90 with some features from Fortran2003 and using some interfaces based on C-routines. During pre-processing GRACE observations were processed using the C++ based gravity field analysis software GRavity field Object Oriented Programming System (GROOPS). As CLM3.5 output is organized in the Network Common Data Format (NetCDF), postprocessing is mainly based on Climate Data Operators (CDO) scripts.

In this thesis, extensions were implemented to the TerrSysMP-PDAF of Kurtz et al. (2016) in order to realize the assimilation of TWSA derived from GRACE measurements.

- Here, TerrSysMP-PDAF was set up for the land surface component CLM3.5 and for the European COordinated Regional Downscaling EXperiment (CORDEX) area using forcing and surface data sets described in Section 4.1.3. Due to the large study area and the high spatial (12.5 km) and temporal (e.g. daily model output) resolution, the generation of a sufficiently large ensemble of model runs is challenging. Especially the memory space required for 3-hourly atmospheric forcings becomes critical with increasing number of ensemble members. In Section 7.3 the efficient generation of spatially, temporally and cross-correlated noise is described, which currently enables data assimilation experiments with up to 128 ensemble members.

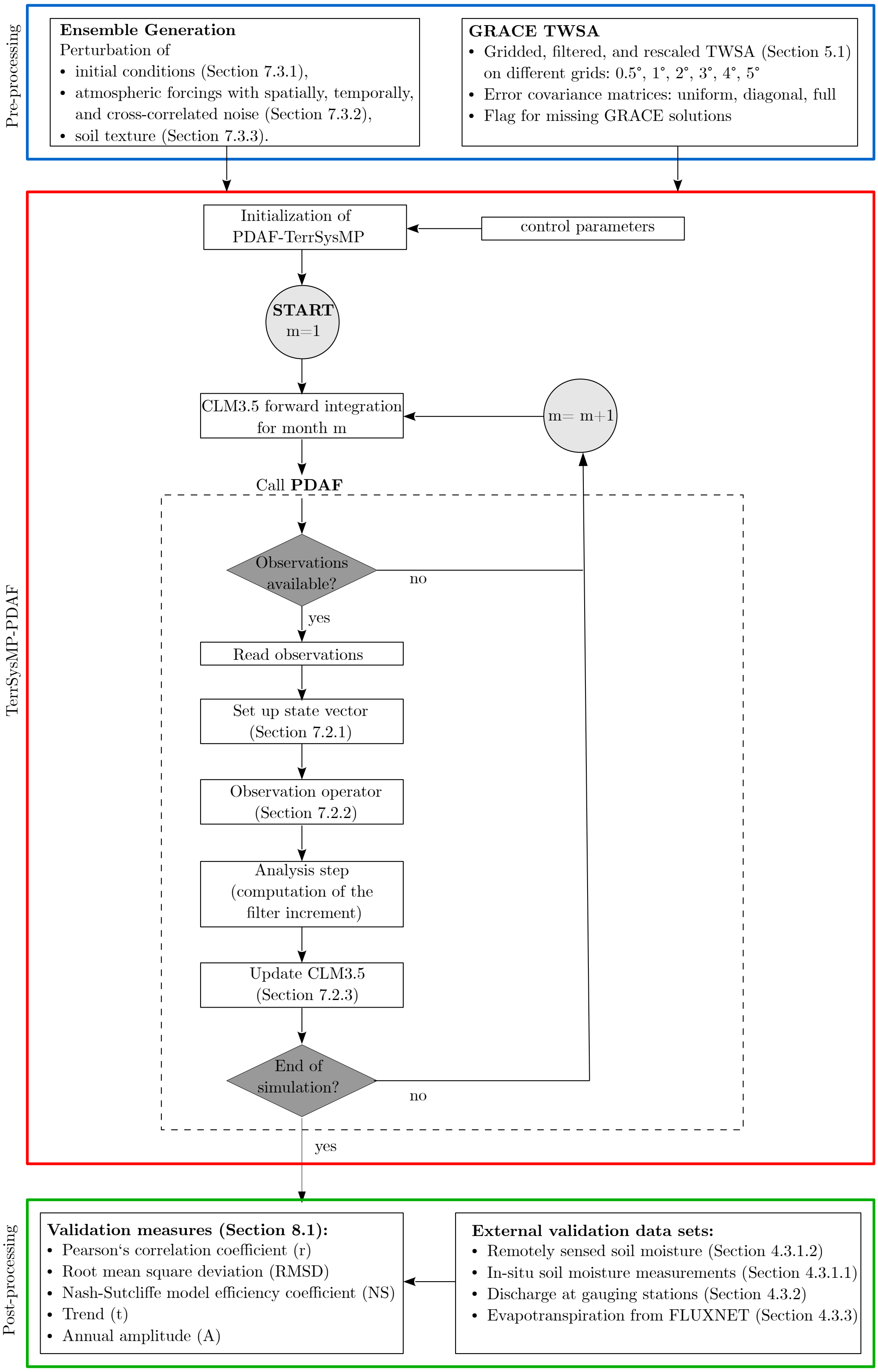


Figure 7.1: Flowchart showing the general set-up of the data assimilation experiments realized within this thesis including (i) the pre-processing of model input and observations, (ii) the data assimilation framework TerrSysMP-PDAF, and (iii) the validation of the performance of the assimilated model.

- In Kurtz et al. (2016) assimilation is limited to soil moisture and pressure data (not assimilated here). In this thesis, the framework was extended to the assimilation of TWSA, which affects the definition of the state vector, the observation operator, and the update of the model states (Section 7.2).
- The framework of Kurtz et al. (2016) requires observation files at constant time intervals. However, GRACE data shall be assimilated at monthly time scale, i.e. at irregular intervals between 28 days and 31 days. With this objective, a flexible assimilation interval was introduced, which can also deal with missing GRACE solutions (Section 7.2.3). Furthermore, TerrSysMP-PDAF was extended for temporal smoothing of modeled total water storage (TWS) at user-specific intervals (Section 7.2.1). This means, that also daily or weekly GRACE solutions can be assimilated.
- Finally, the option of using full observation error covariance matrices was added to all global and local PDAF filter options.

This chapter is organized as follows. First, the actual set-up and implementation of TerrSysMP-PDAF is described (Section 7.1). Then, interface routines are defined for connecting CLM3.5, GRACE-derived total water storage anomalies (TWSA), and PDAF (Section 7.2). These interface routines include the definition of the state vector, the mapping of model states to the observations, and the update of the model variables. Finally, the derivation of a set of perturbations of initial data, forcing data and land surface characteristics is presented (Section 7.3).

7.1 The Assimilation Framework TerrSysMP-PDAF

TerrSysMP is connected to PDAF via memory based coupling, i.e. data is exchanged via main memory and not via the input/output files of the model. This leads to a significantly faster framework with one single executable, which performs both ensemble integration and data assimilation.

The parallel communication between TerrSysMP and the filter routines is established by three parallel communicators: the model communicator, the coupling communicator, and the filter communicator (Figure 7.2). Each ensemble member has an individual model communicator, examplarily shown in Figure 7.2 for the case of three ensemble members using four processors each. On the model sub-domain (model ‘clumps’), the coupling communicator exchanges data between corresponding parts of each ensemble member. The coupling communicator is used before and after the assimilation step and collects/distributes data from/to all ensemble members and, thus, realizes communication between the model and the filter. Finally, the filter communicator is in charge of running the filter algorithms on the processors of the first ensemble member, while all other processors remain idle during this step.

The interface between TerrSysMP and PDAF was realized by Kurtz et al. (2016) by implementing extensions, that call PDAF routines. For this, a driver routine was created which controls initialization, temporal propagation, and finalization of the data assimilation framework:

1. Initialization of the Message Passing Interface (MPI)
2. Initialization of the parallel communication by PDAF
3. Initialization of TerrSysMP (land surface data, atmospheric grid, etc.)

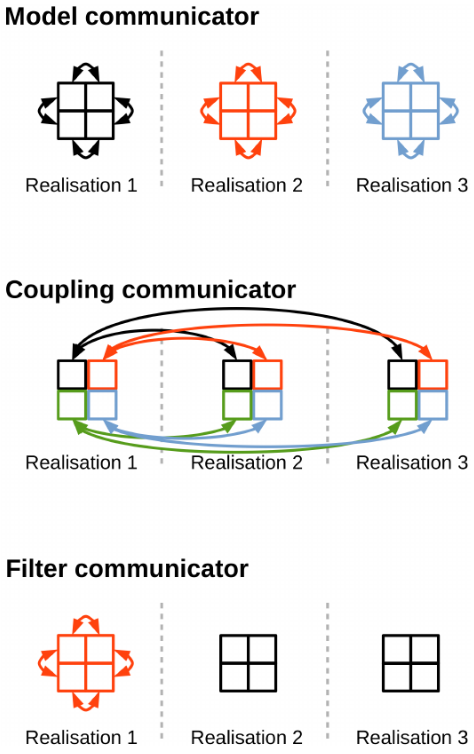


Figure 7.2: The three communicators of the parallel data assimilation framework (PDAF) are shown for the case of three ensemble members running on four processors each. Colors indicate the individual ensemble members, and arrows the parallel communication between different processors. (Figure from Kurtz et al., 2016)

4. Initialization of PDAF variables (model state vector, observation vector, filter specific variables, etc.)
5. Time loop over observation files:
    (a) Propagation of TerrSysMP until the next available observation
    (b) Analysis step by PDAF, i.e. the actual assimilation of the observations
    (c) Update of relevant model variables in TerrSysMP
6. Finalization of PDAF, TerrSysMP, and MPI

In steps 1 and 2, the parallel communication is initialized. In step 3, model control variables (initial conditions, model outputs, etc.) and surface data sets (grid, topography, soil properties, etc.) are read from different files for each ensemble member. In step 4, parameters defining computational resources, filter algorithm (filter type, assimilation interval, and observation files), and timing information are read from an input file. Then, a time loop over the observation files is carried out, where the model is advanced until the next observation time, followed by the actual assimilation of the observations, and finalized by the update of the current model state. After finishing all assimilation cycles, all data structures are de-allocated and the framework is finalized.

The filter algorithm of the TerrSysMP-PDAF framework is governed by the following parameters obtained from input files or command line options:

- **Number of ensemble members**: The true state of the model is approximated by the ensemble mean and its uncertainty is defined by the ensemble covariance matrix

(Section 6.1). If the ensemble size is not sufficiently large, sampling errors arise. Generally speaking, the number of ensemble members and their skill in picturing the main directions of the error covariance determine the stability of the model error covariance matrix. However, a trade-off must be found between affordable computational resources and ensemble size. Additionally to the ensemble size, the number of processors for each ensemble member is defined.

- **Observation files**: Within this thesis, an option was added that allows to provide gridded observations to TerrSysMP-PDAF in NetCDF format and to optionally include error information, i.e. diagonal or full error covariance matrices. Furthermore, the time interval until the next available observation can be provided.

- **Filter type**: For the assimilation of TWSA, interfaces for the following filter algorithms of TerrSysMP-PDAF were implemented within this thesis: the Ensemble Kalman Filter (EnKF), the Ensemble Transform Kalman Filter (ETKF), the Error Subspace Transform Kalman Filter (ESTKF), the Singular Evolutive Extended Kalman (SEEK) filter, and the Singular Evolutive Interpolated Kalman (SEIK) filter. Localized variants are available for all filter types except for the SEEK filter.

- **Forgetting factor**: A forgetting factor (which corresponds to covariance inflation in the case of the EnKF) can be applied to reduce the systematic underestimation of the model variance, thereby increasing the filter stability. Either a fixed forgetting factor can be provided to TerrSysMP-PDAF or the forgetting factor is estimated dynamically from the variance of the model ensemble and the observations (Vetra-Carvalho et al., 2018). Here, only experiments with a fixed forgetting factor were performed.

- **Localization type and radius**: Finite ensemble sizes introduce spurious correlations between variables over long spatial distances. Localization is a method for truncating long-range correlations by neglecting observations beyond a certain distance to a model grid point. PDAF offers different localization types by weighting the correlations within a certain radius using a uniform weight, exponential decrease, or a 5th order polynomial.

- **Time-variable data assimilation interval**: Via an input file a fixed data assimilation interval is provided to PDAF. However, here, information from the observation files update the assimilation interval after each analysis step to account for the irregular temporal spacing of the GRACE observations.

- **Smoother option**: The smoother option of PDAF stores the model ensemble of a selected number of time steps. In the case of GRACE observations, this would make it necessary to stop the propagation of the model at least every day to fill the smoother ensemble. As this is not efficient, an alternative solution was implemented here, which computes running averages of the model state variables over user-specific time spans (e.g. one month for the standard GRACE solutions) and, subsequently, uses the averaged quantities for setting up the state vector in the analysis step.

The assimilation algorithm is controlled by user-supplied routines that are called during the analysis step. They depend on the filter algorithm and include the definition of the state vector, the observation vector, the observation covariance matrix, the observation operator, and the model update. Due to domain decomposition of CLM3.5, MPI communication is necessary e.g. to apply the observation operator (Section 7.2.3).

7.2 Interface for Assimilating Total Water Storage Anomalies

7.2.1 State Vector

For each grid cell, TWS is composed by 23 water storage compartments, i.e. 10 levels ($i = 1, ..., 10$) of soil liquid water $w_{liq,i}$, 10 levels of soil ice $w_{ice,i}$, snow water W_{sno}, canopy water W_{can} and water of the unconfined aquifer W_{aqu}.

During first data assimilation experiments, the ensemble of N_e model states ($\mathbf{X}_t^n$, $n = 1, ..., N_e$) was set up using all 23 water storage compartments separately. First, soil liquid water was introduced layer-wise for all grid cells, starting with the first soil layer $w_{liq,1}$ and ending with the 10th layer $w_{liq,10}$. Then, soil ice was added layer-wise, and finally snow water W_{sno}, canopy water W_{can} and water of the unconfined aquifer W_{aqu}. All components of the state vector have the unit mm. As the study area consists of 99 395 grid cells over land, the complete state vector has 23×99 395=2 286 085 entries. Due to domain decomposition, the model state vector is set up for all model 'clumps' separately, e.g. in the case of 8 processors per ensemble member each model 'clump' contains a state vector with about 286 000 entries.

However, first experiments showed that introducing soil liquid water and soil ice into the state vector separately leads to inconsistencies of the model after the analysis step. The reason for this is that due to different temperature distribution of individual ensemble members, different ensemble members can have (this happens only in few cases) completely different partitioning into soil liquid water and soil ice for selected layers and grid cells. As a result, after the analysis step, the distribution of soil ice and soil liquid water of a grid cell at a specific layer may disagree with other variables. Ultimately, this causes the model to crash.

Therefore, as an alternative and standard case in this thesis, the sum of soil liquid water and soil ice $w_{liq,1} + w_{ice,1}$ of each soil layer was used to set up the state vector instead of treating both variables separately. The size of the state vector is reduced correspondingly to 13×99 395=1 292 135 entries.

7.2.2 Observation Files

During pre-processing, observation files were created, which contain GRACE derived TWSA on the rotated CLM3.5 grid together with complementary information. A regular grid with longitude λ and latitude ϕ was designed in the rotated coordinate system of CLM3.5. Corresponding geographic coordinates λ_g and ϕ_g were obtained from the rotated coordinates according to

$$\phi_g = \arcsin\left(\sin\phi \sin\phi_g^N + \cos\phi\cos\lambda\cos\phi_g^N\right), \tag{7.1}$$

$$\lambda_g = \arctan\left(\frac{\cos\phi\sin\phi}{\sin\phi_g^N\cos\phi\cos\lambda - \sin\phi\cos\phi_g^N}\right), \tag{7.2}$$

where the coordinates of the rotated North Pole are $\lambda_g^N = 39.25°$ and $\phi_g^N = -162°$. GRACE derived spherical harmonics were then evaluated for the geographic coordinates λ_g and ϕ_g.

If no error is specified by the observation file, the error covariance matrix is diagonal and the standard deviation is provided by a command line option. Alternatively, the observation file contains a vector of the diagonal entries of the covariance matrix or a fully occupied covariance matrix derived according to Section 5.1.5. Additional variables provide the observation grid space and specify the model time steps until the next observation.

7.2.3 Observation Operator

GRACE observations pose particular challenges for the observation operator, which relates the model states to the observations. While GRACE observations are usually available as monthly solution with a spatial resolution of few hundred kilometers, CLM3.5 over Europe runs at hourly time steps at a 12.5 km grid. This temporal and spatial resolution mismatch has been addressed by previous studies (Sections 7.2.3.1 and 7.2.3.2). In the following subsections, an overview on different observation operators for GRACE observations is given along with the presentation of the observation operator implemented in this thesis.

7.2.3.1 Temporal Aggregation

Hourly simulation time steps of CLM3.5 have to be matched to monthly averages of TWSA from GRACE. This temporal resolution mismatch is addressed when setting up the state vector, computing the analysis increment, and applying the analysis increment to the model. Previous studies (Table 3.1) developed different approaches, which are displayed in Figure 7.3 and discussed in the following.

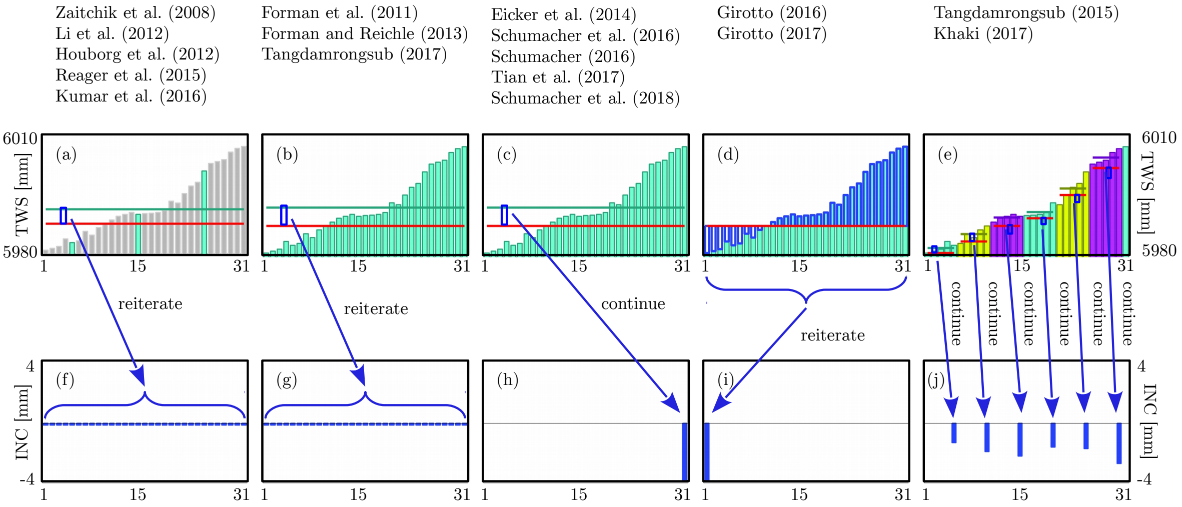


Figure 7.3: Calculation (a–e) and application (f–j) of the analysis increments (INC) as realized by the previous data assimilation studies from Table 3.1. Daily model states given in total water storage (TWS), which are used for setting up the monthly state vector, are shown in cyan and the corresponding temporal mean of the model states is indicated by corresponding vertical lines. The red vertical line indicates the observations.

Technically, the assimilation increment can be calculated at arbitrary points in time within the assimilation interval. In their initial study, Zaitchik et al. (2008) estimated a monthly increment from the mean of the model states at the 5th, 15th, and 25th day of a month to mimic the GRACE observation frequency. Several studies followed this approach (Li et al., 2012; Houborg et al., 2012; Reager et al., 2015; Kumar et al., 2016). Yet, it is questionable whether the choice of three specific days of the month better matches the monthly GRACE solution than an average over all days of the month as later on used by Forman et al. (2012); Forman and Reichle (2013); Eicker et al. (2014); Schumacher et al. (2016); Schumacher (2016); Tian et al. (2017); Schumacher et al. (2018) (case (b) and (c) in Figure 7.3). In contrast, Girotto et al. (2016) and Girotto et al. (2017) computed increments at each day of the month. In doing so, the instantaneous structure of the model as represented by the ensemble covariance matrix is exploited. Subsequently, a mean monthly increment was computed from the daily increments. Finally, Tangdamrongsub et al. (2015) and Khaki et al. (2017b) modified the approach of Forman et al. (2012) and Eicker et al. (2014) by interpolating monthly GRACE solutions to a time series with 5-day intervals. Then, the mean model state of a 5-day assimilation interval was used to compute the corresponding increment. However, interpolation introduces an additional error source to the GRACE observations. Furthermore, successive observations are temporally correlated, which means that the same information is introduced multiple times into the model.

The assimilation increment can be applied to the model in various ways. The early studies by Zaitchik et al. (2008) and Forman et al. (2012) divided the computed monthly increment by the number of days in the month, thus obtaining a uniform daily increment. Then, they reintegrated the model from the beginning of the current month adding the computed daily increment at each day of the month, thus distributing the monthly increment equally over the month (Figure 7.3 a, b). One disadvantage of this approach is that the model needs to be integrated twice for each month, which doubles the computing time. Another drawback is that the daily update with the same increment violates the requirement of the data assimilation algorithm that observation errors must be uncorrelated in time. Furthermore, during the first half of the month the updated state is still closer to forecasted TWS than to TWS from the analysis step, and the resulting monthly average of the second iteration is not consistent with TWS from the analysis step. Yet, in this way a smooth TWS time series is obtained with only small jumps between two successive months.

In contrast, Eicker et al. (2014) applied the whole monthly increment to the last day of the month, followed by the model integration over the next month. This approach was also used by Schumacher et al. (2016), Schumacher (2016), Tian et al. (2017), and Schumacher et al. (2018) (Figure 7.3 h). Eicker et al. (2014) adjusted the previous days of the current month by shifting them according to the monthly increment. This procedure ensures consistency between the monthly mean of updated daily values and the monthly mean from the analysis step. Tian et al. (2017) used a more sophisticated approach and updated the daily model states of the current month from information about the temporal correlation between the daily model state vectors. Similarly, Girotto et al. (2016) and Girotto et al. (2017) applied the whole monthly increment to the first day of the month and then reintegrated the model for the current month (Figure 7.3 i). Thus, updated daily values are obtained. Girotto et al. (2016) found that their approach of computing and applying the analysis increment performs slightly better than the approach of Zaitchik et al. (2008) when validating against root zone soil moisture and groundwater. It must be kept in mind that discontinuities are introduced between two successive months for all of these approaches (Figure 7.3 h, i).

Finally, Tangdamrongsub et al. (2015) and Khaki et al. (2017b) updated the last day of the 5-day assimilation interval and then continued model integration for the next 5 days. They argued that the 5-day interval is a good compromise for preserving the ensemble spread and simultaneously obtaining a smooth time series. However, as mentioned previously this approach violates the assumption of uncorrelated observation errors.

In this thesis, the approach of Eicker et al. (2014) was chosen as it is computationally efficient and allows for including other observation types easily into the data assimilation system. As the smoother options available in PDAF require constant assimilation intervals, here, temporal aggregation was realized by computing moving averages of soil liquid water $w_{liq,i}$, soil ice $w_{ice,i}$, snow water W_{sno}, canopy water W_{can} and water of the unconfined aquifer W_{aqu}. The moving averages of the TWS compartments were updated after each (hourly) model time step, and set to zero at the beginning of the next month. Observation files provide information on the number of model steps until the next available observation. The temporally averaged TWS components were used to set up the state vector (see Section 7.2.1). The analysis increments were applied to the model states of the last time step of the current month (according to Section 7.2.4), which were subsequently used to start the integration of the next month. Sub-monthly output of the current month was shifted by the monthly increment.

7.2.3.2 Spatial Aggregation

The model state vector holds 13 temporally averaged compartments of TWS for each of the 12.5 km × 12.5 km model grid cells (Section 7.2.1). The model state vector is transformed into observation space by the mapping operator, which includes (i) vertical aggregation of all TWS compartments, and (ii) horizontal averaging of the model grid cells to the coarser observation grid (which has a grid size of 0.5°, 1°, 2°, 3°, 4°, or 5°).

One challenge with respect to the horizontal averaging is the decomposition of the model domain into 'clumps', which run on different processors. Typically, data assimilation experiments were performed using 4 to 8 'clumps' depending on the available computing resources (see Section 8.2.3.1). Figure 7.4 and Figure 7.5 show the case of two model 'clumps' indicated by different colors. Model grid cells are distributed to 'clumps' row-wise. Observations were assigned to the 'clump' with most model grid cells within the observation grid cell, i.e the observation is saved on the selected 'clump', not assimilated.

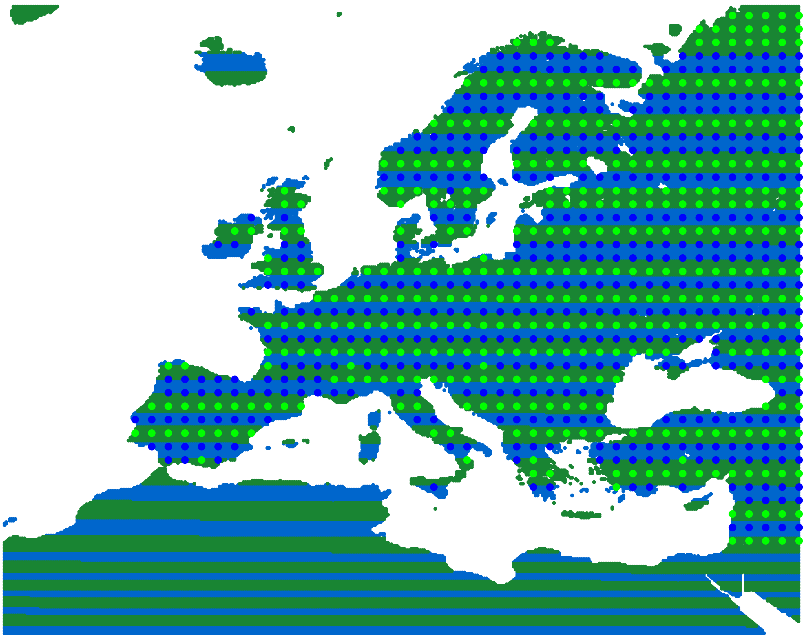


Figure 7.4: Model grid cells are grouped exemplarily in two 'clumps' that run on different processors. The observations (here given of a 1° grid) are assigned to the 'clump', which contributes most model grid cells to the observation grid cell.

Now, the observation operator has to realize averaging of TWS from model grid cells across different 'clumps'. First, for all model domains the sum of all TWS compartments was computed for each model grid cell separately. Second, TWS from all model grid cells within a defined distance in longitude and latitude direction to a specific observation was summed up for each 'clump' separately (Figure 7.5 a). Third, MPI communication was used to compute the total sum over all model 'clumps' for each observation grid cell. This sum was then divided by the number of contributing model grid cells leading to vertically aggregated and horizontally averaged TWS estimates for each observation grid cell (Figure 7.5 b). In the following, only observations grid cells were considered, which are supported by at least half of the number of possible model grid cells, i.e. in the case of a $1^\circ \times 1^\circ$ observation grid $((1^\circ)^2/(0.11^\circ)^2)/2 = 41$ model grid cells (0.11° is the resolution of the CLM grid).

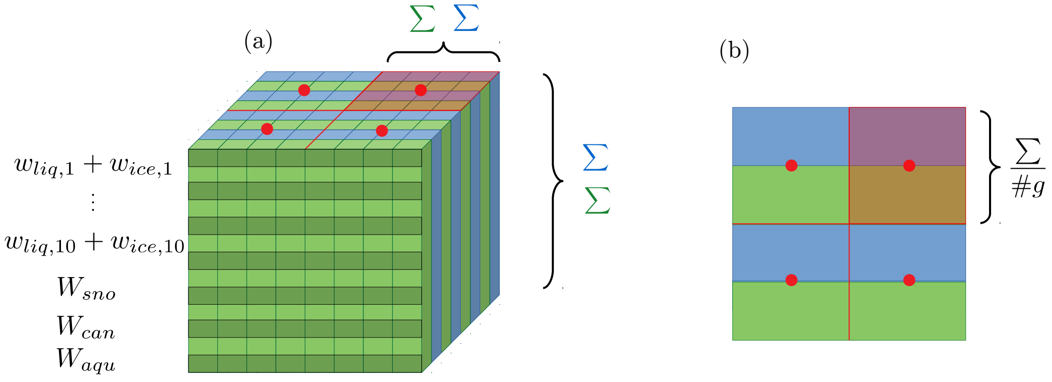


Figure 7.5: The mapping operator includes vertical aggregation over all TWS compartments and horizontal averaging of all model grid cells within a certain distance to the observations (red dots). Domain decomposition of the model into two 'clumps' is represented by green and blue color and requires (a) adding up TWS from all grid cells of a specific 'clumps' separately, followed (b) by the aggregation of TWS from different 'clumps', and the computation of the spatial average by dividing the obtained sum by the number of grid cells $\#g$.

7.2.4 Model Update

Generally speaking, the variables composing the model state vector are updated by the assimilation increments, which are obtained from the analysis step of the specific filter algorithm. In an ideal case, the update of the model state would be completely governed by the filter algorithm. However, the filter has no knowledge about the physical requirements of the model, which makes constraints necessary to guarantee the overall consistency of the model. In this regard two aspects play a role: (i) the updated model states need to be realistic, and (ii) model variables directly related to the state vector variables have to the updated correspondingly. An example for the first aspect is the elimination of negative values for water storage variables. An example for the second aspect is the update of soil liquid water, which involves also changes in the volumetric soil water content.

Previous studies often neglect increments of TWS compartments with small contribution (e.g., Zaitchik et al., 2008, Girotto et al., 2016) and argue that the increments would be spurious. Furthermore, Zaitchik et al. (2008) introduced a rule-based scheme to update snow fields in a physically reasonable way. Similarly, Tangdamrongsub et al. (2015) updated the relevant three TWS compartments of the model by adjusting, first, the major store, and only in case of remaining increments the two other stores.

Here, as few restrictions as possible shall be applied to the filter increments. In the following, the application of the assimilation increments to the model is described under consideration of necessary constraints. Updated TWS is obtained from the sum of all updated water storage compartments.

7.2.4.1 Soil Liquid Water W_{liq} and Soil Ice W_{ice}

Soil liquid water $w_{liq,i}$ and soil ice $w_{ice,i}$ of 10 soil levels ($i = 1, ..., 10$) were either introduced separately into the state vector or summed up layer-wise (typical case). In both cases, the increment was limited to a certain percentage b_{max} of the current value in order to avoid unrealistic and extremely abrupt changes. If the absolute value of the filter increment is larger than $b_{max}w_i$, with w_i representing $w_{liq,i}$ or $w_{ice,i}$, or in the aggregated case $w_{liq,i}+w_{ice,i}$, the increment is adjusted to

$$inc = \mathrm{sgn}(inc) b_{max} w_i, \tag{7.3}$$

where sgn is the signum function, and b_{max} is provided from the input file, which specifies the filter options. A typically value is $b_{max}=0.5$, which means that the increment can have a maximum amount of half of the current value.

In the case of aggregated $w_{liq,i}$ and $w_{ice,i}$, the update was computed considering the actual ratio of the two components according to

$$w_{liq,i} = w_{liq,i} + inc \frac{w_{liq,i}}{w_{liq,i} + w_{ice,i}}, \tag{7.4}$$

$$w_{ice,i} = w_{ice,i} + inc \frac{w_{ice,i}}{w_{liq,i} + w_{ice,i}}. \tag{7.5}$$

Eq. (7.3) ensures that the updated variables are positive. A lower limit of 0.01 mm is required for $w_{liq,i}$ as described in Section 4.1.2. Additionally, the upper limit of $w_{liq,i}$, $w_{ice,i}$, and $w_{liq,i}+w_{ice,i}$ is constrained to 1043 mm, which is also specified by the model. As model physics requires lower layers to have larger values of $w_{liq,i} + w_{ice,i}$ than the upper layer, in the (rare) case of $w_{liq,i} + w_{ice,i}$ being smaller than $w_{liq,i-1} + w_{ice,i-1}$, the difference is added to $w_{liq,i}$ and $w_{ice,i}$ using the same formulas as before (Eq. 7.4 and 7.5). Finally, the volumetric soil water content θ_i was adjusted according to Eq. (4.6) using updated $w_{liq,i}$ and $w_{ice,i}$.

7.2.4.2 Snow Water W_{sno}

Updating snow water W_{sno} is challenging due to multiple affected model variables like snow depth, the number of snow layers, and the thickness of snow layers. Furthermore, W_{sno} represents the total amount of snow water, which is distributed to a varying number of snow layers (Section 4.1.2).

In general, if the increment of snow water inc_{sno} is non-zero two cases may arise: either a model grid cell has zero snow water or a model grid cell has non-zero snow water. In the case of zero snow water, updated W_{sno} was set to inc_{sno} and the corresponding snow depth ΔZ was computed from the bulk density ρ_{sno} [kg m^{-3}] of newly fallen snow according to

$$\Delta Z = \frac{W_{sno}}{\rho_{sno}}, \tag{7.6}$$

with

$$\rho_{snw} = \begin{cases} 50 + 1.7(17)^{1.5} & T_{atm} > T_f + 2 \\ 50 + 1.7(T_{atm} - T_f + 15)^{1.5} & T_f - 15 < T_{atm} \leq T_f + 2 \\ 50 & T_{atm} \leq T_f - 15 \end{cases} \tag{7.7}$$

where T_{atm} [K] is the atmospheric temperature, and T_f [K] is the freezing temperature of water (Oleson et al., 2004). In the case of non-zero snow water, ΔZ was adjusted by the same factor as W_{sno} according to

$$\Delta Z = k_{sno} \Delta Z, \tag{7.8}$$

with

$$k_{sno} = \frac{W_{sno} + inc_{sno}}{W_{sno}}. \tag{7.9}$$

W_{sno} was restricted to the interval between 0 mm and 1000 mm before computing snow depth ΔZ.

In case that at least one snow layer exists, liquid water content $w_{liq,i}$, ice content $w_{ice,i}$, layer thickness Δz_i, node depth z_i, and layer interface depth $z_{h,i}$ are also adjusted by multiplying with k_{sno}.

Finally, the number of snow layers needs to be adjusted. If ΔZ is smaller than 10 mm the number of snow layers is set to zero. If the number of snow layers is zero, but ΔZ is larger than 10 mm a new snow layer is initialized according to Eq. (4.4).

7.2.4.3 Canopy Water W_{can}

Canopy water was updated using the increment obtained from the filter algorithm. As only constraint, negative values were set to zero.

7.2.4.4 Water of the Unconfined Aquifer W_{aqu}

Water of the unconfined aquifer W_{aqu} reaches its prescribed maximum value of 5 m for most grid cells over Europe (Section 4.1.2) and changes slowly with maximum values between 5 and 10 mm/month. For the update, Eq. (7.3) with $b_{max} = 0.001$ was applied, which corresponds to maximum changes of 5 mm. Additionally, lower and upper boundaries were set to 4000 mm and 5000 mm, respectively. The upper boundary is the maximum defined by the model code and the lower boundary corresponds approximately to the minimum value obtained for a simulation over 9 years without data assimilation.

7.3 Generating an Ensemble of Model Runs with CLM3.5

Ensemble based data assimilation algorithms (Chapter 6.1) obtain model error information from the spread of an ensemble of model states. Besides a simplified representation of reality due to model physics and discretization, three additional sources of model errors arise, the uncertainty of (i) empirical model parameters, (ii) atmospheric forcing data, and (iii) land surface data sets. Producing a set of physically realistic model runs is still a field of research and typically subject of sensitivity studies (e.g., Forman and Margulis, 2010; Li et al., 2013; Göhler, 2013). Based on the experience of previous studies (Reichle et al., 2007, 2010; Han et al., 2012, 2014), here, the ensemble was generated by perturbing initial conditions, atmospheric forcings, and land surface characteristics.

Figure 7.6 shows the ensemble spread of TWS in November 2002 for a model run starting with perturbed initial conditions in January 2002. Perturbing only the atmospheric forcings (Figure 7.6 a) results into an ensemble spread smaller than 1 cm, whereas the perturbation of soil parameters (Figure 7.6 b) or of both (Figure 7.6 c) leads to a spread between 1 cm and 2.5 cm. This means, that the model is more confident over Europe than the rescaled GRACE TWSA, which have a standard deviation of about 2.5 cm to 3 cm over Europe (Figure 5.9 d).

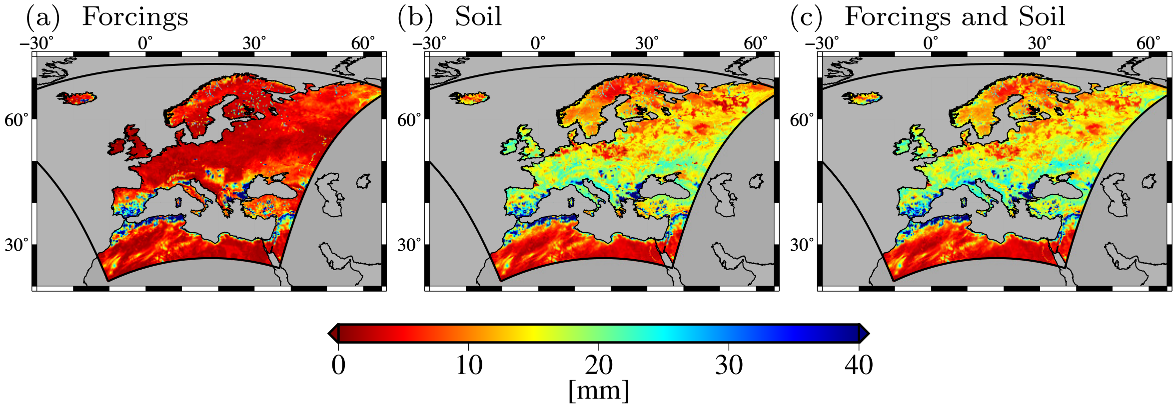


Figure 7.6: Ensemble spread of TWS in November 2002 when (a) perturbing atmospheric forcings only, (b) perturbing land surface characteristics only, and (c) perturbing both forcings and land surface characteristics.

7.3.1 Initial Conditions

During the spin-up period, the initial conditions of a model are stabilized (Rahman and Lu, 2015). The required spin-up period depends on atmospheric forcing and surface conditions. The spin-up of CLM3.5 was realized by running the model 80 years with the same atmospheric forcings from the year 2002. Then, the model reaches an equilibrium state, i.e. water and energy fluxes change less then 0.1 % from one year to the next year over Europe.

Each ensemble member of CLM3.5 was started from different initial conditions. In this scope, CLM3.5 was first run for one ensemble member for the time period 2002 to 2010 using initial conditions obtained after the spin-up phase. An ensemble of initial conditions was subsequently generated from January of each year using every second day. Starting the ensemble of model runs in January 2002, the ensemble spread shrinks until mid 2002, and subsequently remains constant with some variability related to the seasonal signal.

7.3.2 Atmospheric Forcings

Four forcing variables were perturbed with spatially, temporally, and cross-correlated random fields according to Table 7.1:

- PRECTmms: Surface precipitation [mm/s]
- TBOT: Temperatur at lowest atmospheric level [K]
- FSDS: Incident solar radiation [W/m2]
- FLDS: Incident long-wave radiation [W/m2]

The magnitude of perturbations was chosen similar to Reichle et al. (2010) and Han et al. (2014) after tuning experiments. Normally distributed multiplicative perturbations with a mean value of 1 and a standard deviation of 0.3 were applied to surface precipitation and solar short-wave radiation fields, whereas zero-mean normally distributed additive perturbations were added to temperature (standard deviation 2 K) and long-wave radiation (standard deviation 30 W/m^2) forcings. Unrealistic forcing data were avoided by truncating the noise to minimum and maximum values shown in Column 4 of Table 7.1.

Spatial correlation between two grid cells $r_{i,j}$ was generated by assuming an isotropic correlation structure depending on the distance $d_{i,j}$ between the two grid points,

$$r_{i,j} = \exp\left(-\frac{d_{i,j}}{L}\right), \tag{7.10}$$

with the spatial correlation length L given in Column 5 of Table 7.1. To reduce computational costs, the spatial correlations were computed for a grid size of 100×100 and then projected to the 424×436=184864 grid cells of the forcing data via interpolation and by adding Gaussian noise with zero mean and a standard deviation of 0.2 at each grid point. Then, the spatial correlations were applied to the normally distributed noise for each field separately.

Temporal correlation was achieved using a first-order autoregressive model following Evensen (2009),

$$\mathbf{s}_t = \rho \mathbf{s}_{t-1} + \sqrt{1-\rho^2}\mathbf{w}_t, \tag{7.11}$$

where $\mathbf{w}_t$ is the 184864×1 vector of spatially correlated random numbers, and $\mathbf{s}_t$ and $\mathbf{s}_{t-1}$ are the perturbation fields from the current time step and the previous time step, respectively. The temporal persistence parameter ρ determines the temporal correlation of the stochastic forcings according

$$\rho = 1 - \frac{\Delta t}{\tau},$$

where the length of the model time step Δt equals 3 hours, and the decorrelation time τ was chosen to 24 hours in line with Reichle et al. (2010).

Additionally, error cross correlations between the four forcing variables were imposed in line with Reichle et al. (2010) and Han et al. (2014) (Column 6 of Table 7.1). Introducing correlation between variables ensures that the forcing fields have a realistic balance, e.g. a positive perturbation of air temperature is related to a positive perturbation of incoming short-wave radiation and a negative perturbation of incoming long-wave radiation. Finally, long wave radiation was restricted to be larger or equal to zero.

Table 7.1: The atmospheric forcing data of CLM3.5 is perturbed by noise defined through the parameters of the additive and multiplicative perturbation fields including cross correlations given in this table. Perturbations are truncated to a certain interval provided in column 4.

Variables	**Noise**	**Standard deviation**	**Minimum, Maximum**	**Spatial correlation scale**	**Cross Correlation of perturbations in**			
					P	**T**	**SR**	**LR**
Precipitation	Mult.	0.3	[0.3, 1.7]	80 km	1	0	-0.8	0.5
Temperature	Add.	2 K	[-5, 5]	250 km	0	1	0.4	0.4
Solar radiation	Mult.	0.3	[0.3, 1.7]	250 km	-0.8	0.4	1	-0.5
Long wave rad.	Add.	$30\,W/m^2$	[-70, 70]	250 km	0.5	0.4	-0.5	1

Each of the up to 128 ensemble members needs a different time series of perturbed forcings. Technically, in order to save storage, perturbation of the forcing fields was realized by providing only one additional file to TerrSysMP. This perturbation file is read by each ensemble member starting at a different point in time, thus, ensuring consistent and individual perturbations.

7.3.3 Soil Texture

Thermal and hydrological properties of CLM3.5 are determined from the percentages of clay and sand, which are provided to the model in form of a soil map (Section 4.1.3.2, Figure 4.4). For all soil levels the same properties are applied. In line with Han et al. (2014), spatially uniform noise in the range of $\pm 10\,\%$ was added to the Food and Agriculture Organization / United Nations Educational, Scientific and Cultural Organization (FAO/UNESCO) soil map. Constraints were applied to ensure clay and sand percentages and the corresponding sum having values between 0 and $100\,\%$. In the case that the sum of both variables exceeds the prescribed range, half of the surplus or deficit is removed from each variable. Furthermore, for both variables a minimum of $5\,\%$ was imposed in line with the original data set.

Chapter 8

Data Assimilation Experiments

The assimilation of total water storage anomalies (TWSA) from the Gravity Recovery and Climate Experiment (GRACE) mission into the Community Land Model version 3.5 (CLM3.5) was performed for the European COordinated Regional Downscaling EXperiment (CORDEX) area (Section 4.1.3, Figure 4.3). The study domain includes 26 large European river basins, which are aggregated for evaluation purposes into seven major catchments with different climatic conditions (Figure 8.1). The border of each basin is associated with the most downstream discharge gauging station, which has two advantages: (i) the validation of basin-averaged runoff simulated by CLM3.5 using measured discharge from gauging stations is straight forward, and (ii) coastal areas, where attenuation of filtered GRACE-derived TWSA is high due to leakage from the ocean, are excluded from basin-wise comparisons of TWSA simulated by CLM3.5 and TWSA observed by GRACE.

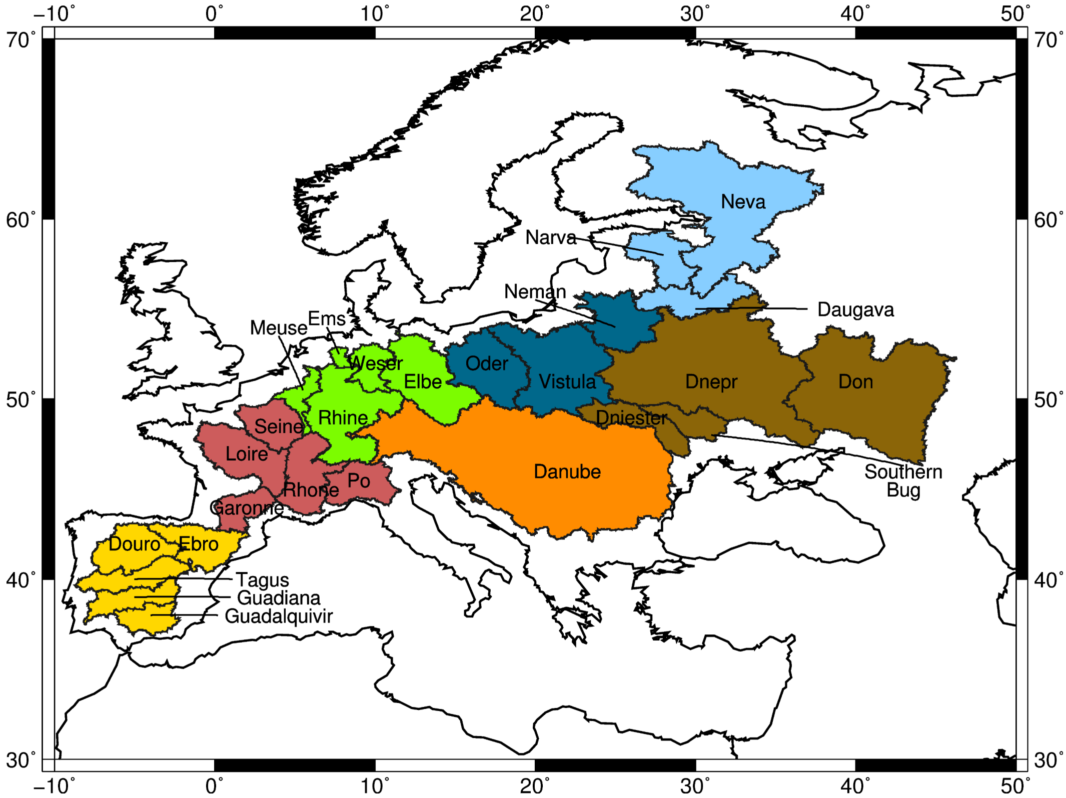


Figure 8.1: The study area of this thesis covers 26 European river basins, which are aggregated into seven major catchments as indicated by the different colors.

The size of the aggregated catchments ranges from 351 000 km^2 for the Iberian Peninsular (IP) to 953 000 km^2 for the rivers of the Black Sea (BL) catchment:

- Iberian Peninsular (**IP**, 351 000 km^2): Ebro, Duero, Tagus, Guadiana, Guadalquivir,
- France-Italy (**FI**, 393 000 km^2): Garonne, Loire, Seine, Rhone, Po,
- North Sea (**NS** 359 000km^2): Meuse, Rhine, Ems, Weser, Elbe,
- Danube (**DB**, 807 000 km^2),
- Baltic Sea 1 (**BS1**, 385 000 km^2): Oder, Vistula, Neman,
- Baltic Sea 2 (**BS2**, 402 000 km^2): Daugava, Narva, Neva,
- Black Sea (**BL**, 953 000 km^2): Dniester, Southern Bug, Dnepr, Don.

Climatic conditions vary over Europe (Peel et al., 2007). Western Europe is dominated by the Gulf Stream, which leads to an Oceanic climate with narrow annual temperature ranges and precipitation throughout the year. The Mediterranean climate of Southern Europe is characterized by rainy winters and hot and dry summers. Eastern Europe has a Continental climate with cold winters, hot summers, and moderate precipitation.

Assimilation of GRACE-derived TWSA into the CLM3.5 was realized within the Terrestrial Systems Modeling Platform - Parallel Data Assimilation Framework (TerrSysMP-PDAF). The data assimilation strategy can be specified via a number of control parameters, e.g. filter algorithm, ensemble size, observation error model, etc. For the assimilation of GRACE observations, previous studies (Table 3.1) focused on individual control parameters applying different hydrological models and study regions. So far, no standard procedure exists for assimilating GRACE data into hydrological models. Control parameters that were found to be relevant by previous studies were tested here for CLM3.5 over Europe. One major aim of this thesis was a thorough assessment of the importance of the individual control parameters.

- The **filter type** defines the algorithm applied in the update step (Section 6.1). While Schumacher et al. (2016) noticed a rather small impact of different filter types on the performance of data assimilation, Khaki et al. (2017a) found different performance for individual filter types. In this thesis, the Ensemble Kalman Filter (EnKF) as the most common filter was used and compared to the Error Subspace Transform Kalman Filter (ESTKF), which is a computationally more efficient filter with favorable properties (Section 6.1.5).
- The **ensemble size** must be large enough for a reasonable representation of the model error covariance in order to minimize sampling errors (see Section 6.2). Previous data assimilation experiments using GRACE data chose between 20 and 30 ensemble members (Zaitchik et al., 2008; Forman et al., 2012; Eicker et al., 2014; Schumacher et al., 2016; Girotto et al., 2016). However, in this thesis, experiments with 8 to 128 ensemble members were evaluated (Section 8.2.3).
- The **forgetting factor** counteracts the underestimation of the model state errors due to undersampling (Section 6.2.2). A typical forgetting factor of $\rho = 0.7$ corresponds to an inflation of the model ensemble by the factor $r = 1/\sqrt{\rho} = 1.2$. The impact of different forgetting factors on the assimilation results is analyzed in Section 8.2.5.

- **Localization** suppresses spurious correlations caused by ensemble undersampling and long range correlations present in the full GRACE error covariance matrix. Khaki et al. (2017b) found improvements of up to 25% in terms of root mean square deviation (RMSD) when using local analysis instead of global analysis. Both, EnKF and ESTKF are available with localization options in PDAF, leading to the Local Ensemble Kalman Filter (LEnKF) and to the Local Error Subspace Transform Kalman Filter (LESTKF). An exponentially decreasing weight function is applied, where the slope of the function is governed by the radius R_e, at which the weight $1/e$ is reached. Additionally, the cutoff radius R_c, at which grid cells obtain zero weight, has to be chosen. The optimal choice of R_e and R_c ensures enough common observations with the neighboring grid cell to avoid blocking of the filter update (Section 8.2.5). In this thesis, tuning experiments were performed to choose R_e and R_c.

- Influence of the **grid space of the observations** has been studied previously by Khaki et al. (2017b) for the World Wide Water Resources Assessment (W3RA) model running at a 1° grid over Australia. Applying localization, they found the best performance of data assimilation for an observation grid space of 3°. In Sections 8.2.6 and 8.3.2 the optimal observation grid space for assimilating GRACE data into the high-resolution Community Land Model version 3.5 (CLM3.5) model is investigated.

- Previous studies showed that correlations present in the GRACE **observation error covariance matrix** cannot be neglected (Schumacher et al., 2016; Tangdamrongsub et al., 2017). Especially when assimilating GRACE data given on a grid with small grid space neglecting correlations would correspond to undue trust into the observations. The impact of different error assumptions is treated in Sections 8.2.4 and 8.2.6.

Basic characteristics of the data assimilation framework were assessed within synthetic experiments (Section 8.2). Influence from ensemble size, observation errors, and from the data assimilation strategy are reported and placed into findings of previous studies. By contrasting the assimilation run against the synthetic truth, the effect of data assimilation on different compartments was analyzed. Synthetic experiments were also applied to assess the influence from biases in model forcings and the influence from phase shifts between model and observations. Finally, the synthetic experiments provide information on the optimal assimilation strategy, which was then applied to the real-case scenario.

The real-case scenario (Section 8.3) provides insights into the impact of GRACE assimilation on different water cycle related model variables. In particular, the sensitivity of predicted TWSA, soil moisture, river discharge, and evapotranspiration was assessed by validating against observations-based data sets (Section 4.3). Furthermore, the assimilation of GRACE observations into CLM resulted into a high-resolution reanalysis of TWSA, which was used to investigate the representation of extreme events.

For both, synthetic and real-case experiments, ensemble runs start in open-loop mode in 2002 and data assimilation experiments start in January 2003. The longest data assimilation runs cover the time span 2003 to 2010. Ensembles of model runs were generated as described in Section 7.3. GRACE observations over Westafrica are not considered. The model runs at hourly time steps and outputs are either daily or monthly averaged states of selected variables:

- total water storage (TWS) and its components soil liquid water $w_{liq,j}$ and soil ice $w_{ice,j}$ at $j = 1, ..., 10$ layers, snow water W_{sno}, canopy water W_{can}, and water in the unconfined aquifer W_{aqu},

- soil moisture θ_j as volumetric water content at $j = 1, ..., 10$ layers,

- runoff q including runoff from the surface q_{over}, runoff from glaciers wetlands and lakes q_{rgwl}, and subsurface drainage q_{drai},
- actual evapotranspiration E_a including evaporation from the ground E_{soil}, evaporation from vegetation E_{vege}, and transpiration from vegetation E_{vegt}.

For a single ensemble member, one year of monthly TWS output including its components results into 0.21 GB of data in binary Network Common Data Format (NetCDF). Daily output requires storage capacities of 6.5 GB per year. Thus, daily output over an 8 year time span requires about 3.3 TB disk space for a typical experiment with 64 ensemble members. During post-processing storage space was reduced by merging output files and by saving only the ensemble mean and standard deviation of the ensemble. The necessary operation on the NetCDF model output was performed using Climate Data Operators (CDO) scripts. Finally, the data assimilation experiments were evaluated based on the metrics summarized in Section 8.1.

8.1 Validation Metrics

Model output from data assimilation experiments were validated either against the synthetic truth (Section 8.2) or against observation-based data sets (Section 8.3). In the following, validation metrics that are common either in geodesy or hydrology are introduced. Validation metrics were either computed for each grid cell of the European CORDEX region (excluding North-Africa) or for selected catchments introduced in Figure 8.1.

The Root Mean Squared Deviation (RMSD) is a measure of accuracy of a modeled time series (M) with respect to a reference (or observation) time series (RF). Small RMSD values stand for a small bias and variance between two time series. The RMSD is defined as the square root of the averaged squared differences between two time series x^M and x^{RF},

$$RMSD = \sqrt{\frac{\sum_{t=1}^{T}\left(x_t^M - x_t^{RF}\right)^2}{T}}. \tag{8.1}$$

Obviously, large differences between x^M and x^{RF} have a higher weight on the RMSD than small differences.

The Pearson's correlation coefficient measures the linear relationship between two time series. The correlation coefficient is defined as

$$\rho = \frac{\sigma_{M,RF}}{\sigma_M \sigma_{RF}}, \tag{8.2}$$

where σ_M and σ_{RF} denote the standard deviations of the time series x^M and x^{RF}, respectively. Furthermore, $\sigma_{M,RF}$ is the covariance between the two time series. The correlation coefficient ranges from -1 to 1. A value of 1 denotes a perfect linear relationship between x^M and x^{RF}, wheres a value of 0 means that there is no linear relationship. Negative correlation coefficients indicate anti-correlation which means that x^M decreases while x^{RF} increases. Water cycle related variables often have a distinct annual cycle, which naturally leads to large correlation coefficients. In order to determine correlations beyond the seasonal cycle and trends, the following mathematical model was subtracted from the time series:

$$x_s = a + bt + c\sin\left(\frac{2\pi}{T}t\right) + d\cos\left(\frac{2\pi}{T}t\right) + e\sin\left(\frac{2\pi}{T/2}t\right) + f\cos\left(\frac{2\pi}{T/2}t\right), \tag{8.3}$$

where t denotes time, T is the annual period, and the variables a to f are estimated from the data. The correlation coefficient between de-seasoned and de-trended time series is further on denoted as residual correlation coefficient.

Differences in amplitudes and phase shifts between hydrological models and observation-based data sets are common. Annual amplitudes are derived from the parameters c and d of Eq. 8.3. The relative annual amplitude between time series from model and observations is computed according to

$$A_{rel} = \frac{A^{M}}{A^{RF}}. \tag{8.4}$$

A relative annual amplitude of 1 means perfect agreement of modeled amplitude and observed amplitude. A value smaller than 1 means that the model underestimates the annual amplitude and a value larger than 1 means that the model overestimates the annual amplitude. The trend of the respective time series is provided by parameter the b of Eq. 8.3.

In the context of river discharge it is common to measure the skill of a model in simulating true discharge by the Nash-Sutcliffe (NS) coefficient. The NS coefficient is computed for modeled discharge x^{M} and observed discharge x^{RF} according to

$$NS = 1 - \frac{\sum_t \left(\sqrt{x_t^{RF}} - \sqrt{x_t^{M}}\right)^2}{\sum_t \left(\sqrt{x_t^{RF}} - \sqrt{\overline{x_t^{M}}}\right)^2}, \tag{8.5}$$

where $\overline{x_t^{M}}$ is the temporal mean of observed discharge. Applying the root of discharge ensures that high discharge values do not obtain too much weight. A value of 1 indicates perfect agreement between observed and modeled discharge, values between 0 and 1 mean that the model better simulates discharge than the mean of observed discharge. In the case of negative NS coefficients, the mean of observed discharge better describes the actual discharge than the model.

8.2 Assimilation of Synthetic Total Water Storage Observations

8.2.1 Twin Experiments

Observation System Simulation Experiments (OSSEs) are a valuable tool for investigating the impact of specific observation types on data assimilation results. Generally, a free model run is defined as truth or reference run (RF) from which, subsequently, synthetic observations are derived. The synthetic observations perturbed with suitable errors are assimilated into a different model. Then, the results from the assimilated model (DA) are compared to RF. In the case of a twin experiment RF and DA are based on the same model. Twin experiments with synthetic GRACE observations were set up by Forman and Reichle (2013) for the Mackenzie River basin and by Schumacher et al. (2016) for the Mississippi River basin in order to investigate the impact of observation grid space and error models. In this thesis, twin experiments were set up to study the influence from a number of different control parameters that define the data assimilation strategy, e.g., filter type, ensemble size, observation error model, etc. These experiments allow to quantify the relevance of the individual control parameters and to identify the most appropriate settings for assimilating GRACE data into CLM3.5. The optimal set of control paramters was then applied when assimilating real GRACE data into CLM3.5 (Section 8.3).

8.2.1.1 Set-Up of the Twin Experiments

Twin experiments require setting up a reference or true run (RF) and an ensemble of perturbed model runs (Figure 8.2). RF was generated using the atmospheric forcing data set composed by the COnsortium for Small Scale MOdelling REAnalysis at 6 km horizonal resolution (COSMO-REA6) and Princeton Meteorological Forcings (Section 4.1.3.1) and soil texture from ESDAC (Section 4.1.3.2). An equilibrium state was achieved after 80 years of spin-up. Then the model was run for the year 2002 to obtain initial states for 2003. Ensemble runs of the perturbed model were performed using perturbed atmospheric data from the Weather Research and Forecasting (WRF) model and perturbed soil texture from the Food and Agriculture Organization (FAO, Sections 4.1.3.2, 7.3.2, 7.3.3). Perturbed initial conditions were obtained after 80 years of spin-up (Section 7.3.1). The model was run for all ensemble members for the year 2002 and the ensemble spread was found to stabilize after about 6 months. Ensemble runs starting in 2003 were either performed in open-loop mode (OL, i.e. no data assimilation) or in data assimilation mode (DA).

Synthetic TWS observations were derived from TWS output based on RF. In a first step, the temporal mean of TWS of RF was adjusted to the temporal mean of the ensemble mean of OL. Using 64 ensemble members, the difference between the ensemble mean of the perturbed model and the original unperturbed model is negligible. Adjusting the temporal mean mimics missing information about absolute TWS, as it is the case for real GRACE observations. In the next step, TWS of RF was spatially averaged to selected observation grids (between 0.5° and 5°). Finally, GRACE-like observation errors were added by correlating random Gaussian noise generated for each individual month using the full GRACE error covariance matrix of January 2007 (Figure 5.9). The resulting (perturbed) synthetic observations were introduced into the data assimilation framework TerrSysMP-PDAF and contribute to the assimilated model run (DA).

The data assimilation framework allows for assuming different observation error models. Colored or correlated noise is represented by the full covariance matrix that was also used for perturbing the synthetic observations. As simplifications the main diagonal of the GRACE-like covariance matrix can be used (white noise) or uniform noise can be assumed. Please note that for all twin experiments we distinguish between simulated observation errors (used for perturbing the synthetic observations) and assumed observation errors (used as error information in the data assimilation framework).

Finally, output from OL and DA was validated against RF for selected model states. If not specified otherwise, validation was performed for monthly averaged model states. Model output, however, can be generated arbitrarily at hourly, daily, or monthly time steps.

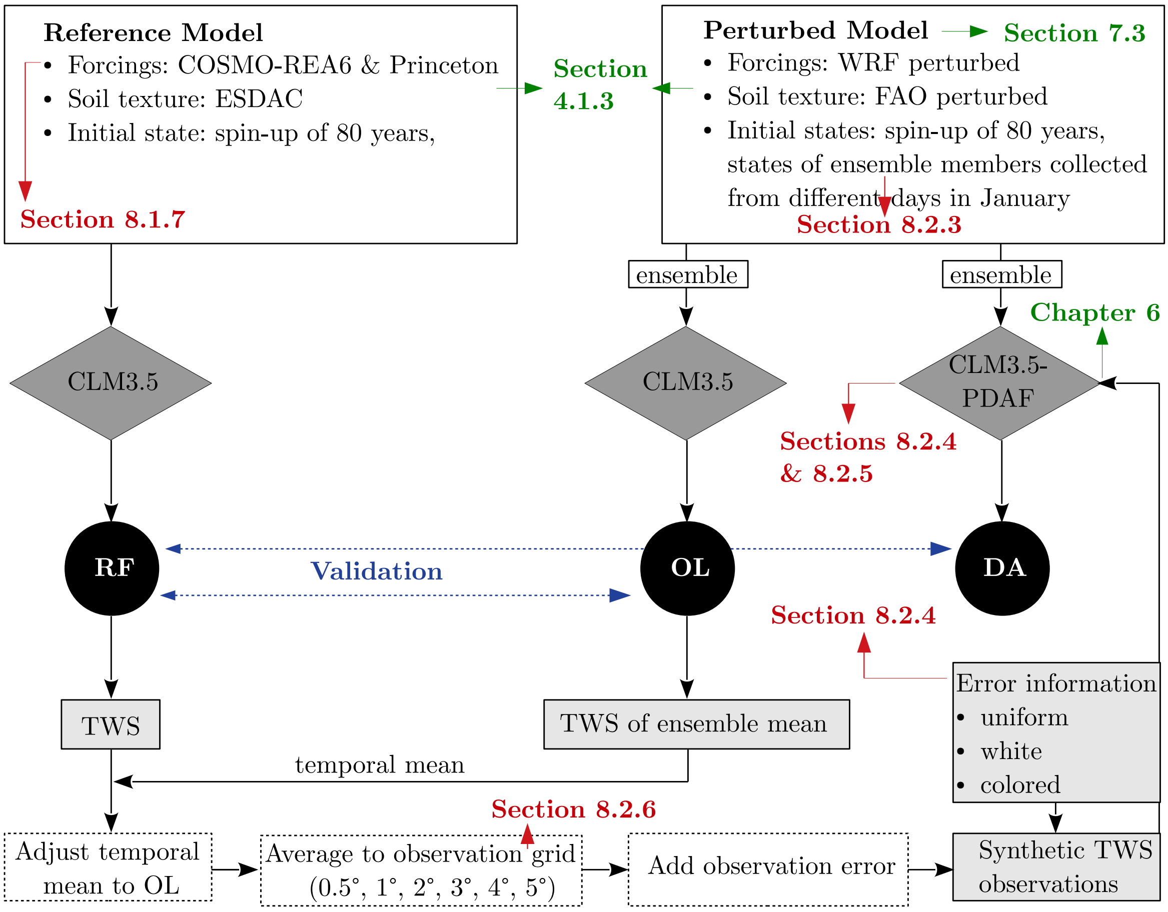
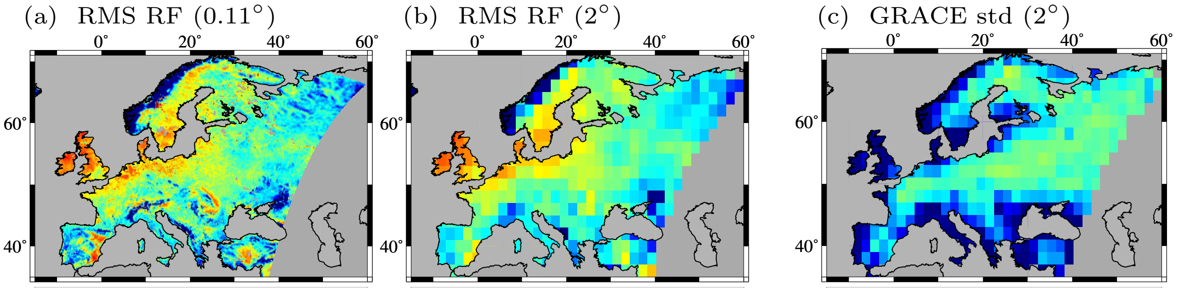


Figure 8.2: Flowchart visualizing a twin experiment, in which synthetic observations are derived from a reference CLM3.5 model run (RF) and assimilated into a perturbed CLM3.5 model run to obtain the assimilated CLM3.5 model (DA). Additionally, an open-loop run (OL) of the perturbed model is performed. Green arrows indicate sections where the corresponding aspects are discussed theoretically. Red arrows indicate Sections of related data assimilation experiments.

8.2.1.2 Starting Point of the Twin Experiments

The RMS variability of TWS computed from monthly outputs of RF and OL is shown in Figure 8.3, for the native CLM grid resolution of 0.11° (a and d) and for a typical 2° resolution of the observations grid (b and e). Overall, RF has a higher variability than OL especially in Eastern Europe, which is mainly due to the underestimated annual amplitude of TWS from OL by up to 50 % compared to RF. Differences between OL and RL are mainly attributed to the two atmospheric forcing data sets, which differ in particular over Eastern Europe. The huge impact of different forcing data sets on hydrological modeling has been assessed by previous studies (e.g., Berg et al., 2003; Decharme and Douville, 2006; Wang and Zeng, 2011; Nasonova et al., 2011; Houborg et al., 2012; Poméon et al., 2017).

Errors for perturbing the synthetic observations (Figure 8.3 c) were propagated from the full error covariance matrix of January 2007 of the ITSG-Grace2016 level-2 solution (see Section 5.1.5.2). Errors of the OL run were derived for each month from the ensemble spread. Here, a temporal average of the ensemble spread for 2003 to 2006 is shown (Figure 8.3 f). The standard deviation derived from GRACE amounts to about 25 mm to 30 mm inside of the continent and reaches values larger than 50 mm at the coast. In comparison, the temporal average of the ensemble spread of TWS from OL is smaller with 15 mm to 20 mm. However, the ensemble spread varies with the seasonal signal, and is larger for storage minima and maxima in autumn and spring than during summer and winter time (see Section 8.2.2.2).


(a) RMS RF (0.11°) (b) RMS RF (2°) (c) GRACE std (2°)

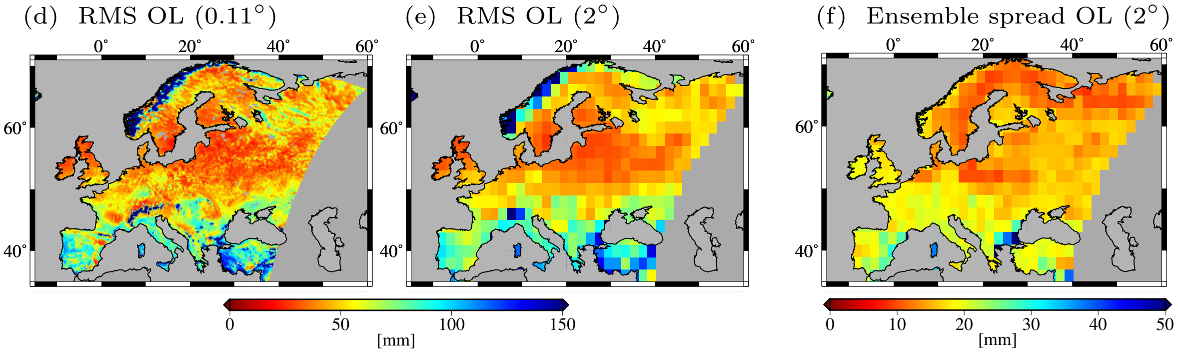

(d) RMS OL (0.11°) (e) RMS OL (2°) (f) Ensemble spread OL (2°)

Figure 8.3: The RMS variability of TWS simulated by the reference run (RF) and by the open-loop run (OL) during the time period 2003 to 2006 is shown for the native CLM grid resolution of 0.11° (a, d) and for a typical resolution of the observations of 2° (b, e). (c) Errors of the synthetic observations were propagated from the full error covariance matrix of the GRACE solution of January 2007 considering also leakage and rescaling covariance matrices (see Section 5.1.5.2). (f) The average error of the perturbed model run during the study period was computed from the temporal mean of the ensemble spread.

8.2.1.3 Outline of the Twin Experiments

Assimilation results are considerably influenced by the control parameters of the data assimilation framework. Twin experiments aim at (i) investigating the relevance of the individual control parameters, and (ii) determining the most appropriate settings of the control parameters. Within this scope, a number of data assimilation experiments were performed.

First, a simple experiment without perturbing the synthetic observations was set up in order to obtain a general picture of the effect of data assimilation on different storage compartments (Section 8.2.2). Further on, the effect of the ensemble size was characterized and the corresponding computational costs were derived from scaling experiments (Section 8.2.3). Results from different filter algorithms were analyzed assuming either white or correlated observation noise (Section 8.2.4). Taking the previous experiments into account, a localized data assimilation algorithm was tuned regarding the localization radius, the shape of the localization function, and the strength of covariance inflation (Section 8.2.5). Finally, the impact of different observation grids was investigated considering 0.5° to 5° grids (Section 8.2.6).

Additionally, experiments with manipulated forcing data allowed for characterizing the influence of biases in precipitation forcings (Section 8.2.8). Impacts from the very common problem of phase shifts between hydrological models and GRACE observation were investigated by assimilating phase-shifted synthetic observations (Section 8.2.9).

Please note that Kalman filters do not necessarily improve the target model everywhere. Indeed, the distribution of the assimilation increment is governed by correlations between observation grid cells, model grid cells, and cross-correlations between model and observation grid cells (see Section 6.1). This complex correlation structure is the core of the Kalman filter and can lead to unexpected results.

All twin experiments were performed for the time span 2003 to 2006. In general, 64 ensemble members were applied. Validation measures were computed for individual river catchments and for the whole study area excluding North Africa.

8.2.2 Influence of Data Assimilation on the TWS Compartments

This introductory experiment aims at providing a general picture of the impact of data assimilation on the individual TWS compartments and on regional and seasonal patterns. In this scope, the most simple and straightforward settings were chosen for the data assimilation framework. Unperturbed observations were assimilated assuming uniform observation errors of 2 cm. The temporal mean of the observations was not adjusted to enable comparison between the absolute values of TWS compartments simulated by RF and DA. Further settings of the DA experiment are reported in Table 8.1.

Table 8.1: Experiment 1 – Influence of the assimilation increments on TWS compartments

Filter algorithm	EnKF	**Ensemble size**	64
Inflation factor	0.7	**Observations**	RF, not perturbed, 2°
Localization	–	**Assumed errors**	uniform (2 cm)

Sensitivity of the TWS compartments to data assimilation is expected to depend on the respective magnitudes and ensemble spreads. The impact of assimilation is expected to be larger in Eastern European catchments than in Western and Central Europe, as here differences between OL and RF are larger.

The first analysis step of the data assimilation framework takes place after running the model for January 2003. RF shows smaller scale features than the history output (HO) before assimilation due to the higher resolution soil map (Figure 8.4). Figure 8.4 shows major differences of monthly averaged TWS simulated by RF and HO for North-Eastern Europe, Scandinavia, and close to the Black Sea. Assimilation increments (Figure 8.4 c) indicate water abstraction (negative or red increments) for the Iberian Peninsula, which was simulated too wet by HO compared to RF. Likewise, water is abstracted north of the Black Sea. In contrast, boreal regions of Finland and Russia were simulated too dry by HO, which results in water injection (positive or blue increments). Over Central Europe, TWS from RF is larger than TWS from HO resulting in small negative increments. All in all, in this first assimilation time step a water column of about 1 cm thickness is abstracted from the model.

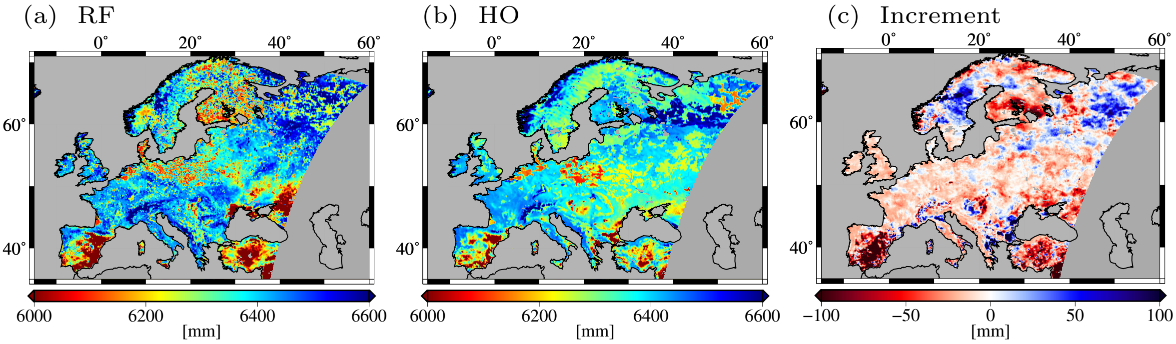


Figure 8.4: Absolute TWS in January 2003 shown for (a) the monthly mean of the reference run (RF) and (b) the monthly mean of the history files (HO). The corresponding monthly increment, which is added to HO to obtain the monthly mean of the updated model (DA), is displayed in (c). Blue color indicates water injection, and red color indicates water abstraction.

8.2.2.1 Evaluation of TWS from OL and DA

The improvement of TWS simulated by the assimilated model (DA) over TWS simulated by the open-loop run (OL) was assessed by computing the correlation coefficient for each grid cell with respect to RF, considering the study period 2003 to 2006 (Figure 8.5 a, b). On the average, data assimilation increases the correlation coefficient from 0.63 for OL to 0.68 for DA. However, the improvement of the model shows strong regional differences (Figure 8.5 c). Degradation in terms of correlation coefficients are identified for the Iberian peninsula. Largest improvements are found for the Baltic states, where correlation coefficients increase from about 0.3 to about 0.8.

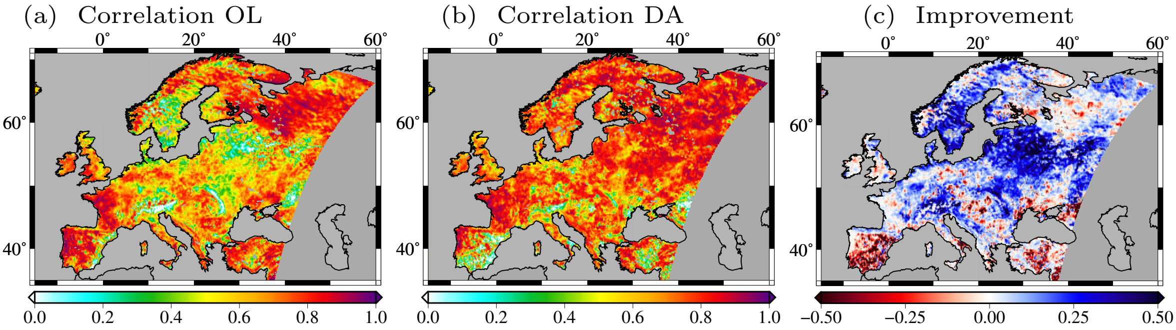


Figure 8.5: Correlation between TWS simulated by (a) the open-loop run (OL) and the reference run RF and (b) the assimilation run (DA) and RF. In (c), the improvement of the correlation coefficient through data assimilation is shown in absolute numbers. Blue (red) color in indicates that DA has a larger (smaller) correlation coefficient than OL.

Regional differences in the performance of DA were investigated closer by analyzing TWS and the corresponding assimilation increments for the major European catchments defined in Figure 8.1. Generally, in all catchments TWS reaches its maximum between January and April and its minimum between August and September (Figure 8.6). Please note that all plots have the same relative scaling (not the same absolute values) of the y-axis. In general, DA (blue) is closer to RF (black) than OL (green) as data assimilation naturally pulls the model towards the observations. The ensemble spread of TWS reduces from about 20 mm to 25 mm for OL to 10 mm to 20 mm for DA. This behavior is typical as all ensemble members are drawn towards the same observations. However, a stable ensemble spread is maintained over the whole study period. As in this experiment absolute values of TWS were assimilated, in particular the offset between the perturbed model and RF is reduced, which is obvious in the catchment of the Iberian peninsula (Figure 8.6 a).

It is straightforward that catchments with large differences between OL and RF show largest assimilation increments and, thus, largest improvements through data assimilation. This was also shown e.g. by Forman et al. (2012) for the Mackenzie River basin and by Schumacher et al. (2016) for the Mississippi basin. Especially in the Eastern European catchments the annual amplitudes increase through data assimilation and better match the annual amplitude simulated by RF (Figure 8.6 e – g).

On the average, every month 6 mm of water are injected or abstracted from the model. However, assimilation increments have seasonal patterns, e.g. the increments are largest in spring and smallest in summer.

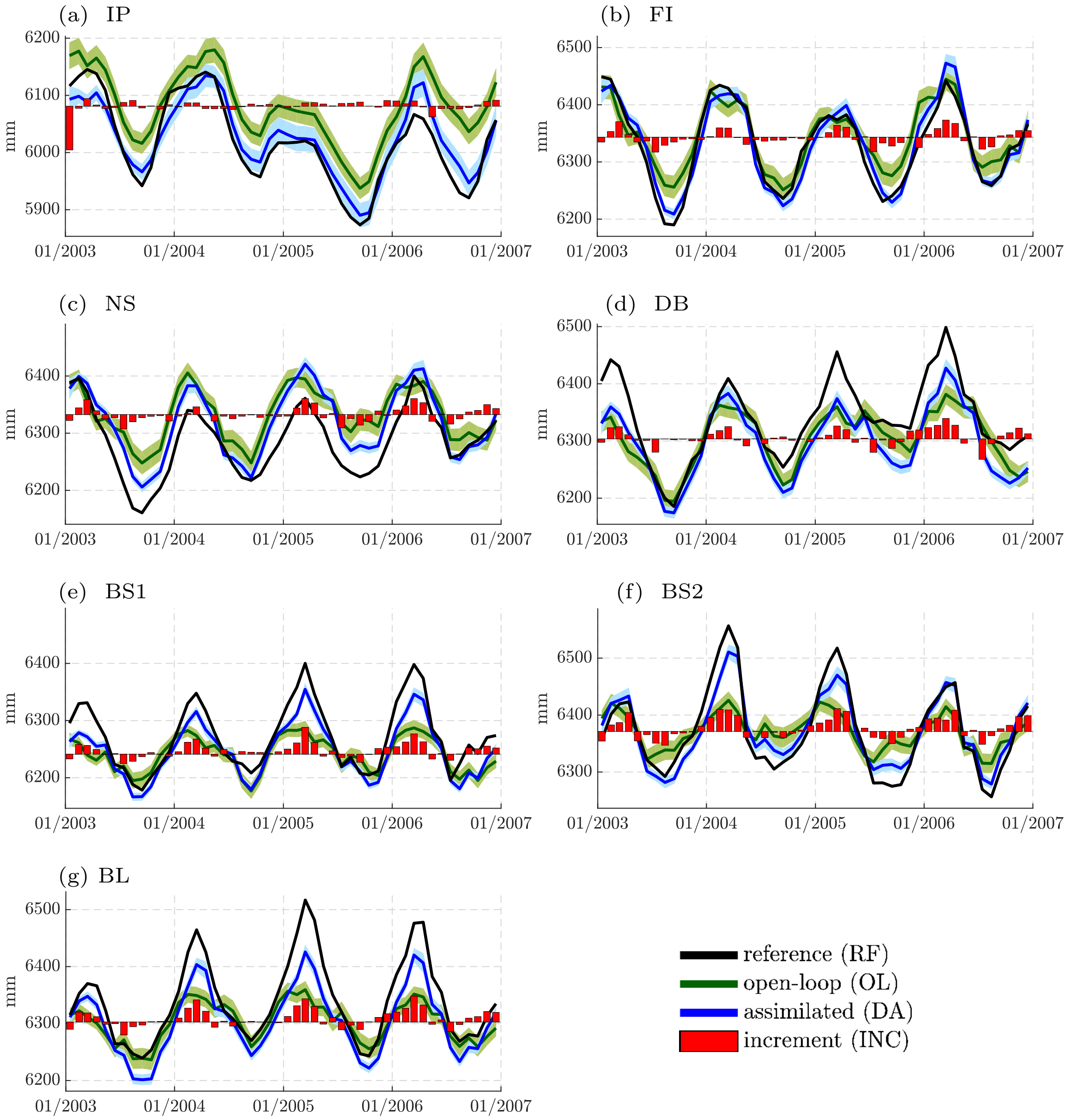
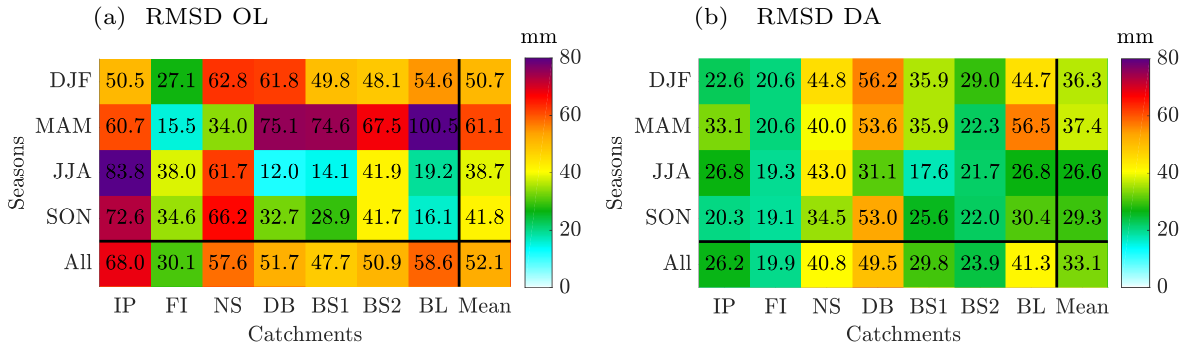


Figure 8.6: Monthly averaged TWSA are shown for the reference run (RF, black), the ensemble mean of the open-loop run (OL, green), and the ensemble mean of the assimilation run (DA, blue) together with the corresponding standard deviations derived from the ensemble spread for the 7 major European catchments (defined at the beginning of Chapter 8). For DA, the standard deviation was computed each month before the filter analysis step. Assimilation increments are drawn in red. The y-axis covers 350 mm in each subfigure.

8.2.2.2 Evaluation of Seasonal Patterns

Forman et al. (2012) found seasonal patterns of assimilation increments when assimilating GRACE into the CLSM set up for the Mackenzie River basin. Positive increments were identified for subsurface water in spring and summer and negative increments for snow water in winter. In contrast, assimilation increments computed for the same model set up for USA (Girotto et al., 2016) showed no seasonal patterns, most likely because the climatology of the observations was matched to the model before assimilation.

Here, seasonal patterns of the data assimilation procedure were investigated by computing the RMSD of OL versus RF and DA versus RF for each season and each catchment separately (Figure 8.7). The last row provides the RMSD considering all seasons, and the last column provides the mean RMSD over all catchments. Please note that here the RMSD is computed for the time series of absolute TWS without removing the temporal mean of the time series.

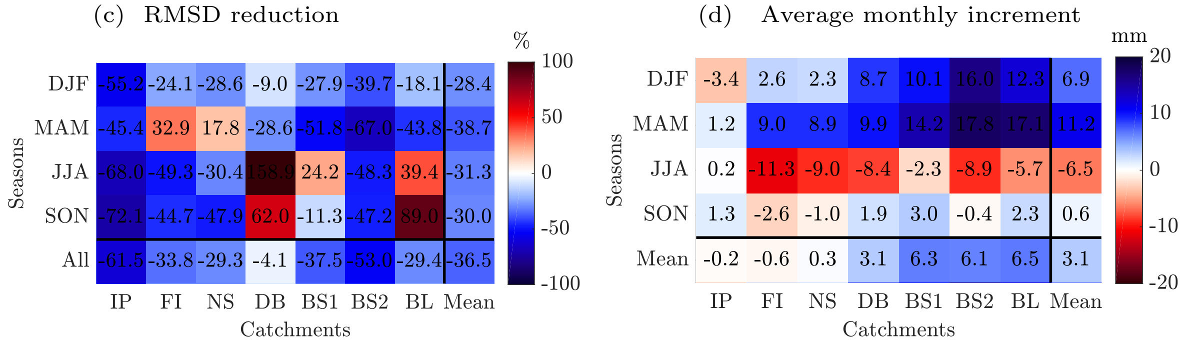


(a) RMSD OL

Seasons	IP	FI	NS	DB	BS1	BS2	BL	Mean
DJF	50.5	27.1	62.8	61.8	49.8	48.1	54.6	50.7
MAM	60.7	15.5	34.0	75.1	74.6	67.5	100.5	61.1
JJA	83.8	38.0	61.7	12.0	14.1	41.9	19.2	38.7
SON	72.6	34.6	66.2	32.7	28.9	41.7	16.1	41.8
All	68.0	30.1	57.6	51.7	47.7	50.9	58.6	52.1

(b) RMSD DA

Seasons	IP	FI	NS	DB	BS1	BS2	BL	Mean
DJF	22.6	20.6	44.8	56.2	35.9	29.0	44.7	36.3
MAM	33.1	20.6	40.0	53.6	35.9	22.3	56.5	37.4
JJA	26.8	19.3	43.0	31.1	17.6	21.7	26.8	26.6
SON	20.3	19.1	34.5	53.0	25.6	22.0	30.4	29.3
All	26.2	19.9	40.8	49.5	29.8	23.9	41.3	33.1



(c) RMSD reduction

Seasons	IP	FI	NS	DB	BS1	BS2	BL	Mean
DJF	-55.2	-24.1	-28.6	-9.0	-27.9	-39.7	-18.1	-28.4
MAM	-45.4	32.9	17.8	-28.6	-51.8	-67.0	-43.8	-38.7
JJA	-68.0	-49.3	-30.4	158.9	24.2	-48.3	39.4	-31.3
SON	-72.1	-44.7	-47.9	62.0	-11.3	-47.2	89.0	-30.0
All	-61.5	-33.8	-29.3	-4.1	-37.5	-53.0	-29.4	-36.5

(d) Average monthly increment

Seasons	IP	FI	NS	DB	BS1	BS2	BL	Mean
DJF	-3.4	2.6	2.3	8.7	10.1	16.0	12.3	6.9
MAM	1.2	9.0	8.9	9.9	14.2	17.8	17.1	11.2
JJA	0.2	-11.3	-9.0	-8.4	-2.3	-8.9	-5.7	-6.5
SON	1.3	-2.6	-1.0	1.9	3.0	-0.4	2.3	0.6
Mean	-0.2	-0.6	0.3	3.1	6.3	6.1	6.5	3.1

Figure 8.7: The RMSD of TWS from (a) the open-loop run (OL) and the reference run (RF), and (b) the assimilation run (DA) and RF is shown for four seasons and seven major European catchments. Additionally, the mean over all seasons is displayed in the last row, and the mean over all catchments in the last column. In (c), the reduction of the RMSD through data assimilation is shown in percent, with blue indicating improvement (decrease) of the RMSD and red indicating degradation (increase) of the RMSD. In (d), the assimilation increment is shown for each season, with blue indicating water injection and red water abstraction.

Averaged over all catchments, the RMSD between OL and RF reaches values of about 40 mm for June to August (JJA, summer) and for September to November (SON, autumn). For December to February (DJF, winter), the RMSD between OL and RF amounts to 51 mm. The largest RMSD is found for March to May (MAM, spring) with 61 mm. The RMSD between DA and RF amounts to 27 − 37 mm averaged over all catchments. That means, considering all catchment, data assimilation reduces the RMSD significantly by about 40 % in spring and by about 30 % during the other seasons. Yet, in summer in 3 out of 7 catchments the RMSD increases after data assimilation. It is notable that in these catchments OL has already a comparably small RMSD, which may result into overfitting of the model towards the observations during data assimilation.

An overall large improvement is obtained for IP (62 %) and BS2 (53 %) for different reasons. For IP, a bias between OL and RF is removed during data assimilation (see Figure 8.6). In contrast, for BS2 the annual amplitude is adjusted notably. DB is the catchment with the smallest improvement in terms of RMSD (4 %). This is due to large peaks in winter occuring for RF that cannot be reproduced neither by OL nor by DA.

Assimilation increments (Figure 8.7 d) are positive in winter and spring (water injection, blue) and negative in summer (water abstraction, red). In autumn increments are about zero. This means that OL is too dry in summer and too wet in winter and spring. The long-term mass imbalance is given by the bottom row. In Western Europe the mass balance of the model is preserved, whereas in Eastern Europe on the average 6 mm of water are injected each month. In Western Europe approximately the same amount of water is injected in spring and subtracted in summer. In contrast, in Eastern Europe 10 mm to 18 mm of water are injected during winter and spring. By this injection the underestimated annual maximum of TWS is adjusted (see Figure 8.6).

8.2.2.3 Evaluation of TWS Compartments

Previous studies showed that the assimilation of GRACE data does not only improve TWS, but also has the potential to improve individual storage compartments. For instance, Khaki et al. (2018) found that assimilating GRACE could effectively improve the representation of groundwater and soil water storage of the W3RA over Iran.

In the case of CLM, assimilation increments were applied to the TWS compartments soil liquid water $w_{liq,j}$ and soil ice $w_{ice,j}$ at $j = 1, ..., 10$ layers, snow water W_{sno}, canopy water W_{can}, and water in the unconfined aquifer W_{aqu}. As the temporal mean of the synthetic observations was not adjusted in this experiment, the TWS compartments of RF can be used as reference states. Figure 8.8 shows the impact of DA on the individual TWS compartments exemplarily for the catchment BS1, which is characterized by significant contributions from snow and ice in winter.

The impact of data assimilation on the TWS components varies with the seasons (Figure 8.8). Differences in soil liquid water simulated by OL and RF are particularly large in spring. Between March and April a strong and sudden increase of soil liquid water happens due to precipitation and melting. While this increase amounts to about 60 mm for RF, it reaches only 30 mm for OL, most likely due to underestimated snow water. As expected, soil ice and snow water do not play any role during summer. However, in winter soil ice contributes up to 6 mm and snow water up to 10 mm to the total increment, which is at least as much as the increment of soil liquid water ($<$ 10 mm). The increments of soil liquid water, soil ice, and snow water have the same sign. Contributions of canopy increments to the total increment are small compared to the total increment, but they are frequently as large as canopy water itself. Aquifer water varies only slightly for RF, and increments close to zero.

The impact of data assimilation on the TWS components also varies regionally. Figure 8.9 shows the RMSD, the residual correlation after de-trending and de-seasoning, and the relative annual amplitude computed for OL versus RF (green) and for DA versus RF (blue) for all storage compartments and river catchments. As stated in Section 8.2.2.2, the RMSD of TWS is reduced in all basins through data assimilation. Yet, for soil liquid water the RMSD increases in 3 out of 7 catchments with particular large values for IP. For IP, aquifer water simulated by RF is about 50 mm smaller than aquifer water simulated by OL, which results into a bias in TWS. This bias also impacts the increment of soil liquid water, which is already

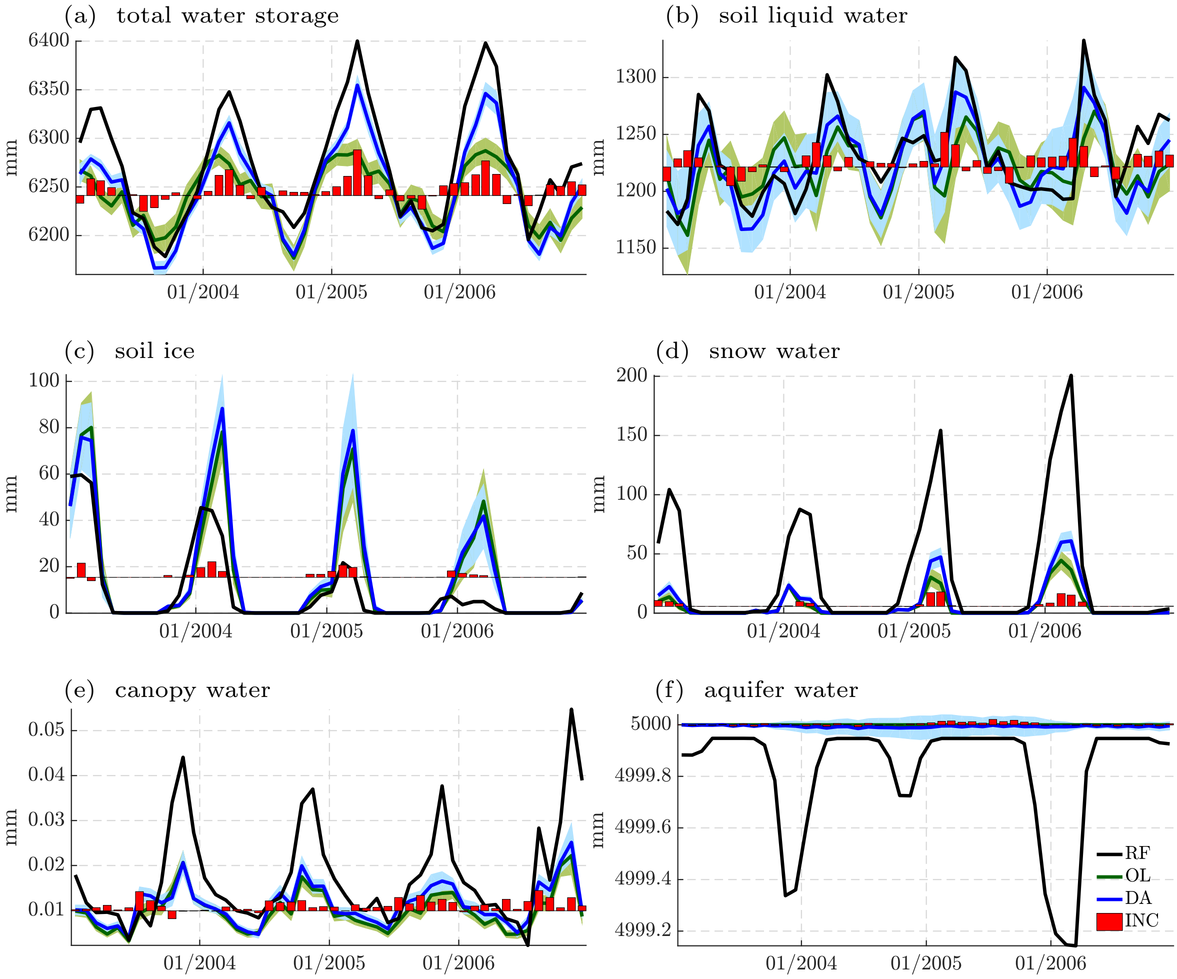


Figure 8.8: TWS and its components in the catchment BS1 (Western rivers discharging into the Baltic sea) are shown for the reference run (RF), the open-loop run (OL), and the assimilation run (DA). Uncertainties are indicated by the error bands. The red bars represent the assimilation increments of each time step.

well represented by OL and then degrades through assimilation.

For soil ice the RMSD increases in 4 out of 7 basins. It should be remembered here, that the model state vector contains the sum of soil liquid water and soil ice in order to maintain the data assimilation system stable. The update is then distributed proportionally to soil liquid water and soil ice (Section 7.2.4). Apparently, this approach can lead to unrealistic increments for soil ice and should be modified in future.

The residual correlation of TWS becomes larger in all catchments except for IP, and even increases by about 50 % for BS1 and BS2 reaching values up to 0.8. For most catchments (except IP and BL) soil liquid water also yields larger residual correlation after data assimilation. Likewise to the results for the RMSD, data assimilation has no positive impact on soil ice in terms of residual correlation.

The annual amplitude of TWS simulated by OL is underestimated (relative annual amplitude smaller than 1) in particular in Eastern European catchments, i.e. the amplitude of OL only reaches half of the amplitude of RF. Data assimilation increases the annual amplitude of TWS in all catchments towards 1, which would correspond to optimal agreement with RF. For Eastern European catchments the annual amplitude was even increased from 40 % of the reference amplitude to 80 %. Likewise, the annual amplitudes of soil liquid water and soil ice are increased via data assimilation in all catchments but one. Interestingly, BS2 is the only catchment where OL overestimates the annual amplitude of soil liquid water, while DA leads to an underestimation and reduces the relative amplitude from 1.7 to 0.7. However, in this basin soil liquid water and soil ice show no distinct seasonal signal. For soil ice, large differences in the annual amplitude of RF and OL arise. This is most likely due to differences in the temperature forcings that lead to a larger portion of frozen water for RF.

It is worth noting that snow water and canopy water are both closer to RF after data assimilation in terms of RMSD and annual amplitude. For snow water also residual correlations increase. Especially canopy water represents only a fractional amount of TWS and TWS variability, but it is still improved through data assimilation. These findings contradict the assumptions of previous studies, which excluded small storage compartments to avoid spurious increments (e.g., Zaitchik et al., 2008; Girotto et al., 2016).

8.2.2.4 Evaluation of Individual Soil Layers

The 10 individual soil layers simulated by CLM3.5 have different thicknesses, thus, different water storage capacities and, therefore, different contributions to TWS. In the following, the sensitivity of the individual soil layers to data assimilation is quantified. This is relevant regarding the validation with independent soil moisture data.

The first soil layer is the surface layer with a thickness of only 1.4 cm. The tenth soil layer is the bottom layer of the soil column with a thickness of 1.4 m. Figure 8.10 shows the contribution of each layer to total soil liquid water (first row) and to total soil ice (third row) exemplarily for catchment BS1. Numbers provide absolute values, while the percentage with respect to the column sum is color coded. For instance, the fifths layer of soil liquid water contributes 31 mm or about 3 % to the total amount of soil liquid water. Additionally, the absolute assimilation increment of each layer is displayed for soil liquid water (second row) and for soil ice (forth row). All results were temporally averaged over the assimilation period 2003 to 2006.

Obviously, W_{liq} increases from the top layer to the bottom layer. Please be aware that this is primarily due to increasing layer thickness and not necessarily due to a wetter soil. The bottom layer contains more than 30 % of total soil water, the top layer less than 1 %. The size of the increments roughly follows the magnitude of W_{liq}. In contrast, W_{ice} and the corresponding

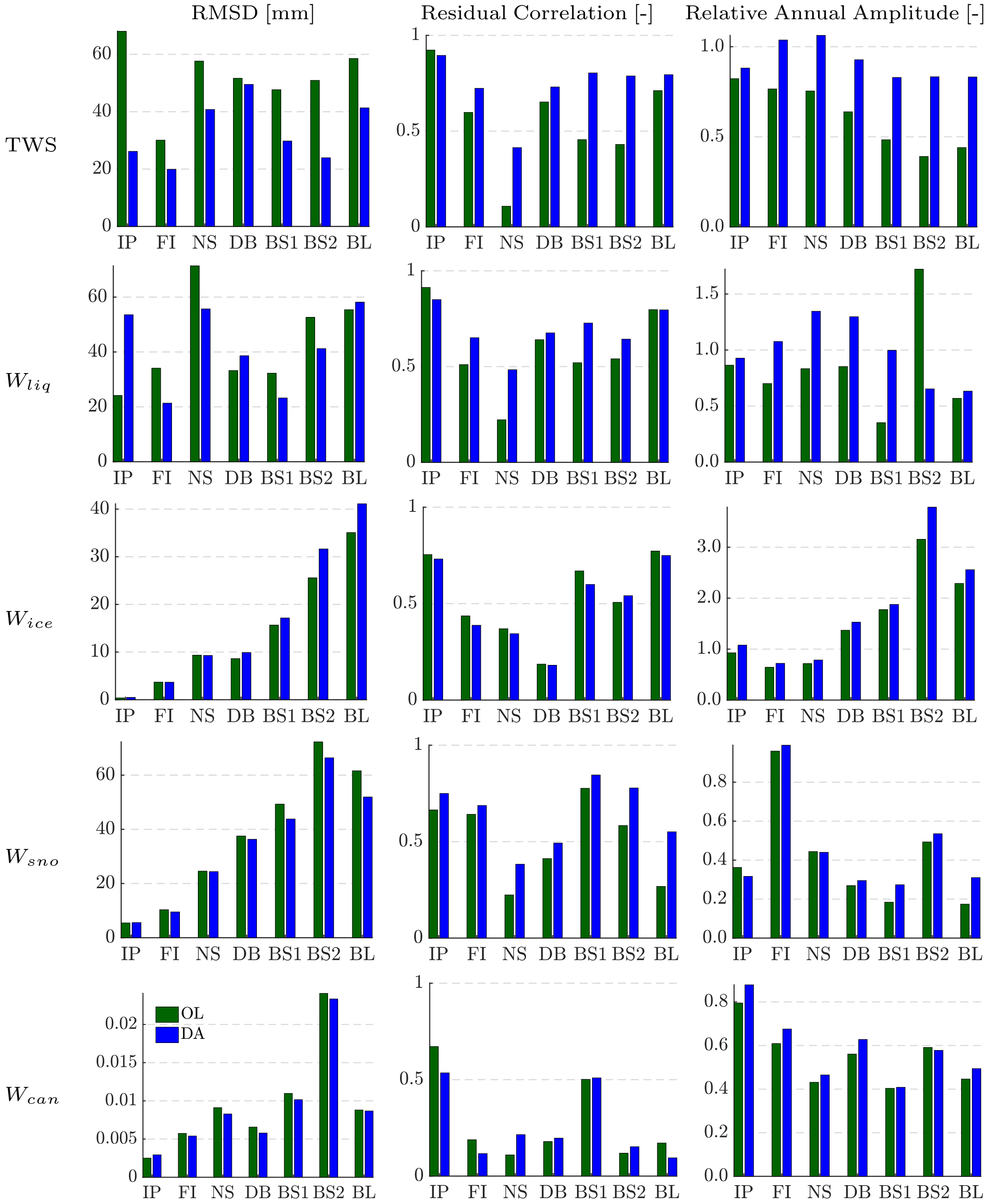
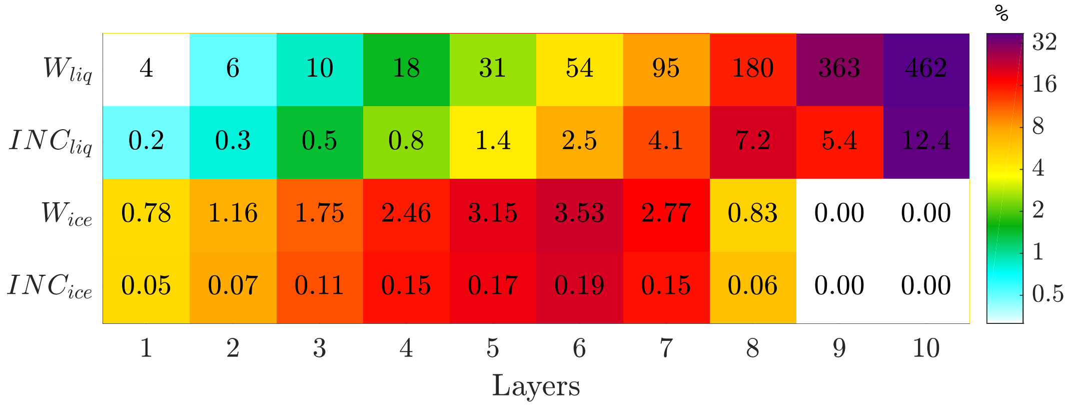


Figure 8.9: Root mean square deviation (RMSD), residual correlation after de-trending and de-seasoning, and relative annual amplitude of TWS and its compartments are shown for OL versus RF (green) and for DA versus RF (blue), for the major European catchments (OL: open-loop run, RF: reference run, DA: assimilation run).

assimilation increments increase from layer 1 to 6, then the increments decrease, and starting from layer 9 (below 1.3 m depth) no frozen water is contained in the soil. Largest assimilation increments of W_{ice} are obtained for layers 4 to 7 with values between 0.15 mm and 0.19 mm. In the top layer, the increment of soil ice amounts to one fifth of the increment of soil liquid water. For Western and Central European catchments, increments of W_{ice} are smaller especially for the lower soil layers.



	1	2	3	4	5	6	7	8	9	10
W_{liq}	4	6	10	18	31	54	95	180	363	462
INC_{liq}	0.2	0.3	0.5	0.8	1.4	2.5	4.1	7.2	5.4	12.4
W_{ice}	0.78	1.16	1.75	2.46	3.15	3.53	2.77	0.83	0.00	0.00
INC_{ice}	0.05	0.07	0.11	0.15	0.17	0.19	0.15	0.06	0.00	0.00

Figure 8.10: Individual levels of soil liquid water W_{liq} and soil ice W_{ice} of the assimilated CLM model and the corresponding increments are displayed as averages over the time span 2003 to 2006 for the catchment $BS1$. The color coding shows the contribution to the total amount of W_{liq} and W_{ice} in percent.

8.2.2.5 Influence of the Temporal Mean on DA

In this experiment, synthetic observations were assimilated into the perturbed model without adjusting the temporal mean of the observations. Hence, direct validation of changes in the absolute values of the TWS components before and after data assimilation was possible using RF. However, in the real world GRACE only provides TWSA and not absolute TWS. Thus, it is only possible to improve the variability of TWS and not its absolute value. In the following, the impact of adjusting the long term mean of synthetic TWS observations on the data assimilation results is assessed.

Figure 8.11 illustrates the effect of assimilating synthetic observations without adjusted temporal mean (DA1, blue) and with adjusted temporal mean (DA2, yellow). Please note that validation measures are now computed for centered time series and not for absolute values and, thus, differ from the results shown in Figure 8.9.

Generally, adjusting the temporal mean changes the RMSD of TWS simulated by DA by about 10%. No clear improvement or deterioration can be determined. A small positive effect of adjusted temporal mean is found for the RMSD of soil ice. Without adjusting the temporal mean, the RMSD of soil ice increases in all basins, while the RMSD decreases in some basins when adjusting the temporal mean. In contrast, for snow water the RMSD is slightly larger when adjusting the mean but still improves with respect to OL. When not adjusting the temporal mean, a huge increase of the RMSD was identified for soil liquid water in the catchment IP (Figure 8.9). This increase in RMSD was caused by a bias in WA. Adjusting the temporal mean of RF, removes the bias between RF and OL and, thus, improves the RMSD of soil liquid water.

For TWS, residual correlation increases in all basins (except for BS1) slightly when adjusting the temporal mean. In line with the RMSD, residual correlation of snow water slightly

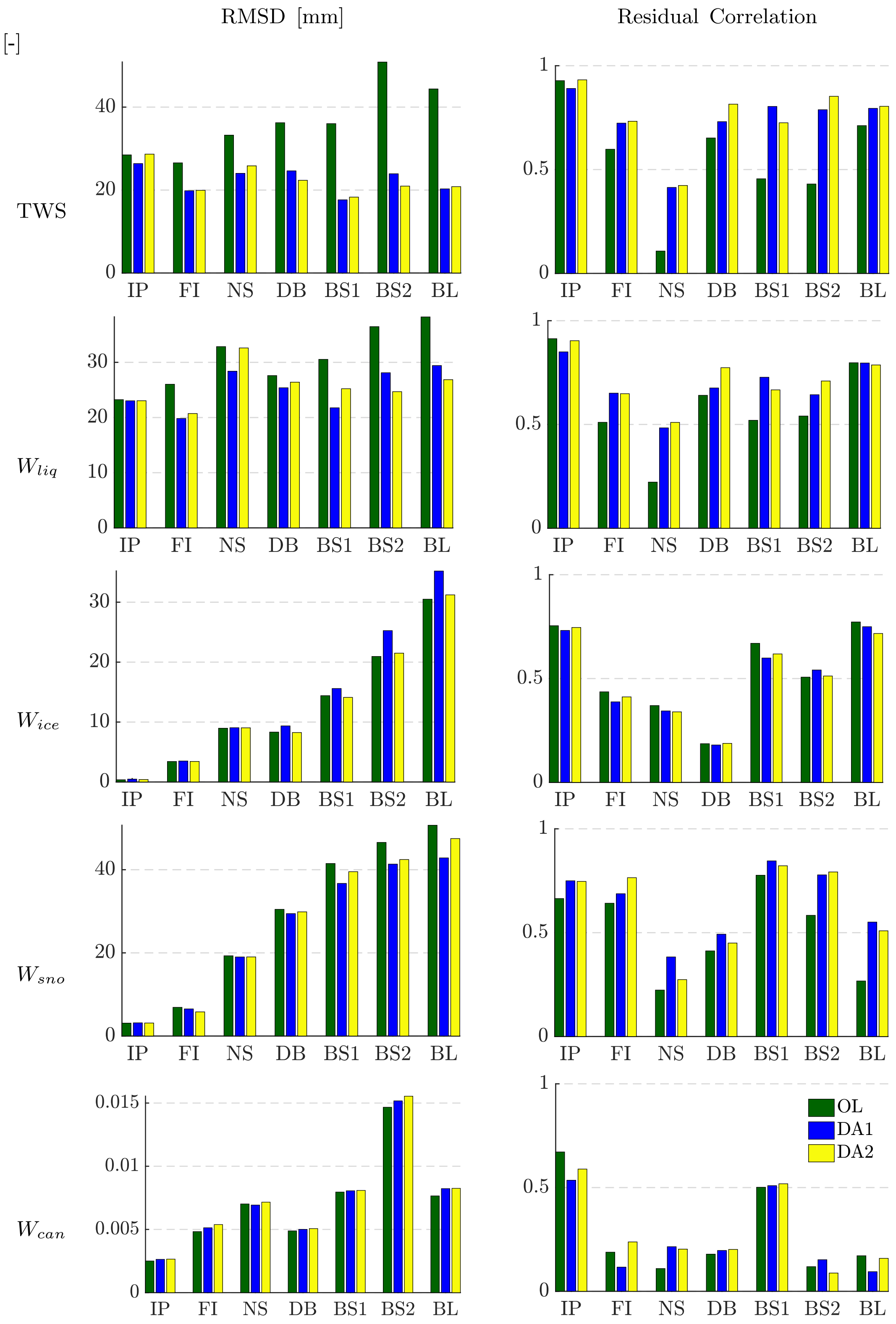


Figure 8.11: Root mean square deviation (RMSD) and residual correlation after de-trending and de-seasoning of TWSA and its compartments are shown for OL versus RF (OL, green), for DA with adjusted temporal mean versus RF (DA1, blue), and for DA without adjusted temporal mean versus RF (DA2, yellow). (OL: open-loop run, RF: reference run, DA: assimilation run).

degrades for DA2 compared to DA1 (except for FI). Still, with respect to OL, residual correlation of snow water improves by more than 0.2 for BS2 and BL. For canopy water, adjusting the temporal mean leads to an ambiguous picture. While the RMSD tends to degrade for DA2 compared to DA1, residual correlation becomes better for 5 out of 7 catchments. In summary, adjusting the temporal mean has no clear positive or negative effect on the variability of the TWS components.

All in all, this introductory experiment showed that assimilating TWS improves the overall representation of TWS of the model, without improving all individual compartments of TWS. Assimilation increments and the performance of DA vary regionally and seasonally.

8.2.3 Influence of the Ensemble Size

Insufficiently large ensembles cause sampling problems like (i) underestimation of the ensemble spread, (ii) filter divergence, and (iii) errors in estimated correlations (Chapter 6). Previous data assimilation experiments using GRACE data chose between 20 and 30 ensemble members (Zaitchik et al., 2008; Forman et al., 2012; Eicker et al., 2014; Schumacher et al., 2016; Girotto et al., 2016). Here, the necessary ensemble size was identified by performing experiments with 8 to 128 ensemble members. The experiments were set up with the classical Ensemble Kalman Filter (EnKF), perturbed observations, and a fully occupied error covariance matrix (Table 8.2). Starting from a certain ensemble size, results are expected to change only negligibly when further increasing the ensemble.

Table 8.2: Experiment 2 – Influence of the ensemble size

Filter algorithm	EnKF	**Ensemble size**	8, 16, 32, 48, 64, 128
Inflation factor	0.7	**Observations**	RF, perturbed, 2°
Localization	–	**Assumed errors**	correlated

Experiments with 8 to 64 (the compute cluster has 32 cores per node) ensemble members are evaluated regarding RMSD, residual correlation, and relative annual amplitude with respect to RF (Figure 8.12). Generally, 8 and 16 ensemble members are determined as not sufficient, as most measures improve further for 32 ensemble members. The average RMSD over all catchments amounts to 41 mm for 8 ensemble members, to 40 mm for 16 ensemble members, and to 36 mm for 32 or more ensemble members. However, for the individual catchments the validation measures still change slightly when using 32, 48, or 64 ensemble members, most likely due to changes in the correlation structure.

Averaged over all catchments, residual correlation amounts to 0.61 for 32 ensemble members and to 0.64 for 64 ensemble members. The relative annual amplitude changes only slightly with the ensemble size. For an ensemble size of 128 members, a two-year run was performed because of limited computational resources. For the defined European catchments the RMSD decreased by 2 mm to 5 mm when using 128 ensemble members instead of 64 ensemble members for 6 out of 7 catchments.

All in all, as a trade-off between computational costs and optimal sampling the following experiments are performed using 64 ensemble members.

8.2.3.1 Computational Costs

The optimal configuration of the available computational resources was determined by scaling experiments.

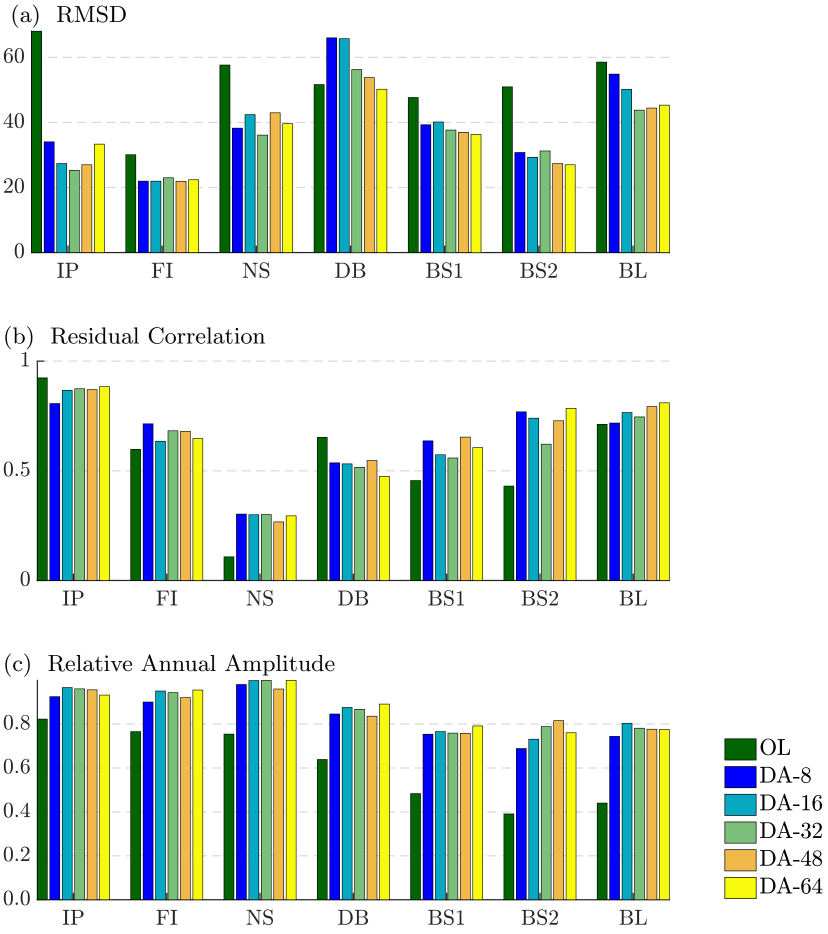


Figure 8.12: Root mean square deviation (RMSD), residual correlation after de-trending and de-seasoning, and relative annual amplitudes of TWS for OL versus RF (green), and for DA simulating 8 to 64 ensemble members versus RF (OL: open-loop run, RF: reference run, DA: assimilation run).

Detailed scaling experiments for TerrSysMP-PDAF, running the coupled CLM3.5-ParFlow model, were performed by Kurtz et al. (2016) on the supercomputer JUQUEEN located at Forschungszentrum Jülich (Germany). In this thesis, scaling experiments with CLM3.5 were carried out on the cluster of the APMG group at the Institute of Geodesy and Geoinformation (Bonn University, Germany). The cluster consists of 44 nodes with 32 cores (each of them having 64 GB RAM), and 12 nodes with 48 cores (each of them having 512 GB RAM). Here, only the small nodes with 32 cores were used.

In a first experiment, one realization of CLM3.5 was run on 1 to 32 cores producing monthly outputs of TWS and its components during one year without data assimilation. This experiment identifies the optimal number of cores that should be used for each ensemble member. It should be noted, that the execution time also depends on the assigned atmospheric forcings and parameter sets, and on the amount of data that is written to the disk (Kurtz et al., 2016). Parallel efficiency $E(n_p)$ for N_p processors was computed as a measure for the speed-up when using more processors according to

$$E(N_p) = \frac{T_1}{T(N_p)}, \tag{8.6}$$

where T_1 is the execution time for 1 processor, and $T(N_p)$ the execution time for N_p processors. Parallel efficiency stays larger than 0.8 when using up to 8 processors and decreases to 0.58 for 32 processors (Figure 8.13 a, black). Correspondingly, the CPU time (i.e. the execution time summed up for all processors) increases with the number of processors from 11 h for 1 CPU, to 12.5 h for 8 CPUs, and to 20 hours for 32 CPUs (Figure 8.13 b, blue). Based on these results, for future DA experiments 4 to 8 cores per ensemble member were used.

In a second experiment the scaling behavior for ensemble size 8 to 64 was tested for one year, using 8 cores per ensemble member and monthly data assimilation. The parallel efficiency remains close to 1 (Figure 8.13 b, black). Increasing the number of ensemble members from 8 to 64 increases the total computation time per ensemble member by 30 minutes. This corresponds to 4 minutes per CPU when using 8 CPUs per ensemble member.

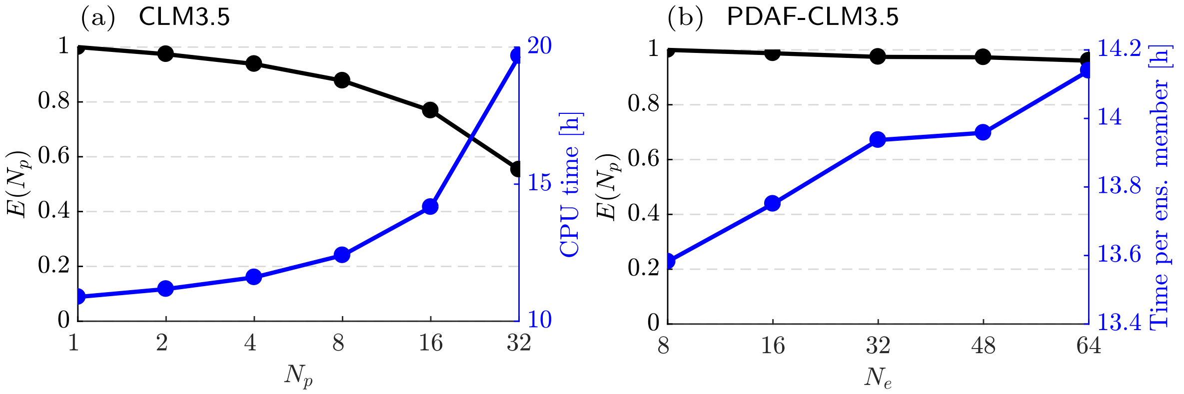


Figure 8.13: The performance of (a) stand-alone CLM3.5 and (b) PDAF-CLM3.5 is shown in terms of parallel efficiency $E(p)$. In (a) one ensemble member was run on a different number of processors N_p (1 to 32), and N_p is plotted against $E(p)$ and against the total CPU time. i.e. the execution time summed up for all processors. In (b) each ensemble member was run on 8 processors and the number of ensemble members (N_e) was increased from 8 to 64. On the left y-axis E_p is displayed and on the right y-axis the computation time is normalized by the number of ensemble members.

A typical experiment with 64 ensemble members using 8 processors for each of them requires 512 cores or 16 nodes, respectively. Typical data assimilation experiments then have a runtime of less than 2 h per year. This is also valid when using other global filter variants. When applying localization with a large localization radius, the computation time during the analysis step increases by a few minutes.

8.2.4 Influence of Assimilation Algorithm and Observation Error Model

The performance the individual Kalman filter algorithms was investigated for two global filters, EnKF and ESTKF, and one localized filter, the Localized Error Subspace Transform Kalman Filter (LESTKF). Here, two scenarios were assessed, (i) taking into account the full observation error covariance matrix of gridded TWSA (colored or correlated noise) and (ii) neglecting correlations (white noise). Correlated and white noise of gridded TWSA can be either derived from the full error covariance matrix of spherical harmonic coefficients or from the diagonal part.

Only minor differences between the two global filters are expected. In contrast, localization should significantly improve the results as remaining long-range error correlations in the GRACE data and spurious correlations between the model grid cells are suppressed. Intuitively, one would expect the most realistic error model to lead to the best result. However, in the case of GRACE observations neighboring observation grid cells are highly correlated. Thus, assuming white noise for gridded TWSA overestimates the information content of the GRACE observations, and leads to larger assimilation increments towards the observations than when assuming correlated noise. This means that including the full observation error covariance matrix of gridded TWSA into the assimilation process can be expected to lead to more realistic, but not necessarily better validation measures, when comparing to RF.

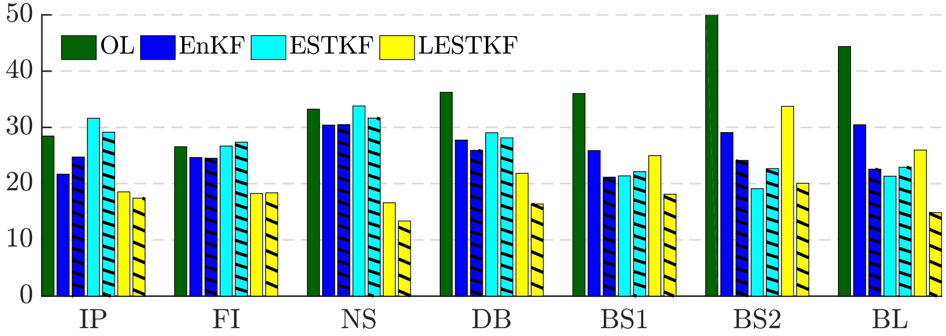
Previous studies found negligible differences when using distributed errors of gridded TWSA instead of uniform errors (Kumar et al., 2016). In contrast, significant differences arised when taking into account correlated noise of gridded TWSA instead of white noise (Schumacher et al., 2016). Schumacher et al. (2016) found that in the case of white noise, GRACE observations had stronger impact, but this did not necessarily improve the validation metrics. A later study by Tangdamrongsub et al. (2017) found the assimilated model better matching the GRACE observations when assuming white noise, but overfitting groundwater well observations. Regarding the choice of the assimiliaton algorithm, Schumacher et al. (2016) attributed only minor impact to three tested different algorithms. However, Khaki et al. (2017a) found significant differences for eight types of Kalman and Particle filters with the EnKF belonging to the most promising filters. Furthermore, Khaki et al. (2017b) attributed improvements in the RMSD, which were larger than 50 % with respect to in-situ groundwater measurements, to localization.

In the following experiments, synthetic TWSA observations were assimilated on a 2° grid perturbed with correlated noise (Table 8.3). If not stated otherwise, observation errors were propagated from the full error covariance matrices of SH coefficients. For LESTKF, a localization radius of 6° was chosen. Observations within the localization radius were weighted using an exponential decay function reaching $1/e$ at a distance of 3°.

Table 8.3: Experiment 3 – Influence of the filter algorithm and the assumed observation error model

Filter algorithm	EnKF, ESTKF, LESTKF	**Ensemble size**	64
Inflation factor	0.7	**Observations**	RF, perturbed, 2°
Localization	6° (3°)	**Assumed errors**	white, correlated

Figure 8.14 shows the performance of the filter algorithms in terms of RMSD and residual correlation. Solid bars indicate correlated observation noise assumed in the filter algorithm and hatched bars indicate white noise.


(a) RMSD [mm]

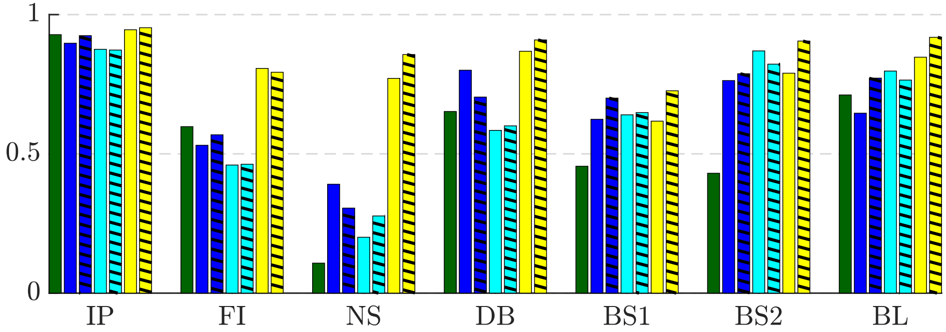

(b) Residual Correlation [–]

Figure 8.14: Root mean square deviation (RMSD) and residual correlation after de-trending and de-seasoning of TWSA for OL versus RF, and for DA versus RF considering different data assimilation scenarios (OL: open-loop run, RF: reference run, DA: assimilation run). Solid bars indicate correlated observation noise and hatched bars indicate white observation noise.

Data assimilation results are more sensitive to localization than to different global filter variants (Figure 8.14). Furthermore, correlated observation errors lead to larger differences between the two global filters than white noise. In fact, the RMSD between assimilation results from EnKF and ESTKF differs by up to 30% for correlated noise and by up to 5% for white noise. Residual correlation differs by up to 0.2 for correlated noise and by up to 0.1 for white noise, respectively. Possibly, the EnKF has problems in representing the complex structure of correlated noise by perturbing the observations. Overall, the EnKF performs better over Western and Central Europe while the ESTKF performs better over Eastern Europe, where differences between model and synthetic observations are larger (Figure 8.3).

In Western European catchments (IP, FI, NS, DB), the local filter variant LESTKF clearly outperforms the global filters and reduces the RMSD by up to 50% compared to OL. In Eastern European catchments (BS1, BS2, BL) the local filter leads to the best results for white observation noise with an average RMSD of 17 mm and residual correlation of 0.85. In contrast, in the case of correlated observation errors LESTKF performs similar as EnKF and worse than ESTKF in Eastern European basins. One can speculate that in Eastern Europe remote observations have a positive impact on the data assimilation result, which is reduced through localization.

The previous experiments were related to full error covariances of SH coefficients and assessed effects of neglecting error correlations of the gridded TWSA. An additional experiment was performed to assess the influence of neglecting correlations on the level of SH coefficients.

Figure 8.15 shows results for ESTKF and LESTKF. Assimilation experiments were again performed assuming correlated noise of gridded TWSA (solid bars) and white noise (hatched bars). Asterisks indicate experiments where correlations between SH coefficients were neglected. The impact is negligible. The error model of the SH coefficients changes the RMSD by less than 1 mm and residual correlation by less than 0.01 for the global filters. Only in the case of LESTKF and correlated gridded observation errors the RMSD decreases by up to 4 mm and residual correlation increases by up to 0.04 when neglecting correlations between SH coefficients. However, the impact from neglecting correlations between SH coefficients might be different for observation covariance matrices of other months or in other regions of the Earth. Therefore, for the following experiments the full error covariance matrix of spherical harmonic coefficients is used.

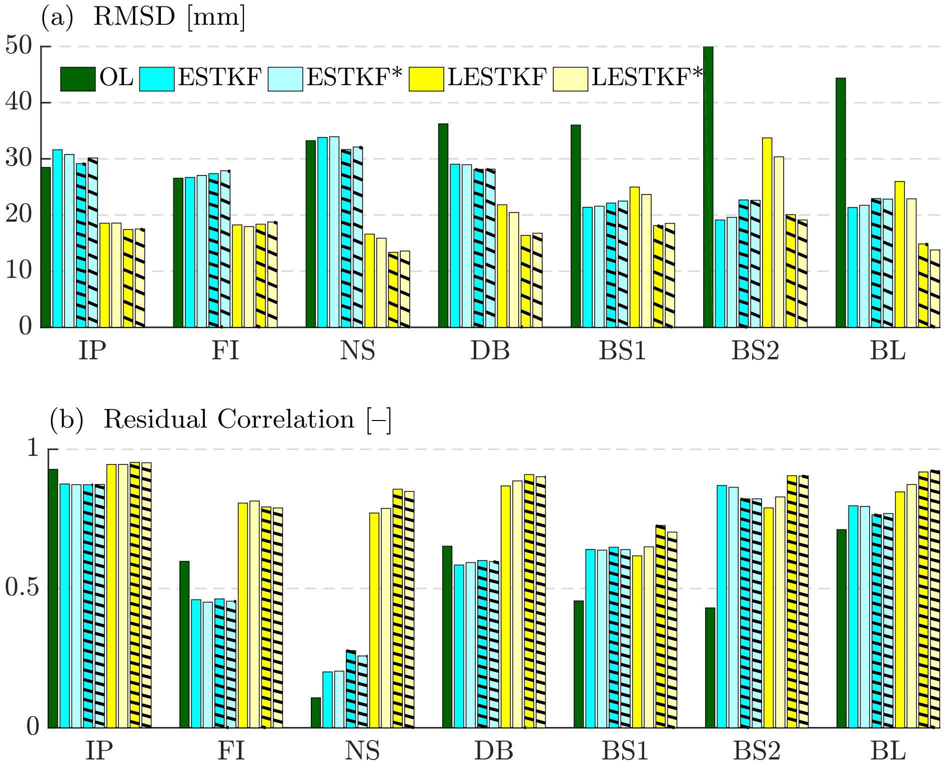


Figure 8.15: Root mean square deviation (RMSD) and residual correlation after de-trending and de-seasoning of TWSA for OL versus RF, and for DA versus RF considering different data assimilation scenarios (OL: open-loop run, RF: reference run, DA: assimilation run). Asterisks indicate that correlations between spherical harmonic coefficients are neglected. Solid bars indicate correlated observation noise between grid cells and hatched bars indicate white observation noise between grid cells.

Influences from filter algorithm and observation error model on the assimilation results can also be visualized by the assimilation increments (Figure 8.16). Figure 8.16 shows (a) monthly mean TWS in July 2004 before the assimilation step, (b) observations, (c) differences between model and observations, and (d) to (i) different assimilation increments. Before data assimilation, the model is wetter than the synthetic observations over Eastern Europe and the Iberian peninsular (blue color in (c)), and drier than the synthetic observations over Central Europe (red color in (c)).

Assimilation increments of global and local filters differ significantly (Figure 8.16 d – i). The local algorithm clearly distinguishes between regions of lacking water and regions of too much water. Clear regional patterns arise with positive increments (blue) over Central Europe and negative increments (red) elsewhere. In contrast, for the global filters negative increments

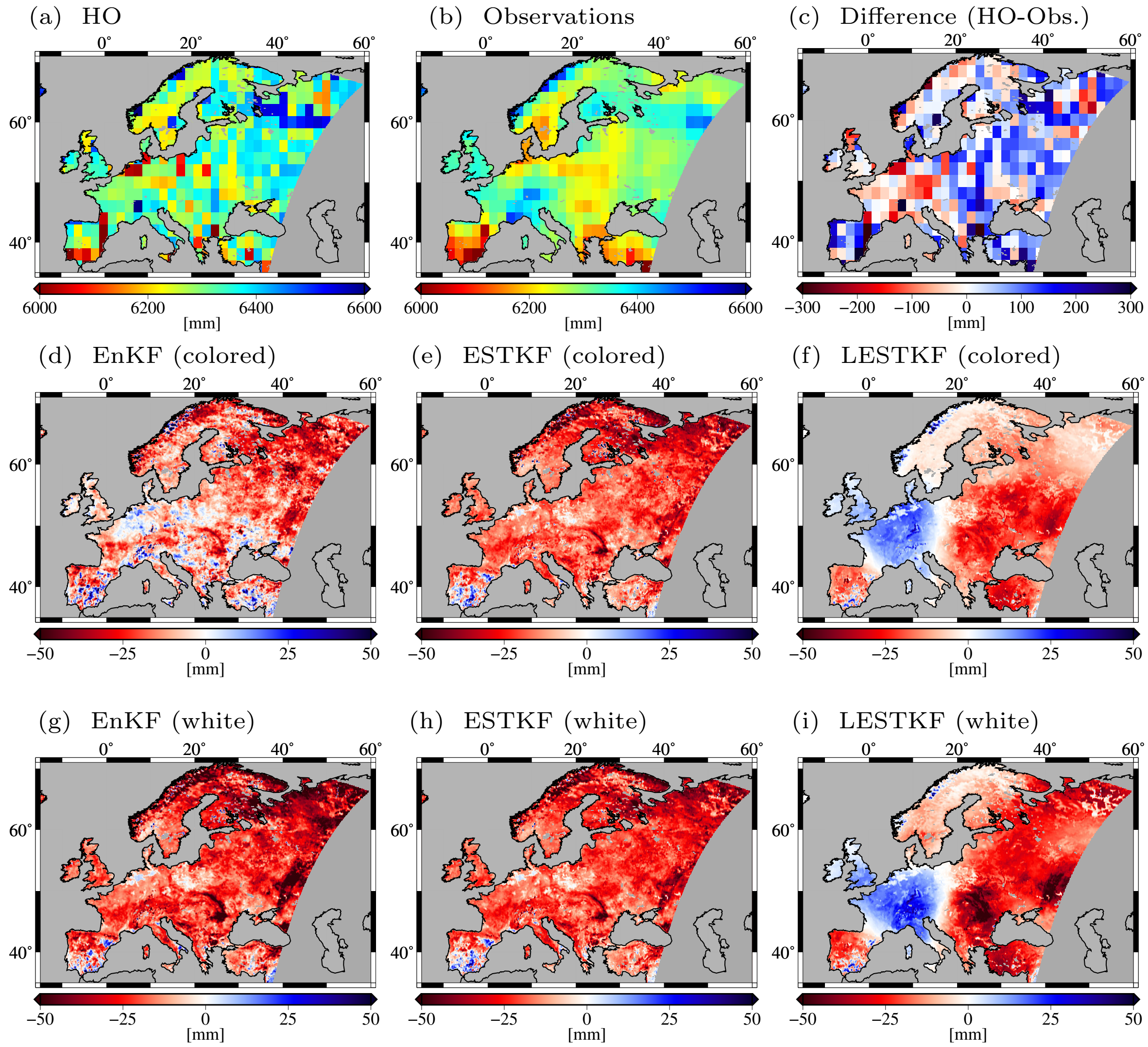


Figure 8.16: (a) Monthly averaged total water storage (TWS) of the ensemble mean before data assimilation (H0) in July 2004, (b) the corresponding synthetic observations, and (c) differences between (a) and (b) with blue indicating larger TWS simulated by HO than by the observations. Assimilation increments of the different filter algorithms assuming colored observation noise are shown in (d) to (f), and assimilation increments in the case of white noise are shown in (g) to (i). Blue indicates that water is introduced into the model, red indicates water abstraction.

prevail over whole Europe in this assimilation step.

Increments of EnKF and ESTKF are similar with EnKF showing slightly larger values. Negative increments prevail over Europe with values larger than 50 mm over Eastern Europe and Scandinavia. Positive increments arise in the South of the Iberian pensinsular. Correlated noise leads to smaller increments than white noise. Applying EnKF with correlated noise results in different signs of the increments over parts of Central Europe in comparison to the other global filter variants.

For the local filter, minor differences regarding the magnitude of the increments and the location of the maxima are found when comparing white and colored noise. For white noise, the maxima of water intrusion are situated over the Alpes, while for colored noise positive increments are rather uniformly distributed over Central Europe. Increments over the permafrost region in North-Eastern Europe are close to zero for colored noise and reach -25 mm for white noise.

Absolute assimilation increments give an impression about the monthly violation of water balance of the model through data assimilation. Figure 8.17 shows the temporal average (2003 to 2006) of the absolute values of assimilation increments. Applying global filters, between 10 – 30 mm of water are introduced into the model or subtracted from the model every month. In contrast, for LESTKF and colored noise, average absolute assimilation increments are much smaller, with values between 5 – 12 mm per month. The spatial patterns of the increments remain similar for all filter types and are closely related to the variability of TWS (Figure 8.3 d).

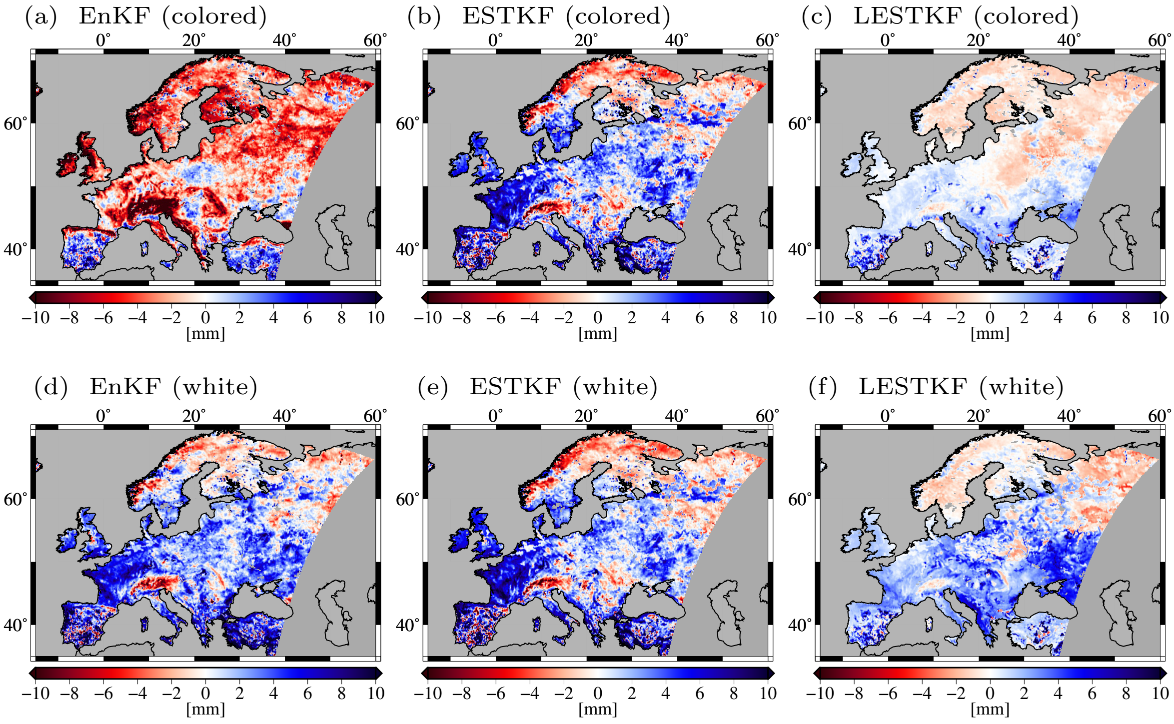
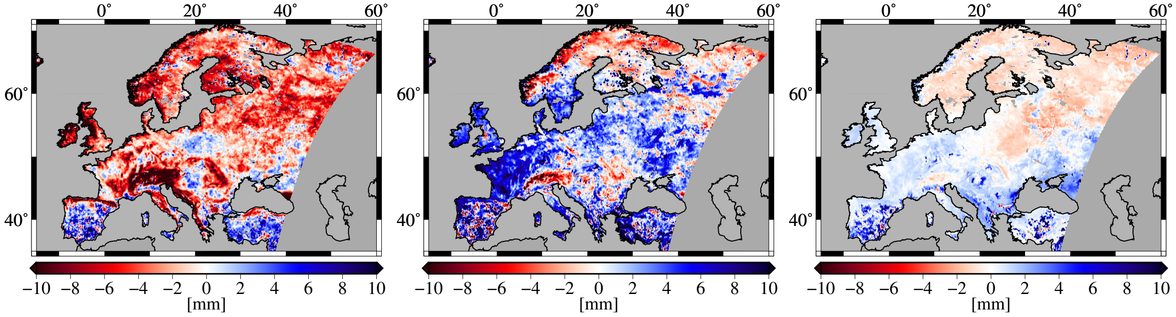


Figure 8.17: Average absolute values of monthly assimilation increments of TWS are shown in the case of colored noise (a – c) and in the case of white noise (d – f). The figures do not distinguish between water intrusion and water abstraction. Blue indicates large changes of TWS due to assimilation and red indicates small changes of TWS.

Violation of the water balance on the long term can be assessed by summing up the monthly increments over the whole assimilation period (Figure 8.18). If positive and negative assimilation increments balance each other, the sum of the monthly assimilation increments is zero, and the long-term water balance is maintained. For the global filters the monthly violation of the water balance is between -10 – 10 mm. This means, over the study period of four years up to 0.5 m of water are injected or abstracted. Interestingly, EnKF applied with correlated noise shows distinctly different patterns than the other global variants. EnKF with white noise and the two ESTKF variants show positive increments for most parts over Europe. In contrast, EnKF with colored noise shows negative increments nearly everywhere over Europe except for the Iberian pensinsular. One explanation could be deficiencies of the EnKF algorithm in representing the complex error correlation structure in the case of correlated noise.

On the long term the water balance is well maintained by LESTKF. In the case of white noise, the violation of the water balance is between -3 – 5 mm per month. In the case of correlated noise, on the average 2 mm of water is introduced or abstracted from the model each month. This corresponds to 5 cm of water during the study period of 4 years.

(a) EnKF (colored) (b) ESTKF (colored) (c) LESTKF (colored)



(d) EnKF (white) (e) ESTKF (white) (f) LESTKF (white)

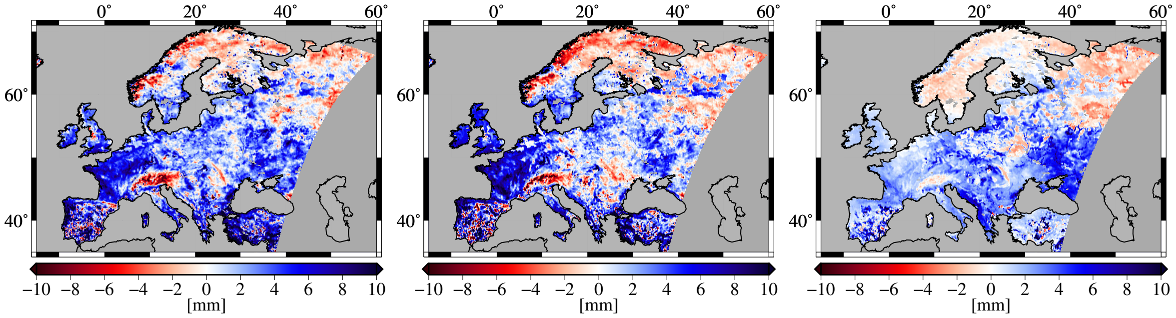


Figure 8.18: Assimilation increments of TWS averaged over the time period 2003 to 2006 are shown in the case of colored noise (a – c) and in the case of white noise (d – f). Blue indicates water injection and red indicates water abstraction on the long term.

8.2.5 Influence of Localization Radius and Inflation Factor

The localization radius and the forgetting factor are tuning parameters of the data assimilation algorithm. Localization avoids the influence of remote observations, which show huge correlations due to the typical GRACE striping patterns. Covariance inflation via the forgetting factor counteracts the underestimation of the model error caused by a limited ensemble size and neglected model errors. The following experiments aim at deriving the optimal parameter values in terms of minimum RMSD and maximum residual correlation, first, for the localization radius and, second, for the forgetting factor.

Khaki et al. (2017b) assessed different localization radii for the assimilation of gridded GRACE observations into W3RA using the Square Root Analysis (SQRA). They validated the results against in-situ groundwater measurements and obtained best results for a 5° localization radius, which reduced the RMSD by more than 50 %. The optimal localization radius was independent from the observation grid space.

The localization radius and the width of the exponential localization function govern the weight w of each observation grid cell depending on the distance d to the model grid cell, according to

$$w = \exp^{-\frac{d}{\lambda}} \, . \tag{8.7}$$

Above, λ is the exponential decay constant, which is typically chosen smaller or equal to the localization radius. Observations outside of the localization radius have zero weight. The optimal localization radius is commonly derived through tuning experiments. If the localization radius is chosen too small, blocking arises in the assimilation increments. If the radius is chosen too big, the advantages of the local algorithm are not optimally used. Table 8.4 gives an overview on the assimilation experiments performed with the LESTKF applying different localization radii and different exponential decay constants (given in brackets).

Table 8.4: Experiment 4 – Influence of the localization radius

Filter algorithm	LESTKF	**Ensemble size**	64
Inflation factor	0.7	**Observations**	RF, perturbed, 2°
Localization	2° (1°), 3° (1.5°), ... , 10° (5°) and 7° (1.25°), 7° (3.5°), 7° (5.75°), 7° (7°)	**Assumed errors**	correlated

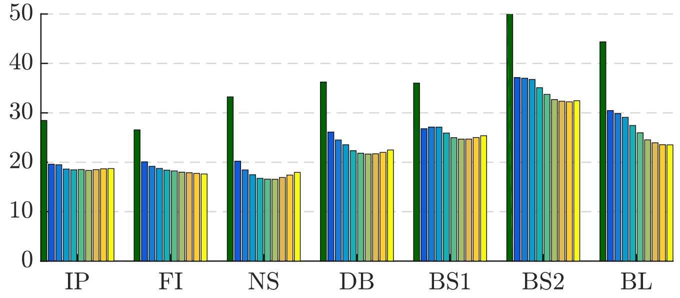
First, the localization radius was iterated from 2° to 10°, and λ was chosen to half of the localization radius (Figure 8.19). Starting from a localization radius of 2°, the RMSD drops steadily until a localization radius of 7°. With few exceptions also residual correlation becomes larger until a localization radius of 7°. For larger localization radii, the performance of the filter becomes slightly better only in BL. This means that we obtain a larger optimal localization radius than Khaki et al. (2017b), who obtained best results for a radius of 5°. This might be due to the higher spatial resolution of the model used in this thesis and or due to differences in the study region.

In a second experiment, the optimal localization radius of 7° was tested with different exponential decay constants, i.e. λ was chosen to 1.25°, 3.5°, 4.75°, and 7°. Moving from 1.25° to 3.5° decreased the RMSD by up to 20 % (Figure 8.20). Choosing the exponential decay constant larger than 3.5° only slightly improved the result over Eastern Europe, and even lead to degradation of Western Europe. Therefore, in the following experiments a localization radius of 7° and a value of $1/e$ at 3.5° was used.

With the next experiments, the data assimilation framework was tuned regarding the forgetting or inflation factor (Section 6.2.2). Choosing the forgetting factor smaller than 1 increases the spread of the model ensemble, thus, increases the uncertainty of the model and, consequently, leads to a higher weight of the observations. Here, experiments with forgetting factors between 1 and 0.5 were performed (Table 8.5).

The impact of covariance inflation increases with growing differences between model and observations, which are largest during TWS maxima in spring and TWS minima in autumn (Figure 8.21). Furthermore, a larger impact of covariance inflation is observed for BS2 compared to DB.

(a) RMSD [mm]



(b) Residual Correlation [–]

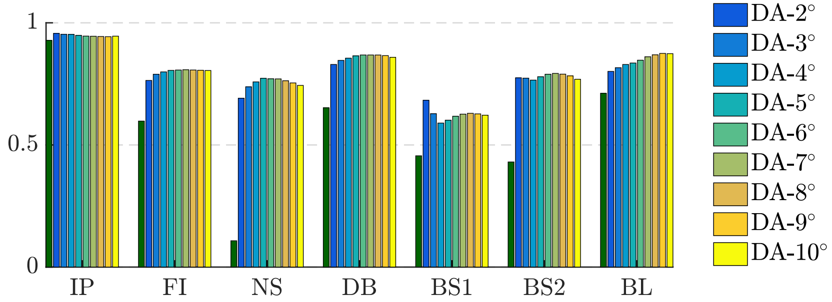
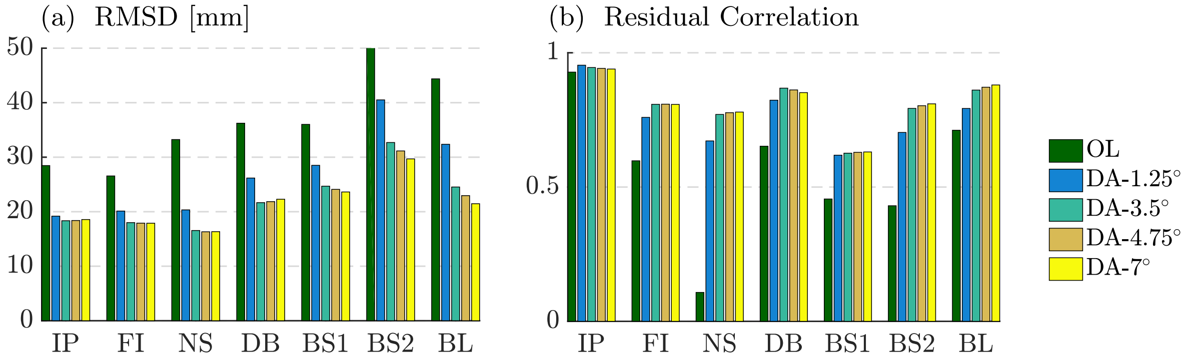


Figure 8.19: Root mean square deviation (RMSD) and residual correlation after de-trending and de-seasoning of TWSA for OL versus RF, and for DA versus RF considering different localization radii (OL: open-loop run, RF: reference run, DA: assimilation run). The exponential decay function was chosen to half of the localization radius.



(a) RMSD [mm]

(b) Residual Correlation

Figure 8.20: Root mean square deviation (RMSD) and residual correlation after de-trending and de-seasoning of TWSA for OL versus RF, and for DA versus RF applying a localization radius of 7° and different exponential decay functions that reach $1/e$ at selected distances (OL: open-loop run, RF: reference run, DA: assimilation run).

Table 8.5: Experiment 5 – Influence of the forgetting factor

Filter algorithm	LESTKF	**Ensemble size**	64
Inflation factor	1, 0.9, ..., 0.5	**Observations**	RF, perturbed, 2°
Localization	7° (3.5°)	**Assumed errors**	correlated

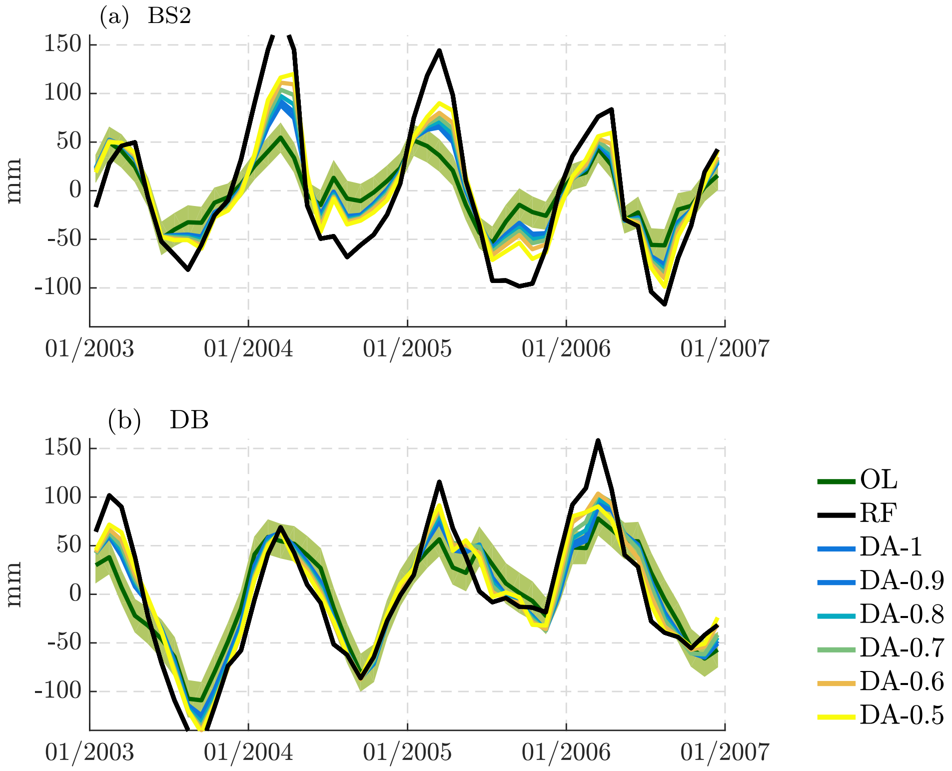


Figure 8.21: TWS simulated by the reference run (RF), the open-loop run (OL), and the assimilation run (DA) using different forgetting factors. Results are shown exemplarily for (a) the catchments in the Eastern part of the Baltic Sea (BS2) and, (b) the Danube catchment (DB). The green shading indicates the ensemble spread of OL.

Down to a forgetting factor of 0.7 the RMSD decreases and residual correlation increases (Figure 8.22). Choosing a smaller forgetting factor leads to improvement mainly over Eastern European basins and to degradation over the Iberian peninsula. However, a forgetting factor smaller than 0.7 leads to unrealistic growth of the model ensemble spread over time, while it remains stable for a forgetting factor of 0.7. Therefore, in the following experiments the forgetting factor was set to 0.7, which corresponds to an inflation of the ensemble of state vectors by 1.2. Khaki et al. (2017b), for instance, inflates the model ensemble by 1.12.

8.2.6 Influence of the Observation Grid

The spatial resolution of GRACE data is limited to about 200 km. Thus, choosing an observation grid size smaller than 2° or 3° is not expected to provide any more information. However, the choice of the observation grid is expected to have a huge influence in the case of white noise as the same information is introduced into the model multiple times. In contrast, for correlated noise similar results are expected for an observation grid size below the spatial resolution of GRACE. As the synthetic observations were not filtered spatially and, thus, represent smaller spatial structures, higher resolution observation grids might lead to better performance here, which will not be the case for real GRACE data.

Forman and Reichle (2013) introduced GRACE-derived TWSA for a different number of subbasins assuming white noise, and found an improved performance when increasing the number of basins. They recommended to assimilate GRACE-derived TWSA on the smallest scale at which errors can be assumed uncorrelated. Later studies took full error correlations between grid cells into account and assimilated GRACE data mostly on a 1° grid (Schumacher et al.,

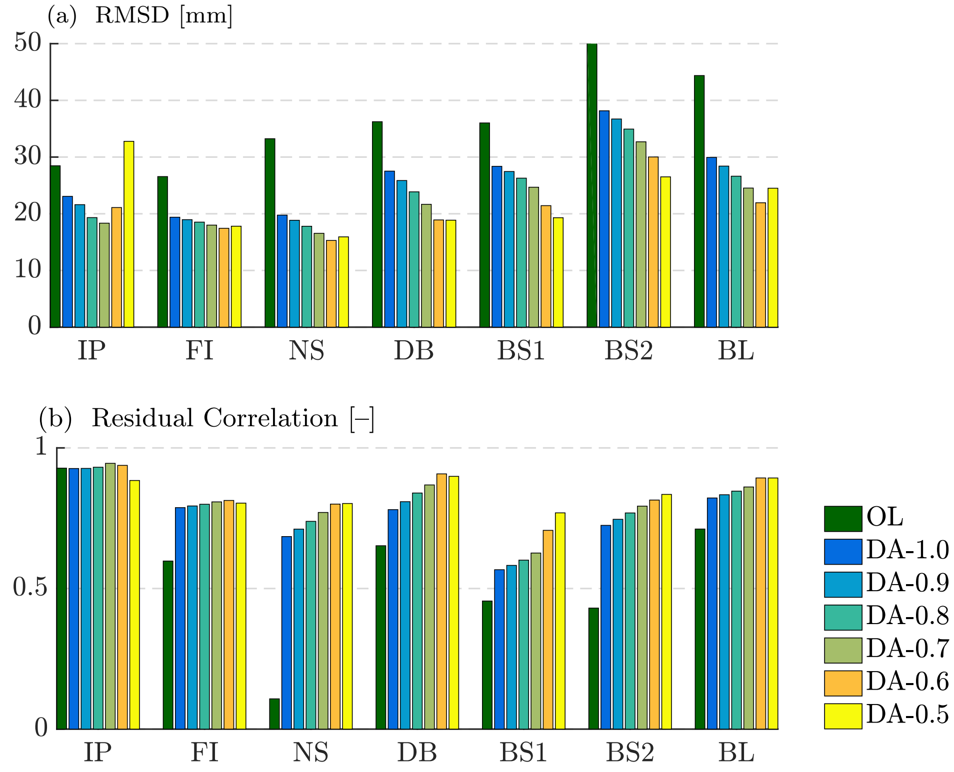


Figure 8.22: Root mean square deviation (RMSD) and residual correlation after de-trending and de-seasoning of TWSA for OL versus RF, and for DA versus RF applying different forgetting factors (OL: open-loop run, RF: reference run, DA: assimilation run).

2018; Tangdamrongsub et al., 2017; Khaki et al., 2017b). In fact, for a global filter, Khaki et al. (2017b) obtained the best result when using a 4° observation grid. For a local filter, best results were achieved for a 3° observation grid.

In this experiment, GRACE data was assimilated on 0.5° to 5° grids assuming (i) colored noise, and (ii) white noise. Data assimilation was performed with the LESTKF applying a localization radius of 7°. For the observation grid of 0.5° a localization radius of 3.5° was chosen for computational efficiency.

Table 8.6: Experiment 6 – Influence of the observation grid

Filter algorithm	LESTKF	**Ensemble size**	64
Inflation factor	0.7	**Observations**	RF, perturbed, 0.5°, 1°, 2°, 3°, 4°, 5°
Localization	7° (3.5°)	**Assumed errors**	white, correlated

In the case of correlated noise, observation grids of 2°, 3°, and 4° perform best in terms of RMSD and residual correlation (Figure 8.23). The 5° grid does not inform the model about smaller scale structures and thus performs worse. Interestingly, observation grids of 0.5° and 1° lead to the largest RMSD and to the smallest residual correlation, and perform even worse than the 5° grid. In the case of the 0.5° grid, this could be related to the smaller localization radius. Averaged over all seven catchments, the 2° and 3° observation grids leads to the best validation measures with an average RMSD of 23 mm and 22 mm, and average residual correlation of 0.82 and 0.81. The results are to some extent in line with Khaki et al. (2017b), who also obtained best performance for the 3° grid, but significantly worse performance for the 2°

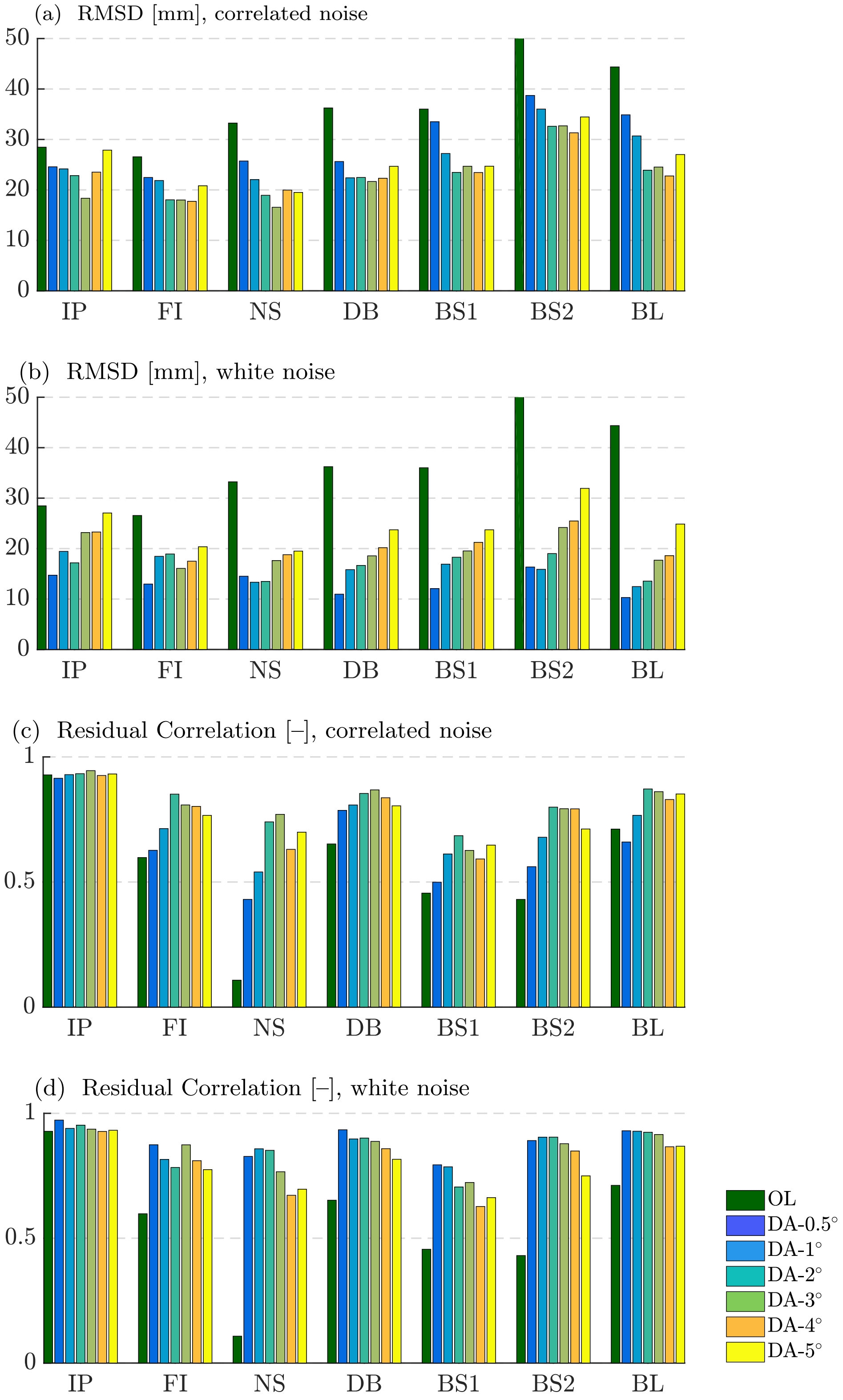


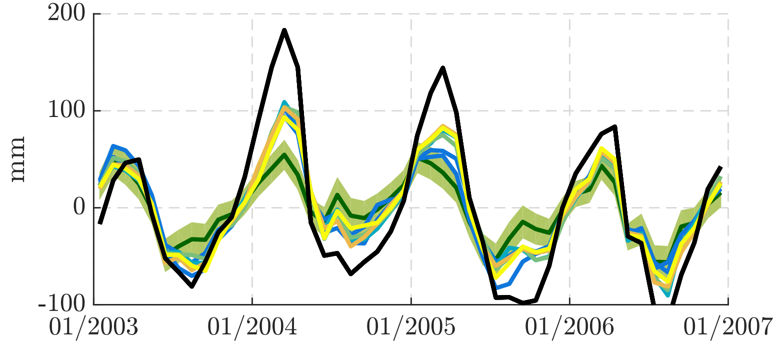
Figure 8.23: Root mean square deviation (RMSD) and residual correlation after de-trending and de-seasoning of TWSA for OL versus RF, and for DA versus RF considering data assimilation scenarios with different observation grids and (a,c) colored observation noise and (b,d) white observation noise (OL: open-loop run, RF: reference run, DA: assimilation run).

grid.

In the case of white noise, differences between the observation grids are much larger (Figure 8.23 b – d), as the finer the grid the more trust is placed in the observations. The RMSD doubles from the 0.5° to the 5° grid and residual correlation decreases.

The effects from correlated and white noise on the TWSA time series are shown exemplarily in Figure 8.24 for the catchment BS2. In the case of correlated noise, the differences between 2° to 5° observation grids are rather small, while 0.5° and 1° grids stand out especially at the beginning of 2005. In the case of white noise, applying the 0.5° grid nearly completely pulls the model towards the observations. Obviously, due to overfitting to the observations the weight of the model is close to zero. In contrast, for 4° and 5° grids, results from white noise are similar to those from correlated noise. This means, data assimilation results performed with correlated noise can be approximated by experiments with white noise, when choosing the observation grid space accordingly. However, some differences can still arise. Based on these findings, the following experiment were performed using a 2° observation grid and considering correlated errors.

(a) correlated noise



(b) white noise

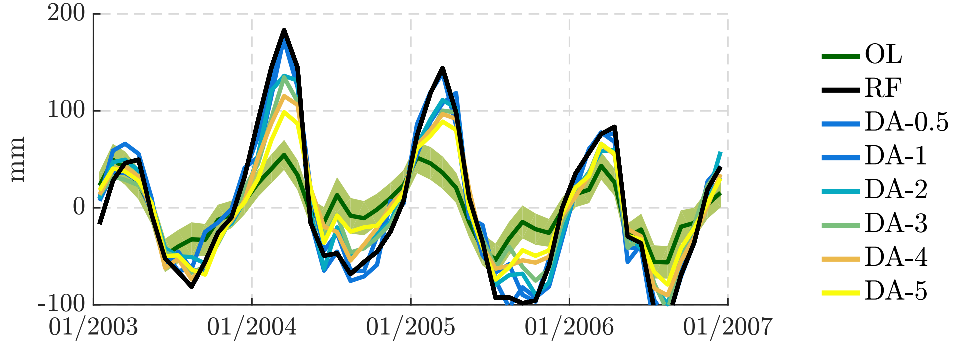


Figure 8.24: Total water storage anomalies (TWSA) simulated by the reference run (RF), by the open-loop (OL) run, and by the assimilation runs (DA) applying synthetic observations on different grids are shown for the catchment BS2.

8.2.7 Influence of Data Assimilation on Sub-Monthly TWS Variability

Improving modeled TWS through data assimilation using monthly TWSA is expected to improve short-term fluxes. As a result, TWS variability at sub-monthly time scales also has the potential to improve. The following experiment was performed with the optimal settings derived above. Synthetic observations on a 2° grid were introduced into the model via the LESTKF assuming correlated errors (Table 8.7).

Table 8.7: Experiment 7 – Influence of data assimilation on sub-monthly TWS variability

Filter algorithm	LESTKF	**Ensemble size**	64
Inflation factor	0.7	**Observations**	RF, perturbed, 2°
Localization	7° (3.5°)	**Assumed errors**	correlated

Submonthly TWS variability was assessed by comparing daily TWSA from RF, OL, and DA after removing (i) the mean of each month and (ii) the mean and the trend of each month (Table 8.8). When removing only the mean of each month, correlation and RMSD improve for all catchments by up to 23.9 % and 15.4 %, respectively. However, especially in Eastern European catchments data assimilation increases the annual amplitude, which results in linear patterns within each month. When removing mean and trend of each month, correlation and RMSD degrade by about 3 % in most catchments and by up to 8.9 % for BL. This means, pure daily TWS variability cannot be improved with respect to RF by assimilating monthly TWS.

Table 8.8: Root mean square deviation (RMSD) and correlation between the reference run and the assimilation run are shown together with the improvement relative to the open-loop run in percent. In columns two and three, the monthly means were removed and in columns four and five the trend of each month was removed additionally.

Catchment	**Mean removed**		**Mean & trend removed**	
	Correlation	**RMSD**	**Correlation**	**RMSD**
IP	0.87 (+0.7 %)	5.73 (+1.2 %)	0.68 (-2.3 %)	4.22 (-3.6 %)
FI	0.75 (+2.3 %)	9.36 (+1.7 %)	0.68 (-2.1 %)	4.82 (-2.9 %)
NS	0.76 (+5.3 %)	8.35 (+3.4 %)	0.59 (-3.1 %)	4.96 (-4.4 %)
DB	0.76 (+8.2 %)	8.27 (+9.0 %)	0.53 (-2.4 %)	5.37 (-2.9 %)
BS1	0.67 (+12.0 %)	9.63 (+7.6 %)	0.55 (-2.8 %)	4.62 (-0.9 %)
BS2	0.65 (+15.9 %)	12.77 (+8.1 %)	0.51 (+0.2 %)	5.88 (-1.2 %)
BL	0.73 (+23.9 %)	9.65 (+15.4 %)	0.52 (-6.8 %)	4.43 (-8.9 %)

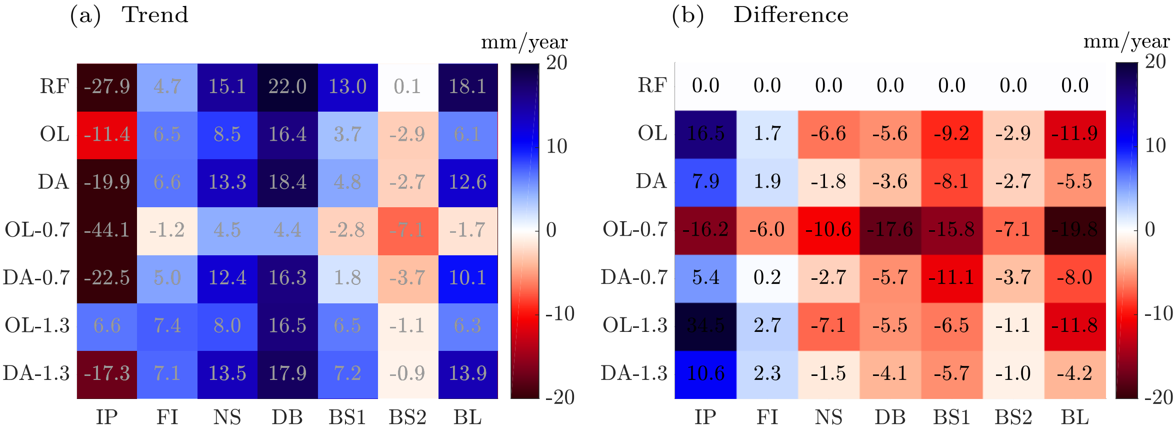
8.2.8 Influence of Biases in Precipitation Forcings

Biases in fluxes create trends in storages. Assimilating GRACE observations is expected to improve the trend of modeled TWS. In the following experiments, CLM was run with biased precipitation forcings. Precipitation was multiplied by 0.7 and 1.3, respectively, expecting, in the first case, to introduce a negative trend in storage and, in the second case, to introduce a positive trend in storage. Data assimilation is expected to adjust the trends towards the trends of RF. The following experiments all start from the same model states in January 2003 with the settings given in Table 8.9.

Table 8.9: Experiment 8 – Influence of biases in precipitation forcings

Filter algorithm	LESTKF	**Ensemble size**	64
Inflation factor	0.7	**Observations**	RF, perturbed, 2°
Localization	7° (3.5°)	**Assumed errors**	correlated
		Biased precipiation	0.7, 1.3

Trends were evaluated for RF, OL, and DA produced with the original forcing data, and for OL and DA produced with biased forcing data (Figure 8.25). Blue color indicates a positive trend (increasing TWS) and red color indicates a negative trend (decreasing TWS).



(a) Trend

	IP	FI	NS	DB	BS1	BS2	BL
RF	-27.9	4.7	15.1	22.0	13.0	0.1	18.1
OL	-11.4	6.5	8.5	16.4	3.7	-2.9	6.1
DA	-19.9	6.6	13.3	18.4	4.8	-2.7	12.6
OL-0.7	-44.1	-1.2	4.5	4.4	-2.8	-7.1	-1.7
DA-0.7	-22.5	5.0	12.4	16.3	1.8	-3.7	10.1
OL-1.3	6.6	7.4	8.0	16.5	6.5	-1.1	6.3
DA-1.3	-17.3	7.1	13.5	17.9	7.2	-0.9	13.9

(b) Difference

	IP	FI	NS	DB	BS1	BS2	BL
RF	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OL	16.5	1.7	-6.6	-5.6	-9.2	-2.9	-11.9
DA	7.9	1.9	-1.8	-3.6	-8.1	-2.7	-5.5
OL-0.7	-16.2	-6.0	-10.6	-17.6	-15.8	-7.1	-19.8
DA-0.7	5.4	0.2	-2.7	-5.7	-11.1	-3.7	-8.0
OL-1.3	34.5	2.7	-7.1	-5.5	-6.5	-1.1	-11.8
DA-1.3	10.6	2.3	-1.5	-4.1	-5.7	-1.0	-4.2

Figure 8.25: (a) Trend [mm/year] in total water storage (TWS) for the reference run (RF), for the open-loop (OL) run, and for data assimilation (DA) experiments with different precipitation forcings. Precipitation forcings were biased by multiplication with 0.7 and 1.3. Blue color indicates a positive trend (increasing TWS) and red color indicates a negative trend (decreasing TWS). (b) shows the difference between RF and the other model variants. Blue color indicates a larger trend compared to RF, and red indicates a smaller trend compared to RF.

Trends from DA agree better with RF than trends from OL (Figure 8.25 b). Biasing precipitation by the factor 0.7 leads to significantly decreased trends in TWS (OL-0.7). In contrast, a factor of 1.3 only slightly increases the trends in TWS (OL-1.3). One explanation could be that reducing precipitation causes drying of the soil, while increased precipitation predominantly affects surface runoff. Trends from DA-0.7 and DA-1.3 are similar to those from DA. This indicates that assimilating TWS observations can indeed help to remove trends introduced by biases precipitation data. Further assessments should consider the impact on other fluxes, e.g., surface runoff and evapotranspiration.

8.2.9 Influence of Phase Shifts between Model and Observations

Hydrological models often lag behind GRACE-derived TWSA (Zhang et al., 2017). In this experiment, a synthetic phase shift between model and observations was introduced by shifting the synthetic observations (RF) by one month, i.e., in January 2003 observations from February 2003 are assimilated, which corresponds to the model lagging behind the observations by one month.

Table 8.10: Experiment 9 – Influence from a phase shift between model and observations

Filter algorithm	LESTKF	**Ensemble size**	64
Inflation factor	0.7	**Observations**	RF, perturbed, 2°, one month ahead
Localization	7° (3.5°)	**Assumed errors**	correlated

The phase shift between shifted RF and the perturbed model was computed by fitting trend, annual, and semiannual signal to the daily modeled TWS time series. In the OL case, a phase shift between 9 and 22 days is obtained for the individual catchments (Table 8.11). After data assimilation, the phase shift reduces to 6 to 15 days. For BS2 no reduction of the phase shift is obtained. However, in the standard case, without shifted observations, data assimilation also changes the phase shift between model and observations by up to 7 days.

In line with previous experiments, data assimilation moved the relative annual amplitudes towards 1 (Table 8.11). Particularly large improvements were found for Eastern European catchments where the relative annual amplitudes increased from values between 0.38 and 0.48 to values between 0.64 and 0.78. Interestingly, improvements of the relative annual amplitude are about 10 % to 20 % larger in the case of shifted observations compared to the standard case.

The results lead to the conclusion that data assimilation can reduce phase shifts between model and observations, but in order to completely resolve this issue parameter calibration or changes in the model structure would be necessary. Furthermore, the improvement of relative annual amplitudes is sensitive to a phase shift between model and observations.

Table 8.11: Phase shift and relative annual amplitudes computed in days for OL versus RF and DA versus RF when introducing synthetic observations that are one month ahead of the perturbed model (OL: open-loop run, RF: reference run, DA: assimilation run).

Catchment		IP	FI	NS	DB	BS1	BS2	BL
Phase shift	OL	22	14	10	19	14	9	14
	UP	15	9	6	11	10	9	9
Relative Amp.	OL	0.89	0.78	0.78	0.65	0.48	0.38	0.44
	UP	1.07	0.92	0.98	0.89	0.75	0.64	0.78

8.2.10 Key Messages of Synthetic Assimilation Experiments

The synthetic twin experiments were performed to investigate the impact of different data assimilation strategies on the assimilation results. Within this scope, the changes of the validation measures RMSD, residual correlation, and relative annual amplitude through data assimilation were assessed. All validation measures were computed with respect to the reference model that was used to simulate the synthetic observations.

As expected, the representation of simulated TWSA generally improved through data assimilation, and, to some extent also the representation of individual TWS components was improved. However, in some cases, internal correlations between the state variables and physical constraints lead to deterioration of the representation of individual TWS components.

Among all parameters defining the assimilation strategy, localization was identified to have the largest potential to improve the assimilation results. Results from the ESTKF and from its local variant LESTKF differed by up to 50 % in terms of RMSD and residual correlation (Figure 8.14). These results demonstrated once again that the treatment of long-range correlations between model grid cells and, in particular, between observation grid cells play an important role when assimilating TWSA.

An only slightly smaller impact on the data assimilation results was obtained for the choice of the observation error model, i.e. assuming colored noise or white noise for gridded TWSA. Interestingly, the local filter was more sensitive to differences in the observation error model than the global filters. For the global filters, the choice of the observation error model caused changes in RMSD and residual correlation of about 10 %. For the LESTKF, the choice of the observation error model lead to changes in RMSD and residual correlation of 20 – 40 %. Although considering correlations between gridded TWSA provides a more realistic error estimates, the agreement with the synthetic truth was not necessarily improved. This phenomena can be explained by the higher weights that are placed on TWSA observations in the filter analysis step, when these observations are assumed to be uncorrelated. Error correlations between spherical harmonic coefficients played a negligible role.

The choice of the localization radius had a smaller impact on the assimilation results than localization itself. Increasing the localization radius from 2° to 10° changed the validation measures by 5 – 20 % (Figure 8.19). The performed experiments also showed that the exponential decay constant must have a certain minimum radius to obtain stable results for correlated observation noise (Figure 8.20). For small localization radii or small exponential decay constants, the assimilation results approach the case of white observation noise. The impact of the forgetting factor had a similar magnitude like the impact of the localization radius. Varying the forgetting factor from 1 to 0.5 changed RMSD and residual correlation by 10 – 30 % (Figure 8.22).

The choice of the observation grid space had a large influence in the case of white noise, with changes of the validation measures of up to 50 % (Figure 8.23). In the case of colored noise, the validation measures still changed by 20 – 40 %, when varying the observation grid space from 5° to 0.5°. In the case of white noise, the performance of the assimilated model decreased with increasing observation grid space. In contrast, in the case of colored noise, best results were obtained for a grid space of 2° and worst results for 5° and 0.5° grids. Using real GRACE data, the differences arising from the observation grid space might be smaller, as the information content of the observations is limited due to the spatial resolution of GRACE derived TWSA.

A small impact on the assimilation results can be attributed to the number of ensemble members. As long as more than 32 ensemble members were used, changes in the validation measures were usually below 10 % and in many catchments even below 5 % (Figure 8.12). Furthermore, the adjustment of the observations to the temporal mean of the perturbed model had a rather small impact (Figure 8.11).

Promising results were obtained for the impact of data assimilation on artificial trends, which can be introduced into the model through biases in precipitation forcings, and on phase shifts of simulated TWSA. Indeed, the twin experiments showed that the assimilation of TWSA has the potential to significantly improve simulated trends in TWSA and to slightly reduce phase shifts between model and observations.

Assimilation increments showed seasonal and regional patterns with varying contribution to the individual storage compartments. The analysis of the assimilation increments proved to be a valuable tool to determine model deficits. Interestingly, on the long term, the assimilation of synthetic TWSA into CLM3.5 only slightly violated the water balance of the model, at least, when applying a local filter algorithm.

In summary, the results of the twin experiments enabled to derive an optimal strategy for assimilating GRACE-derived TWSA into CLM3.5 over Europe. According to these experiments, the LESTKF should be applied with a localization radius of 7° and the exponential decay constant should be chosen to 3.5°. A forgetting factor of 0.7 is suitable. The observation error model should consider correlated observation errors and observations should be assimilated on a 2° grid. 64 ensemble members are an optimal ensemble size, but 32 ensemble members also provide stable results.

8.3 Assimilation of GRACE-Derived Total Water Storage Observations

8.3.1 Set-up of Assimilation Experiments with Real GRACE Data

GRACE-derived TWSA were assimilated into CLM3.5 for the time period 2003 to 2010. GRACE data was processed according to Section 5.1. Following the implications of the synthetic experiments (Section 8.2), the LESTKF was applied with 64 ensemble members taking fully correlated observation errors into account. Furthermore, a localization radius of 7° with an exponential decay constant of 3.5° and a forgetting factor of 0.7 were used. The synthetic observations show smaller scale patterns compared to real GRACE data. Therefore, GRACE data was again assimilated on 1° to 5° grids to confirm the optimal observation grid space derived in Section 8.2.6. Before data assimilation, the temporal mean of GRACE-derived TWSA was adjusted to the temporal mean of the open-loop run (Figure 8.26). After data assimilation, the ensemble mean was compared to GRACE-derived TWSA and validated against independent observation-based data sets.

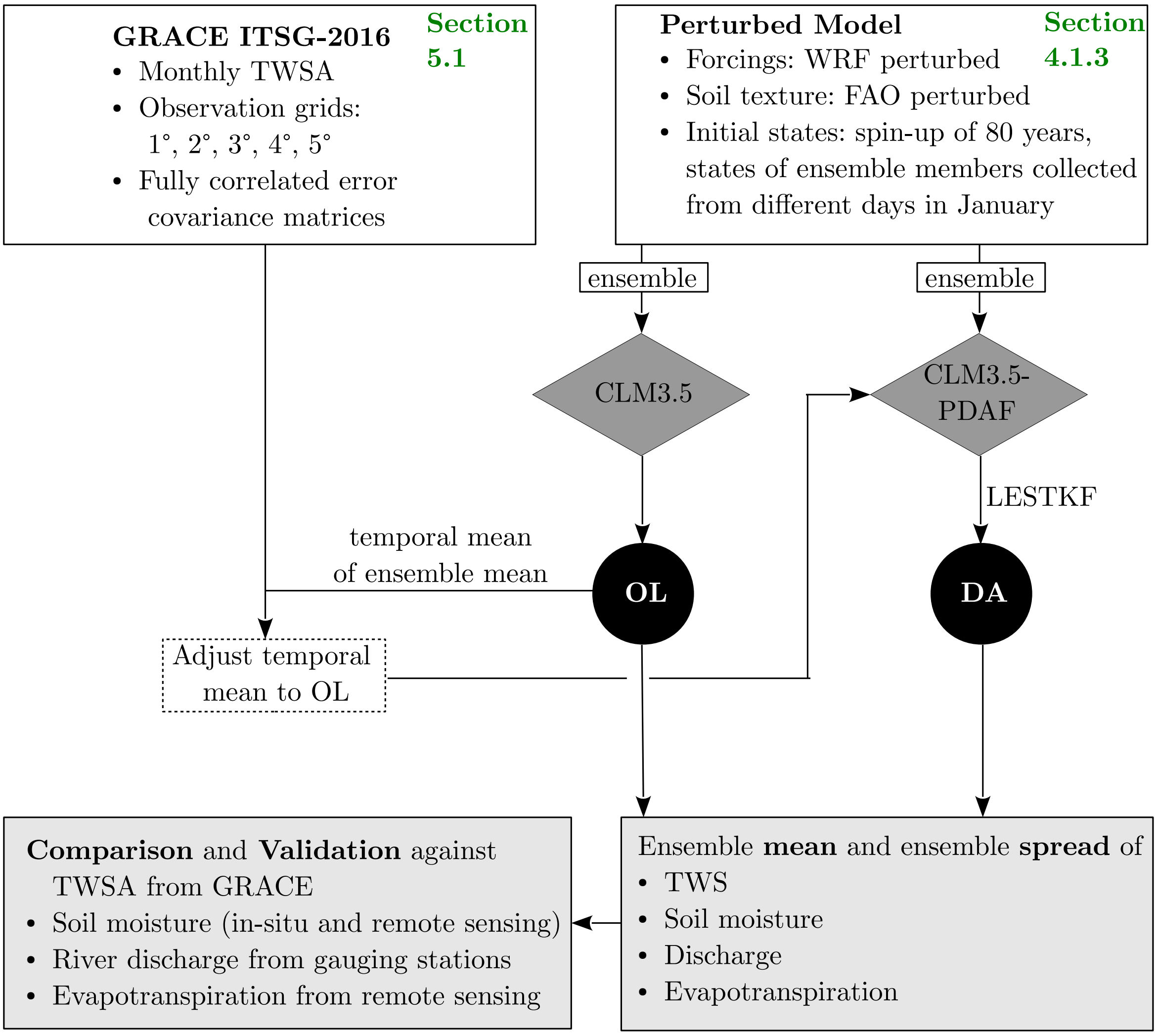


Figure 8.26: Flowchart visualizing the assimilation of GRACE-derived total water storage anomalies (TWSA) into the perturbed CLM3.5 model (DA). Additionally, an open-loop CLM3.5 run (OL) of the perturbed model was performed in order to adjust the temporal mean of the GRACE observations.

First, the impact of assimilating GRACE data on TWS estimates is assessed by comparing the output of the assimilated model to GRACE observations (Section 8.3.2) and by analyzing the assimilation increments (Section 8.3.3). One important aspect of the GRACE assimilation framework is the disaggregation of GRACE data horizontally and vertically, which should draw each TWS component towards its true state. Here, the model states were validated against independent soil moisture observations from in-situ stations and remote sensing (Section 8.3.4). Changes in TWS storage due to data assimilation cause land-surface models to compensate by changes fluxes such as surface infiltration, drainage, runoff, or evapotranspiration. Here, modeled gridded runoff was validated against discharge at gauging stations (Section 8.3.5). Modeled evapotranspiration was validated against a gridded observation-based data set from FLUXNET stations (Section 8.3.6). Furthermore, the impact of data assimilation on the diurnal cycle of evapotranspiration was assessed. Finally, data assimilation results were used to obtain a detailed picture (in terms of temporal and spatial resolution) of the European heat wave 2003 and the European flood 2006.

8.3.2 Comparison of GRACE Data and Output from the Assimilated Model

By design, TWSA from the assimilated model (DA) should be in between the model state before the analysis step (HO) and the GRACE observations. Deviations from this assumption can be caused by close agreement between model and observations, by an extremely small ensemble spread, or by spurious correlations. Please note that HO differs from the open-loop run (OL) due to assimilation experiments that were applied during the previous assimilation steps.

Zaitchik et al. (2008) found a closer agreement of the annual cycle of modeled TWS and GRACE observations after data assimilation. Later studies showed that more than just the seasonal cycle was improved (Reager et al., 2015; Kumar et al., 2016). Recent studies focused on the improvement of TWS trends, which can usually not be represented realistically by models. In this scope, Girotto et al. (2016) assessed modeled trends due to water storage depletion over India. Schumacher et al. (2018) found an improved representation of negative trends in storage due to the Millenium Drought in the Murray-Darling basin. Likewise Khaki et al. (2018) showed a better representation of water storage depletion within Iran due to GRACE assimilation.

Data assimilation implies temporal downscaling of the monthly GRACE observations. Figure 8.27 shows time series of daily averaged TWSA from OL (green) and DA (blue), together with their corresponding ensemble spreads, for the seven major catchments defined in Figure 8.1. Additionally, monthly GRACE-derived TWSA (black) and the corresponding errors as well as the assimilation increments (red) are displayed. The seasonality is captured well by OL for the Western European catchments (IP, FI, NS, DB), while for the Eastern European catchments (BS1, BS2, BL) the annual amplitude is clearly underestimated. Assimilating GRACE data increases the seasonal signal in Eastern European catchments, and also impacts the interannual variability of modeled TWS. Yet, sub-monthly patterns remain mostly the same for OL and DA.

The ensemble spread of catchment-averaged TWSA was computed by averaging the ensemble spread of the individual grid cells of the catchment. Generally, the ensemble spread of DA is between 10 mm and 40 mm depending on catchment and season. For comparison, the ensemble spread of OL varies between 10 – 20 mm. Yet, in case of the Iberian Peninsula (IP), the ensemble spread increases dramatically starting from 2005. The reason is that for 98 % of grid cells on the European continent, the water table lays within the soil column (Section 4.1.2). However, for individual grid cells in the South of the Iberian pensinsula and for some grid cells

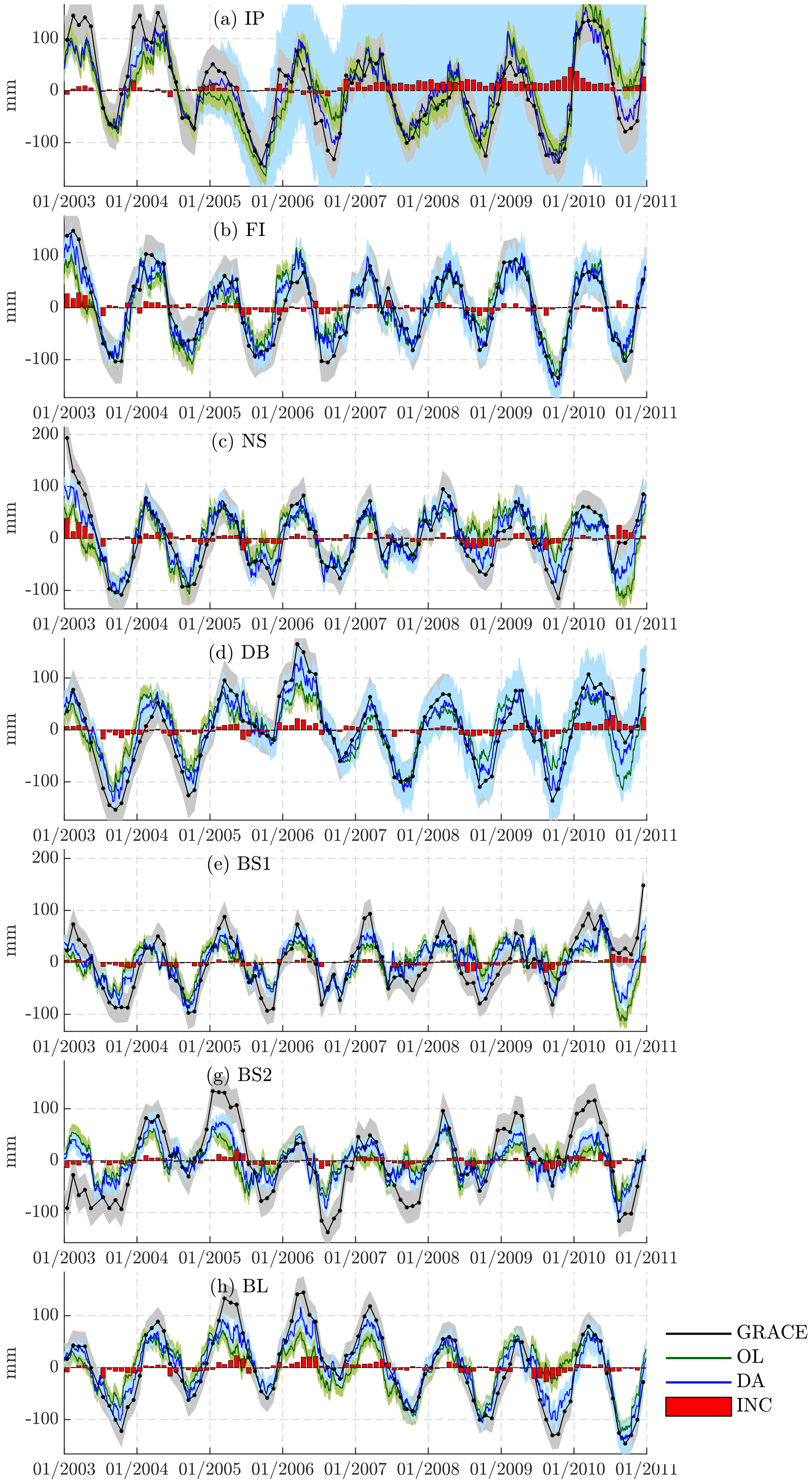
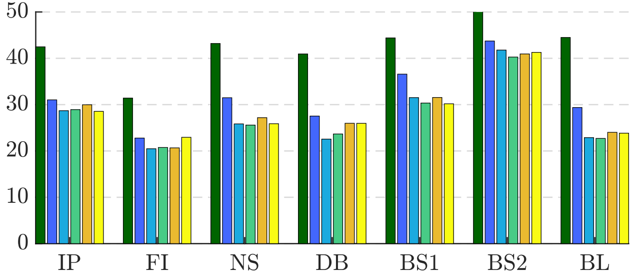


Figure 8.27: Daily averaged TWSA are shown for the ensemble mean of the open-loop run (OL, green) and the ensemble mean of the assimilation run (DA, blue) together with the corresponding ensemble spreads for the 7 major European catchments. GRACE observations and errors are drawn in black and the assimilation increments are drawn in red. The y-axis covers 350 mm in each subfigure.

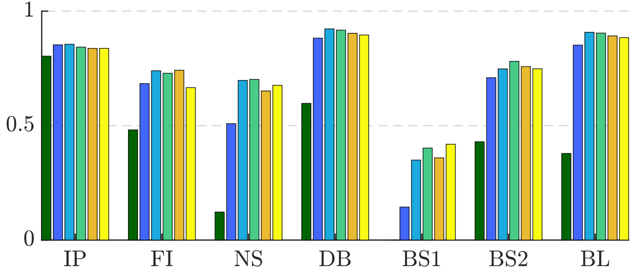
in the Danube catchment close to the Black Sea, the water table is below the soil column. This means that water seeps from the above soil layers into the unconfined aquifer. For specific ensemble members data assimilation leads to conditions in which the soil column starts to run empty. This also explains the large positive increments for the second half of the time series of IP. The problem could possibly be avoided by using a more recent CLM version, by adding more sophisticated constraints to the update step, or by replacing the lower soil layers by a more sophisticated groundwater model, e.g. ParFlow. However, the problem arises for few grid cells only and the resulting ensemble mean still provides a good estimate of catchment averaged TWS.

Modeled TWSA from OL and DA is validated against GRACE-derived TWSA in terms of RMSD, residual correlation, and relative annual amplitude for assimilation scenarios using different observation grids (Figure 8.28). The RMSD is reduced by 25 % to 50 % due to data assimilation with largest improvements for BL, where OL underestimates the annual amplitude by 40 %. For BS1 and BS2, the annual amplitude was underestimated even more, but the improvement from data assimilation is smaller. This can be explained by the smaller ensemble spread of DA for BS1 and BS2, which implies more confidence in the model than in the GRACE observations. In general, the annual amplitude of TWS is increased by data assimilation, but it is still smaller than the annual amplitude of GRACE.

(a) RMSD [mm]



(b) Residual Correlation [–]



(c) Relative Annual Amplitude [–]

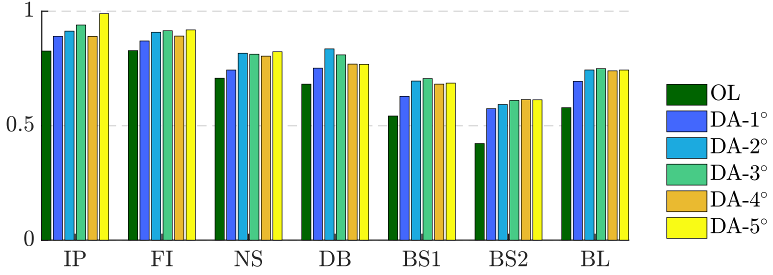
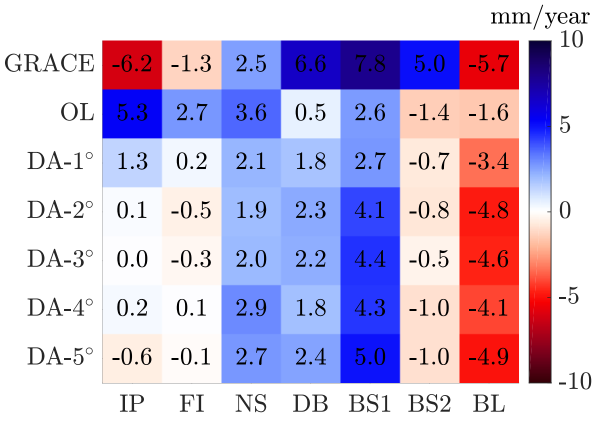


Figure 8.28: Root mean square deviation (RMSD), residual correlation after de-trending and de-seasoning, and relative annual amplitude for the open-loop run (OL) versus GRACE observations, and for the assimilated model (DA) versus GRACE observations, considering assimilation experiments with different observation grids.

In all catchments but IP, data assimilation has a huge contribution to interannual TWS variability. In fact, residual correlation computed from de-trended and de-seasoned time series increases by up to 0.6. While for OL residual correlation is below 0.5, for most basins residual correlation increases to values larger than 0.7 after data assimilation and reaches 0.9 for DB and BL. This improvement also leads to a better representation of extreme events. Examples are the flood in Danube in 2006 (Section 8.3.7.2) or the European heat wave in summer 2003 (Section 8.3.7.1).

Observation grids of 2° and 3° lead to the best validation measures for DA, which is in agreement with the results from the synthetic experiments (Section 8.2.6). Largest improvements are found here when moving from a 1° grid to a 2° grid which is in line with Khaki et al. (2017b). Additionally, Khaki et al. (2017b) found large differences in performance between 3° and 5° observation grids which cannot be confirmed here. One reason might be that Khaki et al. (2017b) used a different localization radius, which also has a strong influence on data assimilation results (Section 8.2.5). All in all, the results of this study show that GRACE adds only few information at spatial scales smaller than 5°.

Trends in TWS over Europe are small compared to regions like India or the High Plain Aquifers. Nevertheless, data assimilation may draw modeled TWS trends towards GRACE. The significance of the results is limited due to the short study period of only 8 years. Figure 8.29 shows TWS trends in the seven study catchments for GRACE, OL, and for different DA scenarios. GRACE observes negative trends in IP, FI, and BL. The negative trend of IP could be due to human influences, such as irrigation. Such influences are not modeled by CLM, which might be one reason for the diverging positive trend of OL. Data assimilation reduces this positive trend. Generally, data assimilation improves the agreement of trends between GRACE and the model. The choice of the observation grid space leads to differences in trends of up to 2.3 mm/year, which can be seen as an indicator for the uncertainty of the trend estimates.



	IP	FI	NS	DB	BS1	BS2	BL
GRACE	-6.2	-1.3	2.5	6.6	7.8	5.0	-5.7
OL	5.3	2.7	3.6	0.5	2.6	-1.4	-1.6
DA-1°	1.3	0.2	2.1	1.8	2.7	-0.7	-3.4
DA-2°	0.1	-0.5	1.9	2.3	4.1	-0.8	-4.8
DA-3°	0.0	-0.3	2.0	2.2	4.4	-0.5	-4.6
DA-4°	0.2	0.1	2.9	1.8	4.3	-1.0	-4.1
DA-5°	-0.6	-0.1	2.7	2.4	5.0	-1.0	-4.9

Figure 8.29: Trends in total water storage anomalies for GRACE data, for the open-loop run (OL), and for data assimilation scenarios (DA) based on different observation grids.

8.3.3 Analysis of Assimilation Increments

Assimilation increments provide evidence about TWS imbalances generated during data assimilation. The temporal mean of the assimilation increments indicates imbalances that are introduced into the model on the long term (Figure 8.30 a). The average magnitude of TWS changes in each assimilation time step is indicated by the temporal mean of the absolute values of the assimilation increments (Figure 8.30 b).

Increments from assimilation experiments with observations given on a 1° grid stand out with negative monthly increments of about -3 mm/month (Figure 8.30 a). In contrast, for observation grids of 2° to 5° mean monthly increments are between -1 mm/month and 1 mm/month for all basins except for IP and DB. Thus, for one year usually less than 10 mm of water is introduced into or removed from the model. For IP, on the average up to 10 mm of water are introduced each month, which is mainly due to grid cells running empty for some ensemble members (see also Figure 8.27). This is also the reason for the slightly larger increments of DB. Obviously, the assimilation increments are useful for detecting instabilities in the model. Absolute values of the assimilation increments are large for catchments in which DA has a large ensemble spread, e.g., for BL the ensemble has a spread of 18 mm, for BS2 a spread of 16 mm, and for BS1 a spread of 14 mm.

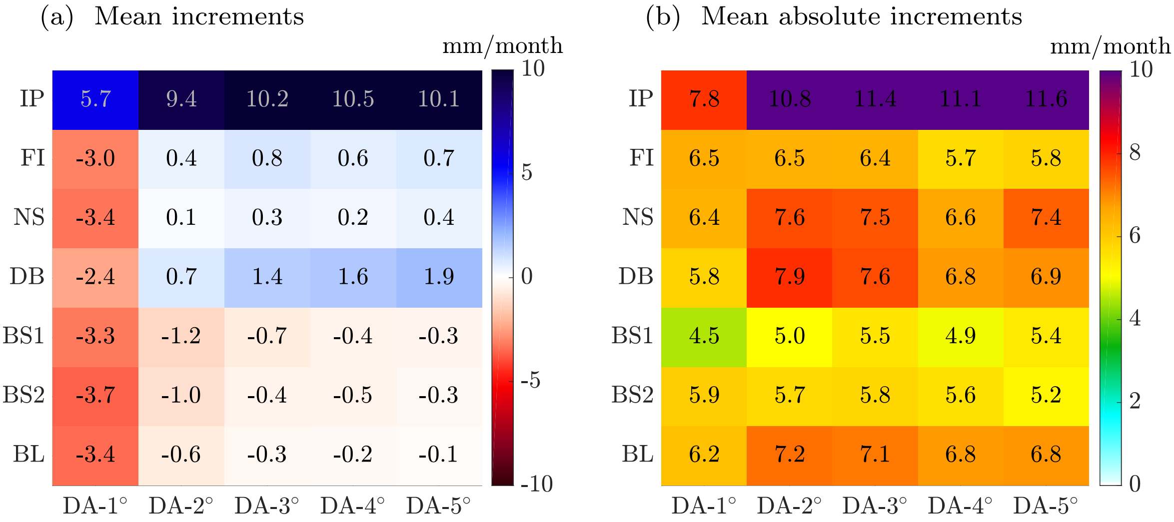


Figure 8.30: (a) Mean monthly TWS increments and (b) mean monthly absolute TWS increments averaged over the study period 2003 to 2010 for the seven major catchments and for data assimilation experiments with different observation grid space.

Assimilation increments are not uniformly distributed over time, but show a seasonal signal (Figure 8.7 d). Figure 8.31 shows average assimilation increments for each month of the year divided into the relevant compartments. Increments to canopy water and aquifer water are too small to be displayed. Usually, during the first half of the year water is introduced into the model, and during the second half of the year water is abstracted from the model. This indicates that the model in OL mode is not able to reproduce the complete range of TWS variability that is seen by GRACE. This might be due to deficiencies in the forcing data or due to deficiencies of the soil storage capacity. A sophisticated groundwater model could help to overcome this problem. For FI and DB nearly the complete assimilation increment is attributed to soil moisture (SOILLIQ). In contrast, for north-eastern catchments, soil ice (SOILICE) and snow water (H2OSNO) also play a role between December and January. The absolute increments of snow are about as large as the increments of soil liquid water for March and April in BS2. For BS2, the increments of soil ice amount to 10 % to 20 % of the increments of snow water.

8.3.4 Validation against Soil-Moisture Observations

Data assimilation primarily impacts modeled soil moisture (Section 8.3.3). Thus, the agreement of the model output with soil moisture observations should improve. However, the comparison of modeled soil moisture and observation-based data sets is challenging. Soil moisture

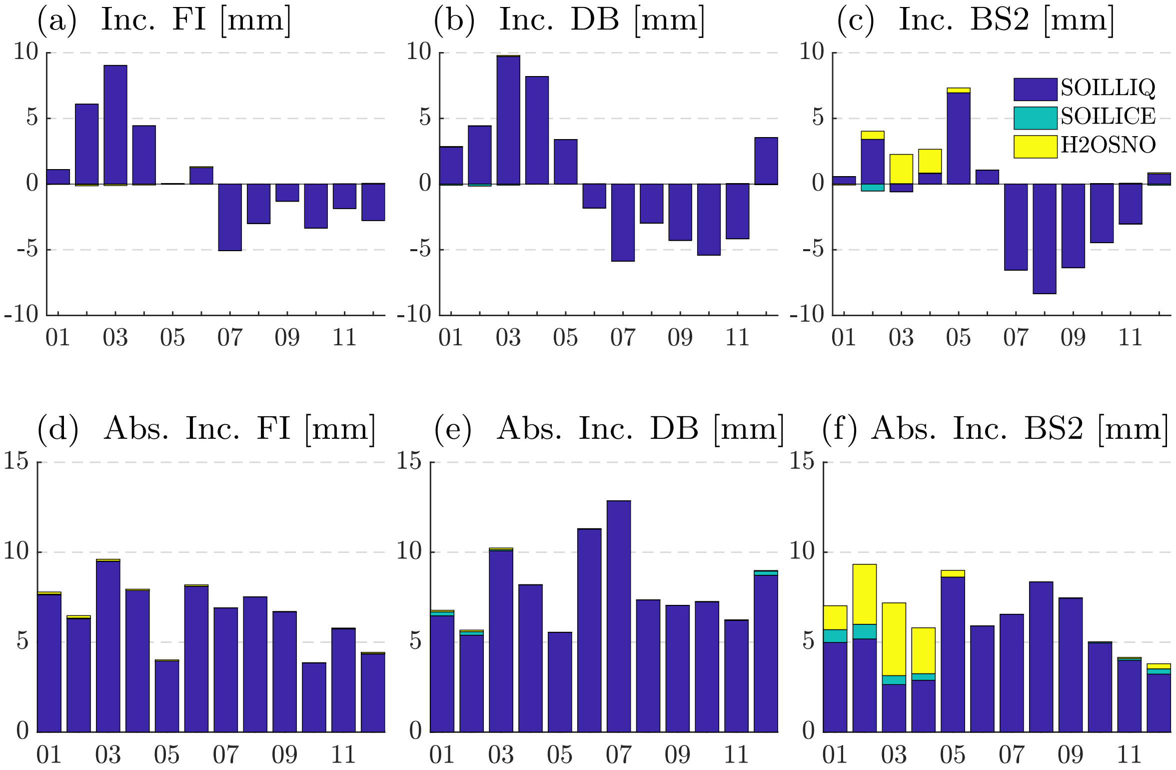


Figure 8.31: Monthly averages (2003 to 2010) of (a – c) assimilation increments and (d – f) absolute assimilation increments of the compartments soil liquid water (SOILLIQ, blue), soil ice (SOILICE, cyan), snow water (H2OSNO, yellow) for three selected different river catchments.

from remote sensing has a good coverage, but is sensitive only to the first few centimeters of the soil. Furthermore, limitations arise for areas covered with snow or vegetation. In contrast, in-situ soil moisture stations provide measurements at different depths, but these measurements are sparse. As in-situ soil moisture measurements are subject to the nearby hydrological conditions, comparisons with modeled soil moisture at 12.5 km spatial resolution is challenging. However, CLM has a relatively high spatial resolution of 12.5 km compared to global models, which usually have a spatial resolution of about 50 km. One advantage of the CLM model is the availability of 10 soil layers, which makes it possible to assign the in-situ observations to the respective depths.

Previous GRACE data assimilation studies show a mixed picture of the validation against soil moisture observations. Generally, data assimilation improved the agreement with in-situ soil moisture measurements only for few stations, while for the majority of stations no changes were obtained (Houborg et al., 2012; Li et al., 2012; Girotto et al., 2016; Kumar et al., 2016). Over Europe, Li et al. (2012) validated a GRACE-assimilated land surface model against in-situ soil moisture observations of the SMOSMANIA network located in South-Western France. They found an improvement of the correlation at a 10 % significance level only for 1 out of 12 stations. With respect to remotely sensed soil moisture, Tian et al. (2017) showed slight degradation of the GRACE assimilating model with respect to SMOS satellite observations.

In the following CLM-DA was validated against 67 in-situ soil moisture stations and against a combined product of remotely sensed soil moisture.

8.3.4.1 In-situ observations from the International Soil Moisture Network

Data assimilation results were validated against monthly time series of 67 soil moisture stations over Europe, which are organized within 17 networks (Section 4.3.1.1). The largest networks are located in North-Western Spain (REMEDHUS, 20 stations), South-Western France (SMOSMANIA, 11 stations), and in Austria (WEGENERNET, 12 station). Each station is compared to the closest CLM model grid cell. Note that the stations cover different time spans from 2 to 8 years and measure at different depths between 0.1 m to 1.5 m (Figure 4.5). Data assimilation produces larger increments for deeper soil layers, so that the impact of DA is expected to be larger for stations measuring deeper soil moisture.

For each station, soil moisture observations were averaged over all available measurement depths. Accordingly, the corresponding closest soil layers of CLM were vertically averaged. However, soil layers of CLM and depth of in-situ observations do not match exactly. This is in particular the case for deeper soil layers as the vertical resolution of the soil layers decreases with depth. To remove biases, the soil moisture time series were normalized to values between 0 and 1 using the maximum and minimum of each time series, i.e. $(sm_i - sm_{min})/(sm_{max} - sm_{min})$. Furthermore, validation measures refer to the particular time periods available for each station.

Figures 8.32, 8.33, and 8.34 show correlation, residual correlation, and RMSD between OL and ISMN observations and DA (2°) and ISMN observations.

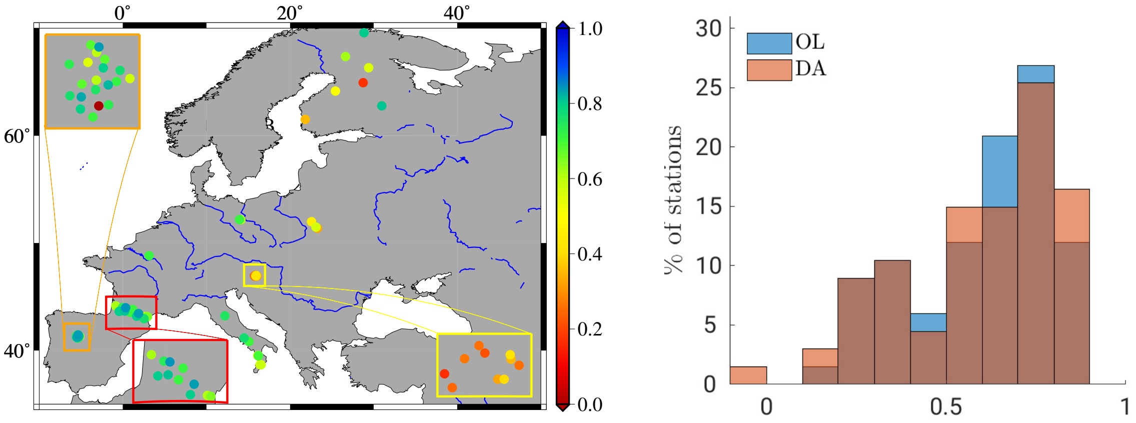
Over Western Europe, the main pattern of in-situ soil moisture stations and DA are in good agreement with correlation coefficients larger than 0.7 for most stations except for WEGENERNET (Figure 8.32 a). In total, 85 % of the stations have a correlation coefficient larger than 0.3, about 60 % of the stations have correlation coefficients larger than 0.6. The corresponding histogram shows that the amount of stations with correlation coefficients larger than 0.5 remains similar for OL and DA, but the number of stations with very high correlations (>0.8) increases from 8 for OL to 11 for DA. Generally, the impact of data assimilation on the quality of modeled soil moisture is rather small. These findings are consistent with the literature discussed above. Only for 15% of the stations, the correlation changes by more than 0.05 (Figure 8.32 b). In particular for the network SMOSMANIA in Southern France, which measures down to a depth of 0.3 m, and for stations in Italy no noticeable changes are induced by data assimilation for all three validation measures.

Residual correlation is larger than 0.3 only for about 40 % of the stations (Figure 8.33 a). This means for 60 % of the stations no significant correlation exists for interannual variability between modeled and observed soil moisture. Surprisingly, residual correlation changes by more than 0.05 for 30 % of the stations and by more than 0.1 for 13 % of the stations due to data assimilation (Figure 8.33 a). Significant improvements at the 0.1 level are determined for the Southern stations of the Spanish network REMEDHUS, for the station next to Paris, and for stations of the Austrian network WEGENERNET. All in all, data assimilation improved 5 stations by more than 0.1 in terms of residual correlation.

For 30 % of the soil moisture stations, the RMSD changes by more than 10 % due to data assimilation (Figure 8.34). Most of these stations belong to the Spanish network REMEDHUS, where the RMSD improves by up to 15% for several stations. Although these stations have only a depth of 30 cm, data assimilation has a surprisingly large effect here.

Assessing experiments of different observation grid space leads to ambiguous results. Residual correlation changes by less than 0.05 when varying the grid space from 5° to 1°. Likewise, for the RMSD no significant impact from the grid space of the observations is obtained. Moreover, no clear relationship between the maximum depth of the soil moisture measurements and the

(a) Correlation of DA versus ISMN



(b) Difference of correlation between OL vs. ISMN and DA vs. ISMN

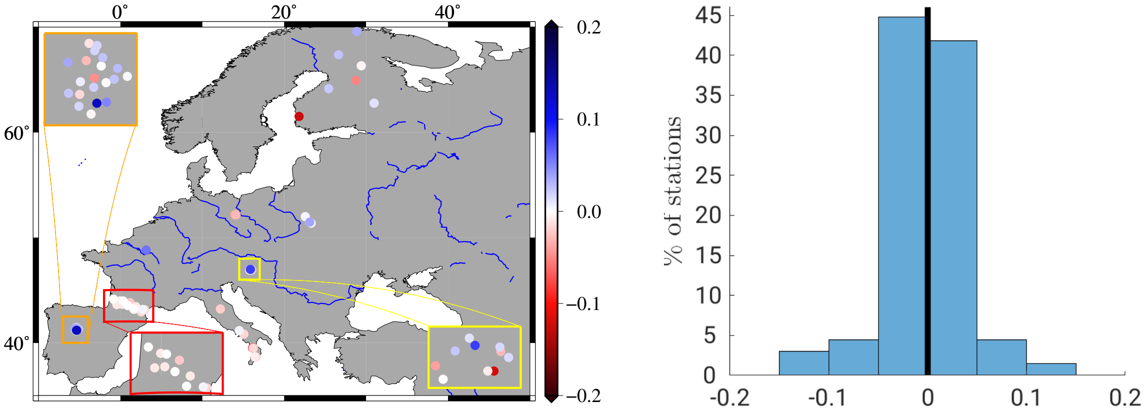
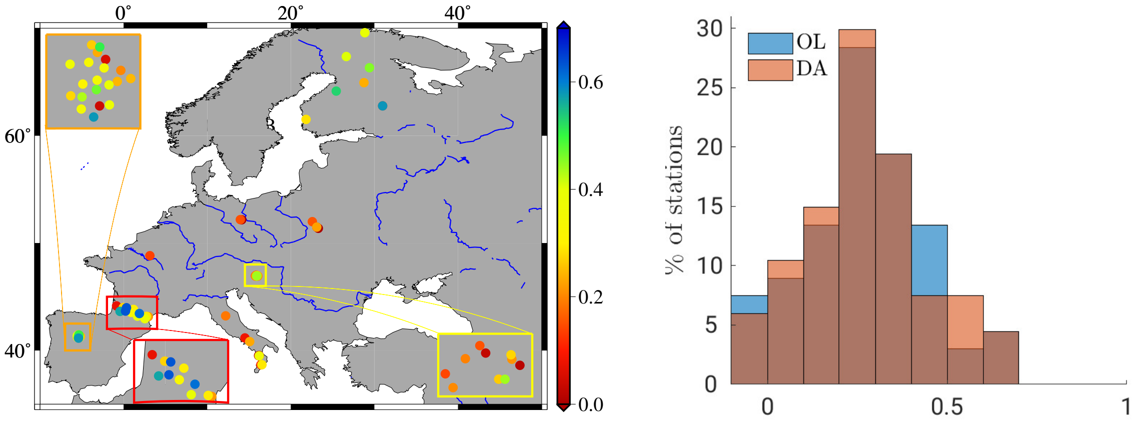


Figure 8.32: (a) Correlation between DA and ISMN soil moisture stations and the corresponding histograms when using DA or OL, respectively (OL: open-loop run, DA: assimilation run). (b) Difference in correlation when computing DA versus ISMN instead of OL versus ISMN, and the corresponding histogram of the differences. Blue color on the left indicates an increase in correlation through data assimilation.

(a) Residual correlation of DA versus ISMN



(b) Difference of residual correlation between OL vs. ISMN and DA vs. ISMN

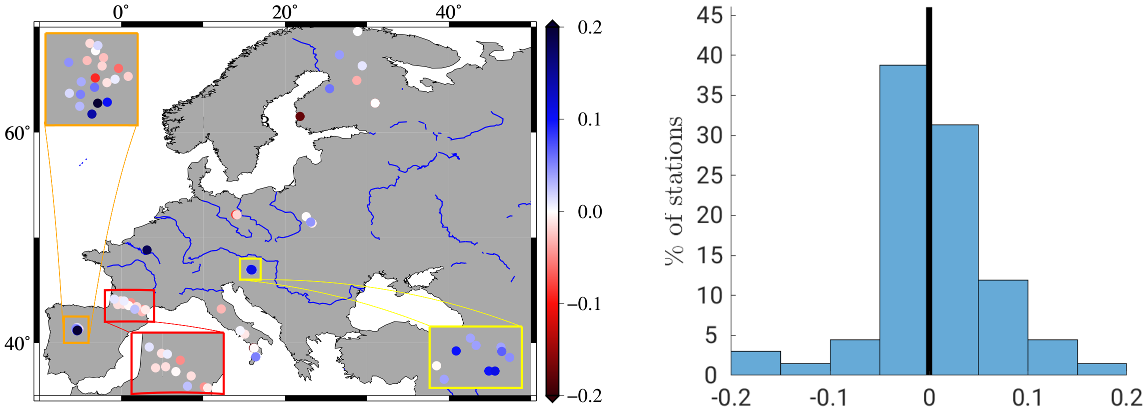
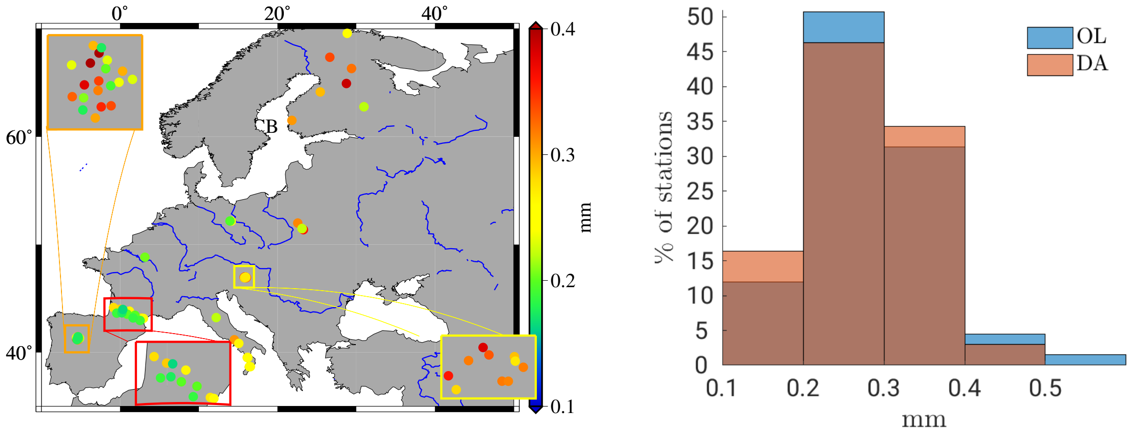


Figure 8.33: (a) Resiual correlation between DA and ISMN soil moisture stations and the corresponding histograms when using DA or OL, respectively (OL: open-loop run, DA: assimilation run). (b) Difference in residual correlation when computing DA versus ISMN instead of OL versus ISMN, and the corresponding histogram of the differences. Blue color on the left indicates an increase in residual correlation through data assimilation.

(a) RMSD of DA versus ISMN



(b) Difference in RMSD between OL vs. ISMN and DA vs. ISMN in %

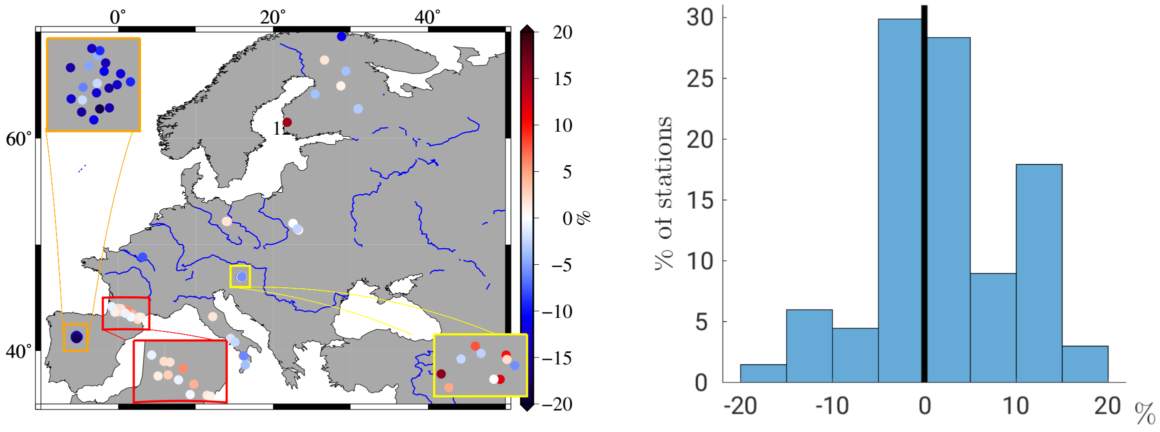


Figure 8.34: (a) Root mean square deviation (RMSD) of DA and ISMN soil moisture stations and the corresponding histograms when using DA or OL, respectively (OL: open-loop run, DA: assimilation run). (b) Difference in RMSD in % when computing DA versus ISMN instead of OL versus ISMN, and the corresponding histogram of the differences. Blue color on the left indicates that the RMSD is reduced through data assimilation.

impact of data assimilation was found. A more detailed analysis could assess each individual depth measured at the in-situ stations instead of the vertical average.

All in all, the ambiguous picture of in-situ soil moisture validation described in literature is confirmed. Data assimilation only impacts a minority of the in-situ stations. From the above results and literature one question arises: Do in in-situ soil moisture stations provide valuable information for validating GRACE data assimilation experiments? Residual correlation and RMSD of close-by stations can lead to very different validation results, as in-situ measurements depend much on local conditions and soil properties. Nevertheless, when assessing differences in the performance of OL and DA, spatial patterns arise with similar results for close-by stations. This gives confidence that in-situ soil moisture stations can provide information about the performance of data assimilation that should not be neglected. This also suggests that GRACE data assimilation is indeed able to improve the representation of soil moisture at least for some regions.

8.3.4.2 Remote sensing observations from ESA-CCI

The European Space Agency (ESA) Climate Change Initiative (CCI) soil moisture product from remote sensing observations (Section 4.3.1.2) provides information about the first few centimeters of soil. This product is compared here to soil moisture averaged over the first two soil layers of CLM, which represent ca. the upper 5 cm of the soil. In Northern latitudes and over the Alps, the availability of observations is restricted to less than 30 % due to snow coverage (Figure 4.6).

Figure 8.35 a shows the residual correlation between the CCI product and CLM-DA. High correlations of about 0.5 are obtained in the very West of Europe and for parts of Eastern Europe, while no correlation exist over Central Europe. Assimilating GRACE observations has no significant positive impact on the performance of surface soil moisture (Figure 8.35 b). In fact, 55 % of the grid cells show a slight decrease in residual correlation which is in line with the findings of Tian et al. (2017). The joint assimilation of GRACE-derived TWS observations and remotely sensed soil moisture as performed by Tian et al. (2017) might lead to a better agreement with in-situ soil moisture measurements.

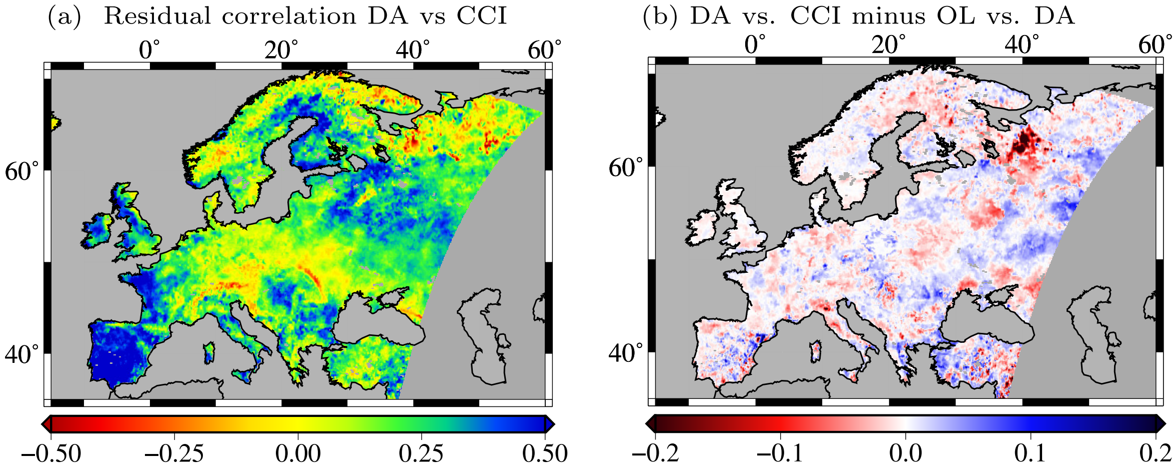


Figure 8.35: (a) Residual correlation between surface soil moisture simulated by the assimilated model (CLM-DA) and the European Space Agency (ESA) Climate Change Initiative (CCI) soil moisture product. (b) Changes in correlation when using DA instead of open-loop model (OL). Blue indicates that residual correlation improves due to data assimilation.

8.3.5 Validation against Discharge Gauges

Although fluxes are not directly updated in the data assimilation framework, they change as a result from updating the state vector. Direct validation of runoff simulations from CLM against discharge measurements from gauging stations is not possible due to the limited temporal availability of such measurements (Section 4.3.2). Here, runoff from CLM was validated against simulated discharge from the rainfall-runoff model GR2M-snow, which was used to extend the observation-based data set from GRDC to cover the study period (Section 4.3.2). Total runoff was obtained from CLM by aggregating subsurface runoff, surface runoff and runoff from lakes, glaciers, and wetlands for all grid cells of a catchment. Before validating the performance of data assimilation against discharge from GR2M-snow, the skill of GR2M-snow must be assessed (Section 8.3.5.1).

Previous GRACE assimilating studies showed a mixed picture when validating against discharge, but generally a larger impact and better improvement was obtained for discharge than for soil moisture. For the Mississippi catchment, Zaitchik et al. (2008) found marginally increased correlations (by 0.01 to 0.02) at 8 out of 10 stations and significant improvements at a 10 % level for 2 stations. Forman et al. (2012) identified an increased bias for the assimilated model and degradation in terms of correlation for 2 out of 6 stations. In Kumar et al. (2016) the impact of data assimilation on discharge was characterized as rather small. Tangdamrongsub et al. (2017) obtained an increase in the correlation coefficients by 0.02 and a reduction of RMSD by about 5%.

8.3.5.1 Quality of Discharge Time Series from GR2M-snow

Discharge time series of GR2M-snow were validated against GRDC data for the 10 years preceding the calibration time span (Figure 4.7). Validation metrics presented in Table 8.12 attest a general good performance to GR2M-snow. The bias of mean discharge is smaller than 2 mm/month except for Po (4.0 mm/month) and Rhone (2.5 mm/month). Nash-Sutcliffe (NS) coefficients reach values between 0.6 and 0.9 for Western and Central European basins. Only for few Eastern European catchments (Narva, Dnepr, Southern Bug and Don) NS coefficients are smaller than 0.3. For de-seasoned and de-trended time series NS coefficients range between 0.2 and 0.6 for most rivers with largest values again for Western and Central Europe. Negative NS coefficients of de-seasoned and de-trended time series (Danube, Don, Neva, Po) indicate changes in the short-term behavior of discharge for the validation period in comparison to the calibration period. Nevertheless, correlations are between 0.7 and 0.8 for de-trended and de-seasoned time series and between 0.8 and 0.9 for the original time series, except for Don and Neva. All in all, GR2M-snow is clearly superior in simulating discharge compared to CLM, as the rainfall-runoff model is specifically designed for this purpose. Therefore, simulations from GR2M-snow are well suited for validating discharge simulated by CLM.

8.3.5.2 Validation of CLM against GR2M

According to GRDC and GR2M-snow, mean monthly discharge over Europe is largest over Central Europe with discharge rates of about 40 mm/month. In contrast, discharge rates for Eastern Europe, for France, and for the Iberian peninsula amount to about 10 mm/month to 20 mm/month only. CLM-OL was found to overestimate discharge, which results into a positive bias with respect to GR2M (Figure 8.36 a). The bias is smallest for the Iberian peninsula with values between 0 – 10 mm/month and reaches values of up to 40 mm/month for Central Europe. Data assimilation reduces the bias by a few millimeters over Eastern Europe (Figure

Table 8.12: Evaluation of simulated discharge from GR2M-snow for the validation period of each river basin: mean values, root mean squared deviation (RMSD), bias of the mean, and Nash-Sutcliffe (NS) coefficients for time series with seasonal cycle and de-seasoned and de-trended (des., det.) time series are computed according to Section 8.1. Due to missing observations no validation could be performed for Seine and Tagus. (*Neman and Vistula are calibrated and validated using E-RUN because of erroneous GRDC data.). The results are taken from Springer et al. (2017).

Catchment	**Mean** $\left[\frac{\text{mm}}{\text{month}}\right]$	**RMSD** $\left[\frac{\text{mm}}{\text{month}}\right]$	**Bias** $\left[\frac{\text{mm}}{\text{month}}\right]$	**NS**	**NS** des., det.
Danube	19.1	4.2	0.4	0.67	-0.34
Daugava	18.0	9.1	0.9	0.60	0.28
Dniester	13.8	5.9	-0.9	0.65	0.38
Dnepr	8.2	4.1	-0.5	0.42	0.22
Don	4.8	2.1	0.2	-0.12	-0.22
Douro	11.0	6.2	-1.0	0.61	0.45
Ebro	14.9	6.2	0.23	0.73	0.25
Elbe	11.9	4.2	-0.1	0.69	0.17
Ems	28.3	9.8	1.6	0.84	0.49
Garonne	33.7	11.2	2.5	0.81	0.38
Narva	16.2	4.0	1.2	0.62	0.42
Guadalquivir	3.4	4.6	-1.2	0.63	0.53
Guadiana	4.0	6.1	-1.8	0.66	0.65
Loire	23.8	7.4	1.8	0.84	0.35
Meuse	35.2	11.0	1.7	0.84	0.50
Neman*	14.6	5.8	-3.2	0.69	0.58
Neva	22.3	3.5	-0.5	0.66	-0.73
Oder	15.3	4.4	0.5	0.59	0.27
Po	61.4	20.4	4.0	0.51	-0.11
Rhine	38.2	7.6	1.1	0.77	0.36
Rhone	51.1	10.6	1.7	0.76	0.45
Southern Bug	5.7	4.5	-0.3	0.21	0.07
Vistula*	15.3	4.8	-1.0	0.54	0.29
Weser	22.5	7.3	0.1	0.83	0.51

8.36 b). In contrast, over Central Europe the bias does not change significantly. On the Iberian peninsula data assimilation leads to an increased bias.

In fact, data assimilation mainly affects one component of the total runoff, namely subsurface drainage. Subsurface drainage is driven by the liquid water content of the lower soil layers, which obtain the largest assimilation increments. In North-Eastern catchments, surface discharge also changes significantly due to changes of the snow cover during winter and spring.

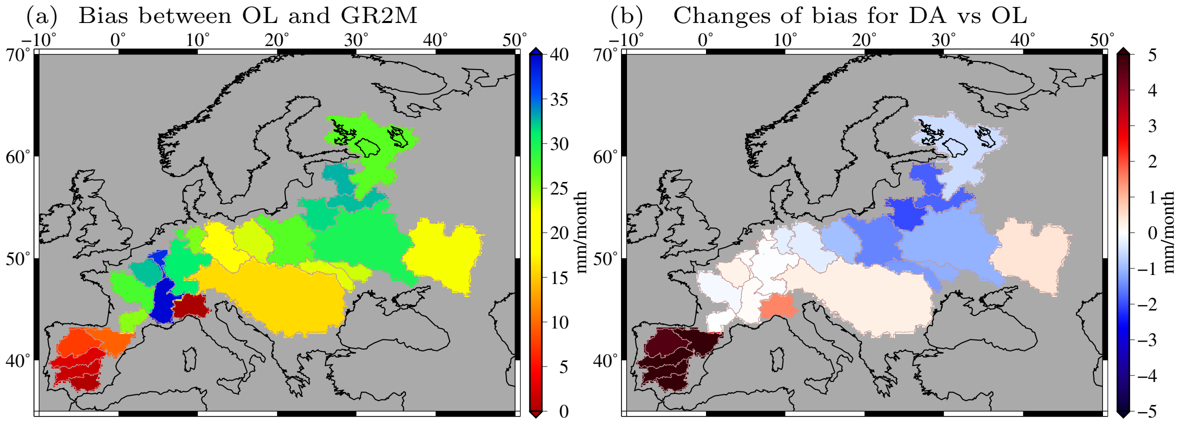


Figure 8.36: (a) Bias between mean runoff from GR2M and the open-loop model (OL) and (b) difference of the biases computed for OL and for the assimilated model (DA).

Runoff from CLM-OL and from GR2M-snow are correlated with correlation coefficients larger than 0.5 for most river basins (Figure 8.37 a). For Oder and Southern Bug, correlation coefficients are smaller than 0.3. In the case of Southern Bug this could be due to the poor performance of GR2M-snow (Table 8.12). Largest correlation coefficients are obtained for the Iberian pensinsula and France. Data assimilation increases correlation coefficients by up to 0.2 except for Northern catchments of the Iberian pensinsula (Figure 8.37 b, c). Particular large improvements are found for Don and Southern Bug.

Residual correlation also shows extremely good performance of runoff from OL on the Iberian pensinsula with correlation coefficients larger than 0.7. In contrast, residual correlation is lower than 0.3 for several catchments over Central Europe, indicating clear deficiencies of CLM in simulating sub-annual variability of runoff. Here, data assimilation improves residual correlation by 0.05 to 0.2. For France, Danube, and Eastern European catchments, residual correlation amounts to 0.3 to 0.5 for CLM-OL and reaches values of 0.5 to 0.7 after data assimilation. In terms of residual correlation, an observation grid space of 2° leads to the best results, with small differences ($< 5\,\%$ in term of the validation measures) to data assimilation results using other observation grids.

In future data assimilation experiments, the calibration of hydraulic parameters may help to improve runoff estimates.

8.3.6 Validation against Evapotranspiration Observations

Total evapotranspiration simulated by CLM was obtained by summing up evaporation from bare soil, and evaporation and transpiration from vegetated areas. Ground evaporation is impacted by the water content of the surface soil layer, which is directly impacted by data assimilation. However, assimilation increments are small for the upper soil layers (Section 8.2.2.4).

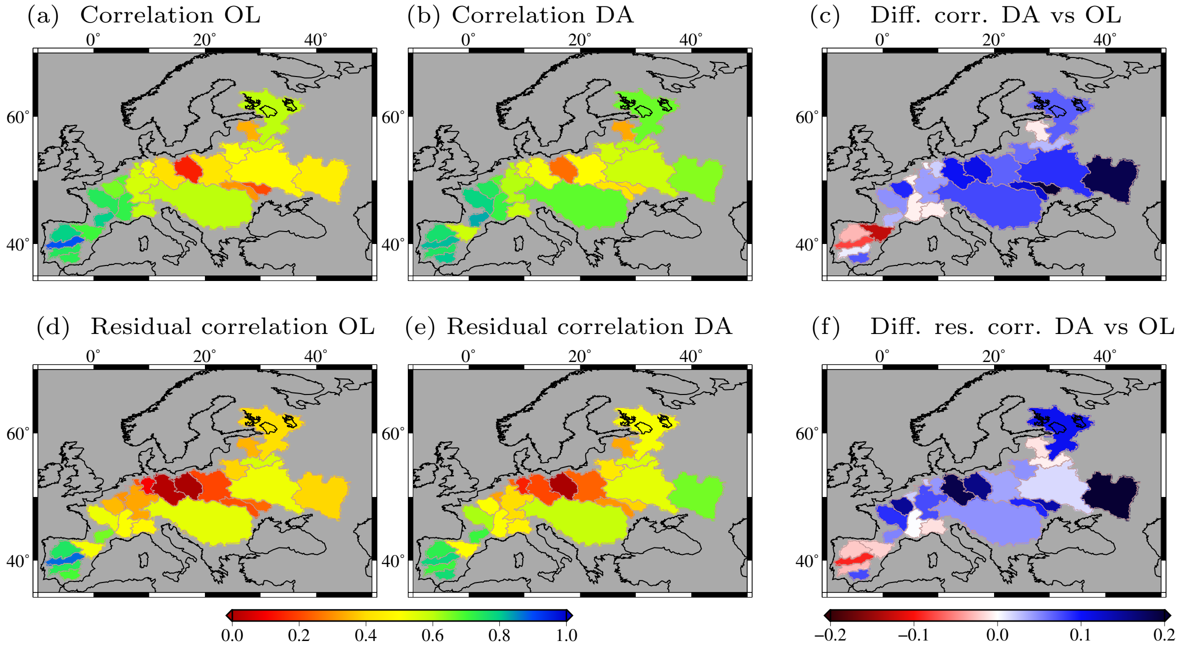


Figure 8.37: (a, b) Correlation and (d, e) residual correlation for runoff from CLM3.5 simulation versus GR2M and (c, f) changes of the correlation coefficients through data assimilation.

Thus, the effect of data assimilation on evapotranspiration is expected to be rather small, especially compared to the overall amount of evapotranspiration. Previous studies showed only small effects and no positive impact of GRACE assimilation on modeled evapotranspiration for the Mississippi basin (Zaitchik et al., 2008), USA (Kumar et al., 2016), and Australia (Tian et al., 2017).

Evapotranspiration has a strong annual cycle with an annual amplitude between 20 mm/month for Southern Europe and up to 50 mm/month for Middle and Eastern Europe. Modeled evapotranspiration from CLM was compared to observation-based data from the Max Planck Institute (MPI, Section 4.3.3) in Jena (Figure 8.38). It should be kept in mind that observation-based data sets of evapotranspiration are usually subject to large uncertainties. However, according to Springer et al. (2014) MPI closes the water balance well over Europe. CLM-OL and MPI have large differences in terms of RMSD especially for Southern Europe with values of up to 20 mm/month (Figure 8.38 a). These are regions were CLM overestimates the annual amplitude of evapotranspiration (Figure 8.38 c). In contrast, in Northern Europe CLM underestimates the annual amplitude in comparison to MPI. Data assimilation leads to a slight increase of the annual amplitude especially in Southern and Eastern Europe (Figure 8.38 d). Here, as a consequence, the RMSD degrades by up to 5 %.

Data assimilation mainly impacts evapotranspiration during summer, when the soil is usually not saturated and potential evaporation is large. Abstraction or injection of water then directly implies less or more water being available for evapotranspiration. Furthermore, evapotranspiration has a pronounced daily cycle depending mainly on solar radiation. Figure 8.39 shows the daily cycle of evapotranspiration averaged over all days of August 2005 for three river basins. Maximum evapotranspiration is reached at about noon and differs significantly for the different regions due to different climate regimes. Largest impact from data assimilation is obtained for IP, where the maximum value of evapotranspiration raises by about 10 %. For FI and BS1, maximum evapotranspiration increases by 1.5 %. This indicates that in August 2005 data assimilation increases the amount of water available for evapotranspiration. For winter months, evapotranspiration does not change due to data assimilation.

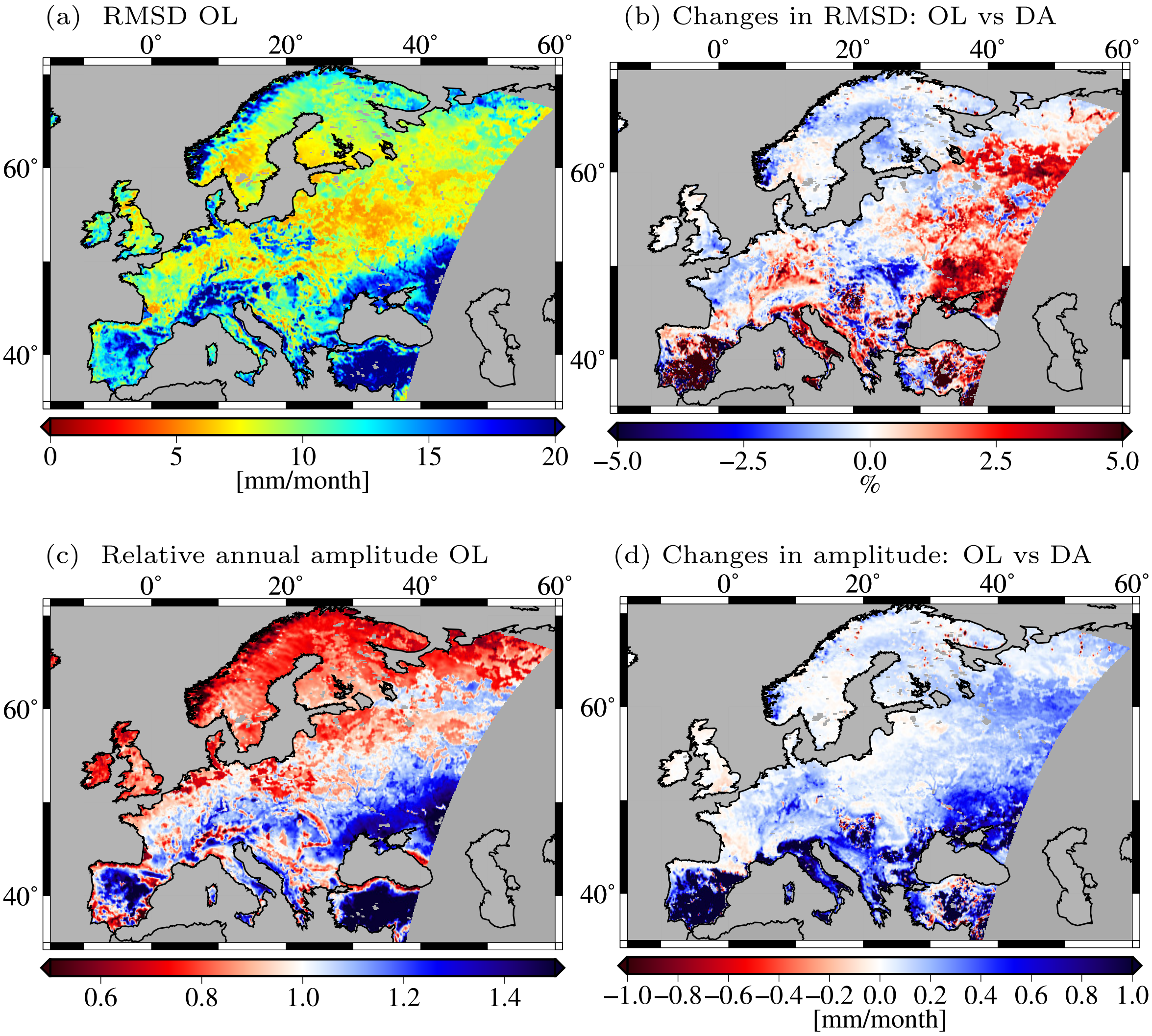


Figure 8.38: (a) Root mean square deviation (RMSD) of evapotranspiration from CLM open-loop (OL) and observation-based data from MPI. (b) Changes in % when computing the RMSD using the assimilated model (DA) instead of OL, with red indicating an increase in RMSD through data assimilation. (c) Relative annual amplitudes of evapotranspiration from OL and MPI with blue color indicating overestimation of the annual amplitude and red color indicating underestimation. (d) Changes in the absolute amplitude for DA versus OL, with blue indicating an increase of the annual amplitude through data assimilation.

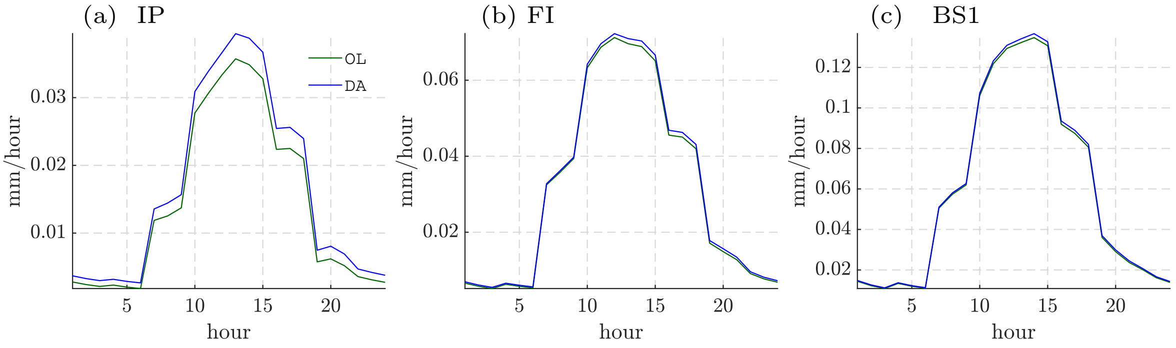


Figure 8.39: Daily cycle of evapotranspiration simulated by the open-loop run (OL, green) and by the assimilation run (DA, blue) in three different catchments averaged over August 2005.

8.3.7 Examples for Spatial and Temporal Downscaling of GRACE Observations

Assimilating GRACE data into CLM over Europe provides a detailed picture of TWS evolution. In particular extreme events are (i) better represented compared to the open-loop run, and (ii) are mapped with more details regarding spatial and temporal resolution compared to GRACE. Previous studies showed that assimilation of GRACE data into hydrological models improves the characterization of droughts (Li et al., 2012; Houborg et al., 2012; Kumar et al., 2016; Schumacher et al., 2018) and floods (Reager et al., 2015).

8.3.7.1 European Heat Wave 2003

The European heat wave in summer 2003 lasted from May to September and hit mainly Western and Central Europe with temperature records and lack of precipitation (Rebetez et al., 2006). Figure 8.40 shows the drop in TWS within 6 months from 15 February to 15 August 2003 for GRACE, OL, and DA. GRACE shows a drop larger than 200 mm in Northern Spain, Western France, Germany, Italy, and the Balkan states. CLM-OL simulates a moderate drop of TWS especially over Central Europe (less than 120 mm). Data assimilation increases the drop in TWS over France, Germany and the Balkan states. Nevertheless, compared to GRACE, CLM-DA still underestimates the effect of the heat wave on TWS. Most likely, the representation of the heat wave could be improved when starting the assimilation run already in the middle of 2002 and/or when adding simultaneous parameter calibration to the assimilation framework. Yet, CLM-DA shows interesting smaller scale structures, e.g. over the Alps, which cannot be represented by GRACE.

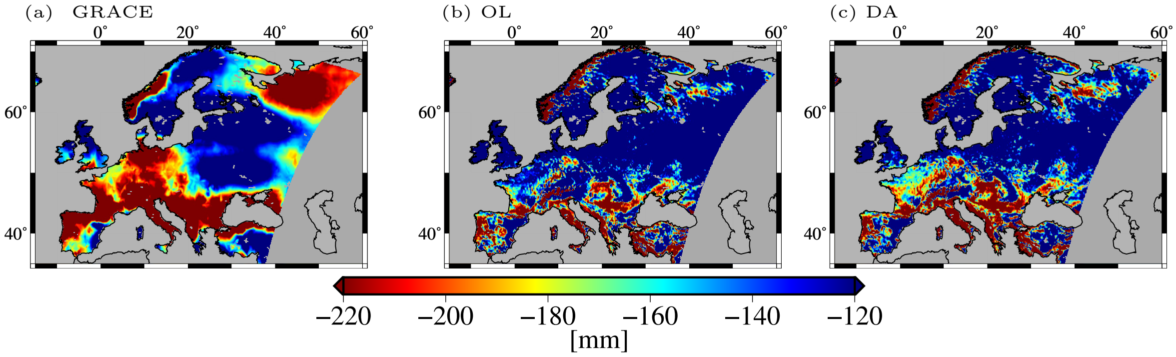


Figure 8.40: European heat wave 2003: Changes in total water storage (TWS) from 15 February to 15 August 2003, for GRACE (only monthly averages available), CLM open-loop, and GRACE assimilating CLM. Red color indicates a stronger drop of TWS than blue color.

8.3.7.2 European Flood 2006

An outstanding spring flood with a 100-year return period occurred in 2006, when the water level of many rivers over Central Europe rose due to heavy precipitation and snow melt (Wachter, 2007). Here, the flood is visualized exemplarily via five snapshots from the beginning of March til the end of April (Figure 8.41). The first column shows TWSA from GRACE for March and April 2006. In March, huge water masses are concentrated over the Balkans and Italy. In April water masses reach the mouth of Danube. OL shows a positive anomaly in TWS along the Danube with small temporal changes and a small spatial extension. Overall, no major flooding is represented by OL. In contrast, DA shows clear positive TWSA along the Western coast of the Balkan states at beginning of March, and decreasing TWS in April. From the middle of March to the beginning of April, TWS increases along the Danube and for Ukraine, which is in line with the GRACE observations. Indeed, GRACE has a relatively large contribution to DA in spring 2006, as indicated by the assimilation increments in Figure 8.27. According to DA, the peak of TWSA is passed by middle of April.

All in all, DA provides a good overview on the evolution of TWS during the flood. For further investigations it would be interesting to include river routing in order to assess simulated discharge together with TWSA.

8.3.8 Key Messages of GRACE Assimilation Experiments

GRACE data was assimilated into CLM over a time span of 8 years (2003 to 2010), creating a high-resolution reanalysis of the evolution of TWS over Europe. In other words, GRACE data was downscaled temporally and spatially with vertical stratification allowing for detailed analysis of the hydrological conditions in different compartments. To my knowledge, no data set with comparable resolution exists over Europe.

Assimilating GRACE observations improved the annual amplitudes of modeled TWS especially over Eastern Europe. Moreover, the assimilation of GRACE data made a significant contribution to interannual variability of TWS and improved residual correlation by up to 50 % for the individual catchments. These results indicate that the assimilation of GRACE data has a huge potential to lead to a better representation of extreme events in CLM3.5. As expected, data assimilation drew modeled trends in TWS towards GRACE, which means that the assimilated model could also be helpful for climate projections at more regional scales than GRACE is able to resolve.

On the long-term, the water balance of the model was only slightly perturbed by GRACE assimilation for most regions. Typical values were about $1 - 6\,\mathrm{cm}$ of water introduced or abstracted from the model over the whole time period. Yet, data assimilation introduced some problems for grid cells, where the water table was situated below the soil column. Starting from a certain point in time, these grid cells and neighboring grid cells started to run empty during one month and to fill up later on again. Applying further restrictions to the assimilation increments, using a more recent CLM version, or replacing the lower soil layers by a more sophisticated groundwater model could help to solve this problem.

Largest assimilation increments were obtained for soil moisture. Yet, validation with remotely sensed soil moisture observations showed no improvement from data assimilation, as the impact of GRACE data assimilation on surface soil moisture was rather small. In-situ soil moisture measurements cover deeper soil layers, but they are difficult to compare to models with grid cells of several kilometers. Nevertheless, data assimilation improved correlation and RMSD at some stations significantly, but also led to some degradation. However, at most soil moisture stations the assimilation of GRACE data had only a small impact. Therefore, the assimilation of soil moisture measurements complementary to GRACE-derived TWSA is a promising option.

The assimilation of GRACE-derived TWSA indirectly also impacted water fluxes. Modeled runoff improved in most catchments except for the Iberian peninsula in terms of correlation and residual correlation through data assimilation. Assimilating GRACE-derived TWSA had a particularly large impact on subsurface drainage, which might indicate deficiencies in the model structure in representing processes of the deeper soil. However, this is no surprise as CLM3.5 targets at representing near-surface processes. A better representation of deeper soil processes could be achieved by adding a dedicated groundwater model. The impact of data

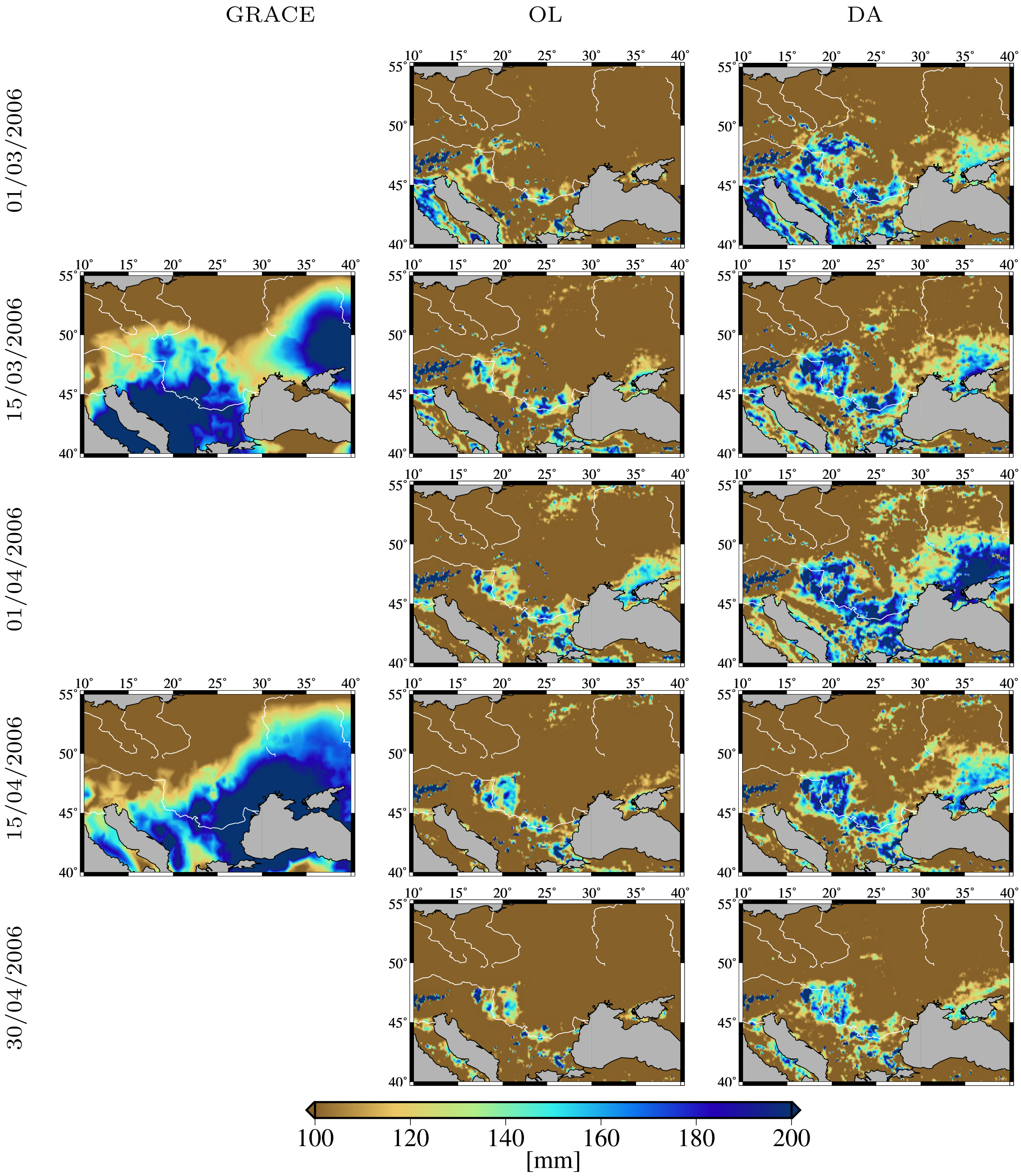


Figure 8.41: Total water storage anomalies (TWSA) showing the situation during the European flood in 2006 for the Danube and the rivers of Ukraine. Five snapshots from beginning of March til end of April are shown for GRACE (only monthly averages available), CLM open-loop, and GRACE assimilating CLM. White lines indicate major rivers.

assimilation on evapotranspiration was smaller than the impact on runoff, and no improvement was obtained with respect to an observation-based data set from upscaled FLUXNET data. Interestingly, a small impact of data assimilation on the daily cycle of evapotranspiration was detected during summer. This indicates that the assimilation of GRACE data can also feedback on the atmosphere.

As exemplarily shown for the European heat wave 2003 and the European flood 2006, GRACE data assimilation proved to have a huge potential for assessing extreme events. While models are often not able to represent such situations realistically, the resolution of GRACE data is often too coarse. In combination, an interesting picture of the spatial and temporal evolution of TWSA in the presence of extreme events was achieved.

The data assimilation framework was optimized using the findings of Section 8.3. However, compared to other studies GRACE errors derived by error propagation starting from the full error covariance matrices are relatively large. It might be interesting to introduce a weighting factor that puts more weight on the GRACE observations and to calibrate the factor against the independent validation data sets.

Chapter 9

Validation of Offline and Assimilated Hydrological Models Using GRACE In-Orbit Residuals

Generally, monthly gridded and filtered GRACE solutions of total water storage anomalies (TWSA) are used for validating hydrological models and for assimilating TWSA into hydrological models. However, gridded maps of TWSA are attenuated due to filtering. Catchment averaged time series of TWSA contain uncertainties caused by leakage effects and are, thus, not directly comparable to hydrological model output.

In this thesis, an approach (named here the residuals approach) to analyze hydrological mass variations directly in-orbit using GRACE level 1B observations was realized (Section 5.2). Therefore, TWSA from hydrological models were converted to simulated K-band observations and reduced from the original measurements, yielding K-band range-rate (KBRR) or K-band range-acceleration (KBRA) residuals.

KBRR and KBRA residuals were computed for different global and local hydrological models according to Section 5.2 and according to Eicker and Springer (2016). K-band residuals allow for i) validating modeled hydrological mass changes using GRACE observations without the downward continuation and filtering steps generally required for validation with gridded gravity field maps, and (ii) assessing the content of high-frequent (daily) signals in total water storage (TWS) variability simulated by hydrological models. Furthermore, the impact of the assimilation of GRACE-derived gridded TWSA into CLM3.5 was analyzed on the level of K-band observations. These investigations represent one step towards the direct assimilation of GRACE Level 1B data, instead of gridded TWSA, into hydrological models.

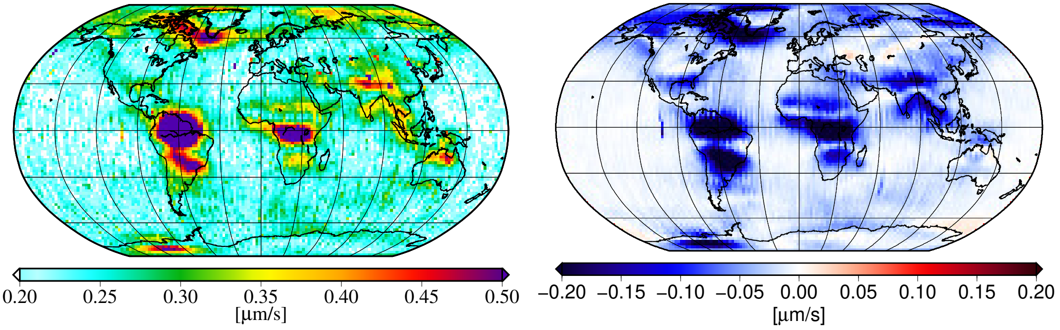
In this thesis, the study of Eicker and Springer (2016) was extended to longer time spans, to local (in addition to global) hydrological models, and to the analysis of KBRA (in addition to KBRR) residuals. Eicker and Springer (2016) evaluated KBRR residuals from three global hydrological models (WGHM, LSDM, GLDAS-NOAH, see Section 4.2) for the years 2006 and 2007 and removed the temporal mean of TWS computed for the time period 2003 to 2008 before computing the residuals. Here, the same global models were investigated for the time span 2003 to 2013, and mean TWS of the period 2003 to 2013 was removed before computing the residuals. For the local model CLM3.5, the time period of the assimilation experiments (2003 to 2010) was considered (Section 8.3).

In line with Eicker and Springer (2016), daily GRACE Kalman solutions from ITSG-Grace2014 (Mayer-Gürr et al., 2014) were considered as additional model for TWSA. The working hypothesis is that GRACE-derived TWSA better explain K-band residuals than TWSA from hydrological models. Thus, GRACE-derived TWSA solutions can serve as a reference, which indicates what kind of reduction of the residuals can currently be achieved. However, it should be kept in mind that GRACE-derived TWSA solutions also contain artifacts beyond hydrological mass changes. TWSA solutions from ITSG-Grace2016 (Section 2.3) lead to similar results, but they are not considered further in this section as the focus is on different hydrological models and not on different GRACE solutions. Nevertheless, for future studies the GRACE-derived TWSA solutions applied as a reference may be replaced by more recent GRACE solutions and the de-aliasing product AOD1B-RL05 by RL06.

Figure 9.1 (a) and (c) show KBRR and KBRA residuals computed from the background models introduced in Section 5.2.2 including the atmosphere and ocean de-aliasing product (AOD1B), but without considering hydrological mass variability. KBRR residuals of AOD1B are large in regions with distinct hydrological signals (e.g., close to the equator). Obviously, when using only the AOD1B product, the residuals contain temporal variations of the gravity field caused by TWS variability in the major river basins. In contrast, KBRA (the derivative of KBRR) residuals of AOD1B are dominated by noise.

Figure 9.1 (b) and (d) show the difference between residuals from AOD1B and residuals when additionally considering TWSA from daily GRACE solutions. Additionally considering TWSA reduces the residuals in particular in regions with large hydrological mass variability (blue color). KBRA residuals resulting from a specific mass anomaly are directly located above the mass anomaly, while KBRR residuals show patterns north and south of the mass anomaly.

(a) KBRR residuals AOD1B (b) KBRR residuals AOD1B vs. GRACE-daily



(c) KBRA residuals AOD1B (d) KBRA residuals AOD1B vs. GRACE-daily

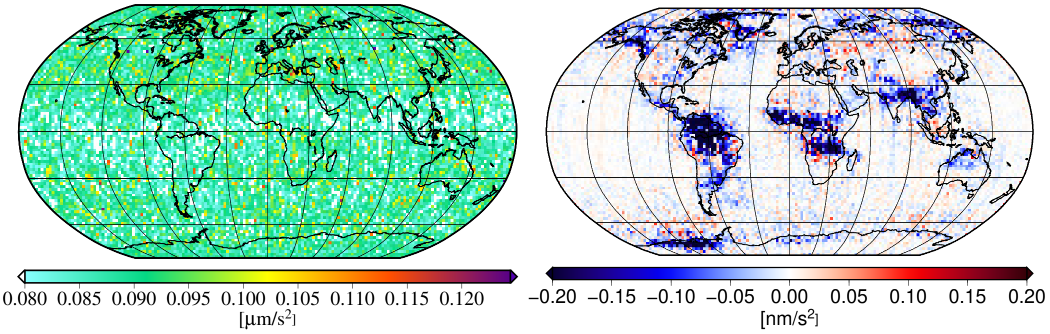


Figure 9.1: Gridded RMS values of the year 2007 on a 2°×2° grid of (a) K-band range-rate (KBRR) residuals of AOD1B-RL05 and (b) differences of KBRR of AOD1B-RL05 and KBRR residuals when additionally considering TWSA from daily GRACE solutions. (c) and (d) show the corresponding results for K-band range-acceleration (KBRA) residuals.

In the following, the residuals approach for validating hydrological mass variability is evaluated first for global hydrological models (WGHM, GLDAS, LSDM). The performance of the individual models in reducing the K-band residuals is assessed globally and for selected regions

(Section 9.1). Then, the short-term (daily) signal content of the global models is analyzed (Section 9.2). Another experiment shows the influence of locally isolated but strong water storage changes on the K-band residuals through the example of Lake Volta (Section 9.3). Finally, CLM-based TWS estimates from the open-loop run, from the GRACE-assimilating run, and the corresponding ensemble spread are investigated on the level of the K-band residuals (Section 9.4).

9.1 In-Orbit Validation of Global Hydrological Models

KBRR and KBRA residuals were computed by removing monthly TWS output from different global hydrological models in addition to the standard background models (tides, etc., as described in Section 5.2) and in addition to AOD1B. Modeled TWSA were developed into spherical harmonics up to degree 180 (WGHM, LSDM, NOAH, see Section 4.2). First, time series of daily and monthly RMS values averaged from the 5-secondly sampling KBRR residuals are shown for the continents. Then, the spatial distribution of the gridded RMS values of the residuals is compared for the three global models. Finally, a more detailed analysis of the residuals over Europe and Westafrica is performed. KBRA residuals are dominated by noise, which is removed to some extent when computing the differences between residuals from different models. For this reason, results of KBRA residuals are shown for selected experiments only.

9.1.1 Time Series of Residuals

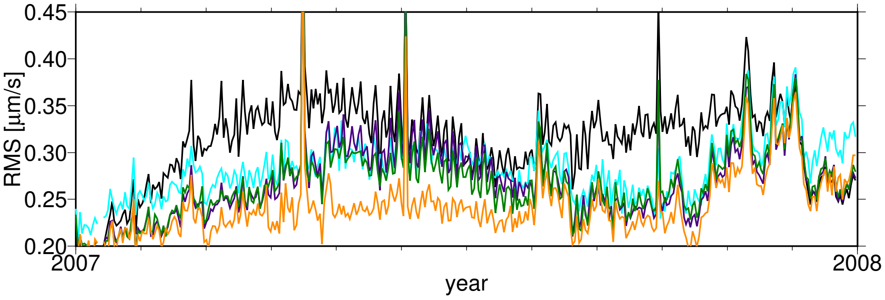
Time series of daily (Figure 9.2 a, 2007) and monthly (Figure 9.2 b, 2003 to 2013) RMS values of KBRR residuals over the continents were computed considering (i) only the AOD1B product in addition to standard background models, (ii) monthly TWS variability of different global hydrological models and AOD1B, and (iii) monthly averages of daily GRACE-derived TWSA solutions of ITSG-Grace2014 and AOD1B (Section 2.3). Gaps in the time series are due to missing GRACE data or arcs corrupted by outliers and vary with time (Section 5.2). The number of valid K-band residuals also varies for each day. Peaks in the time series of daily values could be related to the quality of the AOD1B product or to outliers that were not detected. Interestingly, such peaks in the daily time series do not arise for each year.

KBRR residuals from AOD1B are largest as all signals from hydrological-induced mass variations are included (Figure 9.2). As expected, the residuals are generally reduced relative to AOD1B when additionally considering one of the global hydrological models displayed in cyan (LSDM), purple (NOAH), and green (WGHM). Relative to AOD1B, LSDM leads to a RMS reduction of the residuals of 7.5 % over the continents (Table 9.1). WGHM (11.8 % RMS reduction) and NOAH (10.0 % RMS reduction) perform better than LSDM in reducing the residuals on the continental average.

To place these RMS reductions into the wider context, note that the RMS reduction of the residuals when considering hydrological mass variations is much larger compared to the RMS reduction obtained when moving from AOD1B-RL04 to AOD1B-RL05 (0.79 % RMS reduction for the time span 2006-2007; Eicker and Springer, 2016). The good performance of WGHM in comparison to the other two global models could be due to the modeling of additional storage compartments (groundwater and surface waters) and due to the modeling of human water abstraction. For further comparison, monthly averages of daily GRACE-derived TWSA solutions (yellow line) reduce the RMS of the residuals by 23.5 % relative to AOD1B over the continents. It is obvious that GRACE-derived TWSA solutions better explain the K-band measurements as they were estimated from them and, thus, this comparison is slightly biased. Here, GRACE solutions serve as reference on which (maximum) reduction of the residuals can be achieved. Numbers given here differ slightly from those published in Eicker and Springer (2016) due to the different study periods, but they deliver the same key messages.

9.1.2 Spatial Residual Analysis

A more detailed picture of the skill of the three global hydrological models in explaining K-band measurements is obtained when mapping the residuals to a global grid. Spatial RMS


(a) Daily KBRR

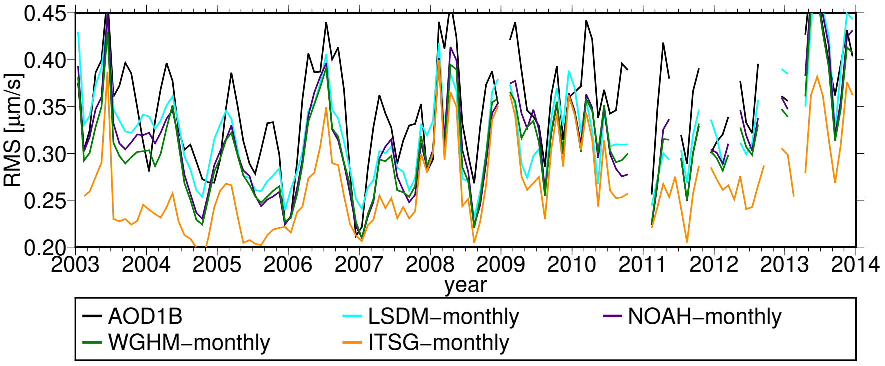

(b) Monthly KBRR

Figure 9.2: Time series of daily (top, year 2007) and monthly (bottom, years 2003 to 2013) RMS values of KBRR residuals over the continents. The black line refers to no hydrological model subtracted in addition to the AOD1B product; cyan/purple/green lines refer to monthly TWSA from different global hydrological models subtracted in addition to AOD1B, and the yellow line refers to the monthly mean of daily GRACE-derived TWSA solutions subtracted in addition to AOD1B.

values of KBRR and KBRA residuals were computed on a 2°×2° grid using the residuals of 2007. Figure 9.3 shows that a significant part of the temporally varying gravity field signals (Figure 9.1 a) is removed from the KBRR residuals in the large river basins. KBRA residuals are not shown here as noise dominates the gridded RMS values.

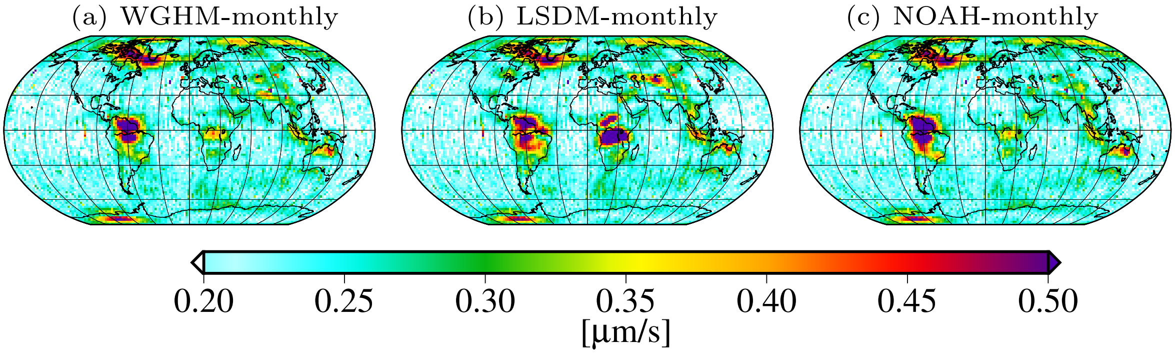

(a) WGHM-monthly (b) LSDM-monthly (c) NOAH-monthly

Figure 9.3: Gridded RMS values of KBRR residuals of the year 2007 on a 2°×2° grid after removing monthly averaged hydrological model output in addition to AOD1B.

As the hydrological models do not perfectly simulate TWS variability, residuals remain large in some regions. The performance of the three global models in explaining K-band observations is compared by computing differences of the gridded RMS values for KBRR residuals and for KBRA residuals (Figure 9.4). Blue color indicates that the first model (i.e., LSDM in Figure 9.4 a, d) performs better in explaining the GRACE Level 1B data and red color indicates that the second model (i.e., WGHM in Figure 9.4 a, d) performs better. Please note the differences in the localization of hydrological signals for KBRR and KBRA residuals. While for KBRA residuals differences between the models are large along the Amazon river, KBRR residuals show largest differences north and south of the river. However, as KBRA residuals are more noisy, KBRR residuals are helpful to investigate also regions with smaller mass variability. Residuals are resolved on a 2°×2° grid, which corresponds to grid cells of about 40 000 km^2. This might allow for validating global hydrological models at smaller spatial scales than it is possible using gridded maps of TWSA. Further analysis is required to define the effective resolution of K-band gridded residuals.

Strong regional differences exist in the skill of the global models to reduce K-band residuals. In particular LSDM shows large differences in the RMS of K-band residuals in comparison to the other two global models. For the year 2007, LSDM shows deficits compared to WGHM and NOAH in equatorial Africa, in North America, and in the Himalayan region. In South America, LSDM performs better along the main course of the Amazon river compared to the other two models. Over Europe and for most parts of Asia, NOAH seems to be superior. Comparing WGHM and NOAH shows better performance of WGHM over Australia and over South America.

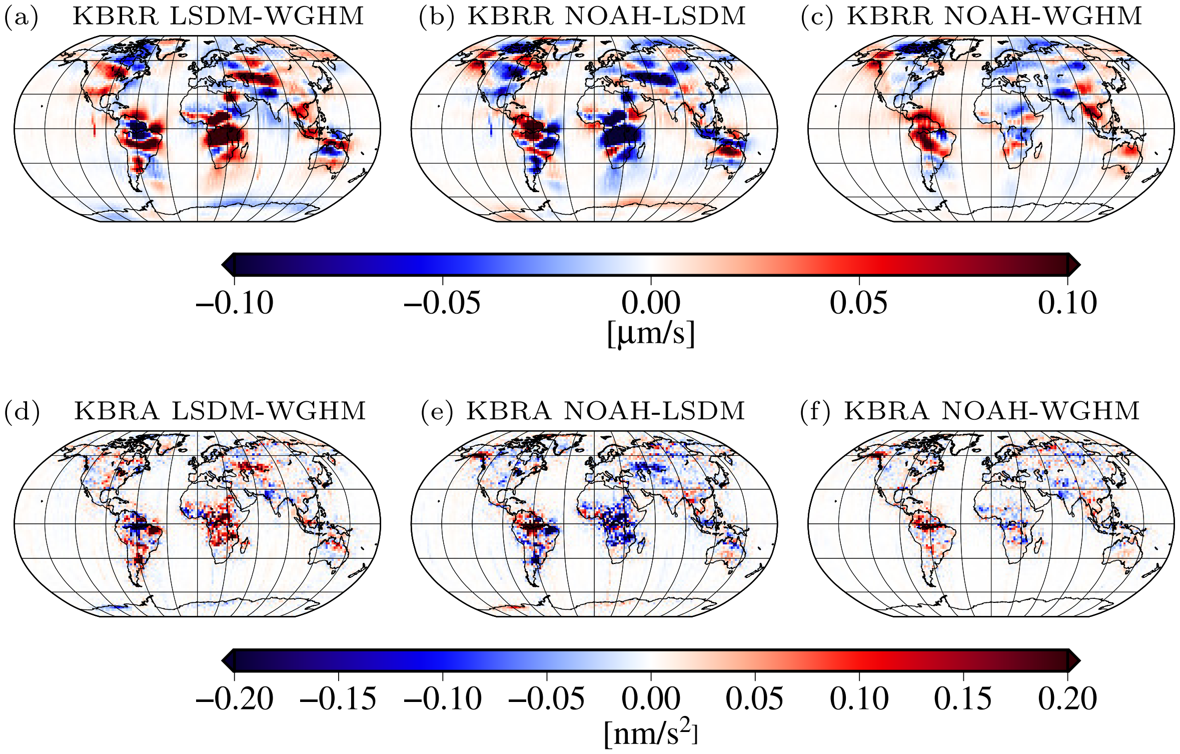


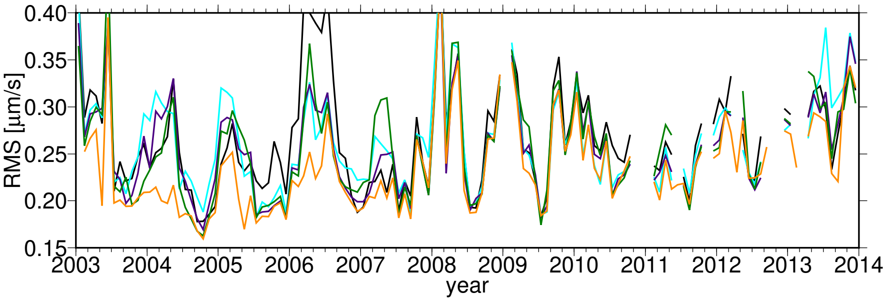
Figure 9.4: Differences of gridded RMS values of KBRR (a – c) and KBRA (d – f) residuals of the year 2007 on a 2°×2° grid between monthly averaged global hydrological model output.

9.1.3 Regional Time Series of Residuals

Regional differences of KBRR residuals between the global hydrological models were investigated in more detail for Central Europe and Westafrica. Time series of monthly RMS values of KBRR residuals (Figure 9.5) and, additionally, time series of monthly TWSA were computed (Figure 9.6). A caveat is that the performance of the models according to KBRR residuals and according to TWS variability cannot be directly compared, as KBRR residuals arise north and south of a mass anomaly.

As expected, RMS reduction of the residuals relative to AOD1B is smaller over Europe than

(a) Europe



(b) Westafrica

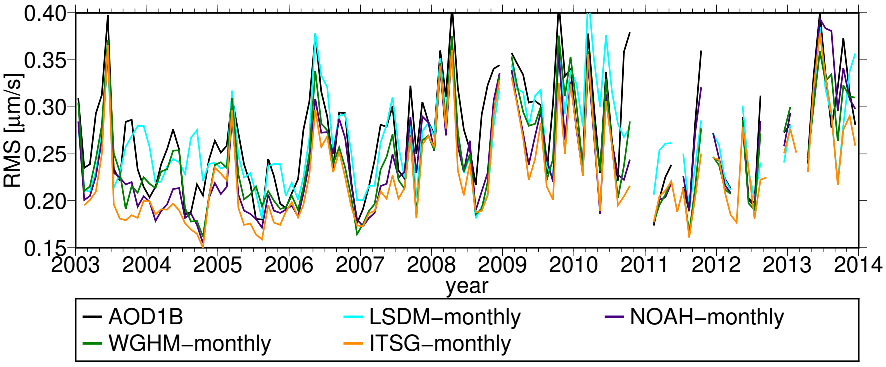
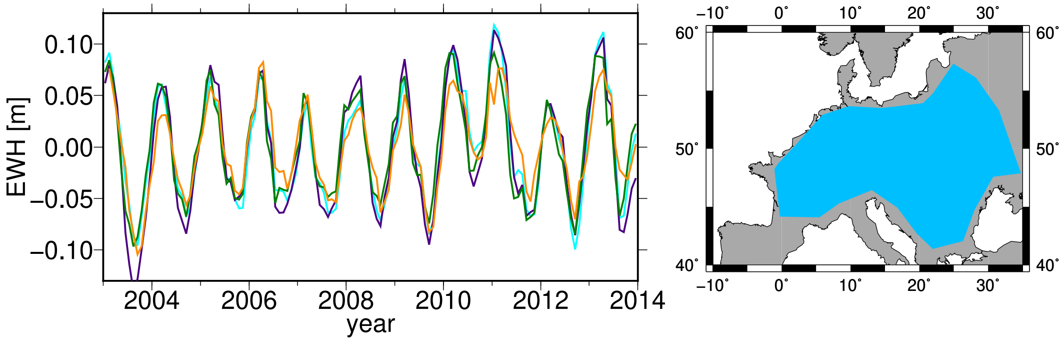


Figure 9.5: Time series of monthly RMS values of KBRR residuals for (a) Central Europe and (b) Westafrica. The outline of the considered regions is shown in Figure 9.5.

(a) Europe



(b) Westafrica

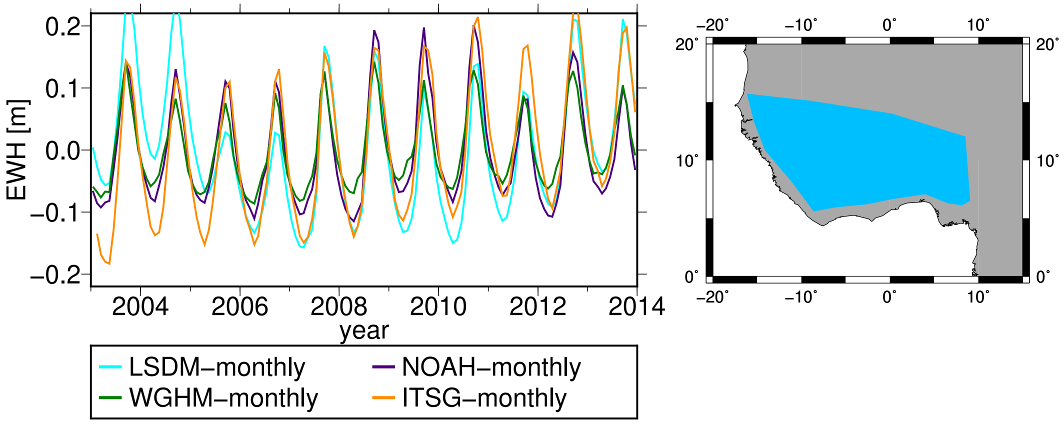


Figure 9.6: Time series of regionally averaged terrestrial water storage changes for the same regions as in Figure 9.5.

the averaged RMS reduction over all continents (Table 9.1). In comparison to regions like e.g. the Amazon, hydrological mass variability over Europe is rather small. NOAH performs best in explaining the KBRR residuals over Europe and reduces the RMS of the residuals by 4.5 % with respect to AOD1B. WGHM and LSDM explain 3.5 % and 1.8% of the KBRR residuals remaining over Europe after considering AOD1B. For comparison, the monthly average of daily GRACE solutions leads to a RMS reduction of 12 % relative to AOD1B. LSDM performs particularly poor in reducing the residuals before the year 2007, and even leads to some increase of the residuals (Figure 9.5).

Surprisingly, TWS estimates from LSDM over Europe are comparable to the other models and after 2007 the RMS of KBRR residuals from LSDM is similar to the other models (Figure 9.6). Further investigations showed that LSDM has some problems in representing trends (e.g., over Westafrica in Figure 9.5, and over the Amazon basin shown in Eicker and Springer, 2016). These problems in representing trends could also effect the RMS of KBRR residuals over Europe.

The year 2006 stands out over Europe. The European flood (see also Section 8.3.7.2) leads to maximum RMS values of KBRR residuals for AOD1B, which are reduced by up to 30 % when additionally considering the global hydrological models (Figure 9.5 a).

TWS time series over Westafrica show larger differences between the individual models than over Europe (Figure 9.6 b), and so do the corresponding RMS values of KBRR residuals (Figure 9.5 b). LSDM shows a strong negative trend in TWS at the beginning of the study period, which is not detected by GRACE or the other models. This trend leads to increased RMS values of KBRR residuals. NOAH and WGHM achieve a RMS reduction of KBRR residuals of 9.8 % and 8.4 % relative to AOD1B, whereas LSDM only reduces the RMS values by 2.7 %.

Since 2008, TWS variability from GRACE shows a positive trend, which is related to water management of the Lake Volta reservoir. This signal is to some extent also represented by WGHM. In Section 9.3, an analysis of the contribution of Lake Volta to the K-band residuals is provided.

Absolute values [μm/s]	Global	Continents	Westafrica	Europe
AOD-RL05	0.3157	0.3548	0.2778	0.2742
LSDM-m.	0.3004	0.3282	0.2704	0.2693
WGHM-m.	0.2981	0.3129	0.2544	0.2645
NOAH-m.	0.3017	0.3195	0.2507	0.2618
ITSG-m.	0.2532	0.2716	0.2346	0.2413

Improvements [%]	Global	Continents	Westafrica	Europe
LSDM-m. vs. AOD-RL05	4.85	7.50	2.66	1.79
WGHM-m. vs. AOD-RL05	5.57	11.81	8.42	3.54
NOAH-m. vs. AOD-RL05	4.43	9.95	9.76	4.52
ITSG-m. vs. AOD-RL05	19.80	23.45	15.55	12.00

Table 9.1: (a) RMS in μm/s of the monthly residuals time series (global, continents-only, Westafrica, Europe) shown in Fig. 9.2 and Fig. 9.5. (b) relative improvements in % when considering total water storage variability from different models in addition to AOD1B.

9.2 In-Orbit Validation of Short-Term Hydrological Variability

In the following, KBRR residuals from daily model output are compared to KBRR residuals from monthly averaged model output. Likewise, KBRR residuals from daily GRACE-derived TWSA are compared to KBRR residuals from daily GRACE-derived TWSA solutions averaged to monthly means. This provides evidence on the short-term information content of modeled TWS variability.

Blue color in Figure 9.7 indicates a smaller RMS of the residuals for daily model output compared to monthly model output. Indeed, for all three global hydrological models a RMS reduction of the residuals (daily versus monthly model output) is achieved with improvements between 0.4-0.7 % over the continents (Table 9.2). This is in the range of the RMS reduction when moving from AOD1B-RL04 to AOD1B-RL05 (0.79 % reduction for the time span 2006-2007; Eicker and Springer, 2016). The daily signal in the models is particularly large over equatorial regions and surprisingly also noticeable over Europe.

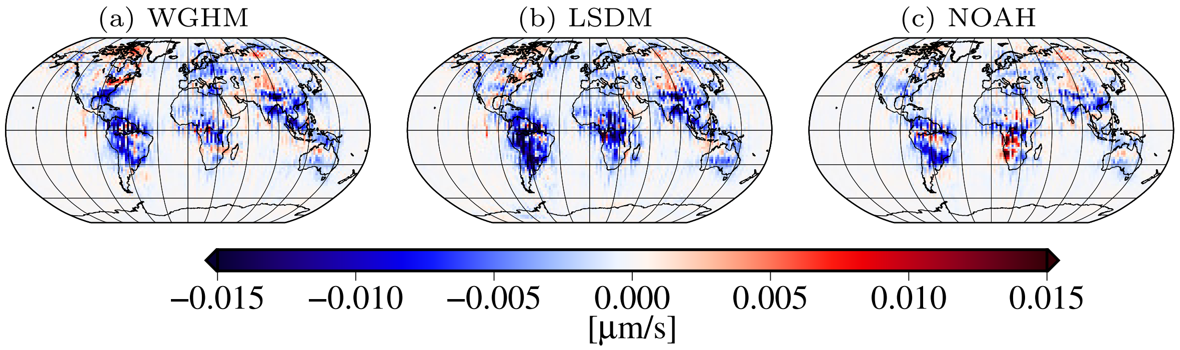

(a) WGHM (b) LSDM (c) NOAH

Figure 9.7: Differences of gridded RMS values of KBRR residuals of the year 2007 on a 2°×2° grid for daily versus monthly model output. Blue color indicates that daily model output further reduces the residuals compared to monthly model output.

Figure 9.8 shows monthly RMS values of KBRR residuals from daily and monthly model output averaged over the continents. Interestingly, large differences between daily and model output is frequently obtained for all three models simultaneously (e.g., January 2004, January 2009, September 2010). This coincidence could be due to specific hydrological conditions that are captured by all three models.

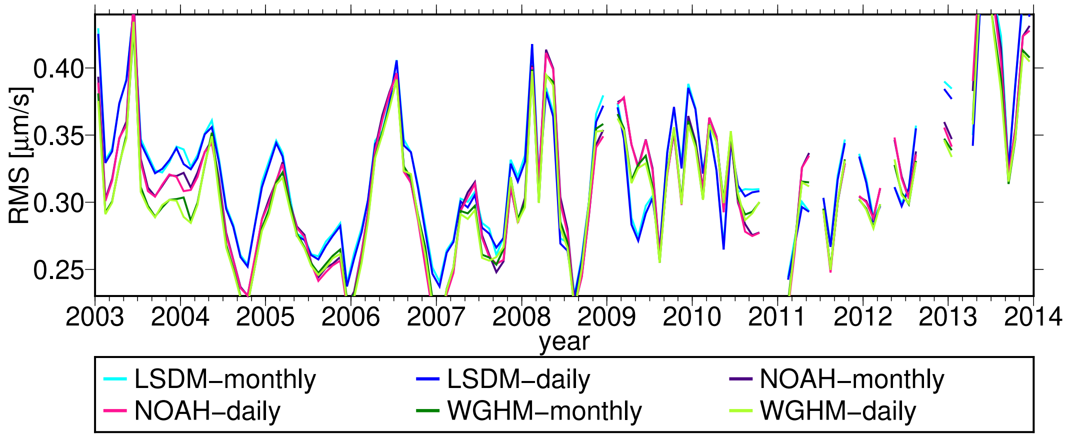


Figure 9.8: Time series of monthly RMS values of continental KBRR residuals for daily and monthly hydrological model output.

All global hydrological models have similar contribution of short-term hydrological signals to KBRR residuals over Europe and reduce the RMS by 0.42-0.45 % for daily versus monthly model output (Table 9.2). Over Europe, the daily GRACE-derived TWSA solutions show an outstanding RMS reduction of 3.5 % with respect to monthly averages of the daily TWSA solutions. Over Westafrica, daily TWS output from NOAH reduces the RMS of KBRR residuals by 0.8 % relative to monthly model output. This is more than the average improvement over the continents and close to the short-term signal content in the GRACE-derived TWSA solutions (reduction of 1.0%). For WGHM and LSDM, RMS reductions of 0.4 % and 0.2 % are achieved over Westafrica when moving from daily to monthly model output.

Improvements [%]	Global	Continents	Westafrica	Europe
LSDM-d. vs. LSDM-m.	0.33	0.73	0.15	0.45
WGHM-d. vs. WGHM-m.	0.20	0.51	0.35	0.45
NOAH-d. vs. NOAH-m.	0.23	0.44	0.80	0.42
ITSG-d. vs. ITSG-m.	1.50	1.95	1.02	3.56

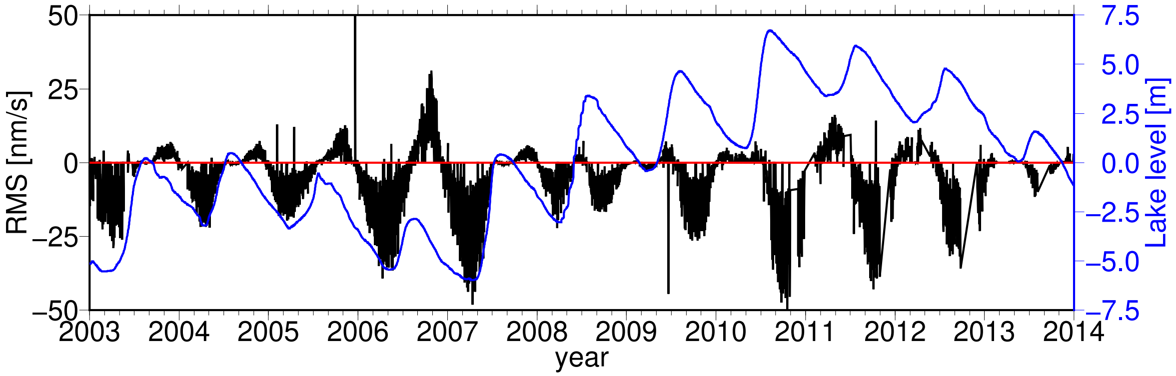
Table 9.2: Relative improvements in % between the RMS of the residuals from daily model versions with respect to the RMS of the residuals from the monthly models (global, continents-only, Westafrica, Europe). The abreviation “d” denotes daily models, “m” stands for the monthly averages of the daily models.

9.3 In-Orbit Validation of Hydrological Signals from Reservoirs

Large effects on K-band residuals can arise from locally isolated but strong waters storage changes. One example is Lake Volta, a managed reservoir with highly variable lake level, which changes by more than 5 m during few months (Figure 9.9). Lake level variations from altimetry (Uebbing et al., 2015) were converted into mass changes assuming a varying lake surface between 4450 km^2 and 9970 km^2 (Tanaka et al., 2002). Mass changes were then represented in spherical harmonic coefficients up to degree and order 360.

Figure 9.9 shows the differences of the RMS of KBRR and KBRA residuals between AOD1B and considering TWS variability of Lake Volta (additionally to AOD1B) together with the lake level. The time series of Lake Volta were computed from all residuals within a surrounding region that accounts for 90 % of the signal induced by Lake Volta. This region extends about 10° north and south of Lake Volta. Positive RMS differences mean that the background atmosphere and ocean modeling better explains the KBRR data than when one adds forward-modeled Lake Volta mass change, and negative numbers suggest that adding the Lake Volta mass correction indeed better explains the data. During the first half of each year, modeling the mass change induced by Lake Volta leads to smaller residuals. Large improvements arise for large changes in lake level.

(a) KBRR: AOD1B versus AOD1B+Volta



(b) KBRA: AOD1B versus AOD1B+Volta

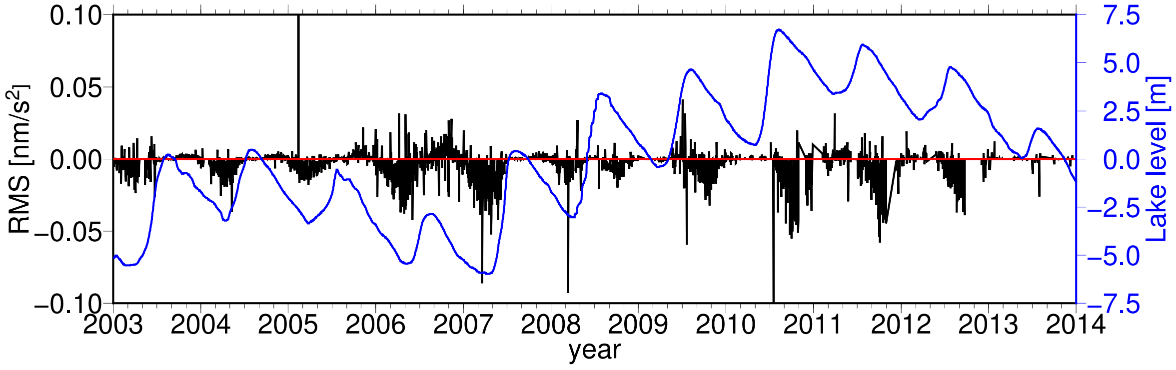
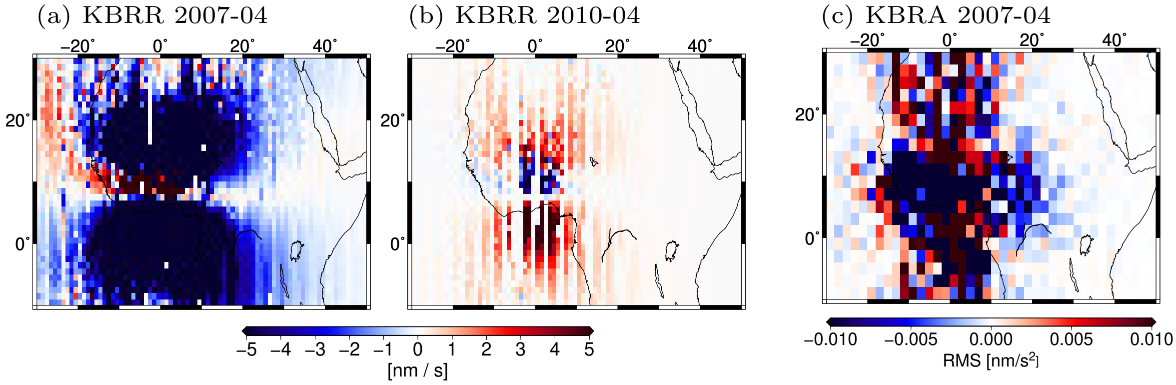


Figure 9.9: Differences between monthly RMS values of (a) KBRR and (b) KBRA residuals from AOD1B and mass changes from lake Volta plus AOD1B (black). The blue line indicates lake level variability.

Snapshots of differences in monthly RMS values of KBRR and KBRA residuals from AOD1B and Lake Volta are shown in Figure 9.10. In April 2007, the RMS of KBRR residuals from Lake Volta is reduced relative to AOD1B north and south of Lake Volta, while the RMS of KBRR residuals is slightly increased in April 2010. RMS values of KBRA residuals are noisy, but also show some improvement directly above and in the surroundings of Lake Volta for April 2007.


(a) KBRR 2007-04

(b) KBRR 2010-04

(c) KBRA 2007-04

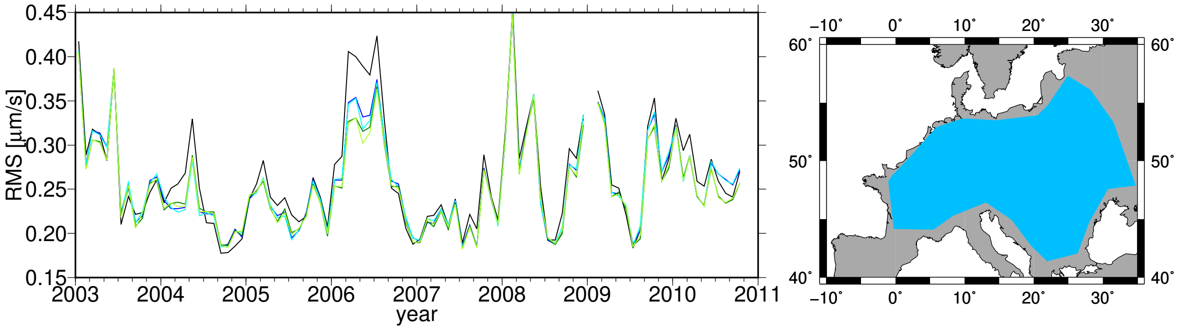
Figure 9.10: Differences of KBRR and KBRA residuals on a 1°×1° grid between AOD1B and mass changes from Lake Volta (additionally considered to AOD1B) for selected months.

9.4 In-Orbit Validation of Data Assimilation Results

Within this thesis, GRACE-derived TWSA were assimilated into the land-surface model CLM3.5 over Europe for the time span 2003 to 2010 leading to a high-resolution reanalysis of TWS (Section 8.3). TWSA from CLM3.5 were converted to spherical harmonic coefficients up to degree and order 240. Hydrological signals outside of Europe were not modeled, but still contribute to the K-band residuals.

Monthly RMS values of KBRR residuals based on the open-loop model (OL) and based on the GRACE assimilating model (DA) are compared in Figure 9.11 for Central Europe and for Eastern Europe. Applying CLM-DA reduces the RMS of the KBRR residuals by 1.6 % compared to CLM-OL over Central Europe and by 3.5 % over Eastern Europe for the time period 2003 to 2010. Particular large improvement is obtained in summer 2006, when an important flood occurred over Europe (Section 8.3.7.2).

(a) Central Europe



(b) Eastern Europe

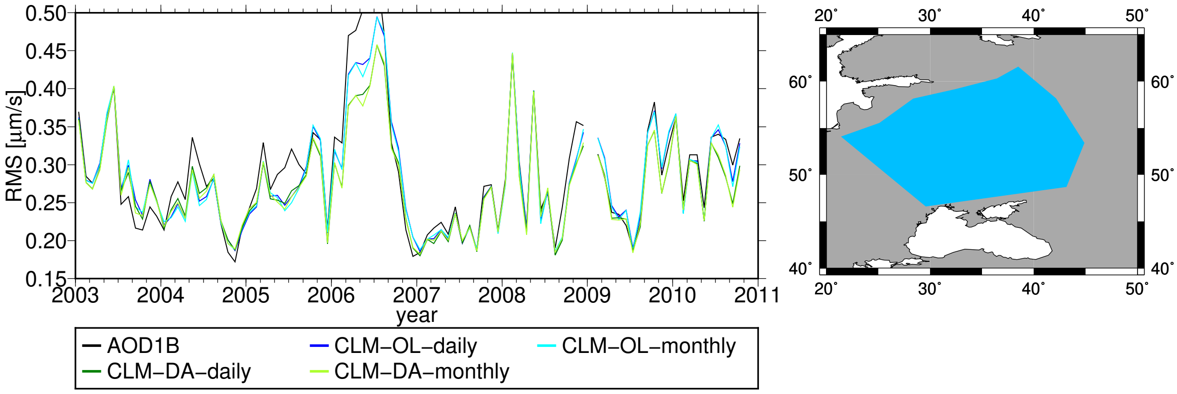


Figure 9.11: Time series of monthly RMS values of KBRR residuals in (a) Central Europe and (b) Eastern Europe. The black line refers to no model subtracted additional to the AOD1B product; blue lines refer to daily and monthly output from the open-loop (OL) version of CLM, and green lines refer to daily and monthly output from the GRACE assimilating (DA) version of CLM.

Figure 9.12 shows the differences of daily RMS values between the ensemble mean of the assimilated model and eight arbitrary ensemble members over Eastern Europe for the first half of the year 2006. The RMS values of the individual ensemble members vary around the ensemble mean, which was computed in spatial domain. The ensemble spread also varies with time. The spread increases with the beginning of the European flood 2006 in March and decreases in June. This means, the computed KBRR residuals indicate increased uncertainty of CLM between March and June 2006.

Figure 9.13 shows differences of RMS values averaged over 2006 for (a) DA versus AOD1B, (b) DA versus OL, and (c) daily output from DA versus monthly output from DA. In 2006,

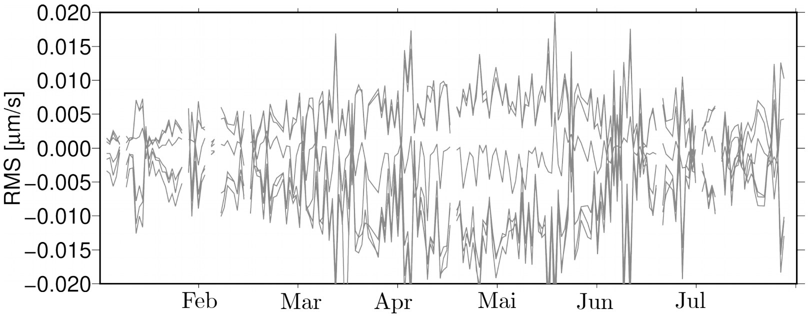


Figure 9.12: Time series of the differences of daily RMS values between the ensemble mean of the assimilated CLM model and eight arbitrary ensemble members for January to July 2006.

CLM-DA reduces the RMS of KBRR residuals relative to AOD1B all over Europe except close to the Black Sea, which means that modeled TWSA from CLM-DA during the year 2006 is indeed realistic. Furthermore, the RMS of KBRR residuals based on DA improves relative to OL north and south of the Black Sea, as well as north and south of regions affected by the major flooding events in 2006. We can conclude that the benefit of data assimilation for improving the representation of extreme events can also be shown on the level of GRACE K-band observations.

When comparing daily versus monthly output from CLM-DA or CLM-OL on the level of KBRR residuals, much smaller changes are obtained than for the global hydrological models (Figure 9.13 c, Figure 9.11). Further investigations are needed to identify the reasons for this discrepancy.

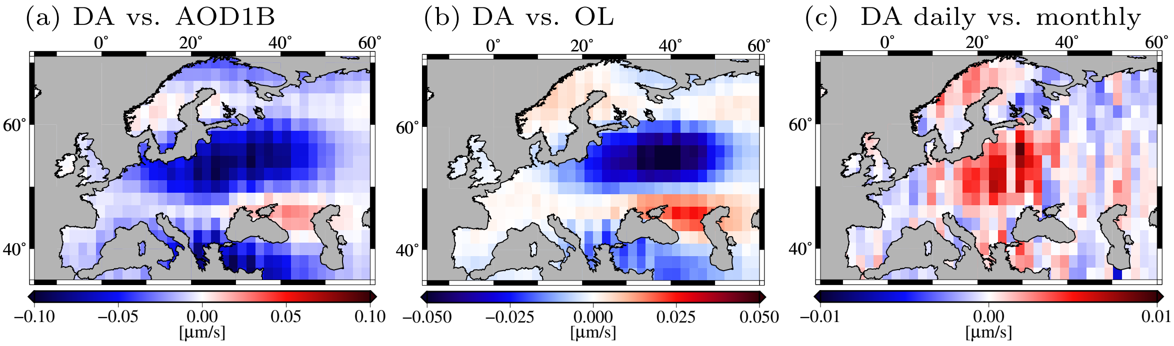


Figure 9.13: Differences of gridded RMS values of KBRR residuals of the year 2006 on a 2°×2° grid for (a) monthly TWS from GRACE assimilating CLM (DA) versus AOD1B, (b) monthly TWS output from DA versus monthly TWS output from the open-loop version of CLM (OL), and (c) daily versus monthly TWS output of DA. Blue color indicates that the first model further reduces the residuals.

9.5 Key Messages of the Analysis of KBRR and KBRA Residuals

Assessing KBRR and KBRA residuals represents a new alternative for validating hydrological model output. The results of this thesis show that evaluating GRACE K-band residuals adds new information that cannot be inferred from the usual TWSA comparisons. Downward continuation and filtering steps generally required for comparing hydrological models to gridded GRACE data are avoided. However, one limitation of the residuals approach is the non-centered localization of KBRR residuals. The computation of KBRA residuals circumvents this problem, but at the cost of increased noise due to the additional differentiation.

Experiments with global hydrological models showed that the models contain information of TWS evolution at time scales smaller than one month. Extreme events, such as the European flood 2006, could be identified in the KBRR residuals. Comparisons of KBRR residuals from CLM-OL and CLM-DA showed the success of data assimilation on the basis of level 1B observations. Furthermore, a reasonable ensemble spread was obtained for KBRR residuals. These results may represent a point of departure for developing future assimilation frameworks that directly use GRACE level 1B observations for assimilation into hydrological models. In this way, the problem of computing a global spherical harmonic solution from time-limited along-track GRACE data could be circumvented, and the full temporal and spatial information content of GRACE observations could be exploited. At the same time, this procedure would potentially inform hydrological models at (sub-) daily time scales.

Chapter 10

Conclusions and Outlook

10.1 Summary

Within this PhD thesis, the Terrestrial Systems Modeling Platform – Parallel Data Assimilation Framework (TerrSysMP-PDAF) was extended to allow for assimilating gridded GRACE derived total water storage anomaly (TWSA) estimates into the land surface model CLM3.5. Assimilation experiments were realized over the European COordinated Regional Downscaling EXperiment (CORDEX) domain and validated against independent observation-based data sets. So far, only few groups worldwide have been working on the assimilation of GRACE data into hydrological models. Generally, the applied hydrological models have a spatial resolution between 0.5° and 1° and model only few vertical layers (Table 3.1). The main result of this thesis is

- a **daily GRACE assimilating total water storage reanalysis (2003 to 2010) at 12.5 km resolution** over the European continent,
- based on **optimized settings** of the data assimilation algorithm,
- validated against **independent observation-based data sets** of storage compartments and water fluxes,
- and analyzed additionally on the level of **GRACE K-band residuals**.

So far, no standard procedure for assimilating GRACE data into hydrological models exists. This thesis includes a detailed review of findings from previous assimilation experiments using GRACE observations. On this basis, synthetic data assimilation experiments were performed with the aim of determining optimal choices for governing parameters of the data assimilation strategy. Thereby, a ranking of the impact of the individual parameters on the assimilation results was obtained. With optimized parameter settings, a run assimilation run with real GRACE data was performed for the time period 2003 to 2010 and validated against observation-based discharge, in-situ and remotely-sensed soil moisture, and evapotranspiration.

All in all, within this thesis, a unique data set of total water storage (TWS) and related water cycle components was created, based on an enhanced understanding about the best way to assimilate GRACE data into hydrological models. Then, a first step towards the direct assimilation of GRACE level 1B data was made by transferring TWS output from the open-loop model and from the assimilated model to the level of K-band range-rate observations.

10.2 Conclusions

Properties of the Data Assimilation Algorithm

This thesis concluded on an optimal way of assimilating GRACE observations into CLM3.5.

Assimilation results are governed by the uncertainties of the observations and the uncertainties of model predictions. In this thesis, emphasis was placed on a thorough representation of the respective error estimates. Starting from the full error covariance matrices of GRACE potential coefficients, full error covariance matrices of gridded, filtered, and rescaled TWSA were derived. In contrast to previous GRACE assimilation studies also leakage and rescaling errors were considered. The model error was represented by the spread of an ensemble of model states generated by perturbing atmospheric forcings with spatially, temporally, and cross-correlated random fields, and by additionally perturbing initial conditions and land surface characteristics. As a result from synthetic experiments, 64 ensemble members were used to obtain a stable ensemble spread, which is a larger ensemble size than applied in previous studies (usually between 20 and 30 ensemble members, e.g., Forman et al., 2012; Girotto et al., 2016; Schumacher et al., 2018).

Main components of the data assimilation framework are the observation operator, the analysis step, and the update step. The observation operator, which relates the model states to the observations, was realized according to previous studies by aggregating all TWS compartments vertically and horizontally to the observation grid and by computing temporal averages. The implementation includes parallelization for the efficient use of compute clusters. Variable assimilation increments (e.g. daily, weekly, monthly, irregularly) were enabled. For the analysis step, user supplied routines of PDAF were implemented and allow to choose between the filter algorithms EnKF, ESTKF, ETKF, SEEK, SEIK and corresponding filter variants. The Localized Error Subspace Transform Kalman Filter (LESTKF) was found to be the most suitable filter algorithm for this study compared to the global filters in terms of an improved description of TWS variability and relatively small assimilation increments. For the application of the assimilation increments during the update step, constraints were developed to guarantee physical consistency of the model.

This thesis confirmed findings of previous studies regarding the importance of taking into account the full covariance matrices of gridded GRACE-derived TWSA (Schumacher et al., 2016) and regarding the positive impact of localization (Girotto et al., 2016; Tangdamrongsub et al., 2017). Indeed, localization had the largest impact on assimilation results leading to an improvement of the validation measures (RMSD, residuals correlation) by up to 50 %. Applying colored versus white noise as error model of the TWS estimates changed the validation measures by 10 % to 40 %. White noise was found to give too much weight to the GRACE observations for observation grids finer than 5°, which emphasized the fact that error correlations should not be neglected. The choice of the localization radius varied the validation measures by 5 % to 20% and the selection of the forgetting factor let to a similar impact. Here, best results were obtained for a localization radius of 7° and a forgetting factor of 0.7, which corresponds to inflating the ensemble spread by the factor 1.12.

One of the key questions in the context of assimilating GRACE data is the optimal choice of the observation grid (Schumacher et al., 2016; Khaki et al., 2017b). Best results were obtained here for a observation grid space of 2° when taking into account correlated observation noise. It is encouraging that the validation measures changed only slightly for different observation grid space, when taking into account correlated observation noise.

This thesis includes an in-depth analysis of the assimilation increments, which provides evidence of seasonal and regional variability of the impact of GRACE assimilation. Assimilating

GRACE data into CLM 3.5 mainly affected soil moisture in the deeper soil layers. In winter, soil ice and snow water also obtained significant assimilation increments (comparable to soil moisture) for Eastern Europe. In all European catchments, water was added to the model during the first half of the year and abstracted from the model during the second half of the year, which indicates deficiencies of CLM3.5 in representing the complete range of TWS variability seen by GRACE.

Assimilating GRACE data into a hydrological model can be seen critically as the water budget is violated by "introducing" or "abstracting" water in the update step. However, I found that when applying the LETKF, the long-term water balance of the model was only slightly impacted by assimilating GRACE-derived TWSA. Over the 8-years study period, only 1 – 6 cm of water were introduced or abstracted from the model for the individual catchments.

Benefit of GRACE Assimilation for Reducing Artificial Trends and Phase Shifts of Modeled TWS

Biases in fluxes cause trends in storages. Experiments with artificial biases in precipitation forcings showed that the assimilation of GRACE data corrects the resulting unrealistic trends in storage. In fact, trends from the assimilated model run with biased forcings were similar to trends from the assimilated model run with unbiased forcings. Phase shifts between model and observations were also reduced by data assimilation. Both aspects suggest that the assimilation of GRACE data might be beneficial in data sparse regions, where hydrological models suffer from low-quality and biased input data.

Benefit of GRACE Assimilation for Variables of the Water Cycle

It was expected that after assimilating GRACE data modeled TWSA fit better to the GRACE observations. Especially, the annual amplitudes of modeled TWSA improved over Eastern Europe. Furthermore, interannual variability was better represented by the assimilated model leading to increases in residual correlation by up to 50 %. I also demonstrated that the assimilated model performed better in explaining GRACE K-band residuals than the open-loop model. However, in contrast to the results from global models, no apparent daily signal was determined in TWS simulated by the CLM3.5 model when evaluating K-band residuals.

The assimilation of GRACE data has the potential to improve also individual storage compartments, but this is not necessarily always the case. Modeled soil moisture was validated against in-situ soil moisture measurements and remotely sensed (surface) soil moisture. The assimilation of GRACE-derived TWSA did not improve the agreement of the model with soil moisture from remote sensing. This was expected as the impact of data assimilation on the upper soil layers, which cover only the first few centimeters, is small. However, large assimilation increments were obtained for soil moisture in deeper soil layers, which can be validated against in-situ measurements. Yet, for 70 % of the in-situ soil moisture stations validation measures did not change significantly due to data assimilation. Generally, it is difficult to compare a point measurement against model grid cells of 12.5 km. Despite these limitations, for more than 10 % of the stations, modeled soil moisture was improved at a 10 % level in terms of residual correlation.

Although data assimilation has a direct impact only on the compartments of TWS, fluxes of water are indirectly affected as more or less water is available after each assimilation step. In fact, data assimilation improved correlations between simulated runoff with respect to discharge from gauging stations in most catchments. Data assimilation mainly impacted the runoff component subsurface drainage. Thus, we can hypothesize that a more sophisticated groundwater model would likely lead to better runoff simulations. Evapotranspiration was only slightly impacted by data assimilation with no positive effect comparing to observation-based data. Interestingly, the amplitude of the daily cycle of evapotranspiration changed in summer 2005 as more water was available due to data assimilation.

Benefit of GRACE Assimilation for the Representation of Extreme Events

Assimilating GRACE-derived TWSA into CLM3.5 can be interpreted as downscaling of coarse monthly GRACE observations to daily high-resolution TWSA with an added vertical stratification. When assessing extreme events, the high-resolution assimilated model adds information on the evolution of TWS at smaller spatial and temporal scales than available from GRACE. Within this thesis, the European heat wave 2003 and the European flood 2006 were analyzed based on the assimilated model. While the open-loop model failed in representing the flood 2006, the evolution of the flood was represented in details by the assimilated model.

10.3 Outlook

Within this thesis, a framework for assimilating GRACE observations into CLM3.5 was optimized regarding the assimilation strategy. However, several aspects can be addressed to further improve the results and to advance applications. The start of the GRACE Follow-On mission in May 2018 ensures that the assimilation of remotely sensed TWSA is going to remain a relevant topic of research.

Study area

At Forschungszentrum Jülich, the CLM3.5 component of TerrSysMP was also set up for a 3 km grid. The GRACE assimilation framework developed within this thesis is transferable to this grid, so that GRACE data could be disaggregated to a 3 km grid over Europe in future experiments. Furthermore, the data assimilation framework is easily transferable to other regions by changing atmospheric forcings and soil maps. The assimilation of GRACE data might be particularly helpful in regions of sparse data coverage to obtain more realistic TWS estimates.

Hydrological Model

The land-surface model CLM3.5 was used here as it is part of the TerrSysMP framework. However, more recent versions of CLM contain improved representations of soil and plant hydrology, snow processes, river modeling and surface water stores. The latest version is CLM5.0, but it is not yet part of TerrSysMP-PDAF. For future investigations, it would be interesting to transfer the data assimilation framework to a more recent CLM version and also to compare the impact of data assimilation on different model versions.

In this thesis, atmospheric forcings were obtained from the Weather Research and Forecasting (WRF) model for the time period 2003 to 2010. In future, the study period should be extended towards the end of the GRACE time span (middle of 2017) or even further towards GRACE-Follow-On.

First experiments using forcings from the COnsortium for Small Scale MOdelling REAnalysis at 6 km horizontal resolution (COSMO-REA6) showed a better agreement of the open-loop model with GRACE-derived TWSA. Furthermore, COSMO-REA6 is available at hourly resolution, while the WRF data set has a resolution of three hours. Here, COSMO-REA6 forcings were available for the time period 2003 to 2006, and they were only used in a reference run for the synthetic experiments. However, in future, WRF forcings should be replaced by COSMO-REA6 forcings to improve the realism of simulated TWS. In this context, it will be interesting to investigate the impact of GRACE data assimilation when using different atmospheric forcing data sets.

The FAO/UNESCO (Food and Agriculture Organization/United Nations Educational, Scientific and Cultural Organization) soil texture data set used in this study has a relatively coarse resolution. Alternative soil texture data sets should be considered, e.g., the data set of the European Soil Data Centre (ESDAC) used here for the reference run of the synthetic experiments.

Within this thesis, only the CLM component of the Terrestrial Systems Modeling Platform (TerrSysMP) was applied. It is tempting to suggest the data assimilation framework should be extended also for coupled applications of CLM and the groundwater component ParFlow (Parallel Flow). However, future studies must be carried out to understand how data assimilation informs the coupled system and how the data assimilation strategy must be optimized. Then, the impact of GRACE data assimilation on groundwater storage could be further assessed. Including the groundwater component could also improve the agreement of updated soil moisture with independent observations. As a final vision, using TerrSysMP-PDAF, GRACE data could be assimilated into a fully coupled system consisting of ParFlow, CLM, and the atmospheric component COSMO together with observations of other key variables of the terrestrial water cycle.

Data Assimilation Framework

The assimilation framework developed within this thesis can be further improved and also provides an environment for further reaching data assimilation experiments.

Currently, soil liquid water and soil ice of each soil layer are added before setting up the state vector. The assimilation increment is then distributed proportionally to guarantee physical consistency of each ensemble member. However, more sophisticated constraints for updating soil liquid water and soil ice independently could lead to better results.

The data assimilation framework can easily be extended for simultaneous parameter calibration, e.g., calibration of soil properties and hydraulic parameters. Furthermore, experiments performed at Forschungszentrum Jülich achieved improvements in modeled surface soil moisture and runoff by assimilating remotely sensed soil moisture (personal communication, Bibi Naz). As my results (Section 8.3.4) showed that GRACE data assimilation does not significantly improve modeled soil moisture, it is possible that in a joint assimilation approach GRACE-derived TWSA and surface soil moisture complement each other. Additionally assimilating river discharge and snow water could further improve the simulated water storages and fluxes.

The data assimilation framework developed within this thesis allows for assimilating GRACE-derived TWSA at arbitrary intervals, i.e., to assimilate daily or weekly GRACE data. However, uncertainties of these non-standard GRACE products are difficult to estimate. So far, experiments with daily GRACE solutions lead to convergence of the model towards the observations. Yet, the assimilation of daily GRACE data is extremely interesting as this is one way to minimize jumps that arise when updating the model every month. Future investigations could compare the assimilation results from GRACE data given at different temporal resolution. Soon, it will be possible to assimilate also TWSA derived from the GRACE Follow-ON mission. In this context, it will be interesting to develop a strategy for dealing with the gap between GRACE and GRACE Follow-On.

The downward-continuation, filtering, and rescaling procedure necessary for computing gridded GRACE-derived TWSA leads to large uncertainties. It would be interesting to directly assimilate GRACE Level 1B data into hydrological models. First tests showed that the assimilated model better explains the K-band range-rate residuals than the open-loop model. However, a sophisticated observation operator must be developed for assimilating GRACE Level 1B observations.

Validation Environment

In this thesis, the performance of data assimilation was validated against in-situ and remotely sensed soil moisture observations, river discharge measurements, and observation-based evapotranspiration. Currently, a comprehensive validation toolbox is developed to also validate against snow depth and snow water observations, gridded runoff estimates, super-conducting gravimeter measurements, and vertical deformation from GNSS height time series. First results showed that the assimilated model better explains the temporal variability in GNSS height time series than the open-loop model or GRACE.

Applications

Data assimilation merges observations and numerical model simulations. In this thesis, data assimilation was used to generate a high-resolution reanalysis of TWS over Europe. I found that modeled water fluxes and individual storages were also improved to some extent by assimilating GRACE data. Thus, the assimilated model provides interesting information on the evolution of different water cycle variables, which might be of use for early warning systems of natural hazards over Europe.

GRACE de-aliasing products aim at removing short-term ($<$ one month) mass redistribution before computing monthly gravity field solutions. As the CLM model runs at hourly time steps, TWSA from the assimilated model provide an interesting option for the use as GRACE de-aliasing product.

Current investigations showed evidence of daily hydrological loading signals in GPS time series over Europe (Springer et al., 2019). The GRACE assimilating CLM model provides a promising option for removing hydrological-induced vertical deformation from daily GPS time series, thus, demasking geophysical processes such as land subsidence and tectonic or volcanic deformation. Indeed, TWSA from the assimilated CLM model might turn out as a great tool for GPS applications, and it could also play a role in improving reference system computations.

Acronyms

3D-Var	**3**-**D**imensional **Var**iational 29
AIUB	**A**stronomical **I**nstitute of the **U**niversity of **B**ern 53
CCI	**C**limate **C**hange **I**nitiative 32
CDO	**C**limate **D**ata **O**perators 84
CF	**C**enter of **F**igure 40
CLM	**C**ommunity **L**and **M**odel 23
CLM3.5	**C**ommunity **L**and **M**odel version **3.5** 23–29
CLSM	**C**atchment **L**and **S**urface **M**odel 18
CM	**C**enter of **M**ass 40
COSMO	**CO**nsortium for **S**mall Scale **MO**delling 24
COSMO-REA6	**COSMO REA**nalysis at **6** km horizonal resolution 29
CRP-GL	**C**entre de **R**echerche **P**ublic - **G**abriel **L**ippmann 29
CSR	University of Texas **C**enter for **S**pace **R**esearch 6
CV	**C**oefficient of **V**ariation 50
d/o	**d**egree and **o**rder 10
DEnKF	**D**eterministic **En**semble **K**alman **F**ilter 20
DLR	**D**eutsches Zentrum für **L**uft- und **R**aumfahrt 6
E-RUN	observational gridded **RUN**off estimates for **E**urope 35
ECV	**E**ssential **C**limate **V**ariables 1
EGSIEM	**E**uropean **G**ravity **S**ervice for **I**mproved **E**mergency **M**anagement 11
EKF	**E**xtended **K**alman **F**ilter 57
EnKF	**En**semble **K**alman **F**ilter 58
EnOI	**En**semble **O**ptimal **I**nterpolation scheme 20
EnSRF	**En**semble **S**quare **R**oot **F**ilter 20
EOP	**E**arth **O**rientation **P**arameters 53
ESA	**E**uropean **S**pace **A**gency 32
ESDAC	**E**uropean **S**oil **D**ata **C**entre 29
ESTKF	**E**rror **S**ubspace **T**ransform **K**alman **F**ilter 60
ETKF	**E**nsemble **T**ransform **K**alman **F**ilter 59
EWH	**E**quivalent **W**ater **H**eight 48
FAO	**F**ood and **A**griculture **O**rganization 29

GES DISC	**G**oddard **E**arth **S**cience **D**ata and **I**nformation **S**ervices **C**enter 31
GFZ	**G**eo**F**orschungs**Z**entrum 6
GIA	**G**lacial **I**sostatic **A**djustment 46
GLDAS	**G**lobal **L**and **D**ata **A**ssimilation **S**ystem 31
GMTED	**G**lobal **M**ultiresolution **T**errain **E**levation **D**ata 29
GPS	**G**lobal **P**ositioning **S**ystem 7, 41
GRACE	**G**ravity **R**ecovery **A**nd **C**limate **E**xperiment 5–11
GRACE-FO	**GRACE** **F**ollow-**O**n 8
GRDC	**G**lobal **R**unoff **D**ata **C**entre 34
GROOPS	**GR**avity field **O**bject **O**riented **P**rogramming **S**ystem 53
GSDS	**G**RACE **S**cience **D**ata **S**ystem 6
GSFC	**G**oddard **S**pace **F**light **C**enter 11
HBV	**H**ydrologiska **B**yrans **V**attenavdelning 34
IERS	**I**nternational **E**arth **R**otation and Reference Systems **S**ervice 53
ISMN	**I**nternational **S**oil **M**oisture **N**etwork 32
JPL	**J**et **P**ropulsion **L**aboratory 6
KBR	**K**-**B**and Microwave **R**anging 6
KBRA	**K**-**B**and **R**ange-**A**cceleration 141
KBRR	**K**-**B**and **R**ange-**R**ate 141
KF	**K**alman **F**ilter 57
LEnKF	**L**ocal **En**semble **K**alman **F**ilter 83
LESTKF	**L**ocal **E**rror **S**ubspace **T**ransform **K**alman **F**ilter 83
LSDM	**L**and **S**urface **D**ischarge **M**odel 31
MODIS	**M**oderate **R**esolution **I**maging **S**pectroradiometer 29
MPI	**M**ax **P**lanck **I**nstitute 35
MPI	**M**essage **P**assing **I**nterface 70
NASA	**N**ational **A**eronautics and **S**pace **A**dministration 6
NetCDF	**Net**work **C**ommon **D**ata **F**ormat 84
NOAH	**N**oah **L**and **S**urface **M**odel 31
NS	**N**ash-**S**utcliffe 85
NWP	**N**umerical **W**eather **P**rediction 5
OASIS3-MCT	**O**cean **A**tmosphere **S**ea **I**ce **S**oil coupling tool 24
OSSE	**O**bservation **S**ystem **S**imulation **E**xperiment 86
ParFlow	**Par**allel **Flow** 24
PCA	**P**rinciple **C**omponent **A**nalysis 15
PDAF	**P**arallel **D**ata **A**ssimilation **F**ramework 70
PDF	**P**robability **D**ensity **F**unction 56
PFT	**P**lant **F**unctional **T**ype 24

RDC	**R**aw **D**ata **C**enter 8
RL	**ReL**ease 53
RMSD	**R**oot **M**ean **S**quared **D**eviation 84
SEIK	**S**ingular **E**volutive **I**nterpolated **K**alman filter 60
SH	**S**pherical **H**armonic 39
SLR	**S**atellite **L**aser **R**anging 41
SQRA	**SQ**uare **R**oot **A**nalysis scheme 20
SST	**S**atellite-to-**S**atellite **T**racking 6
TerrSysMP	**Terr**estrial **Sys**tems **M**odeling **P**latform 70
TWS	**T**otal **W**ater **S**torage 37
TWSA	**T**otal **W**ater **S**torage **A**nomaly 37
UNESCO	**U**nited **N**ations **E**ducational, **S**cientific and **C**ultural **O**rganization 29
VIC	**V**ariable **I**nfiltration **C**apacity 31
W3RA	**W**orld **W**ide **W**ater **R**esources **A**ssessment 20
WGHM	**W**aterGAP **G**lobal **H**ydrology **M**odel 29
WRF	**W**eather **R**esearch and **F**orecasting 29

List of Figures

List of Tables

Bibliography

A, G., J. Wahr, and S. Zhong, 2013: Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: An application to Glacial Isostatic Adjustment in Antarctica and Canada. *Geophysical Journal International*, **192 (2)**, 557–572, doi:10.1093/gji/ggs030.

Abaza, M., C. Garneau, and F. Anctil, 2015: Comparison of Sequential and Variational Streamflow Assimilation Techniques for Short-Term Hydrological Forecasting. *Journal of Hydrologic Engineering*, **20 (2)**, 04014 042, doi:10.1061/(ASCE)HE.1943-5584.0001013.

Agnew, D. C., 2010: 6 Earth Tides. *Geodesy: Treatise on Geophysics*, 163.

Alkama, R., et al., 2010: Global Evaluation of the ISBA-TRIP Continental Hydrological System. Part I: Comparison to GRACE Terrestrial Water Storage Estimates and In Situ River Discharges. *Journal of Hydrometeorology*, **11 (3)**, 583–600, doi:10.1175/2010JHM1211.1.

Andersen, O. B., 2005: Global inter-annual gravity changes from GRACE: Early results. *Geophysical Research Letters*, **32 (1)**, doi:10.1029/2004GL020948.

Andersen, O. B., S. I. Seneviratne, J. Hinderer, and P. Viterbo, 2005: GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophysical Research Letters*, **32 (18)**, n/a–n/a, doi:10.1029/2005GL023574.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230 (1)**, 99–111, doi:10.1016/j.physd.2006.02.011.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, **61 (1)**, 72–83, doi:10.1111/j.1600-0870.2008.00361.x.

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, **127 (12)**, 2741–2758.

Andrew, R., H. Guan, and O. Batelaan, 2017: Estimation of GRACE water storage components by temporal decomposition. *Journal of Hydrology*, **552**, 341–350, doi:10.1016/j.jhydrol.2017.06.016.

Awange, J., M. Gebremichael, E. Forootan, G. Wakbulcho, R. Anyah, V. Ferreira, and T. Alemayehu, 2014: Characterization of Ethiopian mega hydrogeological regimes using GRACE, TRMM and GLDAS datasets. *Advances in Water Resources*, **74**, 64–78, doi:10.1016/j.advwatres.2014.07.012.

Bai, P., X. Liu, and C. Liu, 2017: Improving hydrological simulations by incorporating GRACE data for model calibration. *Journal of Hydrology*, doi:10.1016/j.jhydrol.2017.12.025.

Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation: Ensemble-variational Data Assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143 (703)**, 607–633, doi:10.1002/qj.2982.

Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results. *Monthly Weather Review*, **132 (4)**, 897–914, doi:10.1175/1520-0493(2004) 132<0897:ATVDAS>2.0.CO;2.

Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser, 2003: Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes. *Journal of Geophysical Research: Atmospheres*, **108 (D16)**, doi:10.1029/2002JD003334.

Bettadpur, S., 2012a: GRACE Level-2 Gravity Field Product User Handbook. Tech. rep., Center for Space Research, The University of Texas at Austin.

Bettadpur, S., 2012b: GRACE Product Specification Document. Tech. rep., Center for Space Research, The University of Texas at Austin.

Bettadpur, S., 2012c: UTCSR Level - 2 Processing Standards Documen. Gravity Recovery and Climate Experiment GRACE 327 - 742 (CSR - GR - 12 - xx), Center for Space Research, The University of Texas at Austin.

Bettadpur, S. and C. McCullough, 2017: The Classical Variational Approach. *SpringerLink*, 81–95, doi:10.1007/978-3-319-49941-3_3.

Beutler, G., A. Jäggi, L. Mervart, and U. Meyer, 2010a: The celestial mechanics approach: Application to data of the GRACE mission. *J Geod*, **84 (11)**, 661–681, doi: 10.1007/s00190-010-0402-6.

Beutler, G., A. Jäggi, L. Mervart, and U. Meyer, 2010b: The celestial mechanics approach: Theoretical foundations. *J Geod*, **84 (10)**, 605–624, doi:10.1007/s00190-010-0401-7.

Biancale, R. and A. Bode, 2006: Mean annual and seasonal atmospheric tide models based on 3-hourly and 6-hourly ECMWF surface pressure data. Tech. rep., Deutsches GeoForschungsZentrum GFZ. doi:10.2312/GFZ.b103-06011.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, **129 (3)**, 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bishop, C. H. and D. Hodyss, 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus A*, **61 (1)**, 84–96, doi:10.1111/j.1600-0870.2008. 00371.x.

Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014: The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bull. Amer. Meteor. Soc.*, **95 (9)**, 1431–1443, doi:10.1175/BAMS-D-13-00047.1.

Bollmeyer, C., et al., 2015: Towards a high-resolution regional reanalysis for the European CORDEX domain: High-Resolution Regional Reanalysis for the European CORDEX Domain. *Quarterly Journal of the Royal Meteorological Society*, **141 (686)**, 1–15, doi: 10.1002/qj.2486.

Bonan, G. B., S. Levis, L. Kergoat, and K. W. Oleson, 2002: Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, **16 (2)**, 5–1–5–23, doi:10.1029/2000GB001360.

Bonin, J. A., S. Bettadpur, and B. D. Tapley, 2012: High-frequency signal and noise estimates of CSR GRACE RL04. *Journal of Geodesy*, **86 (12)**, 1165–1177, doi:10.1007/s00190-012-0572-5.

Bosch, W., R. Savcenko, F. Flechtner, C. Dahle, T. Mayer-Gürr, D. Stammer, E. Taguchi, and K.-H. Ilk, 2009: Residual ocean tide signals from satellite altimetry, GRACE gravity fields, and hydrodynamic modelling. *Geophysical Journal International*, **178 (3)**, 1185–1192, doi:10.1111/j.1365-246X.2009.04281.x.

Bruinsma, S., J.-M. Lemoine, R. Biancale, and N. Valès, 2010: CNES/GRGS 10-day gravity field models (release 2) and their evaluation. *Advances in Space Research*, **45 (4)**, 587–601, doi:10.1016/j.asr.2009.10.012.

Burgers, G., P. Jan van Leeuwen, and G. Evensen, 1998: Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, **126 (6)**, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Cai, X., Z.-L. Yang, Y. Xia, M. Huang, H. Wei, L. R. Leung, and M. B. Ek, 2014: Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *Journal of Geophysical Research: Atmospheres*, **119 (24)**, 13,751–13,770, doi:10.1002/2014JD022113.

Carton, J. A., 1991: Effect of seasonal surface freshwater flux on sea surface temperature in the tropical Atlantic Ocean. *Journal of Geophysical Research: Oceans*, **96 (C7)**, 12 593–12 598, doi:10.1029/91JC01256.

Chambers, D. P. and J. A. Bonin, 2012: Evaluation of Release-05 GRACE time-variable gravity coefficients over the Ocean. *Ocean Sci.*, **9**, 2187–2214, doi:10.5194/os-8-859-2012.

Chen, J. L., 2005: Spatial sensitivity of the Gravity Recovery and Climate Experiment (GRACE) time-variable gravity observations. *Journal of Geophysical Research*, **110 (B8)**, doi:10.1029/2004JB003536.

Chen, J. L., M. Rodell, C. R. Wilson, and J. S. Famiglietti, 2005: Low degree spherical harmonic influences on Gravity Recovery and Climate Experiment (GRACE) water storage estimates. *Geophysical Research Letters*, **32 (14)**, n/a–n/a, doi:10.1029/2005GL022964.

Chen, J. L., C. R. Wilson, B. D. Tapley, L. Longuevergne, Z. L. Yang, and B. R. Scanlon, 2010: Recent La Plata basin drought conditions observed by satellite gravimetry. *Journal of Geophysical Research*, **115 (D22)**, doi:10.1029/2010JD014689.

Chen, J. L., C. R. Wilson, B. D. Tapley, Z. L. Yang, and G. Y. Niu, 2009: 2005 drought event in the Amazon River basin as measured by GRACE and estimated by climate models. *Journal of Geophysical Research*, **114 (B5)**, doi:10.1029/2008JB006056.

Chen, Q., Y. Shen, W. Chen, X. Zhang, and H. Hsu, 2016: An improved GRACE monthly gravity field solution by modeling the non-conservative acceleration and attitude observation errors. *J Geod*, **90 (6)**, 503–523, doi:10.1007/s00190-016-0889-6.

Chen, Q., Y. Shen, X. Zhang, H. Hsu, W. Chen, X. Ju, and L. Lou, 2015: Monthly gravity field models derived from GRACE Level 1B data using a modified short-arc approach. *Journal of Geophysical Research: Solid Earth*, **120 (3)**, 1804–1819, doi:10.1002/2014JB011470.

Chen, X., D. Long, Y. Hong, C. Zeng, and D. Yan, 2017: 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 (3)**, 2431–2466, doi:10.1002/2016WR019656.

Cheng, M., B. D. Tapley, and J. C. Ries, 2013a: Deceleration in the Earth's oblateness. *Journal of Geophysical Research: Solid Earth*, **118 (2)**, 740–747, doi:10.1002/jgrb.50058.

Cheng, M. K., J. C. Ries, and B. D. Tapley, 2013b: Geocenter Variations from Analysis of SLR Data. *Reference Frames for Applications in Geosciences*, Z. Altamimi and X. Collilieux, Eds., Springer Berlin Heidelberg, Berlin, Heidelberg, Vol. 138, 19–25, doi:10.1007/978-3-642-32998-2_4.

Chew, C. C. and E. E. Small, 2014: Terrestrial water storage response to the 2012 drought estimated from GPS vertical position anomalies. *Geophysical Research Letters*, **41 (17)**, 6145–6151, doi:10.1002/2014GL061206.

Craig, A., S. Valcke, and L. Coquart, 2017: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0. *Geoscientific Model Development*, **10 (9)**, 3297–3308, doi:https://doi.org/10.5194/gmd-10-3297-2017.

Dach, R., U. Hugentobler, P. Fridez, M. Meindl, and others, 2007: Bernese GPS software version 5.0. *Astronomical Institute, University of Bern*, **640**, 114.

Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H. Neumayer, 2012: GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005. Scientific Technical Report STR12/02 – Data, 21 pp. Revised Edition, January 2013.

Dai, A. and K. E. Trenberth, 2003: New estimates of continental discharge and oceanic freshwater transport. *Proceedings of the Symposium on Observing and Understanding the Variability of Water in Weather and Climate, 83rd Annual American Meteorological Society Meeting, Long Beach, CA*, 1–18.

Danielson, J. J. and D. B. Gesch, 2011: Global multi-resolution terrain elevation data 2010 (GMTED2010). Tech. Rep. 2011-1073, U.S. Geological Survey.

Darcy, H., 1856: *Les Fontaines Publiques de La Ville de Dijon : Exposition et Application Des Principes à Suivre et Des Formules à Employer Dans Les Questions de Distribution d'eau.*

Decharme, B. and H. Douville, 2006: Uncertainties in the GSWP-2 precipitation forcing and their impacts on regional and global hydrological simulations. *Clim Dyn*, **27 (7-8)**, 695–713, doi:10.1007/s00382-006-0160-6.

Desai, S. D., 2002: Observing the pole tide with satellite altimetry. *Journal of Geophysical Research: Oceans*, **107 (C11)**, doi:10.1029/2001JC001224.

Devia, G. K., B. Ganasri, and G. Dwarakish, 2015: A Review on Hydrological Models. *Aquatic Procedia*, **4**, 1001–1007, doi:10.1016/j.aqpro.2015.02.126.

Dewan, T. H., 2015: Societal impacts and vulnerability to floods in Bangladesh and Nepal. *Weather and Climate Extremes*, **7**, 36–42, doi:10.1016/j.wace.2014.11.001.

Dill, R., 2008: Hydrological model LSDM for operational Earth rotation and gravity field variations. Scientific Technical Report STR 08/09, GFZpublic Deutsches GeoForschungsZentrum GFZ.

Dimet, F.-X. L. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A*, **38A (2)**, 97–110, doi:10.1111/j.1600-0870.1986.tb00459.x.

Ditmar, P. and A. A. v. E. van der Sluijs, 2004: A technique for modeling the Earth's gravity field on the basis of satellite accelerations. *Journal of Geodesy*, **78 (1-2)**, 12–33, doi:10.1007/s00190-003-0362-1.

Dobslaw, H., I. Bergmann-Wolf, R. Dill, L. Poropat, and F. Flechtner, 2017a: Product Description Document for AOD1B Release 06. Tech. Rep. GRACE 327-750.

Dobslaw, H., F. Flechtner, I. Bergmann-Wolf, C. Dahle, R. Dill, S. Esselborn, I. Sasgen, and M. Thomas, 2013: Simulating high-frequency atmosphere-ocean mass variability for dealiasing of satellite gravity observations: AOD1B RL05. *Journal of Geophysical Research: Oceans*, **118 (7)**, 3704–3711, doi:10.1002/jgrc.20271.

Dobslaw, H., et al., 2017b: A new high-resolution model of non-tidal atmosphere and ocean mass variability for de-aliasing of satellite gravity observations: AOD1B RL06. *Geophysical Journal International*, **211 (1)**, 263–269, doi:10.1093/gji/ggx302.

Döll, P., H. Douville, A. Güntner, H. Müller Schmied, and Y. Wada, 2016: Modelling Freshwater Resources at the Global Scale: Challenges and Prospects. *Surv Geophys*, **37 (2)**, 195–221, doi:10.1007/s10712-015-9343-1.

Döll, P., M. Fritsche, A. Eicker, and H. Müller Schmied, 2014a: 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 (6)**, 1311–1331, doi:10.1007/s10712-014-9282-2.

Döll, P., H. Hoffmann-Dobrev, F. Portmann, S. Siebert, A. Eicker, M. Rodell, G. Strassberg, and B. Scanlon, 2012: Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, **59-60**, 143–156, doi:10.1016/j.jog.2011.05.001.

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

Döll, P., H. Müller Schmied, C. Schuh, F. T. Portmann, and A. Eicker, 2014b: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, **50 (7)**, 5698–5720, doi:10.1002/2014WR015595.

Dorigo, W., et al., 2017: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, **203**, 185–215, doi:10.1016/j.rse.2017.07.001.

Dorigo, W. A., et al., 2011: The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.*, **15 (5)**, 1675–1698, doi:10.5194/hess-15-1675-2011.

Dorigo, W. A., et al., 2013: Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. *Vadose Zone Journal*, **12 (3)**, doi:10.2136/vzj2012.0097.

Draper, C., R. Reichle, G. D. Lannoy, B. Scarino, C. Draper, R. Reichle, G. D. Lannoy, and B. Scarino, 2015: A Dynamic Approach to Addressing Observation-Minus-Forecast Bias in a Land Surface Skin Temperature Data Assimilation System. http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-14-0087.1, doi:10.1175/JHM-D-14-0087.1.

Duan, Q., et al., 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, **320 (1-2)**, 3–17, doi:10.1016/j.jhydrol.2005.07.031.

Duan, X. J., J. Y. Guo, C. K. Shum, and W. van der Wal, 2009: On the postprocessing removal of correlated errors in GRACE temporal gravity field solutions. *J Geod*, **83 (11)**, 1095, doi:10.1007/s00190-009-0327-0.

Eicker, A., E. Forootan, A. Springer, L. Longuevergne, and J. Kusche, 2016: Does GRACE see the terrestrial water cycle "intensifying"? *Journal of Geophysical Research: Atmospheres*, **121 (2)**, 733–745, doi:10.1002/2015JD023808.

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

Eicker, A. and A. Springer, 2016: Monthly and sub-monthly hydrological variability: In-orbit validation by GRACE level 1B observations. *Journal of Geodesy*, **90 (6)**, 573–584, doi:10.1007/s00190-016-0895-8.

Ek, M. B., 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, **108 (D22)**, doi:10.1029/2002JD003296.

Ellmer, M. and T. Mayer-Gürr, 2017: High precision dynamic orbit integration for spaceborne gravimetry in view of GRACE Follow-on. *Advances in Space Research*, **60 (1)**, 1–13, doi:10.1016/j.asr.2017.04.015.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99 (C5)**, 10 143, doi:10.1029/94JC00572.

Evensen, G., 2009: *Data Assimilation: The Ensemble Kalman Filter*. Springer.

Evensen, G. and P. J. van Leeuwen, 2000: An Ensemble Kalman Smoother for Nonlinear Dynamics. *Mon. Wea. Rev.*, **128 (6)**, 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Famiglietti, J. S., et al., 2011: Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters*, **38 (3)**, doi:10.1029/2010GL046442.

Fang, H., H. K. Beaudoing, M. Rodell, W. L. Teng, and B. E. Vollmer, 2009: Global Land Data Assimilation System (GLDAS) Products, Services and Application from NASA Hydrology Data and Information Services Center (HDISC).

Farrell, W. E., 1972: Deformation of the Earth by surface loads. *Reviews of Geophysics*, **10 (3)**, 761, doi:10.1029/RG010i003p00761.

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

Fenoglio-Marc, L., J. Kusche, and M. Becker, 2006: Mass variation in the Mediterranean Sea from GRACE and its validation by altimetry, steric and hydrologic fields. *Geophysical Research Letters*, **33 (19)**, doi:10.1029/2006GL026851.

Flechtner, F., H. Dobslaw, and E. Fagiolini, 2015: AOD1B Product Description Document for Product Release 05. Tech. Rep. GRACE 327 - 750 (GR - GFZ - AOD - 0001).

Flechtner, F., M. Thomas, and H. Dobslaw, 2010: Improved non-tidal atmospheric and oceanic de-aliasing for GRACE and SLR satellites. *System Earth via Geodetic-Geophysical Space Techniques.*, L. Stroink, V. Mosbrugger, G. Wefer, F. M. Flechtner, T. Gruber, A. Güntner, M. Mandea, M. Rothacher, T. Schöne, and J. Wickert, Eds., Springer, Advanced Technologies in Earth Sciences, 131–142.

Flowerdew, J., 2015: Towards a theory of optimal localisation. *Tellus A: Dynamic Meteorology and Oceanography*, **67 (1)**, 25 257, doi:10.3402/tellusa.v67.25257.

Forman, B. A. and S. A. Margulis, 2010: Assimilation of multiresolution radiation products into a downwelling surface radiation model: 1. Prior ensemble implementation. *Journal of Geophysical Research*, **115 (D22)**, doi:10.1029/2010JD013920.

Forman, B. A. and R. H. Reichle, 2013: The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system: Retrieval and Model Error Scales in Tws Assimilation. *Water Resources Research*, **49 (11)**, 7457–7468, doi:10.1002/2012WR012885.

Forman, B. A., R. H. Reichle, and M. Rodell, 2012: Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resources Research*, **48 (1)**, doi:10.1029/2011WR011239.

Forootan, E., O. Didova, M. Schumacher, J. Kusche, and B. Elsaka, 2014: Comparisons of atmospheric mass variations derived from ECMWF reanalysis and operational fields, over 2003-2011. *J. Geod.*, in-press.

Frappart, F., G. Ramillien, M. Leblanc, S. O. Tweed, M.-P. Bonnet, and P. Maisongrande, 2011: An independent component analysis filtering approach for estimating continental hydrology in the GRACE gravity data. *Remote Sensing of Environment*, **115 (1)**, 187–204, doi:10.1016/j.rse.2010.08.017.

Friedl, M., et al., 2002: Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, **83 (1-2)**, 287–302, doi:10.1016/S0034-4257(02)00078-0.

Gao, H., et al., 2010: Water budget record from variable infiltration capacity (VIC) model. *Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records*, 120–173.

Gardner, A. S., et al., 2013: A Reconciled Estimate of Glacier Contributions to Sea Level Rise: 2003 to 2009. *Science*, **340 (6134)**, 852–857, doi:10.1126/science.1234532.

Gaspari, G. and S. E. Cohn, 2006: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 723–757, doi:10.1002/qj.49712555417.

Gasper, F., K. Goergen, P. Shrestha, M. Sulis, J. Rihani, M. Geimer, and S. Kollet, 2014: Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP v1.0) in a massively parallel supercomputing environment – a case study on JUQUEEN (IBM Blue Gene/Q). *Geoscientific Model Development*, **7 (5)**, 2531–2543, doi:10.5194/gmd-7-2531-2014.

Getirana, A., S. Kumar, M. Girotto, and M. Rodell, 2017: Rivers and Floodplains as Key Components of Global Terrestrial Water Storage Variability: Water Storage in Rivers and Floodplains. *Geophysical Research Letters*, **44 (20)**, 10,359–10,368, doi:10.1002/2017GL074684.

Girotto, M., G. J. M. De Lannoy, R. H. Reichle, and M. Rodell, 2016: Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resources Research*, **52 (5)**, 4164–4183, doi:10.1002/2015WR018417.

Girotto, M., G. J. M. De Lannoy, R. H. Reichle, M. Rodell, C. Draper, S. N. Bhanja, and A. Mukherjee, 2017: Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India. *Geophysical Research Letters*, **44 (9)**, 4107–4115, doi:10.1002/2017GL072994.

Göhler, M., 2013: Sensitivity and Uncertainty Analysis of Environmental Models. Ph.D. thesis, University Jena, Jena.

Golub, G. H. and C. F. V. Loan, 2012: *Matrix Computations*. JHU Press.

Goodliff, M., J. Amezcua, and P. J. Van Leeuwen, 2015: Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity. *Tellus A: Dynamic Meteorology and Oceanography*, **67 (1)**, 26 928, doi:10.3402/tellusa.v67.26928.

Green, T. R., M. Taniguchi, H. Kooi, J. J. Gurdak, D. M. Allen, K. M. Hiscock, H. Treidel, and A. Aureli, 2011: Beneath the surface of global change: Impacts of climate change on groundwater. *Journal of Hydrology*, **405 (3)**, 532–560, doi:10.1016/j.jhydrol.2011.05.002.

Grippa, M., et al., 2011: Land water storage variability over West Africa estimated by Gravity Recovery and Climate Experiment (GRACE) and land surface models. *Water Resources Research*, **47 (5)**, doi:10.1029/2009WR008856.

Gudmundsson, L. and S. I. Seneviratne, 2015: Towards observation-based gridded runoff estimates for Europe. *Hydrology and Earth System Sciences*, **19 (6)**, 2859–2879, doi:10.5194/hess-19-2859-2015.

Gudmundsson, L. and S. I. Seneviratne, 2016: Observation-based gridded runoff estimates for Europe (E-RUN version 1.1). *Earth System Science Data*, **8 (2)**, 279–295, doi:10.5194/essd-8-279-2016.

Gunter, B., J. Ries, S. Bettadpur, and B. Tapley, 2006: A simulation study of the errors of omission and commission for GRACE RL01 gravity fields. *J Geodesy*, **80 (7)**, 341–351, doi:10.1007/s00190-006-0083-3.

Güntner, A., 2008: Improvement of global hydrological models using GRACE data. *Surv. Geophys.*, **29**, 375–397, doi:10.1007/s10712-008-9038-y.

Guo, D., S. Westra, and H. R. Maier, 2017: Impact of evapotranspiration process representation on runoff projections from conceptual rainfall-runoff models. *Water Resources Research*, **53 (1)**, 435–454, doi:10.1002/2016WR019627.

Han, S.-C., H. Kim, I.-Y. Yeo, P. Yeh, T. Oki, K.-W. Seo, D. Alsdorf, and S. B. Luthcke, 2009: Dynamics of surface water storage in the Amazon inferred from measurements of inter-satellite distance change. *Geophysical Research Letters*, **36 (9)**, doi:10.1029/2009GL037910.

Han, S.-C., D. D. Rowlands, S. B. Luthcke, and F. G. Lemoine, 2008: Localized analysis of satellite tracking data for studying time-variable Earth's gravity fields. *Journal of Geophysical Research*, **113 (B6)**, doi:10.1029/2007JB005218.

Han, S.-C., J. Sauber, and S. Luthcke, 2010a: Regional gravity decrease after the 2010 Maule (Chile) earthquake indicates large-scale mass redistribution. *Geophysical Research Letters*, **37 (23)**, n/a–n/a, doi:10.1029/2010GL045449.

Han, S.-C., I.-Y. Yeo, D. Alsdorf, P. Bates, J.-P. Boy, H. Kim, T. Oki, and M. Rodell, 2010b: Movement of Amazon surface water from time-variable satellite gravity measurements and implications for water cycle parameters in land surface models. *Geochemistry Geophysics Geosystems*, **11 (9)**, doi:10.1029/2010GC003214.

Han, X., H.-J. H. Franssen, C. Montzka, and H. Vereecken, 2014: Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations. *Water Resour. Res.*, **50**, 6081–6105, doi:10.1002/2013WR014586.

Han, X., X. Li, H. J. H. Franssen, H. Vereecken, and C. Montzka, 2012: Spatial horizontal correlation characteristics in the land data assimilation of soil moisture. *Hydrol. Earth Syst. Sci.*, **16**, 1349–1363, doi:10.5194/hess-16-1349-2012.

Harris, I., P. Jones, T. Osborn, and D. Lister, 2014: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, **34 (3)**, 623–642, doi:10.1002/joc.3711.

Haylock, M. R., N. Hofstra, A. M. G. K. Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *J. Geophys. Res.*, **113**, D20 119, doi:10.1029/2008JD010201.

Hiederer, R., European Commission, Joint Research Centre, and Institute for Environment and Sustainability, 2013: *Mapping Soil Properties for Europe: Spatial Representation of Soil Database Attributes.* Publications Office, Luxembourg, oCLC: 870627114.

Hofmann-Wellenhof, B. and H. Moritz, 2006: *Physical Geodesy.* Springer Vienna, doi:10.1007/978-3-211-33545-1.

Houborg, R., M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik, 2012: Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resour. Res.*, **48**, W07 525.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, **126 (3)**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Houtekamer, P. L. and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **129 (1)**, 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Houtekamer, P. L. and F. Zhang, 2016: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **144 (12)**, 4489–4532, doi:10.1175/MWR-D-15-0440.1.

Jekeli, C., 1981: Alternative methods to smooth the Earth's gravity field. Tech. Rep. 327, Ohio State University, Columbus, Ohio.

Jekeli, C., 1999: The determination of gravitational potential differences from satellite-to-satellite tracking. *Celestial Mechanics and Dynamical Astronomy*, **75 (2)**, 85–101, doi:10.1023/A:1008313405488.

Jekeli, C., 2017: The Energy Balance Approach. *SpringerLink*, 127–160, doi:10.1007/978-3-319-49941-3_5.

Jin, S. and G. Feng, 2013: Large-scale variations of global groundwater from satellite gravimetry and hydrological models, 2002–2012. *Global and Planetary Change*, **106**, 20–30, doi:10.1016/j.gloplacha.2013.02.008.

Johnson, G. C. and D. P. Chambers, 2013: Ocean bottom pressure seasonal cycles and decadal trends from GRACE Release-05: Ocean circulation implications. *Journal of Geophysical Research: Oceans*, **118 (9)**, 4228–4240, doi:10.1002/jgrc.20307.

Joodaki, G., J. Wahr, and S. Swenson, 2014: Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations. *Water Resources Research*, **50 (3)**, 2679–2692, doi:10.1002/2013WR014633.

Jung, M., et al., 2011: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.*, **116**, G00J07, doi:10.1029/2010JG001566.

Keune, J., F. Gasper, K. Goergen, A. Hense, P. Shrestha, M. Sulis, and S. Kollet, 2016: Studying the influence of groundwater representations on land surface-atmosphere feedbacks during the European heat wave in 2003: Groundwater Feedbacks on 2003 Heat Wave. *Journal of Geophysical Research: Atmospheres*, **121 (22)**, 13,301–13,325, doi:10.1002/2016JD025426.

Khaki, M., E. Forootan, M. Kuhn, J. Awange, A. I. J. M. van Dijk, M. Schumacher, and M. A. Sharifi, 2018: Determining water storage depletion within Iran by assimilating GRACE data into the W3RA hydrological model. *Advances in Water Resources*, **114**, 1–18, doi: 10.1016/j.advwatres.2018.02.008.

Khaki, M., I. Hoteit, M. Kuhn, J. Awange, E. Forootan, A. I. J. M. van Dijk, M. Schumacher, and C. Pattiaratchi, 2017a: Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. *Advances in Water Resources*, **107**, 301–316, doi: 10.1016/j.advwatres.2017.07.001.

Khaki, M., M. Schumacher, E. Forootan, M. Kuhn, J. Awange, and A. van Dijk, 2017b: Accounting for spatial correlation errors in the assimilation of GRACE into hydrological models through localization. *Advances in Water Resources*, **108**, 99–112, doi:10.1016/j.advwatres.2017.07.024.

Kierulf, H. P., H. Steffen, M. J. R. Simpson, M. Lidberg, P. Wu, and H. Wang, 2014: A GPS velocity field for Fennoscandia and a consistent comparison to glacial isostatic adjustment models. *J. Geophys. Res. Solid Earth*, **119 (8)**, 6613–6629, doi:10.1002/2013JB010889.

Kirchgessner, P., J. Tödter, B. Ahrens, and L. Nerger, 2017: The smoother extension of the nonlinear ensemble transform filter. *Tellus A: Dynamic Meteorology and Oceanography*, **69 (1)**, 1327 766, doi:10.1080/16000870.2017.1327766.

Klees, R., E. A. Zapreeva, H. C. Winsemius, and H. H. G. Savenije, 2007: The bias in GRACE estimates of continental water storage variations. *Hydrol. Earth Syst. Sci.*, **11**, 1227–1241, doi:10.5194/hess-11-1227-2007.

Klemann, V. and Z. Martinec, 2009: Contribution of glacial-isostatic adjustment to the geocenter motion. *Tectonophysics*, **511**, 99–108, doi:10.1016/j.tecto.2009.08.031.

Klinger, B. and T. Mayer-Gürr, 2016: The role of accelerometer data calibration within GRACE gravity field recovery: Results from ITSG-Grace2016. *Advances in Space Research*, **58 (9)**, 1597–1609, doi:10.1016/j.asr.2016.08.007.

Koch, K.-R., 1988: *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer-Verlag New York, Inc.

Koch, K.-R., 2007: *Introduction to Bayesian Statistics*. Springer-Verlag Berlin Heidelberg.

Koster, R. and M. J. Suarez, 1996: Energy and water balance calculations in the Mosaic LSM. Tech. Rep. TM-104606, Laboratory for Atmospheres, Goddard Space Flight Center.

Koster, R. D. and M. J. Suarez, 1992: Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *Journal of Geophysical Research*, **97 (D3)**, 2697, doi:10.1029/91JD01696.

Kotlarski, S., et al., 2013: Regional climate modeling on European scales : A joint standard evaluation of the EURO-CORDEX RCM ensemble. *International Conference on Regional Climate - CORDEX 2013*.

Kumar, S. V., et al., 2016: Assimilation of Gridded GRACE Terrestrial Water Storage Estimates in the North American Land Data Assimilation System. *Journal of Hydrometeorology*, **17 (7)**, 1951–1972, doi:10.1175/JHM-D-15-0157.1.

Kurtenbach, E., 2011: Entwicklung eines Kalman-Filters zur Bestimmung kurzzeitiger Variationen des Erdschwerefeldes aus Daten der Satellitenmission GRACE. Ph.D. thesis, Univ. Bonn, Inst. für Geodäsie und Geoinformation.

Kurtenbach, E., A. Eicker, T. Mayer-Gürr, M. Holschneider, M. Hayn, M. Fuhrmann, and J. Kusche, 2012: Improved daily GRACE gravity field solutions using a Kalman smoother. *Journal of Geodynamics*, **59-60**, 39–48, doi:10.1016/j.jog.2012.02.006.

Kurtz, W., G. He, S. J. Kollet, R. M. Maxwell, H. Vereecken, and H.-J. H. Franssen, 2016: TerrSysMP–PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface–subsurface model. *Geosci. Model Dev.*, **9**, 1341–1360.

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

Kusche, J., A. Eicker, E. Forootan, A. Springer, and L. Longuevergne, 2016: Mapping probabilities of extreme continental water storage changes from space gravimetry. *Geophysical Research Letters*, **43 (15)**, 8026–8034, doi:10.1002/2016GL069538.

Kusche, J., R. Schmidt, S. Petrovic, and R. Rietbroek, 2009a: Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *J Geod*, **83 (10)**, 903–913, doi:10.1007/s00190-009-0308-3.

Kusche, J., R. Schmidt, S. Petrovic, and R. Rietbroek, 2009b: Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *Journal of geodesy*, **83 (10)**, 903–913.

Kusche, J. and A. Springer, 2017: Parameter Estimation for Satellite Gravity Field Modeling. *Global Gravity Field Modeling from Satellite-to-Satellite Tracking Data*, M. Naeimi and J. Flury, Eds., Springer.

Landerer, F. W. and S. C. Swenson, 2012: Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, **48 (4)**, n/a–n/a, doi:10.1029/2011WR011453, w04531.

Landerer, F. W., D. N. Wiese, K. Bentel, C. Boening, and M. M. Watkins, 2015: North Atlantic meridional overturning circulation variations from GRACE ocean bottom pressure anomalies. *Geophysical Research Letters*, **42 (19)**, 8114–8121, doi:10.1002/2015GL065730.

Lawrence, D. M., et al., 2011: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, **3 (1)**, n/a–n/a, doi:10.1029/2011MS00045.

Leblanc, M. J., P. Tregoning, G. Ramillien, S. O. Tweed, and A. Fakes, 2009: Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia. *Water Resources Research*, **45 (4)**, doi:10.1029/2008WR007333.

Lemoine, J.-M., S. Bruinsma, S. Loyer, R. Biancale, J.-C. Marty, F. Perosanz, and G. Balmino, 2007: Temporal gravity field models inferred from GRACE data. *Advances in Space Research*, **39 (10)**, 1620–1629, doi:10.1016/j.asr.2007.03.062.

Li, B., D. Toll, X. Zhan, and B. Cosgrove, 2012: Improving estimated soil moisture fields through assimilation of AMSR-E soil moisture retrievals with an ensemble Kalman filter and a mass conservation constraint. *Hydrology and Earth System Sciences*, **16 (1)**, 105–119, doi:10.5194/hess-16-105-2012.

Li, J., et al., 2013: Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrology and Earth System Sciences*, **17 (8)**, 3279–3293, doi:10.5194/hess-17-3279-2013.

Liu, X., 2008: Global gravity field recovery from satellite-to-satellite tracking data with the acceleration approach. *Earth Observation with CHAMP.*

Liu, X., P. Ditmar, C. Siemes, D. C. Slobbe, E. Revtova, R. Klees, R. Riva, and Q. Zhao, 2010: DEOS Mass Transport model (DMT-1) based on GRACE satellite data: Methodology and validation. *Geophysical Journal International*, doi:10.1111/j.1365-246X.2010.04533.x.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. De Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. P. Evans, 2011: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrol. Earth Syst. Sci.*, **15 (2)**, 425–436, doi:10.5194/hess-15-425-2011.

Livneh, B. and D. P. Lettenmaier, 2012: Multi-criteria parameter estimation for the Unified Land Model. *Hydrology and Earth System Sciences*, **16 (8)**, 3029–3048, doi:10.5194/hess-16-3029-2012.

Lo, M.-H., J. S. Famiglietti, P. J.-F. Yeh, and T. H. Syed, 2010: Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data. *Water Resour. Res.*, **46**, W05 517.

Lombard, A., et al., 2007: Estimation of steric sea level variations from combined GRACE and Jason-1 data. *Earth and Planetary Science Letters*, **254 (1-2)**, 194–202, doi:10.1016/j.epsl.2006.11.035.

Long, D., L. Longuevergne, and B. R. Scanlon, 2015: Global analysis of approaches for deriving total water storage changes from GRACE satellites: Global analysis of GRACE total water storage estimates. *Water Resources Research*, **51 (4)**, 2574–2594, doi:10.1002/2014WR016853.

Long, D., Y. Pan, J. Zhou, Y. Chen, X. Hou, Y. Hong, B. R. Scanlon, and L. Longuevergne, 2017: Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models. *Remote Sensing of Environment*, **192**, 198–216, doi:10.1016/j.rse.2017.02.011.

Long, D., et al., 2016: Have GRACE satellites overestimated groundwater depletion in the Northwest India Aquifer? *Scientific Reports*, **6**, doi:10.1038/srep24398.

Longuevergne, L., B. Scanlon, and C. R. Wilson, 2010: GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resour. Res.*, **46**, W11 517, doi:10.1029/2009WR008564.

Lorenz, C., et al., 2014: Large-Scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances. http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-13-0157.1, doi:10.1175/JHM-D-13-0157.1.

Lück, C., J. Kusche, R. Rietbroek, and A. Löcher, 2018: Time-variable gravity fields and ocean mass change from 37 months of kinematic Swarm orbits. *Solid Earth*, **9 (2)**, 323–339, doi:https://doi.org/10.5194/se-9-323-2018.

Luthcke, S. B., D. D. Rowlands, T. J. Sabaka, B. D. Loomis, M. Horwath, and A. A. Arendt, 2014: Gravimetry measurements from space. *Remote Sensing of the Cryosphere*, John Wiley & Sons, Ltd, 231–247, doi:10.1002/9781118368909.ch10.

Luthcke, S. B., T. Sabaka, B. Loomis, A. Arendt, J. McCarthy, and J. Camp, 2013: Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *Journal of Glaciology*, **59 (216)**, 613–631, doi:10.3189/2013JoG12J147.

Lv, M., Z. Ma, X. Yuan, M. Lv, M. Li, and Z. Zheng, 2017: Water budget closure based on GRACE measurements and reconstructed evapotranspiration using GLDAS and water use data for two large densely-populated mid-latitude basins. *Journal of Hydrology*, **547**, 585–599, doi:10.1016/j.jhydrol.2017.02.027.

Mayer-Guerr, T., 2015: The combined satellite gravity field model GOCO05s. *EGU General Assembly Conference Abstracts*, Vol. 17, 12364.

Mayer-Gürr, T., 2006: Gravitationsfeldbestimmung aus der Analyse kurzer Bahnbögen am Beispiel der Satellitenmissionen CHAMP und GRACE. Ph.D. thesis, Universität Bonn.

Mayer-Gürr, T., S. Behzadpour, M. Ellmer, A. Kvas, B. Klinger, and N. Zehentner, 2016: ITSG-Grace2016 - Monthly and Daily Gravity Field Solutions from GRACE. GFZ Data Services, doi:10.5880/icgem.2016.007.

Mayer-Gürr, T., S. Behzadpur, M. Ellmer, A. Kvas, B. Klinger, S. Strasser, and N. Zehentner, 2018: ITSG-Grace2018 - Monthly, Daily and Static Gravity Field Solutions from GRACE. GFZ Data Services, doi:10.5880/ICGEM.2018.003.

Mayer-Gürr, T., N. Zehenter, B. Klinger, and A. Kvas, 2014: ITSG- Grace2014: A new GRACE gravity field release computed in Graz. Presentation at the GRACE Science Team Meeting (GSTM), Pots- dam 09/2014. *Presentation at the GRACE Science Team Meeting (GSTM), Potsdam 09/2014*.

McCarthy, D. and G. Petit, 2004: IERS Conventions 2003. Tech. Rep. No. 32, Verlag des Bundesamts für Kartographie und Geodäsied, Frankfurt am Main.

Mekonnen, M. M. and A. Y. Hoekstra, 2016: Four billion people facing severe water scarcity. *Science Advances*, **2 (2)**, e1500 323, doi:10.1126/sciadv.1500323.

Meyer, U., A. Jäggi, Y. Jean, and G. Beutler, 2016: AIUB-RL02: An improved time-series of monthly gravity fields from GRACE data. *Geophysical Journal International*, **205 (2)**, 1196–1207, doi:10.1093/gji/ggw081.

Meyrath, T., T. van Dam, X. Collilieux, and P. Rebischung, 2017: Seasonal low-degree changes in terrestrial water mass load from global GNSS measurements. *Journal of Geodesy*, **91 (11)**, 1329–1350, doi:10.1007/s00190-017-1028-8.

Miller, R. N., 1994: Perspectives on Advanced Data Assimilation in Strongly Nonlinear Systems. *Data Assimilation*, 195–215, doi:10.1007/978-3-642-78939-7_8.

Minamide, M. and F. Zhang, 2016: Adaptive Observation Error Inflation for Assimilating All-Sky Satellite Radiance. *Mon. Wea. Rev.*, **145 (3)**, 1063–1081, doi:10.1175/MWR-D-16-0257.1.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **130 (11)**, 2791–2808, doi:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.

Mouelhi, S., 2003: Vers une chaîne cohérente de modèles pluie-débit aux pas de temps pluriannue, annuel, mensuel et journalier. Ph.D. thesis, ENGREF (AgroParisTech).

Müller Schmied, H., S. Eisner, D. Franz, M. Wattenbach, F. T. Portmann, M. Flörke, and P. Döll, 2014: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, **18 (9)**, 3511–3538, doi:10.5194/hess-18-3511-2014.

Naeimi, M., I. Koch, A. Khami, and J. Flury, 2018: IfE monthly gravity field solutions using the variational equations. *EGU General Assembly*, Vienna, 8-13 April 2018.

Nasonova, O. N., Y. M. Gusev, and Y. E. Kovalev, 2011: Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components. *Hydrological Processes*, **25 (7)**, 1074–1090, doi:10.1002/hyp.7651.

Nelson, K. and E. Burchfield, 2017: Effects of the Structure of Water Rights on Agricultural Production During Drought: A Spatiotemporal Analysis of California's Central Valley. *Water Resources Research*, doi:10.1002/2017WR020666.

Nerger, L., 2004: Parallel filter algorithms for data assimilation in oceanography. Ph.D. thesis, Universität Bremen.

Nerger, L., S. Danilov, G. Kivman, W. Hiller, and J. Schröter, 2007: Data assimilation with the Ensemble Kalman Filter and the SEIK filter applied to a finite element model of the North Atlantic. *Journal of Marine Systems*, **65 (1)**, 288–298, doi:10.1016/j.jmarsys.2005.06.009.

Nerger, L. and W. Hiller, 2013: Software for Ensemble-based Data Assimilation Systems - Implementation Strategies and Scalability. *Comput. Geosci.*, **55**, 110–118, doi:10.1016/j.cageo.2012.03.026.

Nerger, L., W. Hiller, and J. Schröter, 2005: PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman Filtering. *Use of High Performance Computing in Meteorology: Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology, Reading, UK, 25-29 October 2004/Eds.: Walter Zwieflhofer; Geoge Mozdzynski, Singapore [Ua]: World*, 63–83.

Nerger, L., T. Janjić, J. Schröter, and W. Hiller, 2012: A Unification of Ensemble Square Root Kalman Filters. *Mon. Wea. Rev.*, **140 (7)**, 2335–2345, doi:10.1175/MWR-D-11-00102.1.

Niu, G.-Y. and Z.-L. Yang, 2006: Assessing a land surface model's improvements with GRACE estimates. *Geophysical Research Letters*, **33 (7)**, doi:10.1029/2005GL025555.

Oleson, K., et al., 2013: Technical Description of version 4.5 of the Community Land Model (CLM). Ncar Technical Note NCAR/TN-503+STR, National Center for Atmospheric Research, Boulder, CO.

Oleson, K. W., et al., 2004: Technical Description of the Community Land Model (CLM). NCAR Technical Note, National Center for Atmospheric Research, Boulder, CO.

Oleson, K. W., et al., 2008: Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res.*, **113**, G01 021.

Peel, M. C., B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, **11 (5)**, 1633–1644, doi:https://doi.org/10.5194/hess-11-1633-2007.

Petit, G. and B. Luzum, 2010: IERS Conventions (2010). Tech. Rep. IERS Technical Note ; 36, Frankfurt am Main: Verlag des Bundesamts für Kartographie und Geodäsie, 179 pp pp.

Pham, D. T., 2001: Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. *Mon. Wea. Rev.*, **129 (5)**, 1194–1207, doi:10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2.

Pham, D. T., J. Verron, and M. Christine Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, **16 (3)**, 323–340, doi:10.1016/S0924-7963(97)00109-7.

Pokhrel, Y. N., Y. Fan, G. Miguez-Macho, P. J.-F. Yeh, and S.-C. Han, 2013: The role of groundwater in the Amazon water cycle: 3. Influence on terrestrial water storage computations and comparison with GRACE. *Journal of Geophysical Research: Atmospheres*, **118 (8)**, 3233–3244, doi:10.1002/jgrd.50335.

Poméon, T., D. Jackisch, and B. Diekkrüger, 2017: Evaluating the performance of remotely sensed and reanalysed precipitation data over West Africa using HBV light. *Journal of Hydrology*, **547**, 222–235, doi:10.1016/j.jhydrol.2017.01.055.

Rabier, F., P. Courtier, and M. Ehrendorfer, 1992: Four-Dimensional Data Assimilation: Comparison of Variational and Sequential Algorithms. *Quarterly Journal of the Royal Meteorological Society*, **118 (506)**, 673–713, doi:10.1002/qj.49711850605.

Rahman, M. M. and M. Lu, 2015: Model Spin-Up Behavior for Wet and Dry Basins: A Case Study Using the Xinanjian Model. *Water*, **7**, 4256–4273, doi:10.3390/w7084256.

Rakovec, O., R. Kumar, S. Attinger, and L. Samaniego, 2016: Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, **52 (10)**, 7779–7792, doi:10.1002/2016WR019430.

Ramillien, G., A. Cazenave, and O. Brunau, 2004: Global time variations of hydrological signals from GRACE satellite gravimetry: Global time variations of hydrology from GRACE. *Geophysical Journal International*, **158 (3)**, 813–826, doi:10.1111/j.1365-246X.2004.02328.x.

Ramillien, G., F. Frappart, A. Cazenave, and A. Guntner, 2005: Time variations of land water storage from an inversion of 2 years of GRACE geoids. *Earth and Planetary Science Letters*, **235 (1-2)**, 283–301, doi:10.1016/j.epsl.2005.04.005.

Rangelova, E., W. van der Wal, A. Braun, M. G. Sideris, and P. Wu, 2007: 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*, **112 (F3)**, doi:10.1029/2006JF000615.

Reager, J., A. Thomas, E. Sproles, M. Rodell, H. Beaudoing, B. Li, and J. Famiglietti, 2015: Assimilation of GRACE Terrestrial Water Storage Observations into a Land Surface Model for the Assessment of Regional Flood Potential. *Remote Sensing*, **7 (11)**, 14 663–14 679, doi:10.3390/rs71114663.

Reager, J. T., B. F. Thomas, and J. S. Famiglietti, 2014: River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nature Geoscience*, **7 (8)**, 588–592, doi:10.1038/ngeo2203.

Rebetez, M., H. Mayer, O. Dupont, D. Schindler, K. Gartner, J. P. Kropp, and A. Menzel, 2006: Heat and drought 2003 in Europe: A climate synthesis. *Annals of Forest Science*, **63 (6)**, 569–577, doi:10.1051/forest:2006043.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research: Atmospheres*, **112 (D9)**, doi:10.1029/2006JD008033.

Reichle, R. H., S. V. Kumar, S. P. P. Mahanama, R. D. Koster, and Q. Liu, 2010: Assimilation of Satellite-Derived Skin Temperature Observations into Land Surface Models. *J. Hydrometeor.*, **11**, 1103–1122, doi:10.1175/2010JHM1262.1.

Richards, L. A., 1931: Capillary conduction of liquids through porous mediums. *Physics*, **1 (5)**, 318–333, doi:10.1063/1.1745010.

Rietbroek, R., 2014: Retrieval of sea level and surface loading variations from geodetic observations and model simulations: An integration approach. PhD Thesis, University of Bonn, Germany.

Rietbroek, R., S.-E. Brunnabend, J. Kusche, J. Schröter, and C. Dahle, 2016: Revisiting the contemporary sea-level budget on global and regional scales. *PNAS*, **113 (6)**, 1504–1509, doi:10.1073/pnas.1519132113.

Rietbroek, R., M. Fritsche, S.-E. Brunnabend, I. Daras, J. Kusche, J. Schröter, F. Flechtner, and R. Dietrich, 2012: Global surface mass from a new combination of GRACE, modelled OBP and reprocessed GPS data. *J. Geodyn.*, **59-60**, 64–71, doi:10.1016/j.jog.2011.02.003.

Riley, J. D., M. M. Bennett, and E. McCormick, 1967: Numerical integration of variational equations. *Math. Comp.*, **21 (97)**, 12–17, doi:10.1090/S0025-5718-1967-0228186-1.

Robert, C., E. Blayo, and J. Verron, 2006: Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean. *Ocean Dynamics*, **56 (5-6)**, 624–633, doi:10.1007/s10236-006-0079-9.

Rodell, M., J. S. Famiglietti, D. N. Wiese, J. T. Reager, H. K. Beaudoing, F. W. Landerer, and M.-H. Lo, 2018: Emerging trends in global freshwater availability. *Nature*, **557 (7707)**, 651, doi:10.1038/s41586-018-0123-1.

Rodell, M., I. Velicogna, and J. S. Famiglietti, 2009: Satellite-based estimates of groundwater depletion in India. *Nature*, **460 (7258)**, doi:10.1038/nature08238.

Ross, P. J., 1990: Efficient numerical methods for infiltration using Richards' equation. *Water Resources Research*, **26 (2)**, 279–290, doi:10.1029/WR026i002p00279.

Sakumura, C., S. Bettadpur, and S. Bruinsma, 2014: Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.*, **41**, 1389–1397.

Sakumura, C., S. Bettadpur, H. Save, and C. McCullough, 2016: High-frequency terrestrial water storage signal capture via a regularized sliding window mascon product from GRACE. *Journal of Geophysical Research: Solid Earth*, **121 (5)**, 4014–4030, doi: 10.1002/2016JB012843.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, **98 (12)**, 875–883, doi:10.1175/1520-0493(1970)098<0875:SBFINV>2.3.CO;2.

Savcenko, R. and W. Bosch, 2012: EOT11A - Empirical Ocean Tide Model from Multi-Mission Satellite Altimetry. http://epic.awi.de/36001/.

Save, H., S. Bettadpur, and B. D. Tapley, 2012: Reducing errors in the GRACE gravity solutions using regularization. *Journal of Geodesy*, **86 (9)**, 695–711, doi:10.1007/s00190-012-0548-5.

Save, H., S. Bettadpur, and B. D. Tapley, 2016: High-resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth*, **121 (10)**, 7547–7569, doi:10.1002/2016JB013007.

Scanlon, B. R., Z. Zhang, H. Save, D. N. Wiese, F. W. Landerer, D. Long, L. Longuevergne, and J. Chen, 2016: Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resources Research*, **52 (12)**, 9412–9429, doi:10.1002/2016WR019494.

Scanlon, B. R., et al., 2018: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *PNAS*, **115 (6)**, E1080–E1089, doi:10.1073/pnas.1704665115.

Schewe, J., et al., 2013: Multimodel assessment of water scarcity under climate change. *PNAS*, 201222460, doi:10.1073/pnas.1222460110.

Schmidt, R., et al., 2006: GRACE observations of changes in continental water storage. *Global and Planetary Change*, **50 (1-2)**, 112–126, doi:10.1016/j.gloplacha.2004.11.018.

Schmidt, R., et al., 2007: GRACE Time-Variable Gravity Accuracy Assessment. *Dynamic Planet*, P. Tregoning and C. Rizos, Eds., Springer Berlin Heidelberg, Berlin, Heidelberg, Vol. 130, 237–243, doi:10.1007/978-3-540-49350-1_36.

Schrama, E. J. O., B. Wouters, and D. A. Lavallée, 2007: Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed surface mass variations. *Journal of Geophysical Research: Solid Earth*, **112 (B8)**, doi:10.1029/2006JB004882.

Schumacher, M., 2016: Methods for Assimilating remotely-Sensed Water Storage Changes into Hydrological Models. Ph.D. thesis, Bonn University, Bonn.

Schumacher, M., A. Eicker, J. Kusche, H. M. Schmied, and P. Döll, 2012: International Association of Geodesy Symposia. Springer Berlin Heidelberg.

Schumacher, M., E. Forootan, A. van Dijk, H. Müller Schmied, R. Crosbie, J. Kusche, and P. Döll, 2018: Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model. *Remote Sensing of Environment*, **204**, 212–228, doi:10.1016/j.rse.2017.10.029.

Schumacher, M., J. Kusche, and P. Döll, 2016: A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models. *J. Geod.*

Sheard, B. S., G. Heinzel, K. Danzmann, D. A. Shaddock, W. M. Klipstein, and W. M. Folkner, 2012: Intersatellite laser ranging instrument for the GRACE follow-on mission. *J Geod*, **86 (12)**, 1083–1095, doi:10.1007/s00190-012-0566-3.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *Journal of Climate*, **19 (13)**, 3088–3111, doi:10.1175/JCLI3790.1.

Shrestha, P., M. Sulis, M. Masbou, S. Kollet, and C. Simmer, 2014: A Scale-Consistent Terrestrial Systems Modeling Platform Based on COSMO, CLM, and ParFlow. *Mon. Weather Rev.*, **142**, 3466–3483.

Simon, K. M., R. E. M. Riva, M. Kleinherenbrink, and T. Frederikse, 2018: The glacial isostatic adjustment signal at present day in northern Europe and the British Isles estimated from geodetic observations and geophysical models. *Solid Earth*, **9 (3)**, 777–795, doi:https://doi.org/10.5194/se-9-777-2018.

Skamarock, C., et al., 2008: A Description of the Advanced Research WRF Version 3. doi:10.5065/D68S4MVH.

Solander, K. C., J. T. Reager, Y. Wada, J. S. Famiglietti, and R. S. Middleton, 2017: GRACE satellite observations reveal the severity of recent water over-consumption in the United States. *Scientific Reports*, **7 (1)**, 8723, doi:10.1038/s41598-017-07450-y.

Springer, A., A. Eicker, A. Bettge, J. Kusche, and A. Hense, 2017: Evaluation of the Water Cycle in the European COSMO-REA6 Reanalysis Using GRACE. *Water*, **9 (4)**, 289, doi:10.3390/w9040289.

Springer, A., K. Karegar, J. Kusche, J. Keune, W. Kurtz, and S. Kollet, 2019: Evidence of daily hydrological loading in GPS time series over Europe. *to be submitted to J. Geodesy.*

Springer, A., J. Kusche, K. Hartung, C. Ohlwein, and L. Longuevergne, 2014: New Estimates of Variations in Water Flux and Storage over Europe Based on Regional (Re)Analyses and Multisensor Observations. *J. Hydrometeor.*, doi:10.1175/JHM-D-14-0050.1.

Standish, E., 1998: JPL Planetary and Lunar Ephemeriedes DE406/LE406. http://www.willbell.com/software/jpl.htm.

Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans*, **103 (C4)**, 7755–7766, doi:10.1029/97JC03180.

Su, H., Z.-L. Yang, R. E. Dickinson, C. R. Wilson, and G.-Y. Niu, 2010: Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information. *J. Geophys. Res.*, **115**, D10 104.

Sun, Y., P. Ditmar, and R. Riva, 2017: Statistically optimal estimation of degree-1 and C20 coefficients based on GRACE data and an ocean bottom pressure model. *Geophysical Journal International*, **210 (3)**, 1305–1322, doi:10.1093/gji/ggx241.

Swenson, S., D. Chambers, and J. Wahr, 2008: Estimating geocenter variations from a combination of GRACE and ocean model output. *Journal of Geophysical Research*, **113 (B8)**, doi:10.1029/2007JB005338.

Swenson, S. and J. Wahr, 2006: Post-processing removal of correlated errors in GRACE data. *Geophysical Research Letters*, **33 (8)**, doi:10.1029/2005GL025285.

Swenson, S. and J. Wahr, 2007: Multi-sensor analysis of water storage variations of the Caspian Sea. *Geophysical Research Letters*, **34 (16)**, doi:10.1029/2007GL030733.

Syed, T. H., J. S. Famiglietti, and D. P. Chambers, 2009: GRACE-Based Estimates of Terrestrial Freshwater Discharge from Basin to Continental Scales. *J. Hydrometeor.*, **10**, 22–40, doi:10.1175/2008JHM993.1.

Tanaka, M., T. A. Adjadeh, S. Tanaka, and T. Sugimura, 2002: Water surface area measurement of Lake Volta using SSM/I 37-GHz polarization difference in rainy season. *Advances in Space Research*, **30 (11)**, 2501–2504, doi:10.1016/S0273-1177(02)80320-9.

Tangdamrongsub, N., S. C. Steele-Dunne, B. C. Gunter, P. G. Ditmar, E. H. Sutanudjaja, Y. Sun, T. Xia, and Z. Wang, 2017: Improving estimates of water resources in a semi-arid region by assimilating GRACE data into the PCR-GLOBWB hydrological model. *Hydrology and Earth System Sciences*, **21 (4)**, 2053–2074, doi:https://doi.org/10.5194/hess-21-2053-2017.

Tangdamrongsub, N., S. C. Steele-Dunne, B. C. Gunter, P. G. Ditmar, and A. H. Weerts, 2015: Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin. *Hydrol. Earth Syst. Sci.*, **19**, 2079–2100.

Tapley, B., F. Flechtner, C. Boening, and S. Bettadpur, 2016: GRACE Mission: Status and Prospects. *GRACE Science Team Meeting (GSTM)*, Potsdam, Germany.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins, 2004a: GRACE Measurements of Mass Variability in the Earth System. *Science*, **305**, 503–505, doi:10.1126/science.1099192.

Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber, 2004b: The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, **31 (9)**, doi:10.1029/2004GL019920.

Teutschbein, C., F. Wetterhall, and J. Seibert, 2011: Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale. *Clim Dyn*, **37 (9)**, 2087–2105, doi:10.1007/s00382-010-0979-8.

Tian, S., P. Tregoning, L. J. Renzullo, V. Dijk, A. I. J. M, J. P. Walker, V. R. N. Pauwels, and S. Allgeyer, 2017: Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research*, **53 (3)**, 1820–1840, doi:10.1002/2016WR019641.

Uebbing, B., J. Kusche, and E. Forootan, 2015: Waveform Retracking for Improving Level Estimations From TOPEX/Poseidon, Jason-1, and Jason-2 Altimetry Observations Over African Lakes. *IEEE Transactions on Geoscience and Remote Sensing*, **53 (4)**, 2211–2224, doi:10.1109/TGRS.2014.2357893.

van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning, 2014: 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.

van Leeuwen, P. J. and G. Evensen, 1996: Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. *Mon. Wea. Rev.*, **124 (12)**, 2898–2913, doi:10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.

Vautard, R., et al., 2013: The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Clim Dyn*, **41 (9-10)**, 2555–2575, doi:10.1007/s00382-013-1714-z.

Vetra-Carvalho, S., P. J. van Leeuwen, L. Nerger, A. Barth, M. U. Altaf, P. Brasseur, P. Kirchgessner, and J.-M. Beckers, 2018: State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus A: Dynamic Meteorology and Oceanography.*

Wachter, K., 2007: The Analysis of the Danube Floods 2006. Tech. rep., ICPDR – International Commission for the Protection of the Danube River.

Wahr, J., M. Molenaar, and F. Bryan, 1998: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res.*, **103**, 30 205–30 229, doi:10.1029/98JB02844.

Wahr, J., S. Swenson, and I. Velicogna, 2006: Accuracy of GRACE mass estimates. *Geophys. Res. Lett.*, **33**, L06 401, doi:10.1029/2005GL025305.

Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna, 2004: Time-variable gravity from GRACE: First results. *Geophysical Research Letters*, **31 (11)**, doi:10.1029/2004GL019779.

Wahr, J. M., 1981: The forced nutations of an elliptical, rotating, elastic and oceanless earth. *Geophysical Journal International*, **64 (3)**, 705–727, doi:10.1111/j.1365-246X.1981.tb02691.x.

Wang, A. and X. Zeng, 2011: Sensitivities of terrestrial water cycle simulations to the variations of precipitation and air temperature in China. *Journal of Geophysical Research*, **116 (D2)**, doi:10.1029/2010JD014659.

Wang, C., H. Xu, M. Zhong, and W. Feng, 2015: Monthly gravity field recovery from GRACE orbits and K-band measurements using variational equations approach. *Geodesy and Geodynamics*, **6 (4)**, 253–260, doi:10.1016/j.geog.2015.05.010.

Wang, H., P. Wu, and W. van der Wal, 2008: Using postglacial sea level, crustal velocities and gravity-rate-of-change to constrain the influence of thermal effects on mantle lateral heterogeneities. *Journal of Geodynamics*, **46 (3-5)**, 104–117, doi:10.1016/j.jog.2008.03.003.

Wang, K. and R. E. Dickinson, 2012: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.*, **50**, RG2050, doi:10.1029/2011RG000373.

Wang, L., J. L. Davis, E. M. Hill, and M. E. Tamisiea, 2016: Stochastic filtering for determining gravity variations for decade-long time series of GRACE gravity. *Journal of Geophysical Research: Solid Earth*, **121 (4)**, 2915–2931, doi:10.1002/2015JB012650.

Watkins, M. M., D. N. Wiese, D.-N. Yuan, C. Boening, and F. W. Landerer, 2015: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, **120 (4)**, 2648–2671, doi:10.1002/2014JB011547.

Watkins, M. M. and D.-N. Yuan, 2014: JPL Level-2 Processing Standards Document. Tech. Rep. GRACE 327 - 744 (v 5.1), Jet Propulsion Laboratory California Institute of Technology.

Weigelt, M., 2017: The Acceleration Approach. *SpringerLink*, 97–126, doi:10.1007/978-3-319-49941-3_4.

Werth, S., A. Güntner, S. Petrovic, and R. Schmidt, 2009: Integration of GRACE mass variations into a global hydrological model. *Earth and Planetary Science Letters*, **277 (1-2)**, 166–173, doi:10.1016/j.epsl.2008.10.021.

Whitaker, S., 1986: Flow in porous media I: A theoretical derivation of Darcy's law. *Transp Porous Med*, **1 (1)**, 3–25, doi:10.1007/BF01036523.

Wiese, D. N., F. W. Landerer, and M. M. Watkins, 2016: Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, **52 (9)**, 7490–7502, doi:10.1002/2016WR019344.

Wolff, M., 1969: Direct measurements of the Earth's gravitational potential using a satellite pair. *Journal of Geophysical Research*, **74 (22)**, 5295–5300, doi:10.1029/JB074i022p05295.

Xie, H., L. Longuevergne, C. Ringler, and B. Scanlon, 2012: Calibration and evaluation of a semi-distributed watershed model of sub-Saharan Africa using GRACE data. *Hydrology and Earth System Sciences Discussions*, **9 (2)**, 2071–2120, doi:10.5194/hessd-9-2071-2012.

Yang, S.-C., E. Kalnay, B. Hunt, and N. E. Bowler, 2009: Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Quarterly Journal of the Royal Meteorological Society*, **135 (638)**, 251–262, doi:10.1002/qj.353.

Yang, T., C. Wang, Y. Chen, X. Chen, and Z. Yu, 2015: Climate change and water storage variability over an arid endorheic region. *Journal of Hydrology*, **529**, 330–339, doi:10.1016/j.jhydrol.2015.07.051.

Yang, T., C. Wang, Z. Yu, and F. Xu, 2013: Characterization of spatio-temporal patterns for various GRACE- and GLDAS-born estimates for changes of global terrestrial water storage. *Global and Planetary Change*, **109**, 30–37, doi:10.1016/j.gloplacha.2013.07.005.

Zaitchik, B. F., M. Rodell, and F. Olivera, 2010: Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resources Research*, **46 (6)**, doi:10.1029/2009WR007811.

Zaitchik, B. F., M. Rodell, R. H. Reichle, B. F. Zaitchik, M. Rodell, and R. H. Reichle, 2008: Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin. http://journals.ametsoc.org/doi/abs/10.1175/2007JHM951.1, doi:10.1175/2007JHM951.1.

Zenner, L., I. Bergmann-Wolf, H. Dobslaw, T. Gruber, A. Güntner, M. Wattenbach, S. Esselborn, and R. Dill, 2014: Comparison of daily GRACE gravity field and numerical water storage models for de-aliasing of satellite gravimetry observations. *Surv. Geophys.*

Zenner, L., E. Fagiolini, I. Daras, F. Flechtner, T. Gruber, T. Schmidt, and G. Schwarz, 2012: Non-tital atmospheric and oceanic mass variations and their impact on GRACE data analysis. *J. Geodyn.*, **59-60**, 9–15, doi:10.1016/j.jog.2012.01.010.

Zhang, L., H. Dobslaw, T. Stacke, A. Güntner, R. Dill, and M. Thomas, 2017: Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations. *Hydrology and Earth System Sciences*, **21 (2)**, 821–837, doi:10.5194/hess-21-821-2017.

Zhang, L., H. Dobslaw, and M. Thomas, 2016: Globally gridded terrestrial water storage variations from GRACE satellite gravimetry for hydrometeorological applications. *Geophysical Journal International*, **206 (1)**, 368–378, doi:10.1093/gji/ggw153.

Zhang, X. and S. Jin, 2014: Uncertainties and effects on geocenter motion estimates from global GPS observations. *Advances in Space Research*, **54 (1)**, 59–71, doi:10.1016/j.asr.2014.03.021.

Zhou, H., Z. Luo, and B. Zhong, 2015: WHU-Grace01s: A new temporal gravity field model recovered from GRACE KBRR data alone. *Geodesy and Geodynamics*, **6 (5)**, 316–323, doi:10.1016/j.geog.2015.07.004.

Zhou, H., Z. Luo, Z. Zhou, Q. Li, B. Zhong, and H. Hsu, 2016: A new time series of GRACE monthly gravity field models: HUST-Grace2016. GFZ Data Services, doi:10.5880/ICGEM.2016.009.